Edinburgh Research Explorer

Minimizing Base Station Power Consumption

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IEEE Journal on Selected Areas in Communications

Minimizing Base Station Power Consumption

Hauke Holtkamp, *Member, IEEE,* Gunther Auer, *Member, IEEE,* Samer Bazzi, *Member, IEEE,* and
Harald Haas, *Member, IEEE*

Abstract—We propose a new radio resource management algorithm which aims at minimizing the base station supply power consumption for multi-user MIMO-OFDM. Given a base station power model that establishes a relation between the RF transmit power and the supply power consumption, the algorithm optimizes the trade-off between three basic power-saving mechanisms: antenna adaptation, power control and discontinuous transmission. The algorithm comprises two steps: a) the first step estimates sleep mode duration, resource shares and antenna configuration based on average channel conditions and b) the second step exploits instantaneous channel knowledge at the transmitter for frequency selective time-variant channels. The proposed algorithm finds the number of transmit antennas, the RF transmission power per resource unit and spatial channel, the number of discontinuous transmission time slots, and the multi-user resource allocation, such that supply power consumption is minimized. Simulation results indicate that the proposed algorithm is capable of reducing the supply power consumption by between 25% and 40%, dependend on the system load.

Index Terms—Energy efficiency, green radio, resource allocation, power control, sleep mode, antenna adaptation, discontinuous transmission (DTX), inverse water-filling

I. Introduction

THE POWER consumption of cellular radio networks is constantly increasing due to the growing number of mobile terminals and higher traffic demands, which require a densification of the network topology. Environmental concerns, rising energy costs and a growing need for self-sufficient (power-grid independent) Base Stations (BSs) reinforce efforts to reduce the cellular networks' power consumption [1]. Shortcomings of the state-of-the-art lie in the fact that current BSs are designed to serve peak demands without considering energy efficient off-peak operation. Depending on the time of day or geographic location, a BS may be idle—and thus over-provisioning—for a significant portion of its operation time. In such low-load situations, spectral efficiency can be traded for energy efficiency without reducing the user experience. The importance here is to operate BSs such that they can flexibly adjust to traffic demand. Furthermore, any power-saving operation on the BS side must not negatively affect

the mobile terminal, *e.g.* by increasing its power consumption or reducing the perceived quality of service.

With regard to different power-saving Radio Resource Management (RRM) mechanisms, Power Control (PC) is the most prominent in literature. This is due to the fact that, in addition to reducing power consumption, PC is also beneficial to link adaptation and interference reduction [2], [3]. In [2] Wong *et al.* minimize transmit power for multi-user Orthogonal Frequency Division Multiplexing (OFDM) under rate constraints. Their PC algorithm is computationally complex and does not consider a transmit power constraint or power model. Cui *et al.* analyze the energy efficiency of Multiple-Input Multiple-Output (MIMO) transmissions, being the first to consider the supply power consumption in energy efficiency studies [3]. They show that in terms of energy efficiency, the optimal number of transmit antennas used for MIMO transmission depends on the Signal-to-Noise-Ratio (SNR). Miao *et al.* [4] derive the data rate that maximizes the transmitted information per unit energy (bit per joule). More recently, acknowledging the significance of circuit power consumption, Discontinuous Transmission (DTX) and Antenna Adaptation (AA) have been identified as energy saving RRM techniques. Kim *et al.* [5] establish AA as a MIMO resource allocation problem and adapt the number of transmit antennas on a single link. However, solutions provided for AA on the mobile side cannot be directly applied to the BS, since for the latter multiple links have to be considered. Some works have assessed the energy saving potentials of sleep modes and AA in Long Term Evolution (LTE) BSs, but have applied very crude decision mechanisms [6]–[8], *e.g.* a BS is only allowed to sleep when a cell is empty. To the best of our knowledge, there are no prior works which consider DTX in combination with other power saving RRM mechanism.

The closest related works to Inverse Water-filling, which is introduced in Section IV, have studied the problem of allocating transmit power, user rates and resources per user with suboptimal solutions [3], [9], [10]. Bit capacities, modulation as well as transmission power consumption are investigated in general MIMO point-to-point transmissions [3], but without considering power-allocation to orthogonal channels. A later work assigns power sub-optimally in favor of a fairness constraint [9]. Al-Shatri *et al.* [10] have used Lagrange multipliers to invert water-filling in a rate-maximizing fairness application. We extend their work by relaxing the equal rate assumption among users and present a complete power-minimizing algorithm.

In this paper, we address the issue of overly generic assumptions about underlying hardware by identifying the relevant energy saving schemes for cellular networks on the basis of a

realistic and detailed power model [11]. As a consequence, and unlike previous energy saving studies [2], this work minimizes the *supply power* (or 'AC-plug' power) of the cellular BS rather than the Radio Frequency (RF) transmission power. To provide an applicable mechanism for current cellular systems like 3rd Generation Partnership Project (3GPP) LTE, we present the Resource allocation using Antenna adaptation, Power control and Sleep modes (RAPS) algorithm which reduces the BS supply power consumption of multi-user MIMO-OFDM. Given the channel states and target rates per user, RAPS finds the number of transmit antennas, the number of DTX time slots and the resource and power allocation per user. DTX refers to a micro sleep on the transmitter side, which is sufficiently short so that the regular operation of the mobile is not affected. The RAPS solution is found in two steps: First, PC, DTX and resource allocation are joined into a convex optimization problem which can be efficiently solved. Second, subcarrier and power allocation for a frequency-selective time-variant channel pose a combinatorial problem, which is solved by means of a heuristic solution. To allocate minimal power levels to resources and spatial channels, we derive the Inverse Water-filling algorithm.

The paper is organized as follows: The system and power model is presented in Section II. The global energy efficiency problem is defined in Section III. Section IV presents Step 1 of the two-step RAPS algorithm which determines DTX duration, antenna number, and resource shares on the basis of block fading channels. Step 2 extends these estimates by subcarrier and power allocation in frequency-selective time-varying channels in Section V. Simulation results are presented in Section VI. The paper is concluded in Section VII.

II. SYSTEM AND POWER MODEL

A. System Model

We consider one transmission frame in the downlink of a point-to-multipoint wireless communication system, comprising one serving BS and multiple mobile receivers. The BS transmitter is equipped with M_T antennas and all antennas share the transmit power budget. Mobile receivers have M_R antennas and system resources are shared via Orthogonal Frequency Division Multiple Access (OFDMA) between K users on N subcarriers and T time slots. In total there are TN resources units. OFDMA is used, for example, in 3GPP LTE. A frequency-selective time-variant channel is assumed, with each resource unit characterized by the channel state matrix $\boldsymbol{H}_{n,t,k} \in \mathbb{C}^{M_\mathrm{R} \times M_\mathrm{T}}$, with subcarrier index $n = 1, \ldots, N$, time slot index $t = 1, \ldots, T$, user index $k = 1, \ldots, K$. The vector of spatial channel eigenvalues per resource unit a and user k is $\mathcal{E}_{a,k}$ and its cardinality is $\min\{M_\mathrm{T}, M_\mathrm{R}\}$. The system operates orthogonally such that individual resource units cannot be shared among users. MIMO transmission with a variable number of spatial streams is assumed over the set of resources assigned to each user $\mathcal{A}_k$. The frame structure is illustrated in Fig. 1.

