OFDMA Base Station Power-saving Via Joint Power Control and DTX in Cellular Systems

Hauke Holtkamp

DOCOMO Euro-Labs
D-80687 Munich, Germany
E-mail: h.holtkamp@gmail.com

Harald Haas

Institute for Digital Communications
Joint Research Institute for Signal and Image Processing
The University of Edinburgh, EH9 3JL, Edinburgh, UK
E-mail: h.haas@ed.ac.uk

Abstract—In this paper, we extend the previously proposed power-saving Resource Allocation using Power Control and Sleep (RAPS) algorithm to the interference environment. The RAPS algorithm allocates resources, power and Discontinuous Transmission (DTX) time slots jointly for Orthogonal Frequency Division Multiple Access (OFDMA) downlink resource scheduling such that Base Station (BS) supply power consumption is reduced while providing guaranteed link rates. The RAPS scheduler was previously proposed and studied only as a single-cell mechanism. It is here adapted to and tested in the multi-cell interference setting. The resulting dynamical system is analyzed with respect to convergence, power consumption and load dependence. It is found that the Signal-to-Interference-and-Noise-Ratio (SINR) distribution in the network is greatly improved and average power consumption can be reduced by around 50% in comparison to a Proportional-Fair (PF) scheduling benchmark.

I. INTRODUCTION

With the increasing number and density of cellular BSs, power consumption is becoming a growing concern for network operators [1]. Rising energy costs, the desire for better disaster recovery and the reduction of CO₂ emissions all motivate more energy efficient BSs [2]. While energy efficiency is highly relevant, it still ranks behind behind capacity and reliability in operator priorities. Thus, for an operator, it is not acceptable to increase energy efficiency by degrading user Quality of Service (QoS). However, studies have shown that in many locations and during a significant share of operation time, traffic load is far below cell peak capacity [1]. This opens the opportunity to deliberately reduce cell capacities as long as all user demands can still be fulfilled, thus gaining energy efficiency without affecting the QoS. Exploiting this provisioning gap is the tool for power-saving Radio Resource Management (RRM) mechanisms.

Energy efficient RRM can broadly be divided into two approaches, namely saving through sleep mode and saving through transmission Power Control (PC). Sleep mode approaches can further be separated into long sleep and micro sleep. Long sleep refers to turning devices off completely for durations of minutes and more. This clearly saves most energy as devices’ power consumption can possibly be reduced to zero, but creates issues with regard to QoS, coverage and wake-up triggering. In contrast, micro sleep operates on time scales of milliseconds and can be arranged to be transparent to User Equipments (UEs). Micro sleep is also referred to as DTX because it is more closely related to a short interruption in transmission rather than a deep sleep state. A device in DTX will not turn off entirely and remain available to continue transmission, but has significantly reduced (although not zero) power consumption. The alternative approach, PC, is directly linked to the allocation of subcarriers or Resource Blocks (RBs) when considering OFDMA transmission systems. Most works on OFDMA PC, including this paper, therefore also include resource allocation mechanisms. As we specifically aim at maintaining the QoS, long sleep is not considered in this work.

For a comparison with the state of the art, we proceed to present a literature overview on DTX, PC and resource allocation: DTX in BSs is still a relatively new topic. Frenger et al. [3] propose to use Long Term Evolution (LTE) Multicast-Broadcast Single Frequency Network (MBSFN) subframes to allow DTX in current standards, but do not show how resources should be scheduled to align transmission with the subframes. Abdallah et al. [4] align DTX time slots via a central controller and find that orthogonal time slot assignment to neighboring cells is desirable in terms of rate maximization. In comparison to DTX, PC literature has a longer history as it is also beneficial for link adaptation and interference mitigation. Wong et al. [5] proposed the first, but highly complex transmission power minimizing OFDMA resource and power allocation method. Miao et al. [6] maximize energy efficiency on an individual link while considering circuit consumption. Videv et al. [7] spread signals over unused bandwidth and thus achieve a lower transmission power at unchanged rates. Xu et al. [8] jointly allocate subcarriers, transmission power and antenna chains to minimize device consumption. López-Pérez et al. [9] reduce interference and transmission power via PC and resource allocation, but do not consider DTX. The only consideration of both sleep modes and PC is presented by Domenico [10] who divides traffic into delay tolerant and real-time traffic. Under real-time traffic, transmission power is reduced by bandwidth spreading while under delay-tolerant traffic, the device is put into sleep mode.

