

Energy-Aware Power Control for a Multiple-Relay Cooperative Network using Q -Learning

(Invited Paper)

Farshad Shams,[†] Giacomo Bacci,^{*‡} and Marco Luise^{*‡}

[†] Dept. Computer Science and Engineering, IMT Institute for Advanced Studies, Lucca, Italy

^{*} Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy

[‡] Consorzio Nazionale Internuniversity per le Telecomunicazioni (CNIT), Parma, Italy

Abstract—In this paper, we investigate the power control problem in a cooperative network with multiple wireless transmitters, multiple full-duplex amplify-and-forward relays, and one destination. A game-theory-based power control algorithm is devised to allocate the powers among all active nodes: the source nodes aim at maximizing their energy efficiency, whereas the relays aim at maximizing the network sum-rate. After showing that the proposed game admits multiple pure/mixed-strategy Nash equilibrium points, we formulate a Q -learning-based algorithm to let the active players converge to the best Nash equilibrium point that combines good performance in terms of both energy efficiency and overall data rate, also calling for a low computational burden. Numerical results show that the proposed scheme outperforms Nash bargaining, max-min fairness, and max-rate optimization schemes.

I. INTRODUCTION

New generations of wireless networks aim at using more and more intelligent wireless devices, that should be able to enhance the quality of service (QoS), increase the achievable throughput, and increase the battery life, while not overusing the common scarce wireless network resources, such as spectrum and battery energy. In recent years, the idea of *cooperative diversity* [1] has been considered as a promising technique to significantly enhance the transmission of information. The basic idea is that transmit data can be aided by several intermediate nodes, called *relays*, that retransmit such data to the receiver, using different schemes: either decode-and-forward (DF), or compress-and-forward (CF), or amplify-and-forward (AF), and either half-duplex or full-duplex (please see [2] for further details).

In the literature, there exist many attempts to properly allocate the resources in a relay-aided network. Just to mention a few relevant applications in this field, power control algorithms have been derived for the single-relay (e.g., [3]) and the multiple-relay (e.g., [4]–[6]) scenarios. This problem has been also widely investigated using game theory [7] to model the interactions among the nodes in the network (e.g., [8]–[12]). However, most approaches show a relatively high

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's FP7 under REA Grant agreement no. P10F-GA-2011-302520 GRAND-CRU Game-theoretic Resource Allocation for wireless Networks based on Distributed and Cooperative Relaying Units, and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM# (Grant agreement no. 318306).

computational complexity, which could seriously undermine their applicability. In this contribution, we aim at addressing this drawback, by investigating the problem of power allocation at both source nodes and relay nodes, focusing on the AF strategy for the sake of mathematical tractability, although a similar way of reasoning can be applied to DF and CF as well. The communication is assumed to be full-duplex to enhance the spectral efficiency (SE) of the system. We model the power control problem as a distributed non-cooperative game, in which all nodes act as players that adjust their transmit powers in order to approach an energy efficient data rate vector, provided that each source meets its minimum data rate demand. We show that this game has a mixed-strategy Nash equilibrium, and we propose a Q -learning-based algorithm [13] to achieve it. To the best of our knowledge, this contribution is the first work that considers multiple relays in designing power allocation schemes at *both* source nodes and relay nodes in a *full-duplex* communication mode.

The remainder of the paper is structured as follows. Sect. II contains the formulation of the resource allocation problem as a noncooperative game, whose solution is computed in Sect. III using a reinforcement-learning method. Sect. IV compares the performance of the proposed algorithm with other methods available in the literature, and Sect. V concludes the paper.

II. STATEMENT OF THE PROBLEM

We focus on the uplink of a cooperative relay-aided network, wherein S multiple sources reach the destination through N multiple parallel AF relays working in full-duplex mode, using the links depicted in Fig. 1, where h_{mi} denotes the channel gain between transmitter m and receiver i , and $W_i \sim \mathcal{CN}(0, \sigma_w^2)$ is the additive white Gaussian noise (AWGN) received at node i with power σ_w^2 .

