ANTICIPATORY NETWORKING FOR ENERGY SAVINGS IN 5G SYSTEMS

E. Pollakis⋆*, and S. Stańczak*⋆†

⋆ Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany
† Technische Universität Berlin, Einsteinufer 27, 10587 Berlin, Germany
email: {emmanuel.pollakis,slawomir.stanczak}@hhi.fraunhofer.de

ABSTRACT

In this paper, we devise novel techniques for saving energy in 5G wireless systems. By means of anticipated transmission rates we find user-cell assignments and scheduling policies that help to identify energy-efficient network topologies. In particular, the objective of this paper is to find a user-cell association and rate allocation over time that provides the requested Quality of Service (QoS) to all users while attempting to reduce the energy consumption by identifying the set of active cells consuming the least amount of energy. The used energy consumption model specifically includes the static energy consumption and the dynamic, load-dependent energy consumption of cells. We formalize this problem as a non-convex optimization problem that accounts for the requirements of buffered delay-sensitive applications. We apply relaxation techniques to find feasible anticipated schedules for rate allocation and user-cell assignments with a low complexity algorithm which is amenable to online implementation. We characterize achievable energy savings by means of simulations in a realistic network scenario under realistic traffic patterns.

I. INTRODUCTION

With the advent of the Internet of Things (IoT), billions of new devices will be connected wirelessly to the mobile communication system of the fifth generation (5G). A wide spectrum of use cases entails that 5G will need to support extreme objectives for delay, capacity and energy, which in turn requires a high network adaptability to varying user requirements. Todays networks are operated in a static manner with more or less fixed network configurations providing the maximum quality of service (QoS) at all times. Eventhough such a mode of operation might satisfy delay and capacity requirements, it will inevitably lead to unacceptable high energy consumption. Therefore, it is of utmost importance to develop new mechanisms to support energy savings in both peak hours as well as in off-peak hours.

Most existing energy saving techniques, such as cell deactivation or sleep mode, are designed for stationary users and static user demands. These techniques usually have bad performance in bursty traffic situations. The achievable energy savings are nullified by short time increases of traffic demand in certain areas. Proactive resource allocation and user assignment (PRAUA) is a promising approach to enable 5G to stand up to its high promises by improving

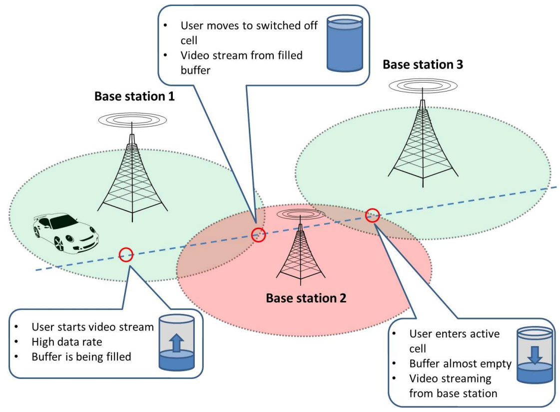

Fig. 1. Toy-example of a buffered delay-sensitive application schedule.

service quality and reducing energy consumption at the same time. In particular, PRAUA helps 'smear' the traffic requirements in time and space allowing for maintaining energy-efficient network configurations over a longer period of time. Thereby, PRAUA exploits the knowledge about users' mobility which can be obtained from side information or estimated with sufficient accuracy due to the high regularity in human mobility [1]. This information along with learned path loss maps [2] can be used to proactively build user-cell assignment and resource allocation schedules that greatly support energy savings in cellular communication systems during off-peak hours. In particular, we develop algorithms that schedule data transmissions for new service applications when it is favorable for energy savings. The developed mechanisms target new service types enabled by the storage capabilities at user devices. In particular we address buffered delay-sensitive applications that include services like stored content streaming (music/video). Such applications require a constant instantaneous data rate where data can be pulled either directly from the access network or from a pre-filled buffer (depicted in Fig. 1). These new service types allow base stations to delay or bring forward the transfer of data to users which is the fundamental concept we exploit for energy savings. Proactive scheduling has been considered in [3] to improve the QoS of users traveling through the service area of several cells. The presented framework plans the resource allocation over a certain time horizon for fixed user-cell assignments to maximize the throughput of users. The authors of [4] propose a predictive framework for video streaming applications to increase the energy-efficiency in wireless networks. The problem is formulated as a mixed integer linear program (MILP) where decisions on multiuser rate allocation, video segment quality, and base station transmit power are jointly optimized. A heuristic multi stage algorithm is used to derive solutions for the MILP problem by first allocating rates to users and then determining the segment quality and active base station set. The reasoning is based on the observation that efficient rate-allocation schemes provide power savings. Other analytical justifications for the perfomance are not given. The work in [5] is most closely related to ours. By proactive resource allocation and video quality decisions the authors reduce the energy consumption of the whole network by solving a mixed integer non-linear problem (MINLP). An algorithm is proposed that decomposes the association and resource allocation problem in a master problem and several sub-problems to make the problem tractable. Thereby, the authors leverage energy costs and video quality taking into account backhaul costs. The resulting integer programs are solved directly by mathematical solvers and the authors argue to achieve decent scalability. However, this is achieved by assuming the allocation of an equal number of resource blocks to all users in the master problem.