In [11], we have recently presented the RAPS algorithm, which allocates OFDMA resources, transmission power and DTX time slots in a multi-user cell under rate guarantees at
reduced device power consumption. This work is extended here to the multi-cell interference system with the following contributions:

- Allocation decisions are based on the UE SINR rather than the Channel State Information (CSI).
- We include an iterative technique which continuously adapts to the dynamical system in contrast to the single iteration previously proposed.
- We derive a power-saving adaptation of a PF scheduler as the state-of-the-art benchmark.
- The effect of power-saving resource allocation is heuristically analyzed in terms of convergence, power consumption, SINR distribution and energy-per-bit and benchmarked against the comparable PF resource allocation mechanism.

The paper is organized as follows. The system model, the multi-cell extensions and the benchmarks are presented in Section II. Simulation results are presented in Section III. The paper is concluded in Section IV.

II. SYSTEM MODEL

The RAPS algorithm [11] reduces power consumption via resource, power and DTX time slot allocation in a cell with multiple UEs while providing a guaranteed rate, $R_k$, to each UE $k$ over each OFDMA frame based on the reported RB CSI. The reduction in supply power consumption is achieved by first computing the joint PC and DTX solution to a real-valued and convex sub-problem and then quantizing the solution to find the integer-valued RB allocation. The allocation is based on CSI values on each RB as reported from the UEs. See Fig. 1 for a block diagram of the RAPS algorithm.

We make the following assumptions about the network. A transmission power limit applies in each time slot, which is determined by the BS hardware. Each RB can only be allocated to one user, i.e. there is no RB sharing. The weighting of the decision on the number of DTX time slots is made on the basis of a linear power model [12]. The linear power model approximates the supply power consumption of LTE BSs as an affine function of the transmission power and provides a reduced power consumption value during DTX, $P_S$.

Assuming flat-fading channels for each user in a Time Division Multiple Access (TDMA) transmission system, a convex optimization problem is used to approximate real-valued resource shares, $\mu_k$, as part of the RAPS algorithm:

$$
\begin{align*}
\text{minimize} & \quad P_{\text{supply}}(r) = \sum_{k=1}^{K} \mu_k (P_0 + \Delta_{PM} P_k(R_k)) \\
& \quad \quad \quad + \mu_{K+1} P_S \\
\text{subject to} & \quad \sum_{k=1}^{K+1} \mu_k = 1 \\
& \quad \mu_k \geq 0 \ \forall k \\
& \quad 0 \leq P_k(R_k) \leq P_{\text{max}},
\end{align*}
$$

(1)

where $P_{\text{supply}}(r)$ is the average cell supply power consumption when providing a rate vector $r$ to $K$ UEs. $P_0$ and $\Delta_{PM}$ are the power consumption while idle and the transmission power load factor, respectively. $P_k(R_k)$, the transmission power required to transmit rate $R_k$ to user $k$, can be determined for up to $4 \times 4$ Multiple-Input Multiple-Output (MIMO) in closed form via inverting the Shannon capacity. Note that in a flat-fading TDMA system, the transmission power per user, $P_1$, is equal to the system transmission power at that time instant. The resource share allocated to DTX is represented by $\mu_{K+1}$.

The solution vector $\mu \in (\mu_1, \ldots, \mu_{K+1})$ provides the supply-power-minimal resource shares that each UE should receive as well as the time share of DTX. Due to the integer nature of OFDMA resource allocation, this result needs to be quantized. The quantized resource shares are used as input to Rate Craving Greedy (RCG) scheduling and power allocation via Margin Adaptation, concluding the RAPS algorithm by returning RB power levels, DTX time slots and UE RB allocation.

In the multicell setting, which is inherently interference limited, applying the CSI without consideration of the interference as the basis for scheduling results in an overestimation of achievable rates. In order to apply the RAPS algorithm effectively in the multicell setting, we propose the following two extensions.

First, instead of calculating achievable rates, $R_k$, based on CSI alone, we take interference into account for scheduling by considering the interference and noise covariance, $F_\eta$, in addition to the CSI matrix, $H$. We assume that as described in [13] on each RB there can exist a flat-fading 2x2 point-to-point MIMO link with capacity

$$
R_{RB} = \log_2 \left| I + HF_\eta^{-1} H^H F_\eta \right| = \sum_{i \in \text{spatial streams}} \log_2 \left( 1 + (\text{eig}(HF_\eta^{-1} H^H)), P_i \right),
$$

(2)
where
\[ F_\eta = I + \sigma_n^2 + \sum_{j \in \text{interferers}} H_j P_j H_j^H \]  
and
\[ H \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{Tx}}} \].

The determinant operator is \(|\cdot|\), \(I\) is the identity matrix, \(\text{eig}()\) is the eigenvalue operation and \(\sigma_n^2\) is the thermal noise power. Each entry of \(H\) is composed of path gain, shadowing gain, antenna gain and a Rayleigh distributed fast fading component.

The achievable rate per UE, \(R_k\), is the sum of all scheduled \(R_k\).