Although cellular wireless networks are generally interference limited, co-channel interference between neighbouring cells is not considered in this work. The main contribution of this work lies in the detailed derivation of the BS supply
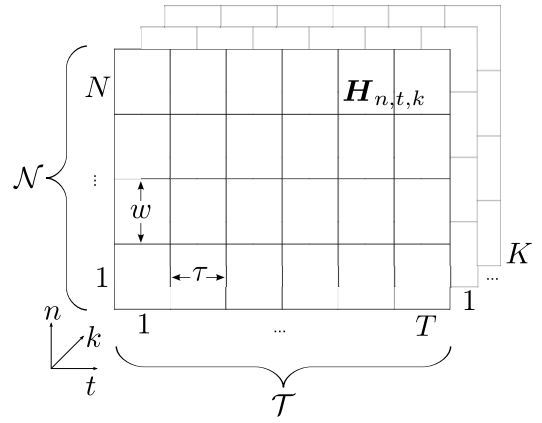

Fig. 1. OFDM frame structure.

power minimization via a power model. The consideration of interference and resulting network dynamics are, therefore, beyond scope of this paper and are left for future work.

B. Power Model

Supply power consumption, *i.e.* the overall device power consumption of the BS, is calculated via a power model. Power consumption models for cellular BSs are established in [11], [12]. The model allows to map RF transmission power to supply power consumption. The power model is constructed from a real BS hardware implementation, and is able to capture the most relevant power consumption effects in commercially available BSs. It provides a simple linear model which is verified by an advanced realistic and detailed hardware model [13]. Therefore, the linear power model provides a sufficient foundation for the BS energy efficiency analysis conducted in this paper. In the power model [11], [12], a RF chain refers to a set of hardware components composed of a small-signal transceiver section, a Power Amplifier (PA) and the antenna interface. The PA consumes a large share of the overall power due to its low efficiency [11]. Each transmit antenna is connected to an RF chain, thus the term AA implies the adaptation of the respective RF chain. Unlike, *e.g.* a shared baseband component, it is assumed that RF chains can be switched off or put to sleep individually when there is no transmission on the respective antenna. The BS consumes less power when fewer RF chains are active. Setting the entire BS to sleep interrupts transmission but further reduces the power consumption. For simplicity, we assume instantaneous and cost-free on-off-switching of hardware, *i.e.* there is no delay or additional power cost incurred by putting a BS into DTX mode or enabling/disabling an RF chain. This assumption is supported by [7], in which a switching time of $30\,\mu\mathrm{s}$ is assumed, which is much smaller than the duration of an OFDMA frame. We, therefore, presume that the additional power penalty due to the settling time of the RF components can be neglected.

Power model parameters are provided in [11] only for a BS with two RF chains. We derive the supply power consumption value when operating with a single RF chain as follows: In [11] power consumption characteristics for a macro BS with three sectors are provided. The power consumption is rated at a
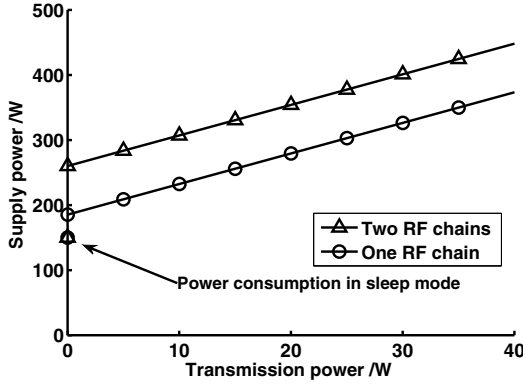
Fig. 2. Supply power model [11].

minimum active consumption P_{0,M_T}, which depends on M_T, a linear transmission power dependence factor Δ_{PM} (the slope) and a power consumption in DTX mode P_S. Transmission power consumption in any time slot t is P_t. Under the assumption of unchanged efficiencies of all components, we use the tables provided in [11] to find the maximum power consumption of a BS operated with a single RF chain and assume Δ_{PM} to be unchanged. Since we presume instantaneous switching capabilities for all components, the sleep mode consumption is the same in both RF chain settings. Lastly, we normalize the power consumption values to a single sector, leading to the power model depicted in Fig. 2. Formally, the supply power consumption is

$$P_{supply}(P_t) = \begin{cases} P_{0,M_T} + \Delta_{PM}P_t, & \text{if } P_t > 0, \\ P_S, & \text{if } P_t = 0. \end{cases} \quad (1)$$

III. GLOBAL PROBLEM STATEMENT

In this section, we deduce power-saving mechanisms from the power model and derive the global OFDMA supply power minimization problem.

A. Energy Saving Strategies

The power model in Fig. 2 reflects three general mechanisms that affect the power consumption of BSs: First, the overall power emitted at the PA P_t affinely relates to supply power consumption (1). Second, if the same BS is operated with fewer RF chains, it consumes less power. Third, if the BS is put into sleep mode, it consumes less power than in the active state with zero transmission power.

These observations lead to the following saving strategies:

PC : Reduce transmission power on each resource unit.
AA : Reduce the number of RF chains.
DTX : Increase the time the BS spends in DTX.

The first and the last strategy are clearly opposing each other as lower transmission powers lead to lower link rates and thus longer transmission duration (for an equal bit-load), whereas it would be beneficial for long DTX to have short transmissions. The second and third strategy are related, as AA can be considered a weak form of DTX which still allows transmission on a subset of antenna elements.
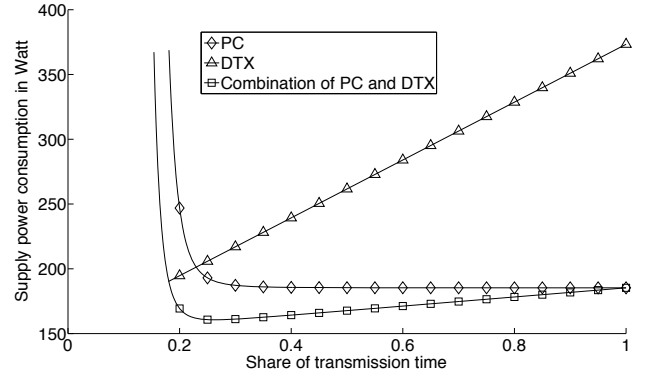

Fig. 3. Supply power consumption for transmission of a target spectral efficiency as a function of time spent transmitting, Φ. The combination of DTX and power control achieves lower power consumption than exclusive operation of either DTX or power control.