In contrast to the above work, we explicitly incorporate the user-cell assignment in the optimization framework and target energy savings with a guaranteed QoS level instead of maximizing the QoS. Furthermore, we use mathematically justified relaxation techniques instead of heuristics to derive solutions ensuring good scalability. In more detail, we exploit the possibility to preload and store data on user devices which will serve as an enabling concept to save energy in the communication system by disengaging certain cells[1]. By delaying the service provision of some users we may avoid to activate cells that are only needed when the traffic demand is of bursty nature. The result is a resource usage that lets users buffer data in high capacity cells whereas it avoids access to cells that are overloaded or switched off for reasons of energy savings. We use the predicted routes of users and the learned path loss coverage maps to find such a user-cell association and rate allocation policy under the exploitation of the users' buffers, i.e., when a user is predicted to pass an area without coverage it will be allocated more resources right before, so that the data can be loaded in the buffer (bridging the coverage hole). To increase the degrees of freedom for the PRAUA by multi-connectivity to multiple cells and to exploit the mutual information received from them we propose to use fountain codes [6]. In the concept of fountain coding a potential infinitely long stream of encoded symbols is generated from a finite set of data symbols. Decoding is possible as soon as a particular amount of code symbols is received error free with no requirement for a consecutive order. The multi-connectivity helps to find better solutions for the problem at hand and increases the robustness of our solution.

[1]Notice, that we are considering only capacity cells for deactivation. Basic coverage for other users and services has to be secured at all times. We therefore assume basic coverage by some legacy network.

In the following we summarize the main contribution of our work:

- We propose an optimization framework that exploits the knowledge of user-cell trajectories and learned path-loss maps. It finds user-cell association and rate schedules that provide the requested QoS of users and reduces the energy consumption of future cellular communication networks.
- Our model for energy consumption is general enough to capture static energy consumption (cooling, basic power conversion etc.) as well as dynamic load-dependent energy consumption.
- We exploit the end user devices' storage capabilities to serve buffered delay-sensitive applications with PRAUA.
- With the introduction of PRAUA we stretch the applicability of cell sleep and switching on/off techniques in the time horizon leading to improved energy savings.
- The use of fountain coding is proposed to improve the QoS of users while being able to deactivate more cells.
- We make use of relaxation techniques that are able to give good solutions to this problem in reasonable time making it amenable for online implementation. Thereby, it has theoretical justification for its good performance.

II. SCENARIO AND SYSTEM MODEL

We consider the downlink of a cellular heterogeneous communication system employing an OFDM-based resource allocation. We are interested in switching off capacity units of the network, e.g. sectors, cells or the entire base station[2]; the corresponding decisions are performed at a central network controller. The set of all cells is denoted by $\mathcal{M} = \{1, 2, ..., M\}$. Each cell i has total number of resource blocks B_i to allocate to its users. There are N users in the system to be served and we denote the set of all users as $\mathcal{N} = \{1, 2, ..., N\}$. The time is divided into K time slots of equal duration Δ_k. For each time slot, the objective is to find a resource allocation and user-cell association. Each user is equipped with a first-in-first-out (FIFO) buffer and we denote the buffer level (in bits) of user j in slot k by $d_j^{(k)} \geq 0$ with $d_j^{(0)} = 0$ (empty buffer at start). In this work we assume a sufficiently large buffer and refer to the technology specific spectral efficiency per resource block of the link from cell i to user j in slot k as $\omega_{i,j}^{(k)}$.