Second, each BS performs the RAPS algorithm on each consecutive OFDMA frame instead of a single execution. As a consequence, estimation errors can be corrected over time and the scheduling decisions of neighboring BSs are incorporated into the current allocation via the change in \(F_\eta\).

These extensions transform the single-cell mechanism into a dynamical system consisting of multiple cells, each of which performs the RAPS algorithm independently. Each user has a rate requirement and reports its SINR to the BS it is connected to. The perceived SINR depends on the transmission powers of all BSs. Thus, the scheduling decisions which the RAPS algorithm makes in one cell affect all other cells by means of the changing interference setting. We consider a system where, first, all UEs report their SINR per RB for an entire OFDMA frame and, then, all BSs simultaneously perform the RAPS allocation over this frame. Note that the BSs do not exchange information. Consequently, the SINR profile reported by each UE is always outdated by one iteration. As transmission power in each time slot is reduced by the RAPS algorithm due to PC or is even set to zero in a DTX slot, interference between BSs is inherently reduced.

In order to assess the performance of the RAPS algorithm in the interference scenario, we have identified Proportional-Fair (PF) scheduling as the closest comparable scheduler. By default, PF scheduling will allocate all resources regardless of requested rates. While it prevents service starvation of users with bad channels, PF is still a rate maximizing scheduler (not a power minimizing scheduler). Therefore, we make the adaptation that PF will stop scheduling more resources to a UE once a rate target has been fulfilled, thus providing guaranteed rates like RAPS while limiting power consumption. In this fashion, many resources will potentially remain unscheduled, reducing the sum bandwidth in each time slot. This is, hence, a bandwidth adapting opportunistic scheduler. Power-saving is achieved by the reduction in transmission power as a consequence of reduced transmission bandwidth at constant power spectral density.

Sequential bandwidth adaptation is taken as a second benchmark. This technique allocates RBs sequentially (not opportunistically) in the time domain UE-by-UE until all rate targets are fulfilled. When the system is under-loaded, some RBs will remain unused, thus reducing the used bandwidth. Due to its negligence of multiuser diversity, this benchmark provides a lower bound on scheduling performance.

To the best of our knowledge, the multi-cell system (like the single-cell system [11]) does not have a closed form solution for the resource, power and DTX allocation as functions of RB SINR and target rates. We therefore inspect the performance of the RAPS algorithm heuristically via Monte Carlo simulation.

### III. Results

We tested a hexagonally arranged macro BS network with a homogeneous UE distribution. The network was tested with the set of parameters shown in Table I. Three tiers of cells were placed around the center cell to generate a calibrated SINR distribution over all resources within the center cell. All cells perform RAPS resource allocation, but only data from the center cell was collected. The simulation comprises sectorization and frequency-selective fading. Unless stated otherwise, the initial power allocation state is an equal distribution of the maximum allowed power over all resources. This creates a high interference initial setting, which the RAPS algorithm improves, as well as a fairer comparison to the benchmarks which also allocate maximum power per RB. Signalling overhead is not considered in this paper. Consideration of signalling RBs prevents some time slots from being scheduled for DTX and is a topic of future study. DTX time slots are allocated sequentially from last-to-first within the OFDMA frame. Supply power consumption is averaged over an entire OFDMA frame.

We begin inspection of the network behavior by means of the SINR distribution over those RBs which were scheduled for transmission by RAPS, see Fig. 2. The SINR distributions are shown for the first ten iterations of the network after startup. As interference is reduced in each iteration, the scheduler assigns lower transmission powers. As a result, the overall SINR distribution of the scheduled RBs remains largely unchanged.

The reduction of interference can be more clearly observed when inspecting the SINR assuming unit power on all RBs in Fig. 3. It can be seen that before the first application of RAPS (Iteration 1), the unit power SINRs are in the region of -20 to +20 dB. With each iteration, the unit power SINR distribution improves, yielding higher unit power SINRs over all RBs. This is caused by the decreased interference through reduced

| TABLE I  
<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Value</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Intersite distance</td>
<td>500 m</td>
</tr>
<tr>
<td>Pathloss model</td>
<td>3GPP UMa, NLOS, shadowing [14]</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>8 dB</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Maximum transmission power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>Thermal noise temperature</td>
<td>290 K</td>
</tr>
<tr>
<td>Interference tiers</td>
<td>2 (19 cells)</td>
</tr>
<tr>
<td>User target rate</td>
<td>2 Mbps</td>
</tr>
<tr>
<td>OFDMA subframes (time slots)</td>
<td>10</td>
</tr>
<tr>
<td>Frequency chunks</td>
<td>50</td>
</tr>
<tr>
<td>Power model [15] (idle; load factor; DTX)</td>
<td>200 W; 3.75; 90 W</td>
</tr>
</tbody>
</table>
transmission powers in interfering cells as a consequence of zero-power DTX resources or reduced-power PC resources. Although there is no communication between the BSs, the RAPS scheduler only selects resources for transmission which have good SINR, thus reducing interference on resources with a low SINR. We conclude that RAPS acts as an interference reduction mechanism. The SINR distribution remains constant after the fifth iteration, reaching a stable network state. SINRs are improved between 20 dB for low SINRs to 50 dB for high SINRs within six iterations.