B. Joint Strategies

For a single link, the energy efficiency trade-off between PC and DTX is illustrated in Fig. 3. Shown is the supply power consumption caused by transmitting a fixed amount of data. Three operation modes are depicted in Fig. 3:

- PC: Only PC but no DTX is available. The transmission power is adjusted depending on the transmission time normalized to the time slot duration, Φ, such that the target bit load is transmitted. The BS consumes idle power P_0 when there is no data transmission. Clearly, the strategy for lowest power consumption in this case is to set $\Phi=1$. Reducing Φ increases the required transmission power, until at $\Phi = 0.18$, $P_{Tx} = P_{max}$.
- DTX: The BS always transmits with full power $P_{Tx} = P_{max}$. The supply power is $P_0 + mP_{max}$ when transmitting or P_S when in DTX mode, yielding an affine function of Φ. At $\Phi = 1$, more data than the target bit load is transmitted. Reducing Φ from $\Phi = 1$ decreases the supply power consumption linearly up to the point where the target bit load is met. Here, the best strategy clearly is to minimize the time transmitting (small Φ), so to maximize the time in DTX $(1-\Phi)$.
- Joint application of PC and DTX: In this mode of operation, the DTX time $(1-\Phi)$ is gradually reduced. The transmission power is adjusted to meet the target bit load within Φ.

Fig. 3 shows that the joint operation of PC and DTX consumes less power than each individual mode of operation with an optimal point at $\Phi=0.25$.

C. Global Problem

We extend the power control and DTX trade-off to MIMO-OFDM serving multiple users over frequency-selective channels. The selection of the number of transmit antennas for AA is made once for the entire frame.

We formulate the global problem statement: Given the channel state matrices $\boldsymbol{H}_{n,t,k}$ on each channel and the vector of target rates per user $\boldsymbol{r} = (R_1, \ldots, R_K)$, we seek:

- the set of resources allocated to each user $\mathcal{A}_k$,

- the power level $P_{a,e}$ per resource unit $a = 1, \ldots, |\mathcal{A}_k|$ and spatial channel $e = 1, \ldots, |\mathcal{E}_{a,k}|$,
- and the number of active transmit antennas M_{T},

such that the supply power consumption is minimized while fulfilling the transmission power constraint $P_{\max}$.

The sum capacity of user k over one transmission frame of duration τ_{frame} is given by:

$$R_k = \frac{w\tau}{\tau_{\mathrm{frame}}} \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e}\mathcal{E}_{a,k}(e)}{N_0 w} \right), \quad (2)$$

with subcarrier bandwidth w in Hz, time slot duration τ in seconds, and noise spectral density N_0 in W/Hz.

The RF transmission power in time slot t is

$$P_t = \sum_{a=1}^{|\mathcal{A}_t|} \sum_{e=1}^{|\mathcal{E}_a|} P_{a,e}, \quad (3)$$

where $\mathcal{A}_t$ is the set of N resources in time slot t and $\mathcal{E}_a$ is the vector of channel eigenvalues on resource a.

Given (1), (2) and (3), the optimization problem that minimizes the supply power consumption is of the form

$$\underset{\mathcal{A}_k \forall k, P_{a,e} \forall a,e, M_{\mathrm{T}}}{\text{minimize}} \quad P_{\mathrm{supply,frame}}(\boldsymbol{r}) =$$

$$\frac{1}{T} \left(\sum_{t=1}^{T_{\mathrm{Active}}} (P_{0,M_{\mathrm{T}}} + \Delta_{\mathrm{PM}} P_t) + \sum_{t=1}^{T_{\mathrm{Sleep}}} P_{\mathrm{S}} \right) \quad (4)$$

subject to $\quad P_t \leq P_{\max}, \quad T_{\mathrm{Active}} + T_{\mathrm{Sleep}} = T,$

$$R_k \leq \frac{w\tau}{\tau_{\mathrm{frame}}} \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e}\mathcal{E}_{a,k}(e)}{N_0 w} \right)$$

with the number of active transmission time slots T_{Active}, and DTX slots T_{Sleep}. This is a set selection problem over the sets $\mathcal{A}_k$, and $\mathcal{E}_{a,k}$, as well as a minimization problem in $P_{a,e}$.

D. Complexity

Dynamic subcarrier allocation is known to be a complex problem for a single time slot in frequency-selective fading channels that can only be solved by suboptimal or computationally expensive algorithms [2], [14], [15]. In this study, we add two additional degrees of freedom by considering AA and DTX, increasing the complexity further. We consequently divide the problem into two steps: First, real-valued estimates of the resource share per user, DTX duration, and number of active RF chains M_{T}, are derived based on simplified system assumptions. Second, the power-minimizing resource allocation over the integer set $\mathcal{A}_k$ and consecutive Inverse Water-filling power allocation are performed.

IV. STEP 1: ANTENNA ADAPTATION, DTX AND RESOURCE ALLOCATION

The first step to solving the global problem is based on a simplification of the system assumptions. This allows defining a convex subproblem with an optimal solution, which is later (sub-optimally) mapped to the solution of the global problem in Step 2, described in Section V.

Instead of time and frequency-selective fading we assume that channel gains on all resources are equal to the center resource unit:

$$\boldsymbol{H}_k = \boldsymbol{H}_{n^c, t^c, k}, \quad (5)$$

where the superscript c signifies the center-most subcarrier and time slot. Step 1 thus assumes a block fading channel per user over $W = Nw$ and τ_{frame}.

We select the center resource unit due to the highest correlation with all other resources. Alternative methods to construct a representative channel state matrix are to take the mean or median of $\boldsymbol{H}_{n,t,k}$ over the OFDMA frame. However, application of the mean or median over a set of MIMO channels were found to result in a channel with lower capacity. See Section VI for a comparison plot between different channel selection methods.

The link capacity is calculated using equal-power precoding and assuming uncorrelated antennas. In contrast to water-filling precoding, equal-power precoding provides a direct relationship between total transmit power and target rate. On block fading channels with real-valued resource sharing, OFDMA and Time Division Multiple Access (TDMA) are equivalent. Without loss of generality, we select resource allocation via TDMA. These simplifications allow to establish a convex optimization problem that can be efficiently solved.

In a block fading multi-user downlink with target rate R_k and equal power precoding, the spectral efficiency for user k is given by

$$C_k = \frac{R_k}{W\mu_k} = \sum_{e=1}^{|\mathcal{E}_k|} \log_2 \left(1 + \frac{P_k}{M_T} \frac{\mathcal{E}_k(e)}{N_0 W} \right), \quad (6)$$

with the vector of channel eigenvalues per user $\mathcal{E}_k$, and transmission power P_k. The resource share $\mu_k \in (0, 1]$ is the normalized (unit-less) representation of the share of time.

