• Multiuser MIMO Tutorial •

Resource Allocation in OFDMA Broadcast Channels

Vienna, Feb 20 - March 2, 2007

Eduard Calvo

SPCOM Group

Dept. of Signal Theory and Communications

Technical University of Catalonia (UPC)

1. Introduction
2. The SISO case
   - Problem Formulation
   - Dual methods
   - Allocation quantization
3. The MISO case
   - A note on complexity
   - Heuristic subcarrier assignment
   - Fixed bandwidth assignment resource allocation
   - Allocation quantization
4. References

Preliminaries and notation

- We focus on the Broadcast Channel (BC) to model the the downlink of a cellular system, possibly MIMO, where the total bandwidth is partitioned into subcarriers via OFDM.
- Notation:
  - W: total transmission bandwidth.
  - P_T: total transmit power constraint.
  - n_T: number of antennas at the Base Station.
  - n_R: number of antennas at each user terminal.
  - $\mathbf{H}_k[n] \in \mathbb{C}^{n_R \times n_T}$: channel matrix of the k-th user on the n-th subcarrier.
  - $\sigma_k^2[n]$: Gaussian noise power of the k-th user on the n-th subcarrier.
  - $R_k[n]$: rate of the k-th user on the n-th subcarrier.
  - R_k: total rate of the k-th user.
  - $p_k[n]$: power allocated to the k-th user on the n-th subcarrier.
  - p_k: total power allocated to the k-th user.

Motivation of the problem

- Precedents:
  - Data flows of different service types share the same physical layer mechanism, leading to different Quality of Service (QoS) constraints (e.g., symbol (bit) error rate, outage probability, minimum rate) among users.
  - Bandwidth is scarce due to license cost or reduced size of unlicensed bands.
  - Other design constraints of either regulatory (e.g., total transmit power) or physical nature (e.g., spectral masks).
- Existing and upcoming wireless cellular systems account for feedback channels for gathering Channel State Information (CSI) of the users at the Base Station (BS).
- How should we allocate the network resources (subcarriers, power, rate) in a CSI-aware manner such that they are used more efficiently than without the feedback information?

"Efficient" resource allocation

- By dynamically adapting the resource allocation to the current network state we can benefit from the following beneficial effects:
  - Multiuser gain (if $K \gg n_T$).
  - Multiplexing gain (if $n_T > 1$).
  - Diversity gain (if either $n_T > 1$ or $n_R > 1$).
- Consequently, two different interpretations can be made of "Efficient Resource Allocation":
  - Use available resources to maximize some network utility function (given some QoS constraints or not).
    E.g., maximize a weighted sum of the users' rates $(\max \sum_k \mu_k R_k)$ given (W, P_T). μ_k is the priority of user k (possibly given by a scheduler obeying higher layer QoS considerations).
  - Use the minimum resources that achieve some pre-specified performance.
    E.g., minimize a weighted sum of the users' powers $(\min \sum_k \lambda_k p_k)$ given W such that each user is guaranteed a pre-specified rate R_k^0 (with a certain QoS or not). λ_k is the price of each Joule intended for the k-th user.

"Efficient" resource allocation

- From the system deployment point of view, both approaches are meaningful and interesting:
  - Network utility maximization can increase spectral efficiency and hence, cell coverage and supported traffic.
  - Resource minimization can reduce intercell interference when neighboring BS's are close (urban scenarios).
- Since both the network utility maximization and the resource minimization optimization criteria *usually* give rise to dual (similar) optimization problems, we will here adopt w.l.o.g. the first approach.

The SISO case

- Consider as a first stepping stone SISO case: $n_T = n_R = 1$, where each subcarrier is assigned to at most one user.
- Let $\mathcal{S}_k$ denote the set of subcarriers allocated to user k. Then,

$$\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset \text{ for } k \neq k', \quad \bigcup_{k=1}^{K} \mathcal{S}_k = \{1, 2, \ldots, N\} \tag{1}$$

- Now, the channel of the k-th user on the n-th tone is the scalar $H_k[n]$, which gives rise to the following channel signal-to-noise ratio $c_k[n] = |H_k[n]|^2/\sigma_k^2[n]$, and power and rate are related by the information theoretic achievable rate as

$$R_k[n] = \log_2(1 + c_k[n]p_k[n]). \tag{2}$$

Problem formulation

- The optimal resource allocation is the solution to the following optimization problem [1]

$$\underset{\{\{p_k[n]\}_{n=1}^{N},\mathcal{S}_k\}_{k=1}^{K}}{\text{maximize}} \quad \sum_{k=1}^{K} \mu_k \sum_{n\in\mathcal{S}_k} R_k[n] \tag{3}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} p_k[n] \leq P_T \tag{4}$$

$$\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset \ \forall k \neq k' \tag{5}$$