In the network, each source node $s \in \mathcal{S} = \{1, \dots, S\}$ needs to meet a minimum end-to-end data-rate demand towards the base-station d , as better detailed below, possibly exploiting (all) relay nodes $\mathcal{N} = \{1, \dots, n, \dots, N\}$. Using an information-theoretic approach, detailed in [14] and omitted here for the sake of brevity, the s th source node's signal-to-interference-plus-noise ratio (SINR) at the destination is

$$\gamma_s = \frac{\varphi_s p_s}{\sigma_w^2 + \sum_{s' \in \mathcal{S}, s' \neq s} \varphi_{s'} p_{s'}} \quad (1)$$

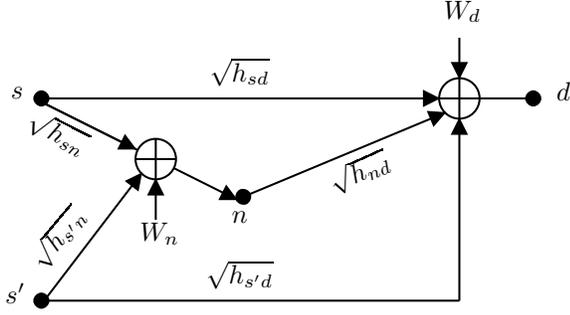


Fig. 1: Relay-aided source-to-destination communication.

where p_m , $p_m \leq \bar{p}_m$, is node m 's transmit power, with \bar{p}_m denoting node m 's maximum power, and

$$\varphi_s = \frac{(\sqrt{h_{sd}} + \sum_{n \in \mathcal{N}} |\alpha_n| \sqrt{h_{sn} h_{nd}})^2}{1 + \sum_{n \in \mathcal{N}} |\alpha_n|^2 h_{nd}} \quad (2)$$

with α_n denoting the scaling factor adopted by relay n to implement the AF strategy in the AWGN scenario, whose amplitude is chosen to satisfy its own power constraint \bar{p}_n as follows:

$$|\alpha_n| = \sqrt{\frac{p_n}{\sigma_w^2 + \sum_{s \in \mathcal{S}} h_{sn} p_s}}. \quad (3)$$

Using (1), we can express the Shannon capacity achievable by source node s as

$$R_s = \Delta f \cdot \log_2(1 + \gamma_s) \quad [\text{b/s}] \quad (4)$$

where Δf denotes the signal bandwidth. Hence, when inspecting (1)–(3), it is straightforward to note the coupling among the powers of all nodes (sources and relays) on the rate achievable by each source node. A joint power control, that consists in finding an optimal vector of transmit powers $[p_1, \dots, p_S]$ and $[p_1, \dots, p_N]$ such that the network performance can be increased, also including QoS constraints on each source node s 's minimum rate \underline{R}_s , is thus highly desirable, and at the same time quite challenging.

To investigate the solution to this problem, we will use the analytical tools of game theory [7], whose aim is to help us predict the behavior of rational agents with conflicting interests competing for some common resources. To this end, we model the interaction among different nodes as the following non-cooperative game:

$$\mathcal{G} = \{\mathcal{M}, \{\mathcal{P}_m\}_{m \in \mathcal{M}}, \{u_m(p_m; \mathbf{p}_{-m})\}_{m \in \mathcal{M}}\} \quad (5)$$

where $\mathcal{M} = \mathcal{S} \cup \mathcal{N}$, with $|\mathcal{M}| = M = S + N$, is the set of all active nodes (both source and relay nodes), that represent the *players* of the game; $u_m(p_m; \mathbf{p}_{-m})$ is the *utility function* of each user $m \in \mathcal{M}$, detailed as follows; and \mathcal{P}_m is the *discrete* set of user m 's transmit power, defined as $\mathcal{P}_m = \{0, \Delta p_m, 2\Delta p_m, \dots, K_m \cdot \Delta p_m\} = \{k\Delta p_m\}_{k=0}^{K_m}$, where $(K_m + 1)$, with $1 \leq K_m < \infty$, denotes the number of power levels (including zero power), and $\Delta p_m = \bar{p}_m / K_m$ is

the power step, with \bar{p}_m denoting user m 's maximum transmit power. Without loss of generality, for the sake of simplicity we assume $K_m = K \forall m \in \mathcal{M}$. Note that, as the number of players M is finite, and the number of actions available to each player $K + 1$ is also finite, \mathcal{G} is called a *finite* game [7].