The transmission power is a function of the target rate, depending on the number of transmit and receive antennas. For the following configurations, (6) reduces to:

1x2 SIMO:

$$P_k(R_k) = \frac{1}{\epsilon_1} \left(2^{\left(\frac{R_k}{W\mu_k} \right)} - 1 \right) \quad (7)$$

2x2 MIMO:

$$P_k(R_k) = \frac{-(\epsilon_1 + \epsilon_2) + \sqrt{(\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1 \epsilon_2 (2^{\frac{R_k}{W\mu_k}} - 1)}}{\epsilon_1 \epsilon_2} \quad (8)$$

where ϵ_i is the i-th member of $\mathcal{E}_k$. These equations can be extended in similar fashion to combinations with up to four transmit or receive antennas. Note that a higher number of antennas would require the general algebraic solution of polynomial equations with degree five or higher, which cannot be found in line with the Abel-Ruffini theorem.

Like on the BS side, the number of RF chains used for reception at the mobile could be adapted for power saving. However, the power-saving benefit of receive AA is much smaller than in transmit AA, where a PA is present in each RF chain. Moreover, multiple receive antennas boost the useful signal power and provide a diversity gain. Therefore, M_{Rx} is assumed to be fixed and set to $M_{\mathrm{Rx}} = 2$ in the following.

Given the transmission power (7), (8), and the power model (1), we derive the supply power consumption for TDMA. With DTX a BS may go to sleep when all K users have been served; the overall average consumption is the weighted sum of power consumed during transmission and DTX time share μ_S:

$$P_{\text{supply}}(\boldsymbol{r}) = \sum_{k=1}^{K} \mu_k \left(P_{0,M_T} + \Delta_{\text{PM}} P_k(R_k) \right) + \mu_S P_S \quad (9)$$

Consequently, we define the optimization problem with $\mu_{K+1} = \mu_S$:

$$\begin{aligned} \underset{(\mu_1,\dots,\mu_{K+1})}{\text{minimize}} \quad & P_{\text{supply}}(\boldsymbol{r}) = \\ & \sum_{k=1}^{K} \mu_k \left(P_{0,M_T} + \Delta_{\text{PM}} P_k(R_k) \right) + \mu_{K+1} P_S \\ \text{subject to} \quad & \sum_{k=1}^{K+1} \mu_k = 1 \\ & \mu_k \geq 0 \ \forall k \\ & 0 \leq P_k(R_k) \leq P_{\max} \end{aligned} \quad (10)$$

The first constraint ensures that all resources are accounted for and upper bounds μ_k. The second constraint encompasses the transmit power budget of the BS and acts as a lower bound on μ_k. Note that due to the block fading assumption and TDMA, the transmission power per user P_k is constant over time and equivalent to P_t in (1).

This problem is convex in its cost function and constraints (proof in Appendix A). It can therefore be solved with available tools like the interior point method [16]. As part of the RAPS algorithm, (10) is solved once for each possible number of transmit antennas. M_T is then selected according to which solution results in lowest supply power consumption.

The solution of the first step yields an estimate for the supply power consumption, the number of transmit antennas M_T, the DTX time share μ_S and the resource share per user μ_k. If a solution for Step 1 cannot be found, Step 2 is not performed and outage occurs. We leave outage handling to a higher system layer mechanism which could be avoided by, *e.g.*, prioritize users and reiterate with reduced system load.

V. STEP 2: SUBCARRIER AND POWER ALLOCATION

This section describes the second step of the RAPS algorithm. In Step 2, the results of Step 1 are mapped back to the global problem to find the subcarrier allocation for each user $\mathcal{A}_k$, the power level per resource and spatial channel $P_{a,e}$ and the number of DTX slots T_{Sleep}.

First, the real-valued resource share μ_k is mapped to the OFDMA resource count per user $m_k \in \mathbb{N}$,

$$m_k = \lceil \mu_k N T \rceil \quad \forall k = 1, \dots, K. \quad (11)$$

Possible rounding effects through the ceiling operation in (11) are compensated for by adjusting the number of DTX time slots

$$T_{\text{Sleep}} = \left\lfloor \frac{T N \mu_{K+1} - K}{N} \right\rfloor = \lfloor T \mu_{K+1} - K/N \rfloor. \quad (12)$$

The remaining time slots are available for transmission,

$$T_{\text{Active}} = T - T_{\text{Sleep}}. \quad (13)$$

The remaining unassigned resources

$$m_{\text{rem}} = N T - \sum_{k=1}^{K} m_k - N T_{\text{Sleep}} \quad (14)$$

are assigned to m_k in a round-robin fashion. After this allocation, it holds that $m_k = |\mathcal{A}_k|$.

Next, the number of assigned resource units per user m_k is equally subdivided into the number of resources per user and time slot $m_{k,t}$ with

$$m_{k,t} = \left\lfloor \frac{m_k}{\sum_{l=1}^{K} m_l} N \right\rfloor. \quad (15)$$

The remaining unassigned resources

$$m_{t,\text{rem}} = N - \sum_{k=1}^{K} m_{k,t} \quad (16)$$

are allocated to different $m_{k,t}$ in a round-robin fashion.

Time slots considered for DTX are assigned statically, starting from the back of the frame. Note that the dynamic selection of DTX slots would create additional opportunities for capacity gains or power savings as they could be assigned to time slots with poor channel states, *e.g.* time slots experiencing deep fades.

A corner case exists when $T N \mu_{K+1} < K$ and (12) becomes negative. This occurs when the DTX share $\mu_{K+1} = \mu_S$ is very small and thus traffic load is high. Due to the high traffic load, it is possible that the target rates cannot be fulfilled within the transmit power constraint, leading to outage for at least one user. Accordingly, if $T N \mu_{K+1} < K$, we set $T_{\text{Active}} = T$ and the resource mapping strategy in (11) is adapted such that

$$m_k = \lfloor \mu_k N T \rfloor \quad (17)$$

which guarantees that $m_k < N T$. The remaining resources are allocated to users as outlined above.

A subcarrier allocation algorithm from Kivanc *et al.* [14] is adopted, which has been shown to work effectively with low complexity. The general idea of the algorithm is as follows: First assign each subcarrier to the user with the best channel. Then start trading subcarriers from users with too many subcarriers to users with too few subcarriers based on a nearest-neighbor evaluation of the channel state. The algorithm is outlined in Algorithm 1 and applied to each time slot t consecutively.

At this stage, the MIMO configuration, the number of DTX time slots, and the subcarrier assignment are determined. Transmit powers are assigned in both spatial and time-frequency domains via an algorithm termed Inverse Water-filling (IWF). The notion of this algorithm is as follows: First, channels are sorted by quality and the water-level is initialized on the best channel, such that the bit load target is fulfilled. Then, in each iteration of the algorithm, the next best channel is added to the set of used channels, thus reducing the water-level in each step. The search is finished once the water-level is lower than the next channel metric to be added. Refer to Appendix B for the derivation.

Let us define for each user the target bit-load

$$B_{\text{target},k} = R_k \tau_{\text{frame}}. \quad (18)$$

Algorithm 1 Adapted Rate Craving Greedy (RCG) algorithm performed on each time slot t. As compared to [14] the cost parameter $h_{n,t,k}$ has been adapted and the absolute value has been added to the search of the nearest neighbor. $\mathcal{A}_{k,t}$ holds the set of subcarriers assigned to user k.