$$\bigcup_{k=1}^{K} \mathcal{S}_k = \{1, 2, \ldots, N\} \tag{6}$$

$$p_k[n] \geq 0 \ 1 \leq k \leq K, \ 1 \leq n \leq N. \tag{7}$$

Problem formulation

- Optimal allocation is the solution to a *combinatorial* problem whose complexity grows as $\mathcal{O}(NK^N)$.
- For some fixed subcarrier assignment the problem (3)-(7) is convex and the optimal solution is given by multilevel waterfilling

$$p_k[n] = \left(\mu_k\theta - 1/c_k[n]\right)^+, \ \forall n, k \tag{8}$$

$$\theta = \frac{P_T + \sum_{k=1}^{K}\sum_{n\in\{\mathcal{S}_k : p_k[n]>0\}} 1/c_k[n]}{\sum_{k=1}^{K}\mu_k|\mathcal{S}_k|}, \tag{9}$$

but we need K^N searches...

Duality gap in a nutshell

- Each constrained optimization problem, such as (3)-(7), has an associated *dual* problem [2].
- If $p^\star$ and $d^\star$ denote the primal and dual optimal values respectively, it follows from duality theory that $d^\star \geq p^\star$, and the duality gap Γ is non-negative,

$$\Gamma = d^\star - p^\star. \tag{10}$$

- For *convex optimization problems* (concave objective and convex constraints) $\Gamma = 0$.
- It is shown in [3] that the non-convex multi-carrier optimization problem (3)-(7) is a problem for which $\Gamma \to 0$ as $N \to +\infty$.
- This implies that dual optimization methods are asymptotically optimal. “Asymptotically” meaning in practice $N \geq 8$ as shown in [4].

Dual optimization

- The Lagrangian of (3)-(7) is

$$\mathcal{L}(\{p_k[n]\}, \{R_k[n]\}, \lambda) = \sum_{k=1}^{K} \sum_{n=1}^{N} R_k[n] - \lambda \left(\sum_{k=1}^{K} \sum_{n=1}^{N} p_k[n] - P_T \right), \quad (11)$$

defined on $\mathcal{D}$, the set of all non-negatives $\{p_k[n]\}$ such that for each n only one $p_k[n]$ is positive for $k = 1 \ldots, K$.

- It has an associated Lagrange dual function of the form

$$g(\lambda) = \max_{\{p_k[n]\} \in \mathcal{D}} \mathcal{L}(\{p_k[n]\}, \{R_k[n]\}, \lambda). \quad (12)$$

Finally, the dual problem of (3)-(7) is

$$d^\star = \min_{\lambda \geq 0} g(\lambda). \quad (13)$$

Dual optimization

- The minimization in (13) can be performed by decomposing

$$g(\lambda) = \sum_{n=1}^{N} g_n(\lambda) + \lambda P_T \equiv \sum_{n=1}^{N} \max_{\{p_k[n]\}\in\mathcal{D}} \left\{ \sum_{k=1}^{K} \mu_k R_k[n] - \lambda p_k[n] \right\} + \lambda P_T, \quad (14)$$

maximizing each $g_n(\lambda)$ for fixed λ, and updating λ using a bisection algorithm.

- If only user k is active in subcarrier n, $g_n(\lambda)$ is maximized by

$$p_k[n] = \Big(\mu_k/(\lambda \ln 2) - 1/c_k[n]\Big)^+ . \quad (15)$$

By searching over all K possible user assignments for subcarrier n, $g_n(\lambda)$ is

$$g_n(\lambda) = \max_k \Big\{ \mu_k \log_2 \Big(1 + c_k[n] \Big(\mu_k/(\lambda \ln 2) - 1/c_k[n]\Big)^+\Big) \quad (16)$$

$$-\lambda \Big(\mu_k/(\lambda \ln 2) - 1/c_k[n]\Big)^+ \Big\} \quad (17)$$

- This yields $\mathcal{O}(NK)$ executions to find the optimal solution, which is such that $p^\star = d^\star$ for $N > 8$ in practice [4].

Allocation quantization (feasible signalling)

- Suppose that optimization has ended and the optimal allocation is $\{p_k^\star[n]\}, \{R_k^\star[n]\}$.
- Since we related $p_k[n]$ and $R_k[n]$ using Shannon capacity, $\{R_k[n]\}$ are continuously-valued real variables.
- In contrast, practical systems have a set of available modulation formats (e.g., M-QAM squared modulations)

$$\mathcal{R} = \{b_0, b_1, b_2, \ldots, b_{|\mathcal{R}|}\}, \tag{18}$$

  where $b_0 = 0$ always (it means no transmission) and b_i are in [bits/symbol].
- Thus, we need to *quantize* (project) the optimal allocation into the set of feasible modulation formats while still satisfying the total power constraint.

