Energy Minimization in HARQ-I Relay-Assisted Networks with Delay-limited Users

Mohamad Maaz, Philippe Mary, Maryline Hélard

HAL Id: hal-01492991
https://hal.science/hal-01492991

Energy Minimization in HARQ-I Relay-Assisted Networks with Delay-Limited Users

Mohamad Maaz, Philippe Mary, *Member, IEEE,* and Maryline Hélard, *Member, IEEE*

Abstract—This paper investigates a practical energy minimization problem for multi-user relay-assisted downlink cellular networks. The system adopts an hybrid-automatic-repeat-request of type I (HARQ-I) protocol and each user has an average delay constraint to be satisfied under a total power constraint for the system. The contribution of this paper lies in several folds: i) By introducing new constraints and new variables in the original problem, the solution is analytically obtained by the dual method. ii) An algorithm that jointly allocates for every user the optimal downlink power at base station (BS) and relay station (RS) and selects the RS (if cooperation is decided) is proposed. iii) Another binary integer programming algorithm that selects the optimal modulation and coding scheme (MCS) and optimizes the number of time-frequency resource blocks (RBs) between users is proposed. In order to corroborate our findings, simulation results in terms of energy-delay tradeoff, average starvation user rates and selected MCS statistics are presented. Our analyses show that relay-assisted strategy improves up to 30 % the energy consumption compared to non-relay-assisted one.

Index Terms—Energy minimization, power allocation, relay selection, HARQ, MCS, convex optimization, integer programming.

I. INTRODUCTION

ENERGY consumption and quality of service (QoS) are two important metrics which will play a key role for evaluating the performance of green wireless cellular networks [1], [2]. Fundamentally, the energy consumption is time and power dependent. Therefore, in non-reliable communications, i.e. non-zero packet error rate (PER), the delay induced by the retransmission mechanism of erroneous packets has an impact on the energy consumption. Moreover, it has been shown that introducing intermediate relays between two communicating nodes enables to transmit over short distances and may enhance the QoS [3]. Hence, jointly exploiting the spatial diversity enabled by relaying schemes [4] and the temporal diversity offered by hybrid-automatic repeat request (HARQ) mechanisms leads to a substantial energy saving. Thus, designing resource allocation strategies that jointly optimize the available physical resources, e.g. power, and considering HARQ mechanisms in relay-assisted networks could be energy efficient for up-to-date and future wireless networks.

M. Maaz is with Orange Gardens (Orange Labs), 44 Avenue de la République, 92320 Châtillon, France. e-mail: mohamad.maaz@orange.com
P. Mary and M. Hélard are with INSA Rennes, IETR UMR 6164, 35708 Rennes, France. Email: firstname.lastname@insa-rennes.fr

Traditional resource allocation policies focusing on the throughput maximization are generally not energy-aware [2], [5]. Moreover, most of the green wireless strategies focus on energy minimization or efficiency under the classical Shannon theory framework, i.e. by considering reliable communications or theoretical outage capacity [6]. Thus, in multi-user (MU) context, minimizing the energy consumption under non reliable communication, i.e. non zero PER, is a challenging task for the scientific community in order to enhance future cellular networks. Moreover, relay-assisted communication has already been identified as a promising technology to provide higher coverage and data-rate for high QoS wireless services [7].

In [8] the authors studied the energy efficiency (EE) for cooperative and non cooperative HARQ schemes by minimizing the energy consumed for a single user with outage probability constraint. In [9], the authors studied the energy efficiency for decode-and-forward (DF), amplify-and-forward (AF) and compress-and-forward (CF) relaying schemes with HARQ under outage probability and delay constraints. Moreover, some recent works have been granted in resource allocation for HARQ-based systems in [10], [11]. The authors have investigated some power-rate allocation strategies for cooperative transmissions in source-relay-destination configurations. In [11], the authors have been interested in the context of cognitive radio, where secondary users (SU) may help primary users (PU) as relays when PUs retransmit the same data packet in HARQ mode. However, they have considered neither energy-efficiency performances nor multi-users resource allocation with delay constraints. It is also convenient to note that an important amount of works in wireless sensor networks (WSN) considers that the energy consumption is of crucial importance, e.g. [12] and [13], [14] for recent contributions on this topic. However, all works above did not consider jointly the resource allocation strategy in MU cellular context with HARQ schemes under delay constraint and non-reliable communications.

In our previous works [15], [16], EE in relay-assisted HARQ scheme have been analyzed in a single user context without considering any resource optimization problem. In particular in [16], we have been interested in the EE analysis for HARQ schemes in single user source-relay-destination framework, when a practical FPGA target is considered. However, no resource allocation problems for optimizing throughput and/or power have been formulated nor solved. The main challenge comes from the need of a closed-form expression of PER with HARQ in order to express the delay requirements [17]. In this work, we propose a new resource allocation algorithm for MU relay-assisted HARQ-I (i.e. considering coded packets

[17]) scheme that aims at allocating powers at base station (BS) and relay stations (RS) and selects the optimal relay that minimizes the energy consumption while satisfying the delay constraint for every user and respecting the total power constraint of the system. In this work, we formulate the energy minimization problem by introducing intermediate variables that enable to efficiently solve this problem thanks to the Lagrange dual optimization. Hence, an energy minimization algorithm that distributes the optimal powers between users is derived. Finally, a binary-integer programming algorithm is proposed aiming to select the optimal modulation and coding scheme (MCS) and number of resource blocks (RB) for every user minimizing the total energy consumed. To the best of our knowledge, this is the first work that jointly considers an HARQ scheme and MCS in multiuser relay-assisted network in order to minimize the overall energy for delay constrained users.

The remaining of the paper is organized as follows: the system model and hypothesis are stated in Section II. Section III presents the proposed energy minimization problem. Section IV deals with the solution of the optimization problem: optimal power, delays at BS and RS and the relay selection strategy. In Section V, the proposed energy minimization algorithm is described. In section VI, a jointly MCS and resource blocks allocation algorithm is proposed while Section VII deals with numerical results. Finally, a conclusion is drawn in Section VIII.

II. SYSTEM DESCRIPTION

A. System model

We consider a downlink multiuser relay-assisted network consisting of a BS, a set of relays $\mathcal{K} = \{1, \cdots K\}$ deployed around the BS and a set of randomly distributed users $\mathcal{M} = \{1, \cdots M\}$ to be served simultaneously. The system is considered adopting the orthogonal frequency division multiple access (OFDMA) technique with N_{RB} RBs to be shared among users. Considering also an HARQ based system, the communication protocol between BS and every user is described as follows:

BS transmits a packet to the m-th user u_m and the k-th selected relay r_k is hearing the transmission. If u_m decodes/does not decode the packet, u_m sends an acknowledgment (ACK) /non-ACK (NACK) to BS, respectively. The ACK/NACK packets are assumed to be received error-free with a negligible delay which is a reasonable assumption considering the low rate of these packets [18]. If r_k successfully decodes and u_m does not, r_k sends an ACK to BS and starts retransmitting the packet until u_m decodes it. If neither r_k nor u_m decodes the packet, BS keeps transmitting the same packet until r_k or u_m receives it.

According to the data packet size and required QoS, the resource scheduler at BS has to decide for every user whether direct or relay-assisted transmission is preferable in order to minimize the energy consumption. Moreover, this latter also depends on MCS and the number of time and frequency RBs. Hence, let us consider that each RB consists of N_s subcarriers and has a duration of N_F OFDMA symbols.
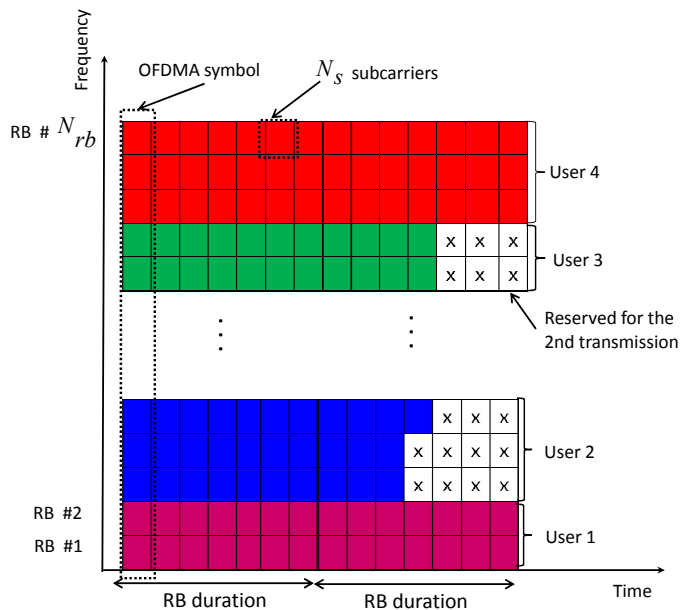

Fig. 1: Resource blocks (RB) sharing scheme between users.

Moreover each symbol has a duration of T_{OFDM} and hence the subcarrier spacing is $\Delta f = 1/T_{OFDM}$. Fig. 1 illustrates the RB sharing process between users. Let us consider that each subcarrier is affected by a centered additive-white Gaussian noise (AWGN) $n \sim \mathcal{CN}(0, \sigma^2)$. Moreover, the small scale channel fading coefficients are independent and identically Rayleigh distributed for different RBs and users. Let us assume that the average received SNR is $\overline{\gamma} = \mathbb{E}[\gamma] = \frac{g}{\sigma^2} p$ where p is the transmitted power and g is the channel gain which is pathloss dependent. Therefore, for a given MCS, the average PER (in fading channel) could be tightly approximated as follows [19]:

$$\Pi(\overline{\gamma}) = \left(1 - \exp\left(-a\overline{\gamma}^b\right)\right)^c \qquad (1)$$

where a, b, and c are curve fitting parameters which depend on the adopted MCS and are given in table I. These parameters have been estimated based on the turbo coding scheme that is adopted in LTE [20].