To account for the different needs demanded by the two classes of users, we will define two different utility functions. The goal of each source node is to trade off its achieved channel capacity with its minimum power consumption. To this end, we define the utility function for source nodes as

$$u_s(p_s; \mathbf{p}_{-s}) = \frac{R_s(\mathbf{p})}{p_s + p_c} \quad \text{s.t.} \quad R_s \geq \underline{R}_s \quad \text{and} \quad p_s > 0 \quad (6)$$

where $\mathbf{p}_{-s} = \mathbf{p} \setminus p_s$ is the power vector of all nodes (including both sources and relays) excluding source s 's power p_s , with \mathbf{p} denoting the $M \times 1$ power vector collecting the transmit powers by all M nodes in the network; $R_s(\mathbf{p})$ is Shannon capacity (4) achievable by transmitter s , in which we explicitly include the dependence on \mathbf{p} ; and $p_c > 0$ is the circuit power, independent of the transmission rate and modeled as in [15], [16]. To explicitly account for the constraints in (6), the payoff of a source node is equal to zero when either $R_s < \underline{R}_s$ or $p_s = 0$.

On the other hand, as the relays are just ancillary nodes that aim at increasing the network performance while not showing significant power-saving constraints, their main purpose is to increase the SE of the system. Hence, we define the utility function for each relay node as

$$u_n(p_n; \mathbf{p}_{-n}) = \sum_{s \in \mathcal{S}} \frac{R_s(\mathbf{p})}{p_c} \quad \text{s.t.} \quad R_s \geq \underline{R}_s \quad \forall s \in \mathcal{S} \quad (7)$$

where p_c is used just to let $u_n(p_n; \mathbf{p}_{-n})$ have the same unit of measure of $u_s(p_s; \mathbf{p}_{-s})$; and $R_s(\mathbf{p})$ is again the Shannon capacity (4) achievable by source s , that depends on p_n according to (1)–(3). Similarly to (6), we assign $u_n(p_n; \mathbf{p}_{-n}) = 0$ when $R_s(p_n) < \underline{R}_s$ for at least one source node s .

A close inspection of the utilities (6) and (7) reveals that including the QoS constraints \underline{R}_s introduces a coupling between the power sets that provide positive utilities for all players $m \in \mathcal{M}$. In the remainder of the paper, we assume that the network setup (that includes channel realizations and set of minimum rate demands) is such that all sources are able to meet their requirements \underline{R}_s , otherwise the problem is declared to be unfeasible, as better detailed in Sect. III. The feasibility of the problem is out of scope of this paper, and left as a future work.

Given the game formulation detailed above, it is apparent that there exists a tradeoff between achieving a high network sum-rate (in terms of achievable Shannon capacity) and maximizing the energy efficiency (EE) of each individual source node. Moreover, the source nodes' power allocations are coupled in a conflicting way, as increasing each source's power level increases its own SINR while generating a higher interference level at both the relays and the destination. Similarly, the relay nodes' power allocation affects the performance

of all sources in a conflicting way, according to the coupling among source-to-relay and relay-to-destination links. To solve the maximization problem¹

$$p_m^* = \arg \max_{p_m \in \mathcal{P}_m} u_m(p_m; \mathbf{p}_{-m}) \quad (8)$$

in a scalable and distributed way, and thus keeping its complexity low, we can make use of the analytical tools of noncooperative game theory [7]. A possible solution to (8) is the *pure-strategy Nash equilibrium*: each player chooses an action $p_m^* \in \mathcal{P}_m$ that is its “best response” (in the sense of utility maximization) to the other players’ choices. Unfortunately, not all games have pure-strategy Nash equilibria. A generalization of this concept is represented by *mixed-strategy Nash equilibria*, which are probabilistic distributions on the set of actions available to each player that maximize each player’s *expected payoff* to the joint probabilistic distribution of all others.