Ensure: $m_{k,t}$ is the target number of subcarriers allocated to each user k, $h_{n,t,k} = |\overline{\boldsymbol{H}_{n,t,k}}|^2$ and $\mathcal{A}_{k,t} \leftarrow \{\}$ for $k = 1, \ldots, K$.

1: **for all** subcarriers n **do**
2:      $k^* \leftarrow \underset{1 \leq k \leq K}{\arg\max} \, h_{n,t,k}$
3:      $\mathcal{A}_{k^*,t} \leftarrow \mathcal{A}_{k^*,t} \cup \{n\}$
4: **end for**
5: **for all** users k such that $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
6:      **while** $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
7:          $l^* \leftarrow \underset{\{l : |\mathcal{A}_{l,t}| < m_{l,t}\}}{\arg\min} \underset{1 \leq n \leq N}{\min} | - h_{n,t,k} + h_{n,t,l}|$
8:          $n^* \leftarrow \underset{1 \leq n \leq N}{\arg\min} | - h_{n,t,k} + h_{n,t,l^*}|$
9:          $\mathcal{A}_{k,t} \leftarrow \mathcal{A}_{k,t}/\{n^*\}, \mathcal{A}_{l^*,t} \leftarrow \mathcal{A}_{l^*,t} \cup \{n^*\}$
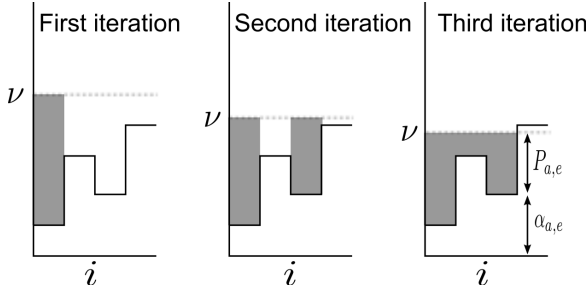10:   **end while**
11: **end for**

Fig. 4. Illustration of Inverse Water-filling (IWF) with three steps. The height of each patch i is given by $\alpha_{a,e} = N_0 w / \mathcal{E}_{a,k}(e)$. The water-level is denoted by ν. The height of the water above each patch is $P_{a,e}$. The first step sets the water-level such that the bit load target is fulfilled on the best patch. The second and third step add a patch, thus reducing the water-level. After the third step, the water-level is below the fourth patch level, thus terminating the algorithm.

To fulfill the target bit-load, the following constraint must be met

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e}\mathcal{E}_{a,k}(e)}{N_0 w} \right) = 0, \quad (19)$$

which assesses the sum capacity over a set of resources and spatial channels.

The water-level ν can be found via an iterative search over the set Ω_k channels that contribute a positive power

$$\log_2(\nu) = \frac{1}{|\Omega_k|} \left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|} \log_2 \left(\frac{\tau \, \mathcal{E}_{a,k}(e)}{N_0 \log(2)} \right) \right). \quad (20)$$

The water-level is largest on the first iteration and decreases on each iteration until it can no longer be decreased.

Using the Lagrangian method detailed in Appendix B, we arrive at a power-level per spatial channel of

$$P_{a,e} = \frac{\nu w\tau}{\log(2)} - \frac{N_0 w}{\mathcal{E}_{a,k}(e)}. \quad (21)$$
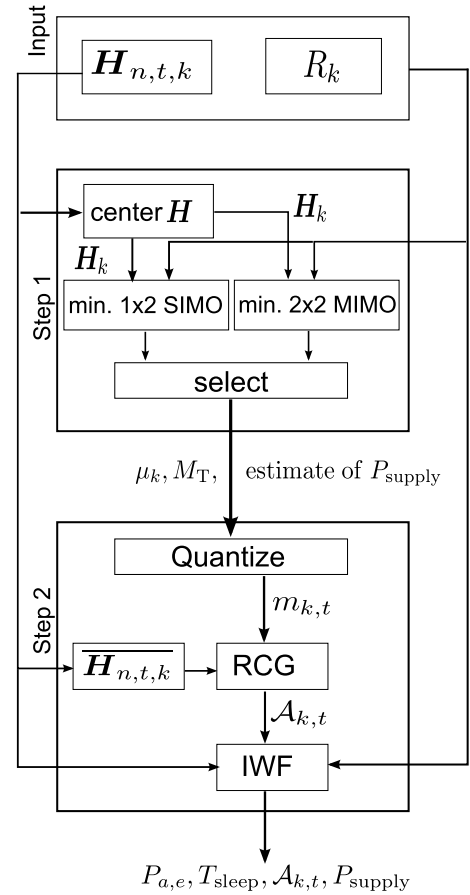

Fig. 5. Outline of the RAPS algorithm.

The inverse water-filling algorithm is illustrated for four channels in Fig. 4. The outcome of Inverse Water-filling as part of RAPS are the transmission power levels $P_{a,e}$ for each resource unit. The supply power consumption after application of RAPS can be found by summation of transmission powers in each time slot and application of the supply power model (1). The entire RAPS algorithm is outlined in Fig. 5.

VI. RESULTS

In order to assess the performance of the RAPS algorithm Monte Carlo simulations are conducted. The simulations are configured as follows: mobiles are uniformly distributed around the BS on a circle with radius $250\,\text{m}$ and a minimum distance of $40\,\text{m}$ to the BS to avoid peak SNRs. Fading is computed according to the NLOS model described in 3GPP TR25.814 [17] with $8\,\text{dB}$ shadowing standard deviation, and the frequency-selective channel model B5 described in [18] with $3\,\text{m/s}$ mobile velocity. All transmit and receive antennas are assumed to be mutually uncorrelated. Further system parameters are listed in Table I.

A. Benchmarks

The following transmission strategies are evaluated for comparison purposes:

• The maximum BS power consumption is obtained by constant transmission at maximum power, $P_{\text{supply,max}} = P_{0,M_{\text{T}}} + \Delta_{\text{PM}} P_{\text{max}}$.

TABLE I
SYSTEM PARAMETERS

Variable		Value
K	Number of users	10
N	Number of subcarriers $\mathcal{N}$	50
T	Number of time slots in set $\mathcal{T}$	10
M_T	Number of transmit antennas	[1,2]
M_R	Number of receive antennas	2
P_{0,M_T}	Circuit power consumption	185 W/260 W
Δ_{PM}	Load-dependence factor	4.7
P_S	Power consumption in DTX	150 W
P_{max}	Maximum transmission power	46 dBm
τ_{frame}/τ	Duration of frame/time slot	10 ms/1 ms
W/w	System/subcarrier bandwidth	10 MHz/200 kHz
N_0	Noise power spectral density	4×10^{-21} W/Hz

- Bandwidth Adaptation (BA), which finds the minimum number of subcarriers that achieves the rate target. No sleep modes are utilized and all scheduled subcarriers transmit with a transmission power spectral density of P_{max}/N. This benchmark is chosen to represent the power consumption of state-of-the-art BSs which are not capable of DTX or AA.
- DTX where the BS transmits with full power $P_{supply,max}$ and switches to micro sleep once the rate requirements are fulfilled. This benchmark assesses the attainable savings when only DTX is applied.