Let us assume that for user m, the scheduler has to select a certain QAM modulation order Q_m and a coding rate R_m. Therefore, according to the packet length L, the number of required frequency or/and time RBs to be allocated to user m is summarized in Table I and given by

$$\text{RB}_m = \left\lceil \frac{L}{R_m \log_2(Q_m) N_s} \right\rceil. \qquad (2)$$

Moreover, let us assume that user m has a delay constraint $\overline{D}_m$, in time slots, to be satisfied, where a time slot is the duration of one resource block. The possible number of RBs to be allocated to user m varies between 1 and RB_m. Thus, as the number of allocated RBs increases, the delay constraint becomes less stringent and varies between $\overline{D}_m$ and $\text{RB}_m \overline{D}_m$ due to the possibility of multiplexing in frequency. Therefore, for every possible combination of MCS and every selected number of RBs, the scheduler has to allocate a transmission power such that the delay constraint is guaranteed.

TABLE I: Curve fitting parameters a, b and c for each MCS

MCS index	QAM	Coding rate $\times 1024$	a	b	c	# of RBs
1	4	449	17.76	-1.90	4.25	15
2	4	602	17.69	-1.50	3.98	12
3	16	378	69.94	-1.64	3.65	9
4	16	490	43.34	-1.27	3.85	7
5	16	616	50.91	-1.22	2.98	6
6	64	466	76.76	-1.04	3.98	5
7	64	567	125.87	-1.16	2.49	4
8	64	666	174.46	-1.20	1.75	4
9	64	772	232.16	-1.30	0.97	3

B. Delay model

Let us consider that correct detection events at the instants n and $n-1$ are independent at each node. Moreover, we assume that the number of allowed retransmissions, $N_{\max}$, is large enough to neglect the packet loss and hence can be considered as infinite [21]. Let us define $\overline{N}_{0m}^{d}$ as the average delay per received packet between BS and user m in case of direct transmission. The subscripts '0' and 'm' stand respectively for BS and user m while the superscript 'd' stands for direct transmission. By definition of the expectation of a discrete random variable, the average delay per received packet is $\overline{N}_{0m}^{d} = \sum_{n=1}^{+\infty} n q_{0m}^{d}(n)$ [15], [21], where $q_{0m}^{d}(n)$ is the successful decoding probability at instant n which is the probability that u_m does not decode at instants $1, 2, \cdots, n-1$ but decodes at n. Hence, $q_{0m}^{d}(n) = \left(\Pi_{0m}^{d}\right)^{n-1}\left(1 - \Pi_{0m}^{d}\right)$ and using [22, eq. (2) §0.231, pp. 8], we obtain:

$$\overline{N}_{0m}^{d} = \frac{1}{1 - \Pi_{0m}^{d}(\overline{\gamma}_{0m}^{d})} \qquad (3)$$

where $\overline{\gamma}_{0m}^{d} = g_{0m}\frac{p_{0m}^{d}}{\sigma^2}$, g_{0m}, p_{0m}^{d} and Π_{0m}^{d} are respectively the average received SNR, the channel gain, the allocated power and the PER in the $BS - u_m$ link. The average PER Π_{0m}^{d} has a similar form as in (1) and it depends on the curve fitting parameters a, b and c that are provided in Table I for every MCS.

Let us consider that relay r_k has been selected to help user u_m. The allocated powers at BS and r_k are respectively p_{0km}^{c} and p_{km}^{c}. The average channel gains in $BS - u_m$, $BS - r_k$ and $r_k - u_m$ links are respectively g_{0m}^{c}, g_{0km}^{c} and g_{km}^{c}. Therefore, under cooperation between BS and r_k, the average delay induced when BS is transmitting is $\overline{N}_{0km}^{c} = \sum_{n=1}^{+\infty} n q_{0km}^{c}(n)$ with $q_{0km}^{c}(n)$ being the successful decoding probability at u_m or r_k at the instant n due to BS transmissions. The correct (or not) decoding events at u_m and r_k are independent, hence the error probability at any instant is $\Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})$, with $\overline{\gamma}_{0m}^{c} = g_{0m}^{c}\frac{p_{0km}^{c}}{\sigma^2}$ and $\overline{\gamma}_{0km}^{c} = g_{0km}^{c}\frac{p_{0km}^{c}}{\sigma^2}$. Moreover, the error events are independent from one time slot to another; $q_{0km}^{c}(n)$ is hence given by $q_{0km}^{c}(n) = (1 - \Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c}))(\Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c}))^{n-1}$ and using [22, eq. (2) §0.231, pp. 8], we obtain:

$$\overline{N}_{0km}^{c} = \frac{1}{1 - \Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})}. \qquad (4)$$

If at instant n', user u_m has not received the packet but relay r_k did, BS stops transmitting the packet, while r_k continues to forward the packet to u_m. Thus, the average delay induced when r_k transmits is

$\overline{N}_{km}^{c} = \sum_{n'=1}^{\infty} q_{0k}^{c}(n') \times \sum_{n=n'+1}^{\infty}(n - n')q_{km}^{c}(n)$ where $q_{0k}^{c}(n')$ is the successful decoding probability at r_k at n' and considering that u_m has not decoded. Moreover, $q_{km}^{c}(n)$ is the successful decoding probability at u_m due to r_k transmissions at the instant $n > n'$. Hence, $q_{0k}^{c}(n') = \Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})^{n'}\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})^{(n'-1)}\left(1 - \Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})\right)$ and $q_{km}^{c}(n) = \Pi_{km}^{c}(\overline{\gamma}_{km}^{c})^{(n-n')}(1 - \Pi_{km}^{c}(\overline{\gamma}_{km}^{c}))$, with $\overline{\gamma}_{km}^{c} = g_{km}^{c}\frac{p_{km}^{c}}{\sigma^2}$ being the average SNR in the $r_k - u_m$ link. Using [22, eq. (2) §0.231, pp. 8], the average delay when r_k transmits is:

$$\overline{N}_{km}^{c} = \frac{\Pi_{om}^{c}(\overline{\gamma}_{0m}^{c})\left(1 - \Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})\right)}{1 - \Pi_{0m}^{c}(\overline{\gamma}_{0m}^{c})\Pi_{0km}^{c}(\overline{\gamma}_{0km}^{c})}\frac{1}{1 - \Pi_{km}^{c}(\overline{\gamma}_{km}^{c})} \qquad (5)$$

Hence, in case of cooperation with relay r_k, the average delay per successfully received packet at u_m is:

$$\overline{N}_{km} = \overline{N}_{0km}^{c} + \overline{N}_{km}^{c} \qquad (6)$$

C. Energy Consumption model

Let us assume that the static power consumed by circuits for transmitting or receiving a modulated signal at BS, RS and UE are respectively denoted by P_{bs}, P_{rs} and P_{ue}. Let us also denote by E_{tx}^{i} and E_{rx}^{i}, the energy consumed for transmitting and receiving a packet at $i \in \{bs, rs, ue\}$ respectively. Hence, in case of direct transmission with u_m, the consumed energy for transmitting one packet is [12]:

$$f_1(p_{0m}^{d}) = \mathrm{E}_{tx}^{bs} + \mathrm{E}_{rx}^{ue} + \mathrm{E}_{ack} \qquad (7)$$

with $\mathrm{E}_{tx}^{bs} = \eta_m\left(P_{bs} + \beta_{amp}p_{0m}^{d}\right)$ and $\mathrm{E}_{rx}^{ue} = \eta_m P_{ue}$, where $\eta_m = L/(\Delta f \log_2(Q_m)R_m)$ depends on the packet length L, the sub-carrier spacing Δf and the selected MCS that is represented by the modulation order Q_m and the coding rate R_m. The factor β_{amp} stands for the power amplifier (PA) inefficiency factor, i.e. the larger β_{amp}, the larger power consumption due to PA. Moreover, $\mathrm{E}_{ack} = \tau(\mathrm{E}_{tx}^{bs} + \mathrm{E}_{rx}^{ue})$ is the energy consumed for the N/ACK, with τ being the ratio of the number of transmitted bits for ACK and the number of transmitted bits in the data packet. Therefore, $f_1(p_{0m}^{d})$ can be re-written as:

$$f_1(p_{0m}^{d}) = k_1 + k_2 p_{0m}^{d} \qquad (8)$$

with $k_1 = \eta_m(1 + \tau)(P_{bs} + P_{ue})$ and $k_2 = \eta_m(1 + \tau)\beta_{amp}$. Therefore, the average energy consumed per received packet is:

$$\overline{E}_{0m}^{d} = f_1(p_{0m}^{d})\overline{N}_{0m}^{d} \qquad (9)$$

In case of cooperation between BS and r_k that assists the communication with user u_m, the relay r_k and u_m receive a copy of the transmitted packet from BS and then they transmit an N/ACK to BS. The consumed energy when BS transmits one packet can be expressed as:

$$f_2(p_{0km}^{c}) = (1 + 2\tau)E_{tx}^{bs} + (1 + \tau)(E_{rx}^{rs} + E_{rx}^{ue}) \qquad (10)$$