Assumption 1: A reliable estimate of the users' routes and the supported spectral efficiency per resource block along those routes is available at the central controller.

The information for Assumption 1 can be obtained by using side information or estimated mobility trajectories based on the high regularity in human mobility [1] together with techniques to learn path loss maps, e.g., [2].

[2]In the text that follows we will use cells as a placeholder for any type of network element that can be switched on/off independently.

The task of our optimization framework is to provide a schedule of resource allocations satisfying the rate requirements of users while trying to reduce the energy consumption. If a user j is served by cell i in time slot k we denote the effective transmit data rate as $r_{i,j}^{(k)} := b_{i,j}^{(k)} \omega_{i,j}^{(k)}$ where $b_{i,j}^{(k)}$ is the number of resource units allocated to user j by cell i in slot k. We collect the rates allocated by cell i to all users at time k in vector $\boldsymbol{r}_i^{(k)} = \left[r_{i,1}^{(k)}, r_{i,2}^{(k)}, ..., r_{i,N}^{(k)} \right]^T$. We further use $\mathbf{R}^{(k)} = \left[\boldsymbol{r}_1^{(k)}, \boldsymbol{r}_2^{(k)}, ..., \boldsymbol{r}_M^{(k)} \right]$ to refer to all rates allocated over all cells to all users in slot k.

Definition 1 (Instantaneous Cell Load): Given the rate assignment matrix $\mathbf{R}^{(k)}$ for slot k, the load of cell i, denoted by $\rho_i^{(k)} \left(\mathbf{R}^{(k)} \right) \in [0,1]$ or simply $\rho_i^{(k)}$ for notational simplicity, is defined to be the ratio of the number of resource blocks allocated to users served by cell $i \in \mathcal{M}$ in slot k to the total number of resource blocks B_i available at this cell, i.e., $\rho_i^{(k)} = \frac{\sum_{j \in \mathcal{N}} b_{i,j}^{(k)}}{B_i}$.

We use $\boldsymbol{\rho}_i := [\rho_i^{(1)}, \ldots, \rho_i^{(K)}]^T \in [0,1]^K$ to denote the vector of cell loads at cell i for all time slots and denote the collection of all cell loads over time by $\mathbf{P} := [\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_M]^T \in [0,1]^{M \times K}$. A consequence of Definition 1 is the following fact:

Fact 1: The load at cell i satisfies $\rho_i^{(k)} > 0$ if and only if (iff) cell i serves at least one user in slot k.

In other words, $|\boldsymbol{\rho}_i \ \mathbf{1}|_0 = 0$ iff cell i serves no user in all time slots K, where $\mathbf{1} \in \mathbb{R}^K$ is a vector of ones and $|\cdot|_0$ is the l_0-norm. If $|\boldsymbol{\rho}_i \ \mathbf{1}|_0 = 0$ cell i can be deactivated for energy saving reasons.

Remark 1: The l_0-norm counts the non-zero elements of a matrix or vector. For a scalar $x \in \mathbb{R}$, the l_0-norm is defined as $|x|_0 := 1$ if $x \neq 0$ and $|x|_0 := 0$ otherwise.

Buffered delay-sensitive applications

Buffered delay-sensitive application are characterized by a strict per time slot data rate requirement of each user. In more detail, the data transmitted to the user is stored in its buffer from which the delay-sensitive application reads with constant data rate r_j^{min}. If the scheduling algorithm allocates a higher data rate to a user in time slot k, i.e., $r_{i,j}^{(k)} > r_j^{\text{min}}$, then the additional transferred data remains in the users buffer $d_j^{(k)} = d_j^{(k-1)} + \Delta_k (r_{i,j}^{(k)} - r_j^{\text{min}})$ and is read in the next time slot. If the user is not allocated a sufficiently high rate in slot k, i.e., $r_{i,j}^{(k)} < r_j^{\text{min}}$, the buffer level decreases as $d_j^{(k)} = d_j^{(k-1)} - \Delta_k (r_j^{\text{min}} - r_{i,j}^{(k)})$. In every time slot k the aggregate data rate from the buffer and streamed from a cell has to be large enough which yields the constraint[3]

$$\sum_{i \in \mathcal{M}} r_{i,j}^{(k)} + \frac{d_j^{(k-1)}}{\Delta_k} \geq r_j^{\text{min}}. \tag{1}$$

[3]Note, that this definition allows users to be served by multiple cells as well as the buffer in a time slot. In such cases fountain coding is used to implement mutual information combing.