B. Performance Analysis

The channel value selection in (5) serves as the channel gain for the block fading assumption of Step 1. How the channel value selection of three possible alternatives (mean, median center) affects the supply power consumption is examined in Fig. 6. It can be seen that the selection of the mean channel gain results in the highest supply power consumption estimate after Step 1. While the estimate can be improved in Step 2, it is still inferior to the other alternatives. Use of the mean channel states causes Step 1 to underestimate the channel quality. Consequently, too few time slots are selected for DTX in Step 2. Choosing the median provides a better estimate than the mean and after Step 2 the solution matches 'Step 2, center selection'. However, in center channel state selection the Step 1 estimate and the Step 2 solution have both the best match and the lowest supply power consumption. Therefore, as mentioned in Section IV, center channel state selection is chosen in the RAPS algorithm and all further analyses.

Fig. 7 compares the outage probabilities of Step 1 with that of the BA benchmark. Outage refers to the lack of a solution to Step 1; when the user target data rates are too high for the given channel conditions, the convex subproblem has no solution, which causes the algorithm to fail. A reduction of user target data rates based on, *e.g.*, priorities, latency or fairness, is specifically not covered by RAPS. One suggested alternative is to introduce admission control, in the way that some users are denied access, so that the remaining users achieve their target rates. Note that Step 2 has no effect on the outage; if a solution exists after Step 1, then Step 2 can be completed. If a solution does not exist after Step 1, then Step 2 is not performed. Fig. 7 illustrates that RAPS selects two transmit antennas for target link rates above 14 Mbps with
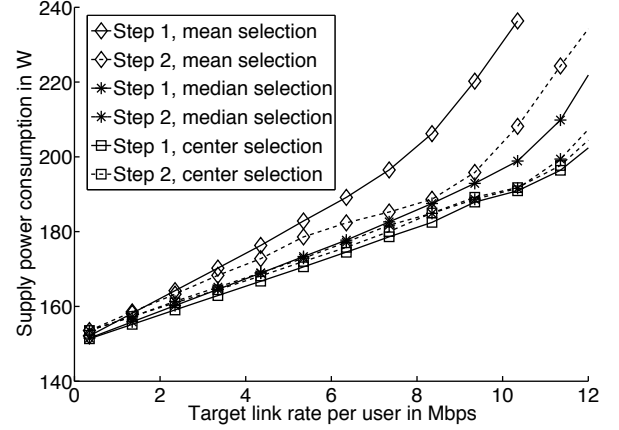

Fig. 6. Performance comparison of different channel state selection alternatives in Step 1.
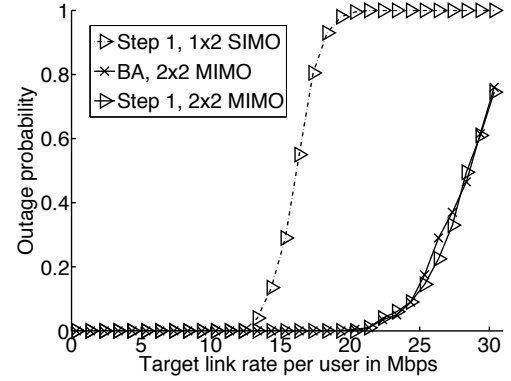

Fig. 7. Outage probability in Step 1 and the BA benchmark.

high probability, as target link rates cannot be achieved with a single transmit antenna. The BA benchmark and the Step 1 MIMO solution have similar outage behavior.

Fig. 8 depicts the average number of DTX time slots over increasing target link rates. The effect of AA (*i.e.*, switching between Single-Input Multiple-Output (SIMO) and MIMO transmission) can clearly be seen in Fig. 8. For low target link rates, a large proportion of all time slots is selected for DTX. For higher target link rates, the number of DTX time slots must be reduced when operating in SIMO mode. For target link rates above 12 Mbps, which approach the SIMO capacity (as described in the previous paragraph), the system switches to MIMO transmission. The added MIMO capacity allows the RAPS scheduler to allocate more time slots to DTX. In the medium load region (around 15 Mbps) the standard deviation is highest, indicating that here the RAPS algorithm strongly varies the number of DTX time slots depending on channel conditions and whether SIMO or MIMO transmission is selected. These variations contribute stronlgy to the additional savings provided by RAPS over the benchmarks. Another observation from Fig. 8 is that it is unlikely that more than two out of ten DTX time slots are scheduled at target link rates above 8 Mbps. This means that for a large range of target rates, no more than two DTX time slots are required to minimize power consumption. This is an important finding for applications of RAPS in established
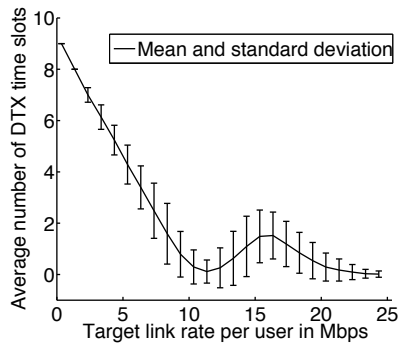
Fig. 8. Average number of DTX time slots over increasing target link rate Error bars portray the standard deviation. Total number of time slots $T = 1$
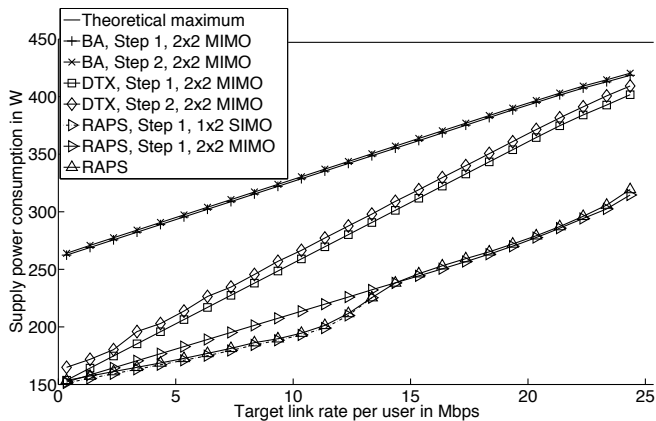

Fig. 9. Supply power consumption for different RRM schemes on block fading and frequency-selective fading channels (comparison of Step 1 and Step 2) for ten users. Overlaps indicate the match between the Step 1 estimate and the Step 2 solution.

systems like LTE, where the number of DTX time slots may be limited due to constraints imposed by the standard.

A comparison of the supply power consumption estimates of Step 1 and Step 2 in Fig. 9 verifies that the estimate taken in Step 1 as input for Step 2 are sufficiently accurate. Although Step 1 is greatly simplified with the assumption of block fading and its output parameters cannot be applied readily to an OFDMA system, it precisely estimates power consumption which is the optimization cost function. The slight difference between the Step 1 estimate and power consumption after Step 2 is caused by quantization loss and resource scheduling. Note that while Step 1 supplies a good estimate of the power consumption, it is not a solution to the original OFDMA scheduling problem, since it does not consider the frequency-selectivity of the channel and does not yield the resource and power allocation.