Hence, the consumed energy when BS transmits can be written as:

$$f_2(p_{0km}^{c}) = k_3 + k_4 p_{0km}^{c} \qquad (11)$$

with $k_3 = \eta_m\left((1 + 2\tau)P_{bs} + (1 + \tau)(P_{rs} + P_{ue})\right)$ and $k_4 = \eta_m(1 + 2\tau)\beta_{amp}$.

$$\sum_{m=1}^{M} \left\{ \frac{\left(1 - \sum_{k=1}^{K} s_{km}\right) f_1(p_{0m}^d)}{1 - \Pi_{0m}^d(p_{0m}^d)} + \sum_{k=1}^{K} s_{km} \left(\frac{f_2(p_{0km}^c)}{1 - \Pi_{0m}^c(p_{0km}^c)\Pi_{0km}^c(p_{0km}^c)} + \right. \right.$$
$$\left. \left. \frac{\Pi_{0m}^c(p_{0km}^c)(1 - \Pi_{0km}^c(p_{0km}^c))}{1 - \Pi_{0m}^c(p_{0km}^c)\Pi_{0km}^c(p_{0km}^c)} \frac{f_3(p_{km}^c)}{1 - \Pi_{km}^c(p_{km}^c)} \right) \right\} \quad (12)$$

Furthermore, when r_k successfully decodes the packet received from BS, it starts retransmitting the packet to u_m. Then N/ACK are sent from u_m to r_k and BS and the consumed energy when r_k is transmitting is:

$$f_3(p_{km}^c) = E_{tx}^{rs} + E_{rx}^{ue} + \tau \left(E_{tx}^{rs} + E_{rx}^{ue} + E_{rx}^{bs} \right) \quad (13)$$
$$= k_5 + k_6 p_{km}^c \quad (14)$$

where $k_5 = \eta_m (\tau P_{bs} + (1 + \tau)(P_{rs} + P_{ue}))$ and $k_6 = \eta_m(1 + \tau)\beta_{amp}$. Hence, the consumed energy per received packet in relay-assisted mode is:

$$\overline{E}_{km}^c = f_2(p_{0km}^c)\overline{N}_{0km}^c + f_3(p_{km}^c)\overline{N}_{km}^c \quad (15)$$

III. PROBLEM FORMULATION

The objective function of the original problem is defined in (12) on the top of the page and has constraints (c_1), (c_2), (c_6) and (c_7) defined in (16) on the next page. The expression in (12) is the overall energy consumed in the cell where the left most term is the energy consumed when a group of users has been selected to communicate directly with BS, i.e. $\sum_{k=1}^{K} s_{km} = 0$. The binary variable s_{km} is equal to 1, if the relay r_k is selected to help the user u_m and zero otherwise. The right most term is the energy consumed in relay-assisted mode, i.e $s_{km} = 1$. The constraint (c_1) ensures that the average delay is less than or equal to $\overline{D}_m$ for every user u_m. (c_2) ensures that only one relay can be selected per user. The constraint (c_6) states that the total allocated power (for BS and relays) is less than or equal to the power constraint P_{tot}. (c_7) ensures the positivity of allocated powers. The solution of this problem over the variables p_{0m}^d, p_{0km}^c, p_{km}^c and the integer variable s_{km} is very challenging due to the combination of non convex functions and binary constraints that make the problem intractable. In order to overcome the difficulty implied by this non linear problem, we consider that the delays $\overline{N}_{0m}^d$, $\overline{N}_{0km}^c$ and $\overline{N}_{km}^c$ $\forall k, m$ are also variables satisfying the new equality constraints in (c_3), (c_4) and (c_5) respectively. Hence, the energy minimization problem under average delay and power constraints can be formulated as in (16). One can notice that problem (12) is equivalent to (16) by just substituting the equality constraints (c_3), (c_4) and (c_5) in the objective function (16). Solving (16) is equivalent to find the optimal vector $\boldsymbol{\lambda_{km}} = \{\lambda_{km} : (k, m) \in \mathcal{K} \times \mathcal{M}\}$ with $\lambda_{km} = [s_{km} \; p_{0m}^d \; p_{0km}^c \; p_{km}^c \; \overline{N}_{0m}^d \; \overline{N}_{0km}^c \; \overline{N}_{km}^c]$. This problem is a mixed combinatorial and continuous optimization problem, which is still very difficult to solve in this form. We first relax the integer constraint on s_{km}, assuming it could be a time sharing factor between 0 and 1, leading to

$\sum_{k=1}^{K} s_{km} \le 1 \; \forall m = \{1, \cdots, M\}$ [23], however the problem in (16) is still not convex.

The Lagrangian $\mathcal{L}$ associated with this problem is expressed in (19), where the PER dependency on transmit powers has been removed in order not to clutter the notation. We refer to the vectors $\boldsymbol{\gamma_m} = [\gamma_1, \cdots, \gamma_M]$, $\boldsymbol{\phi_m} = [\phi_1, \cdots, \phi_M]$, $\boldsymbol{\alpha_{km}} = \{(\alpha_{0m}^d, \alpha_{0km}^c, \alpha_{km}^c) : (k, m) \in \mathcal{K} \times \mathcal{M}\}$ and μ as the Lagrange multipliers respectively associated to the constraints from (c_1) to (c_6), $\mathcal{K}$ and $\mathcal{M}$ stand for the sets of relays and users respectively. A lower bound of the original minimization problem can be found by solving the Lagrange dual problem. The dual $\mathcal{G}$ associated with the primal problem in (16) is defined as:

$$\mathcal{G}(\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{km}}, \boldsymbol{\phi_m}) = \inf_{\boldsymbol{\lambda_{km}}} \mathcal{L}(\boldsymbol{\lambda_{km}}, \boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{km}}, \boldsymbol{\phi_m}) \quad (17)$$

Since the dual function $\mathcal{G}$ is an infimum of a family of affine functions of dual variables $(\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{km}}, \boldsymbol{\phi_m})$, then it is concave even if the problem is not convex [24] and yields to a lower bound of the primal problem for any $\boldsymbol{\gamma_m}$ and $\mu \ge 0$. Therefore, the dual optimization problem of (16) can be expressed as follows:

$$\max_{\boldsymbol{\gamma_m}, \; \mu, \; \boldsymbol{\alpha_{km}}, \; \boldsymbol{\phi_m}} \mathcal{G}(\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{km}}, \boldsymbol{\phi_m}) \quad (18)$$
$$\text{subject to:} \quad \boldsymbol{\gamma_m}, \; \mu \ge 0$$

The primal problem in (16) cannot be proved to be strictly convex for all possible combinations of users and relays. In that sense, solving the dual problem leads to a lower bound of the primal without a zero optimality gap. However, a lower bound remains very interesting since no closed form can be obtained from the primal problem due to the complexity of expressions involved.

IV. PROBLEM SOLUTION

According to (18), it is convenient to firstly find the infimum of the Lagrangian function $\mathcal{L}$ in (19) over the set of primal variables $\boldsymbol{\lambda_{km}}$ and then find the supremum of $\mathcal{G}$ over the dual variables. Thus, the primal variables can be expressed as function of dual variables as follows:

A. Power allocation

Applying the gradient operator over the Lagrangian function $\mathcal{L}$ w.r.t $\overline{N}_{0m}^d$, we get the following:

$$\nabla_{\overline{N}_{0m}^d} \mathcal{L} = \left(1 - \sum_{k=1}^{K} s_{km}\right) \left(f_1\left(p_{0m}^d\right) + \gamma_m\right) - \alpha_{0m}^d \quad (20)$$

$$\min_{\lambda_{km}} \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{km}\right) f_1(p_{0m}^d) \overline{N}_{0m}^d + \sum_{k=1}^{K} s_{km} \left(f_2(p_{0km}^c) \overline{N}_{0km}^c + f_3(p_{km}^c) \overline{N}_{km}^c \right) \right)$$

subject to:

$$(c_1) \quad \left(1 - \sum_{k=1}^{K} s_{km}\right) \overline{N}_{0m}^d + \sum_{k=1}^{K} s_{km} \left(\overline{N}_{0km}^c + \overline{N}_{km}^c\right) \leq \overline{D}_m \forall m \in \{1, \cdots, M\}$$

$$(c_2) \quad s_{km} = \{0,1\} \ \forall (k,m) \in \{1, \cdots, K\} \times \{1, \cdots, M\} \text{ and } \sum_{k=1}^{K} s_{km} = \{0,1\} \ \forall m \in \{1, \cdots, M\}$$

$$(c_3) \quad \overline{N}_{0m}^d = \frac{1}{1 - \Pi_{0m}^d(p_{0m}^d)} \ \forall m \in \{1, \cdots, M\} \tag{16}$$

$$(c_4) \quad \overline{N}_{0km}^c = \frac{1}{1 - \Pi_{0m}^c(p_{0km}^c)\Pi_{0km}^c(p_{0km}^c)} \ \forall (k,m) \in \{1, \cdots, K\} \times \{1, \cdots, M\}$$

$$(c_5) \quad \overline{N}_{km}^c = \frac{\Pi_{0m}^c(p_{0km}^c)(1 - \Pi_{0km}^c(p_{0km}^c))}{(1 - \Pi_{0m}^c(p_{0km}^c)\Pi_{0km}^c(p_{0km}^c))(1 - \Pi_{km}^c(p_{km}^c))} \ \forall (k,m) \in \{1, \cdots, K\} \times \{1, \cdots, M\}$$

$$(c_6) \quad \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{km}\right) p_{0m}^d + \sum_{k=1}^{K} s_{km}(p_{0km}^c + p_{km}^c) \right) \leq P_{tot}$$