Allocation quantization

- Following we describe one possible allocation quantization algorithm (similar to [5]).
- Let us denote by $R_k^q[n]$ the projection of $R_k^\star[n]$ onto $\mathcal{R}$ by the nearest neighbor criterion (a regular slicer) and by $p_k^q[n]$ the updated power allocation.
- $P = \sum_{k=1}^{K} \sum_{n=1}^{N} p_k^q[n]$ is the total transmit power of the quantized rate allocation.
- For a given value of rate $b_i \in \mathcal{R}$ of the k-th user at the n-th subcarrier, define
  - $\Delta_k^+[n]$: the extra power required to increase it to b_{i+1} bits per symbol, (set to $+\infty$ in case $i = |\mathcal{R}|$);
  - $\Delta_k^-[n]$: the extra power obtained when it is decreased to b_{i-1} (set to 0 in case $i = 0$).

Algorithm description

Algorithm *Allocation quantization*

1. **if** $P < P_T$
2. **repeat** $(k', n') = \arg \min_{k,n \,:\, P+\Delta_k^+[n] \leq P_T} \Delta_k^+[n]/\mu_k$
3. Increase $R_{k'}[n']$ one modulation format, update $P = P + \Delta_{k'}^+[n']$, and update $(p_{k'}^q[n'], \Delta_{k'}^+[n'], \Delta_{k'}^-[n'])$.
4. **until** $P + \Delta_k^+[n] > P_T \;\forall n, k$.
5. **else**
6. **repeat** $(k', n') = \arg \max_{k,n \,:\, \Delta_k^-[n]>0} \Delta_k^-[n]/\mu_k$
7. Decrease $R_{k'}[n']$ one modulation format, update $P = P - \Delta_{k'}^-[n']$, and update $(p_{k'}^q[n'], \Delta_{k'}^+[n'], \Delta_{k'}^-[n'])$.
8. **until** $P \leq P_T$

- If there is spare power after quantization, the rates of users with large priorities or small power demands to increase their rate are increased.
- If we need to cut down power instead, we decrease the rates of those users with lowest priorities or very large power returns if their rate is reduced.

The MISO case

- In the MISO case ($n_T > 1$, $n_R = 1$) we have more degrees of freedom at the BS, which bring about a formidable increase in complexity.
  - How many users per subcarrier should we allocate? We can allocate up to $L = \min\{K, n_T\}$ using spatial multiplexing, which results in

$$1 + K + \binom{K}{2} + \ldots + \binom{K}{L} \approx K^L \tag{19}$$

    possibilities per subcarrier.
  - Should we use CSI at BS to *orthogonalize* users? If so, we can still use the model

$$R_k[n] = \log_2(1 + c_k[n]p_k[n]), \tag{20}$$

    where now the channel signal-to-noise ratio depends on $\mathbf{h}_k$ but also on the channels of the rest of users transmitting on the n-th subcarrier.
  - If users are not orthogonalized, then

$$R_k[n] = \log_2(1 + \mathsf{sinr}_k[n]), \tag{21}$$

    where the signal-to-interference-plus-noise ratio $\mathsf{sinr}_k[n]$ couples the powers and usually gives rise to extra non-convexities.

A note on complexity

- Even when the dual methods presented for the SISO case are extended to the MISO case, their complexity order $\mathcal{O}(NK^L)$ can be impractical for large L.
- This fosters the interest on suboptimal allocation schemes based on:
  - Heuristic subcarrier assignment.
  - Resource allocation for fixed subcarrier assignment and orthogonal users.
  - Allocation quantization.

Heuristic subcarrier assignment

- An heuristic subcarrier assignment uses the CSI information $\{\mathbf{h}_k[n]\}$ collected at the BS and the user priorities $\{\mu_k\}$ to attach users to subcarriers.
- Let us define by $\mathcal{S}[n]$, $|\mathcal{S}[n]| \leq L$, the set of users to transmit on subcarrier n.
- E.g., a simple but yet efficient heuristic subcarrier assignment is [5]:

$$\mathcal{S}[n] = \{k_1, k_2, \dots, k_L\}, \tag{22}$$

where

$$k_i = \arg \max_{k \in \{1,\dots,K\} \setminus \bigcup_{j=0}^{i-1} k_j} \mu_k \|\mathbf{h}_k[n]\|^2. \tag{23}$$

- Another popular subcarrier assignment method (for sum-rate maximization) is the Semi Orthogonal user Selection (SUS) algorithm [6], which looks for a set of users $\mathcal{S}[n]$ which are *reasonably* separated in space.