The buffer level of user j at the end of time slot k is therefore described by

$$0 \leq d_j^{(k)} = d_j^{(k-1)} + \sum_{i \in \mathcal{M}} \Delta_k r_{i,j}^{(k)} - \Delta_k r_j^{\text{min}}. \tag{2}$$

Since each base station has only B_i resource units to allocate to users we have the constraint

$$\sum_{j \in \mathcal{N}} \frac{b_{i,j}^{(k)}}{B_i} = \sum_{j \in \mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i \omega_{i,j}^{(k)}} \leq \rho_i^{(k)}. \tag{3}$$

III. PROBLEM STATEMENT

We are now in the position to state the optimization problem that aims at finding the optimal set of active cells, user-cell assignments and rate allocations while consuming the least amount of energy. The objective function $E : [0,1]^{M \times K} \to \mathbb{R}_+$ is a combination of static and dynamic sources of energy consumption. In more detail, each active cell has static energy consumption of e_i per time slot and a load dependent part which is captured by a concave or convex function $f_i(\boldsymbol{\rho}_i) \in \mathbb{R}_+$ or simply f_i. The total network energy consumption in a so called *on/off scheme* is

$$E_{\text{on/off}}(P) = \sum_{i \in \mathcal{M}} K \ e_i \, |\boldsymbol{\rho}_i \ \mathbf{1}|_0 + f_i. \tag{4}$$

The above model assumes that cells are deactivated before the first time slot and stay inactive for all K time slots. The model can easily be adapted to modes of operation where so called micro-sleeps of cells are allowed. In such cases, a cell can be deactivated for a single time slot k in order to save energy and be activated in the next time slot $k+1$. The energy consumption of such a mode of operation is captured by

$$E_{\text{sleep}}(P) = \sum_{i \in \mathcal{M}} \sum_{k=1}^{K} e_i \left| \rho_i^{(k)} \right|_0 + f_i. \tag{5}$$

We refer to the scheme where micro-sleep of cells is allowed as *micro sleep scheme*.

The complete optimization problem for *buffered delay-sensitive applications* with the *on/off scheme* can be composed as

$$\begin{aligned}
\text{min.} \quad & \sum_{i \in \mathcal{M}} K \ e_i \, |\boldsymbol{\rho}_i \ \mathbf{1}|_0 + f_i && \text{(6a)} \\
\text{s. t.:} \quad & \sum_{j \in \mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i \omega_{i,j}^{(k)}} \leq \rho_i^{(k)} & \forall i, k \quad & \text{(6b)} \\
& \sum_{i \in \mathcal{M}} r_{i,j}^{(k)} + \frac{d_j^{(k-1)}}{\Delta_k} \geq r_j^{\text{min}} & \forall j, k \quad & \text{(6c)} \\
& d_j^{(k-1)} + \sum_{i \in \mathcal{M}} \Delta_k r_{i,j}^{(k)} - \Delta_k r_j^{\text{min}} = d_j^{(k)} & \forall j, k \quad & \text{(6d)} \\
& 0 \leq d_j^{(k)} & \forall j, k, \quad & \text{(6e)}
\end{aligned}$$

where the optimization variables are $r_{i,j}^{(k)} \in \mathbb{R}_+$ and $\rho_i^{(k)} \in [0,1]$. Thereby, (6b) assures that cells are not overloaded and

(6c) guarantees that users receive the required instantaneous data rate. Constraint (6d) represents the buffer level increase or decrease at each time slot k.

Problem (6) can be written in a more compact form as

$$\text{min.} \sum_{i\in\mathcal{M}} K\ e_i \left|\boldsymbol{\rho}_i\ \mathbf{1}\right|_0 + f_i \tag{7a}$$

$$\text{s. t.:} \sum_{j\in\mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i\omega_{i,j}^{(k)}} \leq \rho_i^{(k)} \qquad \forall i,k \tag{7b}$$

$$\sum_{l=1}^{k}\left(\sum_{i\in\mathcal{M}} r_{i,j}^{(l)} - r_j^{\min}\right) \geq 0 \qquad \forall j,k, \tag{7c}$$

since the buffer level at the end of time slot k can be stated as the data surplus of the aggregated data transmitted up to time slot k.