To assess the dynamics and convergence of the system and the reduction in power consumption, we next inspect the power consumption of the center cell. In Fig. 4, the center cell average power consumption is shown for 40 independent trials for different initial power allocations. As computing an initial worst-case power allocation over all RBs and cells by brute force is infeasible, we start each trial with one of three power configurations, namely zero power on all RBs, maximum equal power on all RBs and a random allocation over all RBs within the transmission power constraint. In addition, the PF benchmark is applied with a maximum equal power starting condition. It is found that for all of the starting configurations, most RAPS trials converge within four iterations and all RAPS trials converge within ten iterations. Convergence is reached from all three power configurations (max, random and zero power), reinforcing the prediction that the RAPS scheduler converges for any starting configuration in large networks. Although each RAPS trial is performed in an individual simulation with different channel realizations, all trials converge to the same power consumption level of 112 W. This is caused by the following effect: As transmission powers are reduced in each iteration, $P_0$ and $P_S$ begin to dominate the overall consumption. As for the same target rate per UE, two DTX time slots turn out to be the solution, and a total average power consumption of $8P_S + 2P_0 = 112$ W is eventually reached. This consumption level is significantly lower than the upper limit at 350 Watt. This saving is achieved by the exploitation of the provisioning gap between target capacity and available capacity. By only using resources with good channel conditions for transmission, optimal power allocation and DTX, no transmission or device power consumption is wasted. In comparison, the benchmark PF is much less adaptable to the low load and converges to a consumption of 207 W. Overall, RAPS operation allows to reduce power consumption by almost 50% compared to the PF benchmark.

A noticeable effect is the convergence to distinct levels of power consumption. These are owed to the coarse quantization nature of the DTX time slots. A degradation in SINR profiles reported by UEs triggers the RAPS algorithm to allocate one more time slot for transmission, thus increasing power consumption to a higher level. In turn, improved SINRs in a cell lead to one less time slot used for transmission yielding lower power consumption. In this region of operation transmission powers are low compared to other device consumption, emphasizing the effect of DTX time slot selection and creating the stepped pattern (most prominent in ‘RAPS, zero power’). Lastly, we assess the performance of the RAPS algorithm over different cell sum data rates. In Fig. 5, it is found that PF scheduling only achieves marginal gains compared to sequential bandwidth adaptation, a very weak benchmark. The power savings of PF over this weak benchmark are in...
the order of 20 W to 40 W. Savings are larger at higher cell sum rates (i.e., higher traffic load) than at lower cell sum rates. In contrast, RAPS scheduling achieves a reduction in power consumption of up to 50% or 150 W compared to PF scheduling. The savings are almost of constant magnitudes over all cell sum rates tested. Note the missing data point at 32 Mbps for sequential bandwidth adaptation. This is due to the fact that a cell sum rate of 32 Mbps could not be achieved in trials for sequential bandwidth adaptation due to its lack of multi-user diversity exploitation. PF and RAPS can provide these rates due to opportunistic resource allocation.

However, although the PF and RAPS schedulers can provide similar rates, they can do so at significantly differing average power consumption. This contribution is further analyzed by considering the invested energy-per-bit. The same data set as in Fig. 5 is plotted in Fig. 6 on the energy-per-bit metric to emphasize the non-linear relationship between data transmitted and cell power consumption. Generally, at high cell sum rates the energy-per-bit metric is lower than at low cell sum rates for all schedulers. This is not surprising as BSs operate more efficiently at high capacities. However, application of the RAPS algorithm nearly halves the energy-per-bit cost over all cell sum rates, thus doubling efficiency compared to PF scheduling.

IV. Conclusion

In this paper, we have presented the first adaptation of the RAPS algorithm for a cellular system that is subject to co-channel interference. OFDMA SINR is added as an input parameter to provide the basis for RAPS power, DTX and resource allocation. In contrast to the single cell, application of RAPS in the multi-cell setting creates a dynamical system in which the RAPS algorithm has to be executed repeatedly. We find the system to be converging after six iterations. Simulation results indicate that SINR distributions can be improved by more than 20 dB by employing RAPS in OFDMA BSs. A RAPS system is shown to converge independent of the initial power allocation configuration. Without exchanging information between the BSs, cell power consumption is reduced by around 50% compared to the PF scheduler benchmark. The RAPS scheduler is, thus, able to double the energy efficiency of interference limited cellular networks.

REFERENCES