$$(c_7) \quad p_{0m}^d, \ p_{0km}^c, \ p_{km}^c \geq 0$$

$$\mathcal{L} = \sum_{m=1}^{M} \sum_{k=1}^{K} \left(s_{km} \left((f_2(p_{0km}^c) + \gamma_m) \overline{N}_{0km}^c + (f_3(p_{km}^c) + \gamma_m) \overline{N}_{km}^c - (f_1(p_{0m}^d) + \gamma_m) \overline{N}_{0m}^d \right) \right.$$

$$\left. + \alpha_{0km}^c \left(\frac{1}{1 - \Pi_{0m}^c \Pi_{0km}^c} - \overline{N}_{0km}^c \right) + \alpha_{km}^c \left(\frac{\Pi_{0m}^c(1 - \Pi_{0km}^c)}{(1 - \Pi_{0m}^c \Pi_{0km}^c)(1 - \Pi_{km}^c)} - \overline{N}_{km}^c \right) \right)$$

$$+ \sum_{m=1}^{M} \alpha_{0m}^d \left(\frac{1}{1 - \Pi_{0m}^d} - \overline{N}_{0m}^d \right) + \mu \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{km}\right) p_{0m}^d + \sum_{k=1}^{K} s_{km}(p_{0km}^c + p_{km}^c) \right) - \mu P_{tot}$$

$$+ \sum_{m=1}^{M} \phi_m \left(\sum_{k=1}^{K} s_{km} - 1 \right) + \sum_{m=1}^{M} \left((f_1(p_{0m}^d) + \gamma_m) \overline{N}_{0m}^d - \gamma_m \overline{D}_m \right) \tag{19}$$

which nullifies at the optimum delay, i.e. $\nabla_{\overline{N}_{0m}^{d*}} \mathcal{L} = 0$. Therefore, if $\sum_{k=1}^{K} s_{km} = 1$, i.e $\exists (k,m)$ such that $s_{km} = 1$, then $\alpha_{0m}^d = 0$. Moreover if $\sum_{k=1}^{K} s_{km} = 0$, i.e there is no relay that minimizes the overall energy consumed when serving user m, then a direct transmission is preferable and the optimal allocated power p_{0m}^d for direct transmission is given by:

$$p_{0m}^d = \left[\frac{(\alpha_{0m}^d - \gamma_m - k_1)}{k_2} \right]^+ \tag{21}$$

where $[.]^+ \triangleq \max\{0, .\}$.

Moreover, optimizing respectively over the variables $\overline{N}_{0km}^c$ and $\overline{N}_{km}^c$, we get the following:

$$\nabla_{\overline{N}_{0km}^c} \mathcal{L} = s_{km} \left(f_2\left(p_{0km}^c\right) + \gamma_m \right) - \alpha_{0km}^c \tag{22}$$

$$\nabla_{\overline{N}_{km}^c} \mathcal{L} = s_{km} \left(f_3\left(p_{km}^c\right) + \gamma_m \right) - \alpha_{km}^c \tag{23}$$

and then at the optimum point $\nabla_{\overline{N}_{0km}^{c*}} \mathcal{L} = 0$ and $\nabla_{\overline{N}_{km}^{c*}} \mathcal{L} = 0$. Therefore, if $s_{km} = 0$, then $\alpha_{0km}^c = 0$ and $\alpha_{km}^c = 0$. Moreover, let us consider that a relay k is selected to help user

m, then $s_{km} = 1$. Hence, using (22) and (23), the allocated powers when BS or relay k communicating with user m are respectively p_{0km}^c and p_{km}^c and are given by:

$$p_{0km}^c = \left[\frac{(\alpha_{0km}^c - \gamma_m - k_3)}{k_4} \right]^+ \tag{24}$$

$$p_{km}^c = \left[\frac{(\alpha_{km}^c - \gamma_m - k_5)}{k_6} \right]^+ \tag{25}$$

Since, γ_m, k_1, k_3 and k_5 are positive, then $\alpha_{0m}^d \geq \gamma_m + k_1$, $\alpha_{0km}^c \geq \gamma_m + k_3$ and $\alpha_{km}^c \geq \gamma_m + k_5$.

B. Optimal delay

By deriving $\mathcal{L}$ w.r.t to the primal variable p_{0m}^d, we get the following:

$$\nabla_{p_{0m}^d} \mathcal{L} = \left(1 - \sum_{k=1}^{K} s_{km}\right) \left(k_2 \overline{N}_{0m}^d + \mu\right) + \alpha_{0m}^d \nabla_{p_{0m}^d} \mathcal{F}_{0m}^d \tag{26}$$

where $\mathcal{F}_{0m}^d = \frac{1}{1 - \Pi_{0m}^d}$ and at the optimum $\nabla_{p_{0m}^{d*}} \mathcal{L} = 0$. Since the constraint (c_3) has to be satisfied, the value of

$\overline{N}_{0m}^d$ is substituted in (26). Then, α_{0m}^{d*} is obtained by solving numerically $\nabla_{p_{0m}^{d*}} \mathcal{L} = 0$.

Moreover, by applying the gradient of $\mathcal{L}$ over p_{km}^c and p_{0km}^c, we obtain receptively:

$$\nabla_{p_{km}^c} \mathcal{L} = s_{km} \left(k_6 \overline{N}_{km}^c + \mu \right) + \alpha_{km}^c \mathcal{F}_1 \mathcal{F}_2 \nabla_{p_{km}^c} \mathcal{F}_3 \quad (27)$$

$$\nabla_{p_{0km}^c} \mathcal{L} = s_{km} \left(k_4 \overline{N}_{0km}^c + \mu \right) + \alpha_{0km}^c \nabla_{p_{0km}^c} \mathcal{F}_1 \\ + \alpha_{km}^c \nabla_{p_{0km}^c} \left(\mathcal{F}_1 \mathcal{F}_2 \mathcal{F}_3 \right) \quad (28)$$

with

$$\mathcal{F}_1(p_{0km}^c) = \frac{1}{1 - \Pi_{0m}^c \Pi_{0km}^c} \quad (29a)$$

$$\mathcal{F}_2(p_{0km}^c) = \Pi_{0m}^c \left(1 - \Pi_{0km}^c \right) \quad (29b)$$

$$\mathcal{F}_3(p_{km}^c) = \frac{1}{(1 - \Pi_{km}^c)} \quad (29c)$$

where at the optimum $\nabla_{p_{km}^{c*}} \mathcal{L} = 0$ and $\nabla_{p_{0km}^{c*}} \mathcal{L} = 0$. Since the equality constraints (c_4) and (c_5) have to be satisfied, the values of $\overline{N}_{km}^c$ and $\overline{N}_{0km}^c$ are respectively substituted in (27) and (28), and the optimal values of the Lagrange multipliers α_{0km}^{c*} and α_{km}^{c*} are obtained by applying the Newton-Raphson method to the system of two equations $\nabla_{p_{km}^{c*}} \mathcal{L} = 0$ and $\nabla_{p_{0km}^{c*}} \mathcal{L} = 0$.

The addition of the constraints from (c3) to (c5) in (16) allows us to solve the power allocation problem by computing iteratively the Lagrange multipliers α_{0m}^d, α_{0km}^c, α_{km}^c and substituting the values in (21), (24) and (25).

C. Relay selection strategy

Our problem relies also on relay selection, i.e. determining the value of s_{km} that leads to the minimum energy path. Hence, minimizing the primal problem w.r.t. s_{km} [23], we get the following:

$$\nabla_{s_{km}} \mathcal{L} = \mathcal{Z}_{km} + \phi_m \quad (30)$$

with

$$\mathcal{Z}_{km} = (f_3(p_{km}^c) + \gamma_m) \overline{N}_{km}^c + (f_2(p_{0km}^c) + \gamma_m) \overline{N}_{0km}^c \\ - (f_1(p_{0m}^d) + \gamma_m) \overline{N}_{0m}^d \quad (31)$$

and

$$\nabla_{s_{km}} \mathcal{L} |_{\lambda_{km}^*} \begin{cases} = 0, & \text{if } s_{km} \in]0, 1[, \\ < 0, & \text{if } s_{km} = 1. \end{cases} \quad (32)$$

$\mathcal{Z}_{km}$ can be interpreted as a weighted difference between the energy consumption of a relay-assisted communication and a direct communication. Since ϕ_m is common to all relays, only the relay k with the smallest $\mathcal{Z}_{km}$ can be selected to help the user m. Hence, $s_{km} = 1$ if $\mathcal{Z}_{km} < 0$ and $s_{km} = 0$ if $\mathcal{Z}_{km} > 0$. The relay selection strategy can be summarized in Algorithm 1:

Algorithm 1 Relay Selection Procedure

1: **for** $m = 1 : M$ **do**
2:   **if** $\mathcal{Z}_{km} > 0 \quad \forall \; k \in \mathcal{K}$ **then**
3:     $s_{km} = 0$
4:     $(k, m)^* = (0, m)$
5:   **else**
6:     $(k, m)^* = \arg \min_{(k,m)} \mathcal{Z}_{km}$
7:     $s_{km} = 1$
8:   **end if**
9: **end for**

Algorithm 2 Joint Power and Relay Selection algorithm

1: Initialize: $\mu \leftarrow 0$
2: Initialize: $\gamma_m \leftarrow 0 \; \forall m \in \mathcal{M}$
3: **while** $\exists m$ s.t. (c_1) not satisfied **do**
4:   Find α_{0m}^d by solving $\nabla_{p_{0m}^d} \mathcal{L} = 0 \; \forall \; m \in \mathcal{M}$
5:   Find α_{0km}^c and α_{km}^c, $\forall (k,m) \in \mathcal{K} \times \mathcal{M}$, by solving the system of non-linear equations $\nabla_{p_{0km}^c} \mathcal{L} = 0$ and $\nabla_{p_{km}^c} \mathcal{L} = 0$ by applying Newton-Raphson method.
6:   Apply the relay selection strategy in Algorithm 1
7:   Update $\gamma_m(t)$ according to (33)
8: **end while**
9: Substitute $p_{0m}^d, p_{0km}^c, p_{km}^c$ and $s_{km} \; \forall (k,m) \in \mathcal{K} \times \mathcal{M}$ in (c_6)
10: **if** the power constraint (c_6) is satisfied **then**
11:   Finish
12: **else if** problem infeasible **then**
13:   Go to step 17
14: **else**
15:   Update μ according to (34) and go to step 2
16: **end if**
17: **while** Problem is infeasible **do**
18:   Reject $u_m = \arg \max_m \{SR_m : m \in \mathcal{M}\}$
19:   $\mathcal{M} = \mathcal{M} - \{u_m\}$
20:   Return to step 1
21: **end while**

V. ENERGY MINIMIZATION ALGORITHM

As it can be noticed, the optimization problem in (18) is a maximization problem with no constraint and it can be solved w.r.t. $\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{km}}$ with subgradient method.