In a similar way we can state the problem for the *micro sleep scheme* which uses (5) and we obtain

$$\text{min.} \sum_{i\in\mathcal{M}}\sum_{k=1}^{K} e_i \left|\rho_i^{(k)}\right|_0 + f_i \tag{8a}$$

$$\text{s. t.:} \sum_{j\in\mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i\omega_{i,j}^{(k)}} \leq \rho_i^{(k)} \qquad \forall i,k \tag{8b}$$

$$\sum_{l=1}^{k}\left(\sum_{i\in\mathcal{M}} r_{i,j}^{(l)} - r_j^{\min}\right) \geq 0 \qquad \forall j,k, \tag{8c}$$

where the optimization variables are $r_{i,j}^{(k)} \in \mathbb{R}_+$ and $\rho_i^{(k)} \in [0,1]$.

IV. ENERGY SAVINGS OPTIMIZATION

Due to their non-convex objective function Problem 7 and 8 are in general hard to solve. Fortunately, both problems exhibit a structure that can be exploited by low complexity algorithms to find good user-cell association and rate schedules consuming low energy. In more detail, we apply a relaxation of the l_0-norm in combination with the Majorization-Minimization method as proposed in [7]. In the following we derive the algorithm and refer the reader to [7] for details on the complexity and performance evaluation of the algorithm itself.

The objective functions of Problem 7 and 8 are not continuous due to the involved l_0-norm. To obtain an optimization problem that is mathematically tractable, we address the non-continuity problem of the l_0-norm by considering the following relation [8]:

$$\forall_{\boldsymbol{z}\in\mathbb{R}^K}\ |\boldsymbol{z}|_0 = \lim_{\epsilon\to 0}\sum_{k=1}^{K}\frac{\log(1+|z_k|\ \epsilon^{-1})}{\log(1+\epsilon^{-1})}, \tag{9}$$

Thus, (7a) can be equivalently written as

$$\begin{aligned}&\sum_{i\in\mathcal{M}} K\ e_i \left|\boldsymbol{\rho}_i\ \mathbf{1}\right|_0 + f_i \\ &\quad = \lim_{\epsilon\to 0}\sum_{i\in\mathcal{M}} K\ e_i\ \frac{\log(1+\epsilon^{-1}\ \boldsymbol{\rho}_i\ \mathbf{1})}{\log(1+\epsilon^{-1})} + f_i.\end{aligned} \tag{10}$$

Similarly, the objective function of the *micro sleep scheme* (8a) can be equivalently written as

$$\begin{aligned}&\sum_{i\in\mathcal{M}}\sum_{k=1}^{K} e_i \left|\rho_i^{(k)}\right|_0 + f_i \\ &\quad = \lim_{\epsilon\to 0}\sum_{i\in\mathcal{M}}\sum_{k=1}^{K} e_i\ \frac{\log(1+\epsilon^{-1}\ \rho_i^{(k)})}{\log(1+\epsilon^{-1})} + f_i.\end{aligned} \tag{11}$$

We obtain a relaxed version of Problem 7 and 8 by replacing the objective function by the right-hand side of (10) and (11), respectively and fixing $\epsilon > 0$ to a sufficiently small value which yields

$$\text{min.} \sum_{i\in\mathcal{M}} K\ e_i\ \frac{\log(1+\epsilon^{-1}\ \boldsymbol{\rho}_i\ \mathbf{1})}{\log(1+\epsilon^{-1})} + f_i \tag{12a}$$

$$\text{s. t.:} \sum_{j\in\mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i\omega_{i,j}^{(k)}} \leq \rho_i^{(k)} \qquad i\in\mathcal{M} \tag{12b}$$

$$\sum_{l=1}^{k}\sum_{i\in\mathcal{M}}(r_{i,j}^{(l)} - r_j^{\min}) \geq 0 \qquad j\in\mathcal{N} \tag{12c}$$