The performance of RAPS in comparison to the benchmarks is separately analyzed in Fig. 10. An initial observation is that even the supply power consumption of BA is significantly lower than the theoretical maximum for most target link rates. BA with a single transmit antenna always consumes less power than with two antennas, as long as the rate targets with one antenna can be met. AA is thus a valid power-saving mechanism for BA.
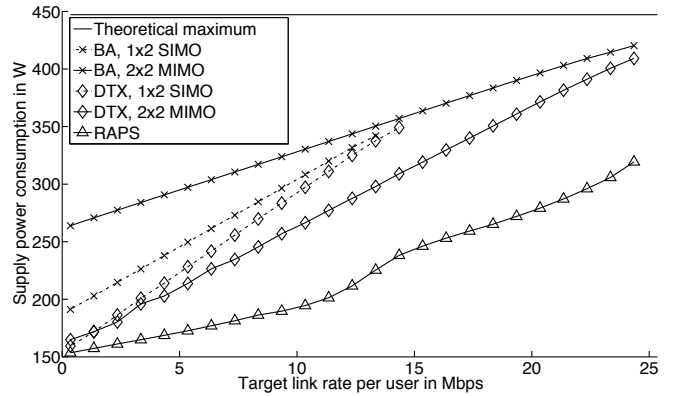

Fig. 10. P_{supply} on frequency-selective fading channels for different RRM schemes and RAPS for ten users. For a bandwidth-adapting BS power consumption is always reduced by AA, while AA is never beneficial for a sleep mode capable BS. For RAPS energy consumption is reduced by AA at low rates. In general, RAPS achieves substantial power savings at all BS loads.

The DTX power consumption is significantly lower than for BA, especially at low target rates, because lower target rates allow the BS to enter DTX for longer periods of time. As the BS load increases, the opportunities of the DTX benchmark to enter sleep mode are reduced, so that DTX power consumption approaches that of BA. Unlike in BA, it is never beneficial for a BS capable of DTX to switch operation to a single antenna, because transmitting for a longer time with a single antenna always consumes more power than a short two-antenna transmission, which allows for a longer DTX duration. (Note that this finding may not apply to other power models.)

RAPS reduces power consumption further than DTX by employing AA at low rates and PC at high rates. Through PC, the slope of the supply power is kept low between 15 and 20 Mbps. At higher rates an upward trend becomes apparent, since link rates only grow logarithmically with the transmission power. In theory, when the BS is at full load, the supply power of all energy saving mechanism will approach maximum power consumption. However, when the BS load is very high, not all users may achieve their target rate and outage occurs (see Fig. 7). In other words, operating the BS with load margins controls outage and allows for large transmission power savings. Power savings of RAPS compared to the state-of-the-art BA range from 102.7 W (24.5%) to 136.9 W (41.4%) depending on the target link rate per user.

In addition to absolute consumption we inspect the energy efficiency of RAPS in Fig. 11. We define energy efficiency $E = P_{\mathrm{supply}}^{-1} \sum_{k=1}^{K} R_k$ in bit/Joule. Observe that all data series are monotonically increasing. Therefore, RAPS does not change the paradigm that a BS is most efficiently operated at peak rates. However, when operating below peak rates, RAPS will always offer the most efficient operation at the requested rate.

VII. Summary and Conclusion

In this paper, we have examined three BS supply power saving mechanisms, namely antenna adaptation, power control and DTX, and evaluated their performance with a base
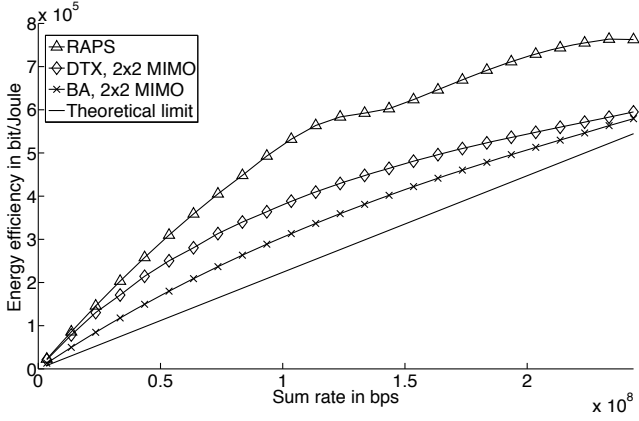
Fig. 11. Energy efficiency as a function of sum rate. Energy efficiency can be increased by the RAPS algorithm. The BS remains most efficient at peak rate.

station power consumption model. The scheduling problem of minimizing BS supply power consumption for the downlink of multiuser MIMO-OFDM was defined. To solve the scheduling problem the RAPS algorithm was proposed, which comprises two steps: first an estimate from a convex subproblem is found assuming constant channel gains on all resource units; then this estimate is applied in the second step to determine the resource and power allocation for each resource unit for a frequency-selective time-variant channel. As part of the second step, we have developed the Inverse Water-filling algorithm which finds the minimal sum power allocation for a target bit-load. Simulation results indicate that RAPS is capable of reducing supply power consumption over all target link rates by 25% to 40%. Savings at low target data rates are mostly attributed to DTX and AA, while at high rates PC is most effective.

While RAPS can be applied to many types of BSs, the reported simulation results clearly depend on the used power model. Future BSs may have different power consumption characteristics. For example, smaller cells can be expected to have lower transmission power compared to circuit consumption. This work is also meant to assist hardware developments in predicting which changes will be most useful for higher-level RRM mechanisms. Further investigations are planned into the effects of hardware switching times and hardware limitations on AA due to the limited operating regions of PAs. Extension of RAPS to the multi-cell interference setting is part of future work.

APPENDIX A
PROOF OF CONVEXITY FOR PROBLEM (10)

The partial second derivative of the MIMO cost function in (10) with respect to μ_k reads as

$$\frac{\partial^2 P_{\text{supply},k}(\boldsymbol{r})}{\partial^2 \mu_k} =$$

$$m \log^2(2) \frac{R_k}{W} \frac{2^{\frac{R_k}{W\mu_k}}\left((\epsilon_1 + \epsilon_2)^2 + 2\epsilon_1\epsilon_2\left(2^{\frac{R_k}{W\mu_k}} - 2\right)\right)}{\mu_k^3\left((\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1\epsilon_2\left(2^{\frac{R_k}{W\mu_k}} - 1\right)\right)^{\frac{3}{2}}} \tag{22}$$

which is non-negative since

$$2\epsilon_1\epsilon_2 2^{\frac{R_k}{W\mu_k}} \geq 2\epsilon_1\epsilon_2. \tag{23}$$

Thus the cost function is convex in μ_k. Convexity of the SIMO cost function and the constraint function can be shown similarly and is omitted here for brevity.