The Lagrange multipliers γ_m and μ are updated according to (33) and (34) respectively. Moreover, β_{γ_m} and β_μ are constant steps belonging to $]0, 1[$. For each μ, the Lagrange multiplier γ_m is updated in parallel for every user m until they satisfy their delay constraints. Thereafter, the optimal Lagrange variables α_{0m}^{d*}, α_{0km}^{c*} and α_{km}^{c*} are obtained as explained in Section IV. Then, the resource scheduler performs the relay selection strategy defined in Algorithm 1 until all users satisfy their delay constraints or reaching a stopping criterion.

Algorithm 2 jointly allocates powers at BS and RS and selects an optimal relay if needed. This algorithm firstly starts by initializing the Lagrange multiplier μ (step 1) that is related to the power constraint. Then, the resource scheduler updates

$$\gamma_m(t+1) = \left[\gamma_m(t) + \beta_{\gamma_m} \left(\left(1 - \sum_{k=1}^{K} s_{km} \right) \overline{N}_{0m}^d + \sum_{k=1}^{K} s_{km} \left(\overline{N}_{0km}^c + \overline{N}_{km}^c \right) - \overline{D}_m \right) \right]^+ \qquad (33)$$

$$\mu(t+1) = \left[\mu(t) + \beta_\mu \left(P_{tot} - \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{km} \right) p_{0m}^d + \sum_{k=1}^{K} s_{km}(p_{0km}^c + p_{km}^c) \right) \right) \right]^+ \qquad (34)$$

for all users in parallel the Lagrange multiplier γ_m (steps 3 to 7) until they satisfy their delay constraints. If $\gamma_m(t) = 0$ for $t > 1$ this means that delay of user m is less than $\overline{D}_m$. For each updated γ_m, the resource scheduler has also to find the variables α_{0m}^{d*}, α_{0km}^{c*} and α_{km}^{c*} that nullify (26), (27) and (28). The algorithm then associates a relay k to a user m or not according to the minimal energy path (step 6). If the power constraint P_{tot} can satisfy all users, the algorithm stops. Otherwise, the Lagrange multiplier μ is updated according to (34) and the resource scheduler restarts allocation process.

In order to better understand how the resource scheduler allocates the power according to the delays and power constraints, let us consider only one user and a relay k is selected to assist this user such as the allocated power is $p_m = p_{0km}^c + p_{km}^c$. Fig. 2 shows the average energy and average delay versus the power p_m for user m. Moreover, the energy curve has a minimum E_{min} at $p_m = p_{E_{min}}$ which corresponds to a delay $\overline{D}_{E_{min}}$. Therefore, the average energy function is strictly decreasing for $p_m < p_{E_{min}}$ and increasing for $p_m > p_{E_{min}}$. Several operating conditions may occur during the power allocation process that can be stated as follows:

- If user m has a delay constraint $\overline{D}_m < \overline{D}_{E_{min}}$, then the required power is such as $p_m > p_{E_{min}}$, since the delay versus p_m is strictly decreasing. Hence, if $p_m \le P_{tot}$, user m is served otherwise, it goes in starvation.
- If $\overline{D}_m \ge \overline{D}_{E_{min}}$, then the required power $p_m \le p_{E_{min}}$. Therefore if $p_{E_{min}} \le P_{tot}$, then the optimal power p_m leading to the minimum energy consumed is the solution of $\overline{N}_{0km}^c + \overline{N}_{km}^c = \overline{D}_{E_{min}}$, i.e $p_m = p_{E_{min}}$ and the consumed energy is E_{min}. In that case, the problem is equivalent to find the minimum of the energy function without considering the delay constraint ($\gamma_m = 0$).
- If $\overline{D}_m \ge \overline{D}_{E_{min}}$ and $p_{E_{min}} > P_{tot}$, the resource scheduler has to allocate a power in $p_m \in]0, p_{E_{min}}[$ and thus the energy function starts increasing as p_m decreases (decreasing p_m means increasing the delay and thus increasing the energy consumption). Hence, μ is updated until $p_m \le P_{tot}$ and until the inequality $\overline{N}_{0km}^c + \overline{N}_{km}^c \le \overline{D}_m$ is satisfied. If such a power can be achieved then the user is served, otherwise, it goes in starvation.

This example also shows that the total consumed energy is at its minimum value if the power constraint is relaxed, i.e. large P_{tot}. As Algorithm 2 updates μ, this signifies that some users have not yet satisfied their delay constraints and hence P_{tot} is not large enough to serve all users. By referring to Fig. 2, powers that satisfy the inequality constraint in (c_1) are on the
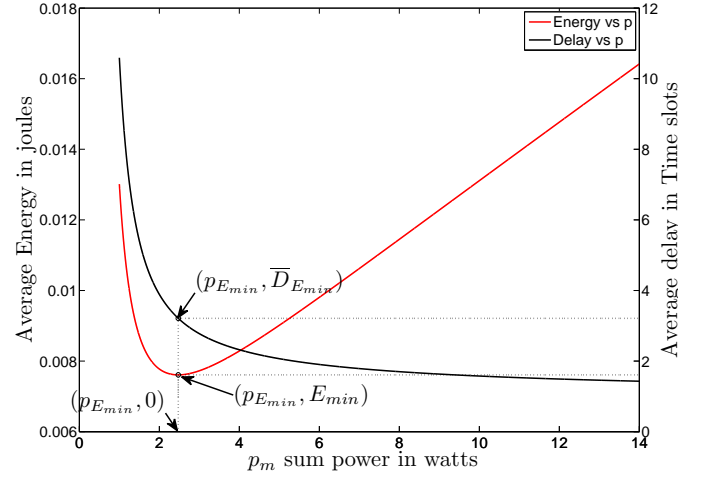

Fig. 2: Energy-Delay versus power $p_m = p_{0km}^c + p_{km}^c$

left of $p_{E_{min}}$ for users that have $\overline{D}_m \ge \overline{D}_{E_{min}}$. Therefore, as μ increases, the system energy consumption increases. If all delay constraints are exactly satisfied but there is no sufficient amount of power, then the problem is infeasible. Hence, the resource scheduler rejects the user that has the greater energy ratio ER_m (step 18) and restarts allocation process. With multiple users, global optimum cannot be ensured because of possible multiple local minima involved in the objective function. However, our algorithm ensures to find at least a local minimum satisfying the constraints.

The energy ratio for each user is defined as the ratio of the average energy consumed in Joules, i.e. E_m^*, and the average delay associated in seconds. If the delay in time slots is N_m^* and the time slot duration in second is $N_F/\Delta f$ (resource block duration), then ER_m is expressed in Watts and can be written as:

$$\text{ER}_m = \frac{\Delta f \times E_m^*}{N_F \times N_m^*} \qquad (35)$$

Algorithm 2 solves the modified problem in (16), i.e. considering the delays as variables and forcing them to be equal to their expressions in constraints (c3)-(c5). This reformulation allows to iteratively solve the powers to be allocated to each user by computing the Lagrangian multipliers α_{0m}^d, α_{0km}^c and α_{km}^c as discussed in the previous section. Since the problem is non linear, the powers cannot be obtained directly and require an iterative procedure. Moreover, if the problem is infeasible, e.g. the power budget is not high enough to satisfy all the users' delay, the rejection of the user the less energy-efficient guarantees the algorithm to end. This highlights the importance

of monitoring the starvation rate and the energy consumption at the same time.