Heuristic subcarrier assignment

Algorithm *SUS Algorithm*

1. **for** $n = 1 \ldots N$
2. Initialize $(\mathcal{T}_1, i, \mathcal{S}[n]) = (\{1, 2, \ldots, K\}, 1, \emptyset)$.
3. For each $k \in \mathcal{T}_i$ calculate $\mathbf{g}_k[n]$, the component of the channel $\mathbf{h}_k[n]$ orthogonal to the subspace spanned by $\{\mathbf{g}_{(1)}[n], \ldots, \mathbf{g}_{(i-1)}[n]\}$.

$$\mathbf{g}_k[n] = \mathbf{h}_k[n] - \sum_{j=1}^{i-1} \frac{\mathbf{g}_{(j)}^T[n]\mathbf{h}_k[n]}{\|\mathbf{g}_{(j)}[n]\|^2} \mathbf{g}_{(j)}[n]. \tag{24}$$

(when $i = 1$, $\mathbf{g}_{(k)}[n] = \mathbf{h}_k[n]$)

4. Select the i-th user as

$$k_i = \arg\max_{k \in \mathcal{T}_i} \|\mathbf{g}_k[n]\|^2 / \sigma_k^2[n]. \tag{25}$$

Update $(\mathcal{S}[n], \mathbf{g}_{(i)}[n]) = (\mathcal{S}[n] \cup k_i, \mathbf{g}_{k_i}[n])$.

5. **if** $|\mathcal{S}[n]| < L$
6. Drop off prospective users if they are not semi-orthogonal to $\mathbf{g}_{(i)}[n]$.

$$\mathcal{T}_{i+1} = \left\{ k \in \mathcal{T}_i, k \neq k_i \mid \frac{|\mathbf{g}_{(i)}^T[n]\mathbf{h}_k[n]|}{\|\mathbf{g}_{(i)}[n]\| \|\mathbf{h}_k[n]\|} < \epsilon \right\}, \tag{26}$$

and update $i = i + 1$.

7. **if** $|\mathcal{T}_{i+1}| = 0$ Finish
8. **else** Finish.

Heuristic subcarrier assignment

- There exists a tradeoff when choosing ϵ:
  - If it is too low, the multiplexing gain and hence the throughput will be small.
  - If it is too high, poorly separable user will be allocated together, and the throughput will be also small.
- There is no analysis done on the optimal scaling of ϵ with K and L.
- The SUS algorithm can be straightforwardly extended to the weighted sum rate general setting.

Fixed bandwidth assignment resource allocation

- Once the sets $\{\mathcal{S}[n]\}$ have been created, the channel signal-to-noise ratios $\{c_k[n]\}$ have to be computed according to the transmission technique used.
  - For Zero Forcing BeamForming (ZFBF), and assuming w.l.o.g. that $\mathcal{S}[n] = \{1, 2, \ldots, |\mathcal{S}[n]|\}$,

$$c_k[n] = |\mathbf{h}_k[n]^T \mathbf{w}_k[n]|^2 / \sigma_k^2[n], \tag{27}$$

    where $\mathbf{w}_k[n] \in \mathbb{C}^{n_T \times 1}$ is the k-th normalized column of $\mathbf{H}(\mathcal{S}[n])^\dagger (\mathbf{H}(\mathcal{S}[n])\mathbf{H}^\dagger(\mathcal{S}[n]))^{-1}$, and

$$\mathbf{H}(\mathcal{S}[n]) = [\mathbf{h}_1[n] \ \mathbf{h}_2[n] \ldots \mathbf{h}_{|\mathcal{S}[n]|}]^T. \tag{28}$$

  - For Zero Forcing Tomlinson-Harashima Precoding (ZF-THP) [7],

$$c_k[n] = |[\mathbf{G}]_{k,k}[n]|^2 / \sigma_k^2[n], \tag{29}$$

    where $\mathbf{H}(\mathcal{S}[n]) = \mathbf{G}[n]\mathbf{Q}[n]$ with $\mathbf{G}[n]$ upper-triangular and $\mathbf{Q}[n]$ unitary is the QR decomposition.
- Finally, the optimal power allocation has a multilevel waterfilling structure similar to (8)-(9).

Allocation quantization

- For quantization of the optimal allocation, one can use for instance the same algorithm described for the SISO case.

References

- L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms", *IEEE Trans. Comm.*, vol. 52, no. 6, pp. 922-930, June 2004.
- S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- W. Yu and R. Lui, "Dual methods for non-convex spectrum optimization of multi-carrier systems", *IEEE Trans. Comm.*, vol. 54, no. 7, pp. 1310- 1322, July 2006.
- K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems", in *Proc. of IEEE Intl. Symp. Inform. Theory*, Seattle, USA, 2006, pp. 1394-1398.
- E. Calvo and J. R. Fonollosa, "Near-optimal joint power and rate allocation for OFDMA broadcast channels", in *Proc. if IEEE ICASSP*, Honolulu, USA, 2007.
- T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming", *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 528-541, March 2006.
- C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding a multi-antenna and multi-user communications", *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305-1316, July 2004.