for the *on/off scheme*. Accordingly, the we state the relaxed form of the *micro sleep scheme* as

$$\text{min.} \sum_{i\in\mathcal{M}}\sum_{k=1}^{K} e_i\ \frac{\log(1+\epsilon^{-1}\ \rho_i^{(k)})}{\log(1+\epsilon^{-1})} + f_i \tag{13a}$$

$$\text{s. t.:} \sum_{j\in\mathcal{N}} \frac{r_{i,j}^{(k)}}{B_i\omega_{i,j}^{(k)}} \leq \rho_i^{(k)} \qquad i\in\mathcal{M} \tag{13b}$$

$$\sum_{l=1}^{k}\sum_{i\in\mathcal{M}}(r_{i,j}^{(l)} - r_j^{\min}) \geq 0 \qquad j\in\mathcal{N}. \tag{13c}$$

Problem 12 and 13 are still not easy to solve because we need to *minimize* a non-convex function over a convex set. Fortunately, [8] presents an optimization framework based on the majorization-minimization (MM) technique [9] to handle problems of this type. The framework can be used to decrease the value of the objective function in a computationally efficient way. For notational convenience we define the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ to be the set of rate allocations $\mathbf{R} \in \mathbb{R}_+^{N\times M\times K}$ and cell loads $\mathbf{P} \in \mathbb{R}_+^{M\times K}$ satisfying constraints 12b-12c and 13b-13c, respectively. In addition, we define the constant $\hat{e}_i := \frac{e_i}{\log(1+\epsilon^{-1})}$. Applying the MM technique to Problem 12 and 13 yields a fast algorithm that iteratively solves

$$\begin{aligned}&[(\mathbf{R},\mathbf{P})]^{[n+1]} \in \arg\min_{(\mathbf{R},\mathbf{P})\in\mathcal{X}_1} \\ &\quad \sum_{i\in\mathcal{M}}\left(K\ \hat{e}_i\frac{\boldsymbol{\rho}_i\mathbf{1}}{\epsilon+[\boldsymbol{\rho}_i]^{[n]}\mathbf{1}} + \nabla f_i\left([\boldsymbol{\rho}_i]^{[n]}\right)^T\boldsymbol{\rho}_i\right)\end{aligned} \tag{14}$$

for some feasible starting point and where we used the notation $[\cdot]^{[n]}$ to refer to the respective variable in the n-th iteration of the MM algorithm.

Analogous to (14) we arrive at the following iterative

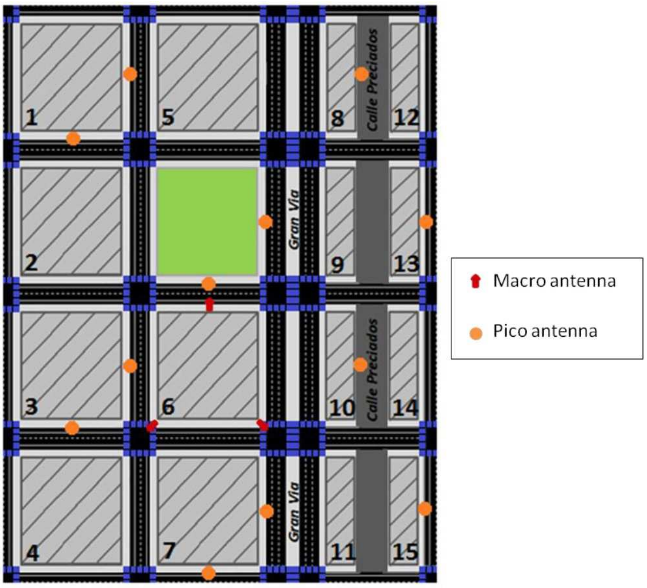

Fig. 2. Network topology. Deployment model for METIS TC2 [10].

algorithm for Problem 13

$$[(\mathbf{R},\mathbf{P})]^{[n+1]} \in \arg \min_{(\mathbf{R},\mathbf{P})\in\mathcal{X}_2} \sum_{i\in\mathcal{M}} \left(\sum_{k=1}^{K} \hat{e}_i \frac{\rho_i^{(k)}}{\epsilon + [\rho_i^{(k)}]^{[n]}} + \nabla f_i \left([\boldsymbol{\rho}_i]^{[n]} \right)^T \boldsymbol{\rho}_i \right) \quad (15)$$

for some feasible starting point.