VI. MCS AND RESOURCE BLOCKS ALLOCATION PROBLEM

In LTE system, the resource allocation manager selects for every user an MCS and a certain number of time/frequency resource blocks in order to satisfy the demanding QoS. We assume that the number of possible MCS is N_{mcs} and the total number of frequency resource blocks is N_{RB}. For a given allocated power policy, the energy consumed by the system depends on MCS and RBs. The minimization of the consumed energy can be formulated in the canonical form as follows:

$$\min_{\mathbf{x}_{mr}^{\mathrm{mcs}}} \sum_{m=1}^{M} \sum_{r=1}^{N_{\mathrm{RB}}} \sum_{\mathrm{mcs}=1}^{N_{\mathrm{mcs}}} \mathrm{x}_{mr}^{\mathrm{mcs}} \mathrm{E}_{mr}^{\mathrm{mcs}}$$

subject to:

$$(c_1) \ \mathrm{x}_{mr}^{\mathrm{mcs}} \in \{0,1\} \ \forall (m,r,\mathrm{mcs}) \in \mathcal{M} \times \mathcal{N}_{\mathrm{RB}} \times \mathcal{N}_{\mathrm{mcs}}$$

$$(c_2) \ \sum_{r=1}^{N_{\mathrm{RB}}} \sum_{\mathrm{mcs}=1}^{N_{\mathrm{mcs}}} \mathrm{x}_{mr}^{\mathrm{mcs}} = 1 \ \forall m \in \mathcal{M}$$

$$(c_3) \ \sum_{m=1}^{M} \sum_{r=1}^{N_{\mathrm{RB}}} \sum_{\mathrm{mcs}=1}^{N_{\mathrm{mcs}}} r \mathrm{x}_{mr}^{\mathrm{mcs}} \leq N_{\mathrm{RB}}$$

$$(c_4) \ \sum_{m=1}^{M} \sum_{r=1}^{N_{\mathrm{rb}}} \sum_{\mathrm{mcs}=1}^{N_{\mathrm{mcs}}} r N_s \mathrm{x}_{mr}^{\mathrm{mcs}} p_{mr}^{\mathrm{mcs}} \leq P_{\mathrm{tot}}$$

$$(36)$$

where $\mathrm{E}_{mr}^{\mathrm{mcs}}$ is the energy consumed for a particular MCS, i.e. mcs, and a number of RBs, i.e. r, p_{mr}^{mcs} is the power allocated per subcarrier for user m and $\mathcal{N}_{\mathrm{RB}}$, $\mathcal{N}_{\mathrm{mcs}}$ are the sets of RBs and MCS respectively. Moreover, N_s is reminded to be the number of subcarriers per RB. The problem in (36) is a binary-integer programming problem, since the variables $\mathrm{x}_{mr}^{\mathrm{mcs}}$ and r in (c_1) and (c_3) are binary and integer respectively. The constraint (c_2) ensures that only one combination of MCS and RB is attributed for every user u_m. The constraint (c_3) ensures that the total number of allocated RBs for all users is less than or equal to N_{RB}. The constraint (c_4) states that the overall allocated power is constrained by P_{tot}.

Integer programming problems are in general NP-difficult to solve. However, some efficient algorithms can be found and perform well in practice. Dedicated solvers to *mixed integer and convex problem* exist in open or professional literature such as Gurobi and MOSEK [25]. In order to efficiently solve problem (36), we used the CVX toolbox in conjunction with an integer programming solver [25], [26] in order to provide for every user, the optimal number of RBs and MCS index leading to the minimum energy consumption. The global procedure iterates the integer programming solver for each allocated power p_{mr}^{mcs} obtained in Algorithm 2 until all users are satisfied or a non feasible solution is declared. In that case, the user with the highest energy consumption goes in starvation and the algorithm restarts allocation. In other words, for a given μ, Algorithm 2 first computes the required powers that satisfy the delay constraints for all users, MCS and number of RBs. Then, for every user with any possible selection of a couple of MCS and number of RBs, the energy and power values found

in Algorithm 2 are used as inputs in the optimization problem in (36). Then, the solver searches for the best number of RBs in *frequency* and the best MCS that minimize the energy consumption under the total power constraint, i.e. (c_4) in (36). If the power constraint cannot be satisfied, the Lagrange multiplier μ is updated and the process described above is repeated.

TABLE II: System model parameters

Parameters	Description	value
f_s	Sampling rate	7.68 MHz
B_w	Bandwidth	5 MHz
Δf	Subcarrier spacing	15 kHz
N_c	Number of subcarriers (FFT size)	512
N_{CP}	Cyclic prefix	128
N_g	Number of Guard subcarriers	211
N_{OFDM}	Number of OFDM symbols in a block	7
N_{RB}	Maximum number of RBs	25
L	Block length	1024
f_c	Carrier frequency	2.6 GHz
N_0	Noise spectral density	-174 dBm/Hz
β_{amp}	Power amplifier inefficiency	1

The computational complexity order of Algorithm 2 depends on the number of relays K, the number of MCS N_{mcs} and the total number of frequency resource blocks N_{RB} and is $\mathcal{O}(K \times N_{\mathrm{mcs}} \times N_{\mathrm{RB}})$. The powers are allocated in parallel for all users and hence only memory space in this dimension, i.e. M, is required. It is also worthwhile noting that each update of the Lagrange multiplier μ involves the execution of the Newton-Raphson algorithm to obtain the multipliers α_{0m}^d, α_{0km}^c and α_{km}^c. However, the complexity of this algorithm is quite low and the convergence really fast (less than 10 iterations). The update of the Lagrange multipliers μ and γ_m are performed through a sub-gradient method, i.e. in (33), (34), which is a very low complexity procedure. Technically, the complexity of problem (36) is in $O\left(2^{M \times N_{\mathrm{mcs}} \times N_{\mathrm{RB}}}\right)$. However, the MOSEK solver that we adopted for solving this problem uses the semidefinite optimization by binary relaxation to get a lower bound on the optimal solution with polynomial complexity [27].

VII. NUMERICAL ANALYSIS

This section investigates the performance of our algorithms in terms of the overall energy consumption [Joule/bit] and the average starved user rate, i.e. the average number of users that cannot satisfy their delay constraint. Our relay-assisted resource allocation algorithms are compared to direct communications. A single circular cell of radius R = 1 km is considered, with 4 circularly distributed relays located at distance R/2 from BS. We also consider that $M = 4$, 8 or 12 users are uniformly distributed in the cell. The resource scheduler can select different MCSs according to Table I, where each packet contains $L = 1024$ information bits. The same path loss model is used for each link, i.e. BS-RS, RS-UE and BS-UE, and the path loss gain in dB between nodes i and j is $l_{ij} = -139.90 - 34.41 \cdot \log_{10}(d_{ij})$, where d_{ij} is the distance between two nodes in km [28]. Moreover, the circuitry power consumption per subcarrier at BS is about 17 dBm and at RS and UE, the circuity power consumption is about 7 dBm. This parameter settings will be used throughout this section unless otherwise mentioned.
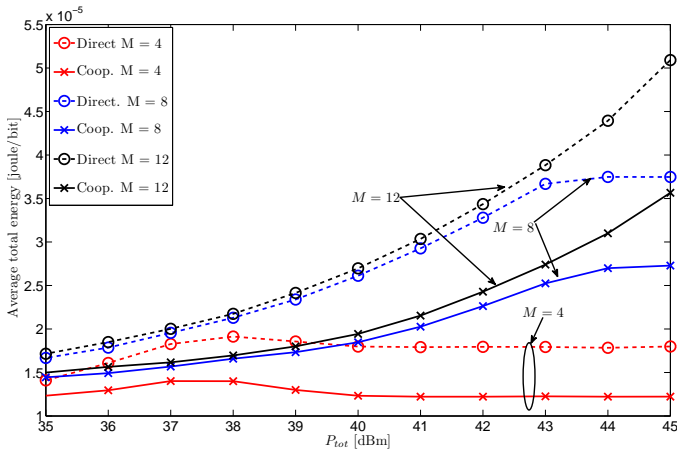
Fig. 3: Average energy consumption in Joule/bit versus P_{tot} in dBm. The number of users is $M = 4$, 8 and 12 and each user has a delay constraint of 3 time slots.
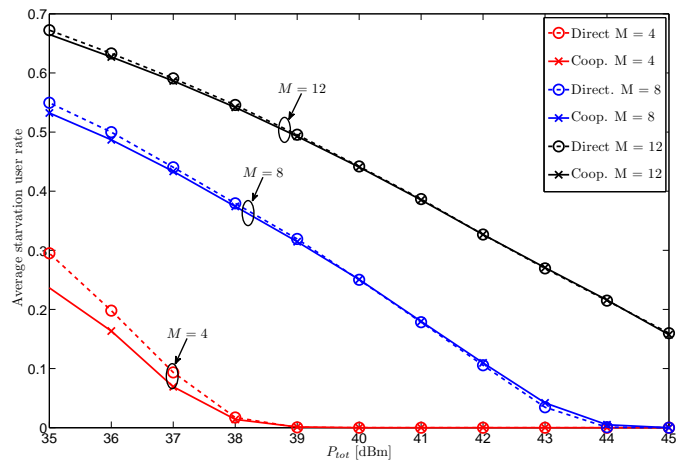

Fig. 4: Average starvation user rate versus P_{tot} in dBm. The number of users is $M = 4$, 8 and 12 and each user has a delay constraint of 3 time slots.