APPENDIX B
INVERSE WATER-FILLING ALGORITHM

Inverse Water-filling over the set of resources $\mathcal{A}_k$ and the vector of channel eigenvalues $\mathcal{E}_{a,k}$ minimizes power consumption over a set of channels while fulfilling a target bit load.

Assuming block-diagonalization precoding, the problem reads

$$\text{minimize} \quad \sum_{a=1}^{|\mathcal{A}_k|}\sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \tag{24}$$

subject to

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{A}_k|}\sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2\left(1 + \frac{P_{a,e}\mathcal{E}_{a,k}(e)}{N_0 w}\right) = 0,$$

$$P_{a,e} \geq 0 \quad \forall a, e,$$

$$\sum_{a=1}^{|\mathcal{A}_k|}\sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \leq P_{\text{max}}.$$

The derivative of the Lagrangian at the minimum is

$$\frac{\partial \mathcal{L}}{\partial P_{a,e}} = 1 - \lambda + \beta - \frac{\nu\, w\tau\, \mathcal{E}_{a,k}(e)}{\log(2)\,(wN_0 + \mathcal{E}_{a,k}(e)P_{a,e})} = 0, \tag{25}$$

with λ the multiplier for the first inequality constraint, β for the second (power-limit) constraint and ν the equality constraint multiplier. The variable λ can be omitted since it acts as a slack variable.

From (25) we arrive at a power-level per spatial channel of

$$P_{a,e} = \frac{\nu w\tau}{\log(2)} - \frac{N_0 w}{\mathcal{E}_{a,k}(e)}, \tag{26}$$

which can be inserted into the equality constraint of (24) to yield the water-level ν

$$\log_2(\nu) = \frac{1}{|\Omega_k|}\left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|}\log_2\left(\frac{\tau\,\mathcal{E}_{a,k}(e)}{N_0\log(2)}\right)\right). \tag{27}$$

The water-level can be found via an iterative search over the vector Ω_k of channels that contribute a positive power. Since the sum power is reduced on each iteration, the power constraint (accounted for by the multiplier β) only needs to be tested after the search is finished.

REFERENCES

[1] A. Fehske, J. Malmodin, G. Biczók, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Commun. Mag.*, 2010.

[2] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[3] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE J. Sel. Areas Commun.*, 2004.

[4] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy Efficient Design in Wireless OFDMA," in *IEEE International Conference on Communications*, 19-23 2008, pp. 3307 –3312.

[5] H. Kim, C.-B. Chae, G. de Veciana, and R. W. Heath, "A Cross-layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009.

[6] M. Hedayati, M. Amirijoo, P. Frenger, and J. Moe, "Reducing Energy Consumption Through Adaptation of Number of Active Radio Units," in *Proc. IEEE VTC 2012-Spring*, 2012.

[7] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proc. IEEE VTC 2011-Spring*, 2011.

[8] R. Wang, J. Thompson, H. Haas, and P. Grant, "Sleep Mode Design for Green Base Stations," *IET Communications*, vol. 5, no. 18, pp. 2606–2616, 2011.

[9] Z. Shen, J. Andrews, and B. Evans, "Optimal power allocation in multiuser OFDM systems," in *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, vol. 1. IEEE, 2003, pp. 337–341.

[10] H. Al-Shatri and T. Weber, "Fair Power Allocation for Sum-Rate Maximization in Multiuser OFDMA," in *Proc. International ITG Workshop on Smart Antennas 2010*, 2010.

[11] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. J. Gonzalez, and C. Desset, "Cellular Energy Efficiency Evaluation Framework," in *Proc. VTC 2011-Spring*, 2011.

[12] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, M. J. Gonzalez, C. Desset, O. Blume, and A. Fehske, "How Much Energy is Needed to Run a Wireless Network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40 –49, Oct 2011.

[13] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. Gonzalez, H. Klessig, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible power modeling of LTE base stations," in *2012 IEEE Wireless Communications and Networking Conference: Mobile and Wireless Networks (IEEE WCNC 2012 Track 3 Mobile & Wireless)*, 2012.

[14] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.

[15] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.

[17] 3GPP, "Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA)," 3GPP TR 25.814 V7.1.0 (2006-09), Sep. 2006. Retrieved Nov. 2, 2009 from www.3gpp.org/ftp/Specs/.

[18] IST-2003-507581 WINNER, "D5.4 v1.0 Final Report on Link Level and System Level Channel Models," Retrieved Apr. 15, 2007, from https://www.ist-winner.org/DeliverableDocuments/, Nov. 2005.

Gunther Auer received the Dipl.-Ing. degree in Electrical Engineering from the Universität Ulm, Germany, in 1996, and the Ph.D. degree from the University of Edinburgh, UK in 2000. From 2000 to 2001 he was a research and teaching assistant with Universität Karlsruhe (TH), Germany. From 2001 to 2012 he is with NTT DOCOMO Euro-Labs, Munich, Germany, where he was team leader and research manager in wireless technologies research. Since February 2012 he is with Ericsson AB, Stockholm, Sweden. His research interests include green radio, heterogeneous networks and multi-carrier based communication systems, in particular medium access, cross-layer design, channel estimation and synchronization techniques.

Samer Bazi (M10) received a B.E. degree in Computer and Communications Engineering from the American University of Beirut (AUB) in 2008, and a M.S. in Communications Engineering from the Technical University of Munich (TUM) in 2010. He is currently a member of DOCOMO Euro-Labs, located in Munich, Germany and an external Ph.D. candidate at the Associate Institute for Signal Processing at TUM. His research interests include signal processing techniques and interference mitigation for multi-cell multi-user MIMO systems in addition to convex optimization for wireless systems.

Harald Haas (SM'98–AM'00–M'03) holds the Chair of Mobile Communications in the Institute for Digital Communications (IDCOM) at the University of Edinburgh, and he currently is the CTO of a university spin-out company pureVLC Ltd. His main research interests are in interference coordination in wireless networks, spatial modulation and optical wireless communication. Prof. Haas holds more than 25 patents. He has published more than 60 journal papers including a Science Article and more than 160 peer-reviewed conference papers. Nine of his papers are invited papers. Since 2007 Prof. Haas has been a Regular High Level Visiting Scientist supported by the Chinese '111 programme' at Beijing University of Posts and Telecommunications (BUPT). He was an invited speaker at the TED Global conference 2011, and his work on optical wireless communication was listed among the '50 best inventions in 2011' in Time Magazine. He recently has been awarded a prestigious Fellowship from the Engineering and Physical Sciences Research Council (EPSRC) in the UK.

Hauke Holtkamp received his B.Sc. degree in electrical engineering at Jacobs University Bremen, Germany, in 2005, and the M.Sc. degree in communication networks from Technical University Munich, Germany, in 2008. He is currently working on the energy efficiency of cellular networks at DOCOMO Euro-Labs GmbH in Munich, Germany, while pursuing a Ph.D. degree in communication networks at the University of Edinburgh, UK.