V. Empirical Evaluation

In this section we present simulation results of the presented buffered delay-sensitive optimization framework for the *on/off scheme* and the *micro sleep scheme*. The numerical evaluation is based on the macro and micro layers of the ”dense urban information society” scenario proposed in the METIS project [10, Sect. 4.2]. The basic layout is depicted in Figure 2 where three macro and 12 pico cells are deployed on rooftops and in the streets, respectively. The main parameters of the simulation scenario are listed in Table I. To obtain the information used in Assumption 1 we use the channel gain data provided by the METIS consortium [11] for the scenario depicted in Figure 2. Additionally, the provided mobility traces for cars are used to generate the traffic in our simulations. We focus on car users only due to their high mobility which will serve well to illustrate the gains achievable with anticipatory resource allocation. The data set for car users provides mobility traces of $N = 420$ different car users in the central deployment and a wrap around model is used to avoid boundary effects. The data set provides mobility traces for a time duration of 3600 seconds for each car user. Further details on the scenario and mobility model can be found in [10].

We normalize the energy consumption achievable by each scheme to the energy consumption of the network topology where all cells are active the whole time. In more detail, we normalize the network energy consumption by $E_{\text{on/off}}(\mathbf{1}) = E_{\text{sleep}}(\mathbf{1}) = K\left(\sum_{i\in\mathcal{M}} e_i + f_i(\mathbf{1})\right)$ and evaluate for our schemes

$$E'_{\text{on/off}}(\mathbf{R}) = \frac{E_{\text{on/off}}(\mathbf{R})}{K\left(\sum_{i\in\mathcal{M}} e_i + f_i(\mathbf{1})\right)} \quad (16)$$

and

$$E'_{\text{sleep}}(\mathbf{R}) = \frac{E_{\text{micro sleep}}(\mathbf{R})}{K\left(\sum_{i\in\mathcal{M}} e_i + f_i(\mathbf{1})\right)}. \quad (17)$$

Table I. Network parameters of the simulation [10, Sect. 4.2].

Parameter	Value
Antenna height of macro cells	53 m
Antenna height of micro cells	10 m
Carrier frequency of macro cells	800 MHz
Carrier frequency of micro cells	2500 MHz
Max. transmit power of macro cells	43 dBm
Max. transmit power of micro cells	30 dBm
e_i of macro cells	400 W
e_i of micro cells	100 W
B_i	100
Noise power spectral density	-145.1 dBm/Hz
ϵ	10^{-2}

In order to show the isolated effective gains from the anticipatory scheduling framework we neglect the dynamic energy consumption of hardware and use $f_i(\boldsymbol{\rho}) = 0,\ \forall i$. The evaluation of the effect of the dynamic energy consumption on the energy savings has been studied in [7] for energy saving algorithms without proactive resource allocation. We expect the effects to carry over to proactive resource allocation schemes and the evaluation of which will be part of future studies.

We evaluate the proposed algorithm for the *on/off scheme* and the *micro sleep scheme* in terms of energy savings capabilities for the described scenario focusing on the influence of the number of time slots the optimization is done for. More precise, we find touples of $(\mathbf{R},\mathbf{P})$ for a different number of time slots $K \in \{40, 80, 120, 240\}$ with time slot duration $\Delta_k = 1s$. The starting point k_0 in the data set where we apply our optimization framework is selected uniformly at random from $\{1, 3600 - K + 1\}$ and the optimization window is selected as $\{k_0, \ldots, k_0 + K - 1\}$.

Figure 3 depicts the achievable normalized network energy consumption with increasing per user data rate requirement for the *on/off scheme* and the *micro sleep scheme*. We observe the intuitive result that with an increasing minimum per user data rate requirement the normalized network energy consumption increases. The more user demand there is in the network the less redundancies are in the system that can be exploited for energy savings. When comparing the two energy savings schemes it can be seen that the energy