Let us consider that each user has a delay constraint $\overline{D}_m = 3 \ \forall \ m$ to be satisfied. Figures 3 and 4 show the overall consumed energy and the average starvation rate w.r.t. the power constraint P_{tot} in dBm respectively. One can notice that as the total power budget increases the starvation rate decreases as expected since more and more users can be served when the power constraint becomes less stringent. For $M = 4$ on Fig. 3, we can observe that the energy consumption starts increasing up to $P_{tot} = 38$ dBm due to the decrease in starvation user rate leading to a higher number of satisfied users and hence a larger energy consumption. At $P_{tot} = 38$ dBm, the energy consumption achieves a maximum and the starvation rate is nearly zero. For higher power budget, the overall energy slightly decreases and finally becomes constant for large values of P_{tot}. When $M = 8, 12$ the same behavior can be observed but at different P_{tot}. A stringent power budget leads to a low energy consumption but to a high number of unsatisfied users. As the available amount of power increases, the energy consumed increases but the number of satisfied users increases also. For instance, for $M = 8$ the energy consumption of direct and cooperative networks increases up to $P_{tot} = 44$ dBm for which the starvation rate reaches zero. While the starvation rate is nearly the same for networks with and without cooperation, this is not the case for energy consumption for which significant gains can be expected with relay-assisted communications. For $M = 4$, relay-assisted scheme can save about 31 % of energy compared to direct transmission and for $P_{tot} = 45$ dBm the energy saving between cooperative and direct transmissions achieves 28 and 29% for $M = 8$ and 12 respectively. It is worth mentioning that when comparing the starvation rate of direct and relay-assisted transmissions, this should be done with a look on the consumed energy for a fixed power constraint. Indeed, the power constraint P_{tot} can be partially or completely consumed, according to the users' position, and does not give any information how much power is consumed. Moreover, since the consumed energy depends on the circuitry power consumption and the required transmission power for all users, relay-assisted network makes a better use of power than non relay-aided system, in average.
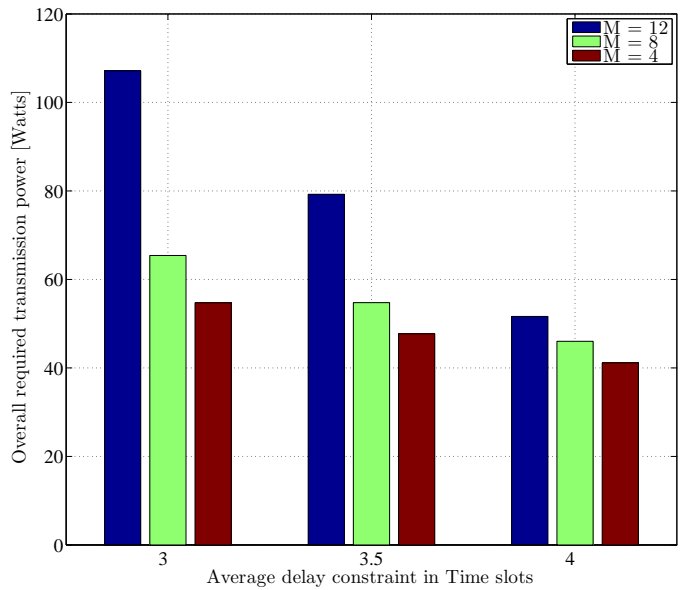

Fig. 5: Average required power in Watts versus the delay constraint in time slots, labeled on the number of users, i.e. $M = 4$, 8 and 12.

Fig. 5 depicts for cooperative transmission the total average required power in Watts (the sum of all optimized powers for all users) as a function of the delay constraint and the number of users, i.e. $M = 4$, 8 and 12. We firstly notice that the required power increases as the number of users increases and this difference is larger for small delay constraint, i.e. $D_m = 3$ TSs, than higher delay constraint, i.e. $D_m = 4$ TSs. Moreover, the required transmission power reduces if the delay constraint becomes loose whatever the number of users considered. The required transmission power for $M = 12$ and stringent delay constraint, e.g. $D_m = 3$, is very high compared to $M = 4$ and 8 since the number of frequency RBs has to be shared among more users and then the delay constraints can only be satisfied by allocating high transmission powers for all users. On the other hand, higher delay constraints can be satisfied by selecting a lower number of resource blocks leading to a lower transmission power.
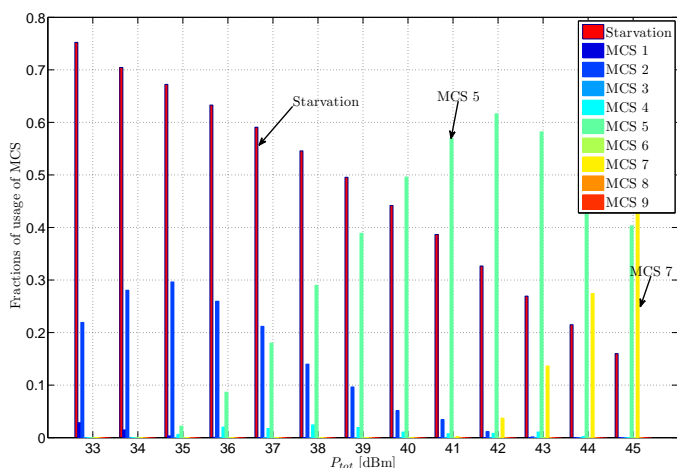
Fig. 6: Fractions of usage of MCS w.r.t. the power constraint in direct networks with $M = 12$. Each user has a delay constraint of 3 time slots.
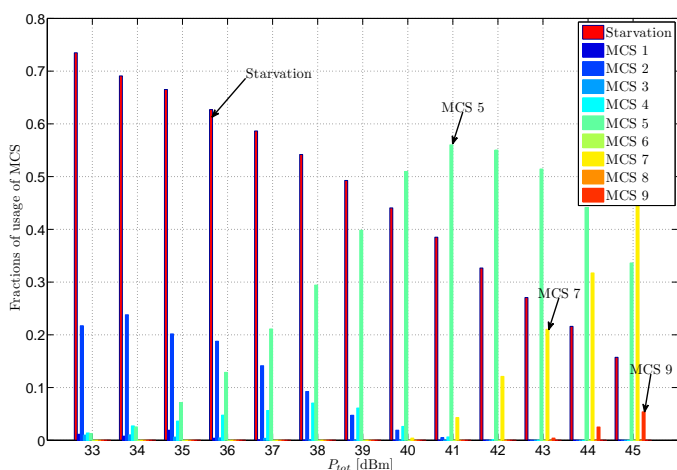

Fig. 7: Fractions of usage of MCS w.r.t. the power constraint in cooperative network with $M = 12$. Each user has a delay constraint of 3 time slots.
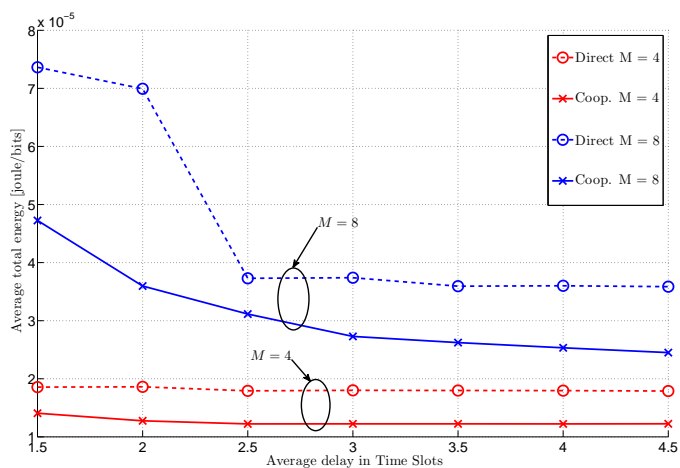

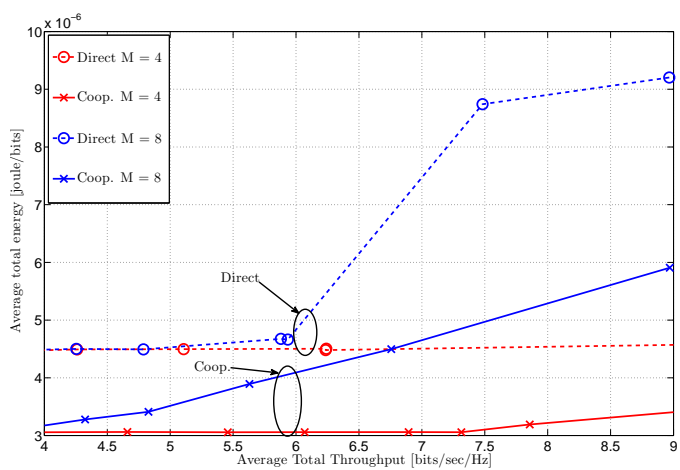Fig. 8: Average consumed energy in joule/bit w.r.t. the delay constraint in Time Slots. The number of users is $M = 4$ and 8.

Fig. 9: Average consumed energy in joule/bit w.r.t. the throughput efficiency in bits/sec/Hz. The number of users is $M = 4$ and 8.

Figures 6 and 7 represent the fraction of usage of different MCS as function of the total power budget, for direct and cooperative transmissions respectively and for $M = 12$. For very stringent power constraint, a large starvation user rate can be observed as it has been previously remarked. Moreover, in this case, the most common MCS is the one with the lowest index, corresponding to a QPSK and a low coding rate inducing a low energy consumption but also a low data rate. As P_{tot} increases, the starvation rate decreases and more users can be satisfied and high MCS indexes start to be used, enabling a higher spectral efficiency. It is also interesting to remark that relay-assisted scheme, Fig. 7, enables the use of the MCS 9, i.e. allowing the highest spectral-efficiency, for $P_{\text{tot}} = 45$ dBm, which is not possible for a cell without relays as it can be inferred from Fig. 6.

Figure 8 depicts the average consumed energy (Joule/bit) w.r.t. the delay constraint in time slots, for $M = 4$ and 8 when there is no constraint on power. From this figure, one can remark that relay-assisted scheme outperforms direct transmissions whatever the delay constraint. For large delay constraints, i.e. 4.5 TS, the energy saving reaches about 27% and 31% for 4 and 8 users respectively. For $M = 4$, the

energy consumed remains relatively constant w.r.t. the delay constraint. This is because the power budget and the number of available RBs are sufficiently important to satisfy the users whatever their delay constraints. On the other hand, for $M = 8$, the overall consumed energy significantly decreases for both direct and cooperative transmissions. For low values of delay constraint, the communication becomes more stringent and then more resource blocks are needed leading to a higher energy consumption. As the delay constraint becomes loose, the resource manager may operate at power values that lead to the minimum energy consumption.

Since the spectral efficiency that the network may achieve is also of great interest, Fig. 9 plots the tradeoff between the average consumed energy and the delivered throughput efficiency (bits/sec/Hz) in the network. As it can be noticed, for $M = 4$, the spectral efficiency increase is almost *energy-free* except for very high spectral-efficiency region and cooperative scheme. In this case indeed, there are enough resource blocks to satisfy all users. Moreover, cooperation enables an energy saving of about 26 % compared to a direct transmission. For $M = 8$, the energy consumption logically increases with the spectral efficiency, since the number of resource blocks

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2654684, IEEE Transactions on Vehicular Technology

TO APPEAR IN IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. XX, NO. XX, XXX 2017
11

becomes not sufficient w.r.t. the number of users and their constraints. However, for $M = 8$ and for a spectral efficiency of 9 bits/sec/Hz, we can see that cooperation enables an energy saving of about 34 % compared to the direct transmission.

VIII. CONCLUSION

In this work, an energy minimization problem for LTE multi-user relay-assisted HARQ network has been investigated. The energy optimization problem has been tackled under total power and individual user delay constraints and considering static and dynamic energy consumption of nodes. Moreover, HARQ-I protocol has been taken into account when optimizing resources. A new energy minimization algorithm that jointly allocates the optimal power at base station and relays has been proposed. In addition to the power allocation, a relay selection procedure is performed to determine whether direct or cooperative transmission is preferable for a particular user from an energy consumption point of view. Moreover, based on the energy minimization algorithm, we proposed a binary and integer programming problem that jointly chooses the optimal MCS and number of RBs for every user according to the total power budget. Our investigations have shown that the proposed algorithms allow to save up to 30% of energy in relay-assisted network compared to direct communications between BS and mobiles. In further works, advanced HARQ mechanisms and multi-cell environments could be investigated.

REFERENCES

[1] M. A. Imran *et al.*, "Most suitable efficiency metrics and utility functions," INFSO-ICT-247733 EARTH, Tech. Rep., Dec. 2011.

[2] O. Blume *et al.*, "Most promising tracks of green network technologies," INFSO-ICT-247733 EARTH, Tech. Rep., Dec. 2011.

[3] M. Dohler, "Virtual Antenna Arrays," Ph.D. dissertation, King's College, London, UK, 2003.

[4] M. Dohler and Y. Li, *Cooperative Communications: Hardware, Channel & PHY*. John Wiley & sons, Ltd, 2010.

[5] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed Interference-Aware Energy-Efficient Power Optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323 –1333, April 2011.

[6] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental Trade-offs on Green Wireless Networks," *IEEE Communications Magazine*, vol. 49, pp. 30–37, June 2011.

[7] M. Salem, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, "Opportunities and Challenges in OFDMA-Based Cellular Relay Networks: A Radio Resource Management Perspective," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2496–2510, Jun 2010.

[8] I. Stanojev, O. Simeone, Y. Bar-Ness, and K. Dong Ho, "Energy efficiency of non-collaborative and collaborative Hybrid-ARQ protocols," *IEEE Transactions Wireless Communications*, vol. 8, no. 1, pp. 326–335, Jan. 2009.

[9] Y. Qi, R. Hoshyar, M. Imran, and R. Tafazolli, "H2-ARQ-Relaying: Spectrum and Energy Efficiency Perspectives," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1547 –1558, Sept. 2011.

[10] S. H. Kim, S. J. Lee, and D. K. Sung, "Rate-Adaptation-Based Cooperative Hybrid-ARQ Relaying Scheme in Rayleigh Block-Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4640–4645, Nov 2011.

[11] M. G. Song, Y. J. Kim, E. Y. Park, and G. H. Im, "Rate Adaptation and Power Allocation for Cognitive Radio Networks with HARQ-Based Primary System," *IEEE Transactions on Communications*, vol. 62, no. 4, pp. 1178–1187, April 2014.

[12] R. Zhang, J.-M. Gorce, and K. Jaffres-Runser, "Low Bound of Energy-Latency Trade-Off of Opportunistic Routing in Multi-Hop Networks," in *ICC '09. IEEE Int. Conf. on Comm., 2009.*, June 2009, pp. 1–6.

[13] H. Lin, L. Wang, and R. Kong, "Energy Efficient Clustering Protocol for Large-Scale Sensor Networks," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7150–7160, Dec 2015.

[14] X. Xu, J. Bao, H. Cao, Y. D. Yao, and S. Hu, "Energy-Efficiency-Based Optimal Relay Selection Scheme With a BER Constraint in Cooperative Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 191–203, Jan 2016.

[15] M. Maaz, P. Mary, and M. Hélard, "Energy Efficiency Analysis in Relay Assisted Hybrid-ARQ Communications," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, 2012, pp. 2263–2268.

[16] M. Maaz, J. Lorandel, P. Mary, J.-C. Prévotet, and M. Hélard, "Energy Efficiency Analysis of Hybrid-ARQ Relay-Assisted Schemes in LTE-Based Systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, 2016.

[17] C. Le Martret, A. Le Duc, S. Marcille, and P. Ciblat, "Analytical Performance Derivation of Hybrid ARQ Schemes at IP Layer," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1305 –1314, May 2012.

[18] B. Maham, A. Behnad, and M. Debbah, "Analysis of Outage Probability and Throughput for Half-Duplex Hybrid-ARQ Relay Channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 7, pp. 3061–3070, 2012.

[19] D. Mackay and C. P. Hesketh, "Performance of Low Density Parity Check Codes as a Function of Actual and Assumed Noise Levels," in *in Electronic Notes in Theoretical Computer Science*. Elsevier, 1998, p. 1645.

[20] 3GPP, "Multiplexing and channel coding," 3GPP TS 36.212 (Release 12), Tech. Rep., 2016.

[21] X. Lagrange, "Throughput of HARQ Protocols on a Block Fading Channel," *IEEE Communications Letters*, vol. 14, no. 3, pp. 257–259, 2010.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Academic Press, mar 2007. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0123736374

[23] D. S. W. Hui, V. K. N. Lau, and W. H. Lam, "Cross-Layer Design for OFDMA Wireless Systems With Heterogeneous Delay Requirements," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2872–2880, August 2007.

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

[25] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[26] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[27] MOSEK, "Mosek modeling manual," Tech. Rep., August 12, 2014.

[28] Y. Chen and K. Hsieh, "A Dual Least-Square Approach of Tuning Optimal Propagation Model for Existing 3G Radio Network," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 6, may 2006, pp. 2942 –2946.

Mohamad Maaz received the B.S. and M.Sc. degrees in Telecommunications and Networking from the Lebanese University (Lebanon) and the University of Western Brittany (France), respectively. In 2013, he obtained his Ph.D. degree in Electrical Engineering from INSA Rennes. During his PhD, he worked on the theoretical performance analysis and resource allocation of cooperative networks.

From Oct. 2013 to June 2015, he worked on an industrial project, at the Institute of Electronics and Telecommunications of Rennes (IETR), for designing new green wireless communication techniques for LTE-A and WiMAX (IEEE-802.16m) based systems. Since Nov. 2015, Dr. MAAZ has been with Orange Labs (Châtillon - France), where he is currently working on millimeter waves with Massive MIMO in 5G wireless networks.

His research interests lie in the areas of wireless communications, statistical signal processing, millimeter waves, optimization theory, cross-layer design, cooperative communications, HARQ protocols and Massive MIMO.

Maryline Hélard received the M.Sc and PhD degrees from INSA Rennes and the Habilitation degree from Rennes 1 University in 1981, 1884 and 2004 respectively. In 1985, she joined France Telecom as a research engineer and since 1991 she has been studying physical layer in the field of digital television and wireless communications. In 2007, she joined the National Institute of Applied Science (INSA) as a professor and she is now the co-director of the Digital Communications department of IETR (Electronics and Telecommunications Institute of Rennes). She is co-author of more than 125 technical papers including 36 journal papers and 30 patents. Her current research interests are in the areas of digital communications such as MIMO techniques, large MIMO, OFDM, MC-CDMA, channel estimation, equalization, spatial modulations and iterative processing applied to wireless communications and more recently to wire communications (ADSL, optical). She was involved in several collaborative research projects including digital television, MC-CDMA techniques, time reversal and spatial modulation.

Philippe Mary (S'06-M'09) received his M.Sc. in Signal Processing and Digital Communications from the University of Nice Sophia-Antipolis and the Dipl. Ing. degree in electrical engineering from the Polytechnic University School of Nice Sophia-Antipolis (France) both in 2004. He received his PhD in electrical engineering from the National Institute of Applied Sciences of Lyon in 2008. During his PhD, he was with France Telecom R&D and he worked on the analytical performance study for mobile communications considering shadowing and fading and multi-user detectors for wireless communications. From 2008 to 2009, he was post-doctoral researcher at ETIS Laboratory in Cergy-Pontoise.

In September 2009, Dr. Mary joined the Digital Communication Systems department of INSA Rennes and IETR laboratory as associate professor. In 2011, he served as TPC chair for the third international workshop on cross-layer design (IWCLD 2011) and he is serving as TPC member of various IEEE conferences, e.g. ICC, Globecom, PIMRC, WCNC, VTC. His research interests include analytical performance analysis and signal processing for digital communications, resource allocation and finite block length communication theory.