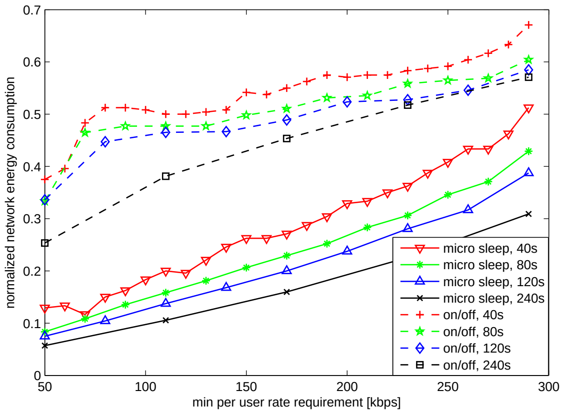

Fig. 3. Normalized network energy consumption of the *micro sleep scheme* and the *on/off scheme* with increasing minimum per user rate requirement and for different optimization windows. Normalization with respect to all cells active the whole time.

savings potentials with the *micro sleep scheme* are larger than the ones with the *on/off scheme*. The reason lies in the nature of the schemes that the *micro sleep scheme* allows the deactivation and activation of cells on a smaller time scale which in turn enables the scheme to save energy even for shorter time periods whereas the *on/off scheme* finds a set of active cells only for the full optimization window. The effect of the size of the optimization window is also evident from Figure 3. We can see that for both schemes the normalized network energy consumption is higher for a smaller optimization window. A reason for this observation can be found in the better chances of larger optimization windows to preallocate resources for more data transmission in advance in order to free some cells from service provision.

VI. CONCLUSION

We have presented an optimization framework that exploits the knowledge of user-cell trajectories and learned path-loss maps to find network topologies with low energy consumption. It finds user-cell association and rate schedules that provide the requested data rate of users and at the same time reduces the energy consumption of cellular communication networks. The used energy consumption is general enough to capture static energy consumption for cooling, basic power conversion etc. as well as dynamic load dependent energy consumption. We exploit the end user devices' storage capabilities to implement buffered delay-sensitive applications with proactive resource allocation and user assignment. Thereby, we stretch the applicability of cell sleep and switching on/off techniques in the time horizon leading to significant energy savings. We have formalized the problem as a non-convex optimization problem and have presented relaxation techniques that are able to give good solutions to this problem in reasonable time making it amenable for online implementation.

We have evaluated the proposed *on/off scheme* and the *micro sleep scheme* in a realistic network deployment with realistic traffic demands and user mobility models. Results for the METIS TC2 deployment show good energy saving potentials for both schemes. Even though the *micro sleep scheme* yields larger energy savings compared to the *on/off scheme*, such a mode of operation might be limiting due to the potential fast switching on and off of individual cells. If such short *micro sleeps* of cells are not permitted then the *on/off scheme* presents a good way to save energy by proactively allocating resources to users. Furthermore, the simulation results show that larger optimization windows lead to larger energy savings.

Acknowledgements

This work has been partly supported by the framework of the research project ComGreen under the grant-number 01ME11010, which is funded by the German Federal Ministry of Economics and Technology (BMWi).

Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

VII. REFERENCES

[1] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[2] M. Kasparick, R. Cavalcante, S. Valentin, S. Stanczak, and M. Yukawa, "Kernel-based adaptive online reconstruction of coverage maps with side information," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[3] H. Abou-zeid, H. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, Dec 2013, pp. 4877–4882.

[4] ——, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," in *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, Jun 2014, pp. 2013–2026.

[5] A. Galanopoulos, G. Iosifidis, A. Argyriou, and L. Tassiulas, "Green video delivery in LTE-based heterogeneous cellular networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, June 2015, pp. 1–9.

[6] D. MacKay, "Fountain codes," *Communications, IEE Proceedings-*, vol. 152, no. 6, pp. 1062–1068, Dec 2005.

[7] E. Pollakis, R. Cavalcante, and S. Stanczak, "Traffic demand-aware topology control for enhanced energy-efficiency of cellular networks,"

EURASIP Journal on Wireless Communications and Networking, 2016, accepted. Preprint available at http://arxiv.org/abs/1503.08627.

[8] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[9] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[10] METIS Project, "Deliverable D6.1: Simulation guidelines," METIS Project (Mobile and wireless communications Enablers for the Twenty-twenty Information Society), Tech. Rep. ICT-317669-METIS/D6.1, 2013.

[11] METIS ray tracing files. Accessed on 20.03.2015. [Online]. Available: https://www.metis2020.com/documents/simulations/