Definition 1: A mixed-strategy Nash equilibrium for a game \mathcal{G} is a M -tuple of vectors $[\sigma_1^*, \dots, \sigma_M^*]$, with $\sigma_m^* \in [0, 1]^{K+1}$, such that, for all $m \in \mathcal{M}$ and all $\sigma_m \in [0, 1]^{K+1}$,

$$\begin{aligned} \sum_{p_m \in \mathcal{P}_m} \sum_{\mathbf{p}_{-m} \in \mathcal{P}_{-m}} \sigma_{-m}^*(\mathbf{p}_{-m}) \sigma_m^*(p_m) u_m(p_m; \mathbf{p}_{-m}) \geq \\ \sum_{p_m \in \mathcal{P}_m} \sum_{\mathbf{p}_{-m} \in \mathcal{P}_{-m}} \sigma_{-m}^*(\mathbf{p}_{-m}) \sigma_m(p_m) u_m(p_m; \mathbf{p}_{-m}), \end{aligned} \quad (9)$$

where $p_m \in \mathcal{P}_m$ is a pure strategy, $\mathcal{P}_{-m} = \times_{i \neq m} \mathcal{P}_i$ is the cartesian product of all strategy sets other than m ’s one, and, likewise, $\sigma_{-m}^*(\mathbf{p}_{-m})$ is the product of probability of the opponents’ joint strategy \mathbf{p}_{-m} , given by $\sigma_{-m}^*(\mathbf{p}_{-m}) = \prod_{i \neq m} \sigma_i^*(p_i)$, where the product stems from the independence of each player’s action with respect to the other ones. ■

Theorem 1 (Nash [17]): In every finite static game \mathcal{G} there exists at least one mixed-strategy Nash equilibrium. The proof makes use of the Brouwer-Kakutani fixed-point theorem and can be found in [7]. ■

III. Q-LEARNING-BASED ALGORITHM

Once the existence of (at least) one mixed-strategy Nash equilibrium in \mathcal{G} is assessed, we now aim at computing it. In general, there does not exist a specific algebraic method to solve mixed-strategy best response equations, and solving such problems is typically NP-hard [18], in particular when there is a huge number of strategies and players. For a two-player game, mixed-strategy Nash equilibria can be computed explicitly (e.g., see [19]). For more complex games, we can resort to *learning methods* [20], that are able to let the players interact so that they can learn about the game and gather information about each other in the course of playing. In particular, we can resort to *reinforcement learning* techniques [21], well-suited for multi-agent systems, where agents know little about the other agents. To cache the results of each player’s experience,

¹Note that, although all relay nodes earn the same payoff for a given power allocation \mathbf{p} , the optimal transmit powers p_n^* for $n \in \mathcal{N}$ are in general not equal, due to the different channel link conditions.

most reinforcement learning methods use *Q-value* functions $Q_m^{t=0}(\mathbf{p})$, computed at time step t for each player m and any power profile $\mathbf{p} = (p_m; \mathbf{p}_{-m}) \in \mathcal{P} = \mathcal{P}_m \times \mathcal{P}_{-m}$, that lead agents to the optimal strategy (i.e., the mixed-strategy Nash equilibrium) [13]. *Q-learning* has been deeply investigated, and possesses a firm foundation in the theory of Markov decision processes. In addition, it is also quite easy to use, and has been widely employed in many fields of application, such as communications and networking [22].

The algorithm proposed in this paper, which adapts the one derived in [23], starts with an initialization of the *Q-values* to $Q_m^{t=0}(p_m; \mathbf{p}_{-m}) = u_m(p_m; \mathbf{p}_{-m})$ for all $m \in \mathcal{M}$ and for all $\mathbf{p} \in \mathcal{P}$. Then, at the beginning of each time step t , each agent m individually updates the table of probabilities $\pi_m^t = [\pi_m^t(\mathbf{p})]$ for all $\mathbf{p} \in \mathcal{P}_m \times \mathcal{P}_{-m}$ using

$$\pi_m^t(p_m; \mathbf{p}_{-m}) = \frac{\exp\{\Omega_m^t(p_m; \mathbf{p}_{-m}) / T_m\}}{\sum_{p_m \in \mathcal{P}_m} \sum_{\mathbf{p}_{-m} \in \mathcal{P}_{-m}} \exp\{\Omega_m^t(p_m; \mathbf{p}_{-m}) / T_m\}} \quad (10)$$

wherein $\Omega_m^t(p_m; \mathbf{p}_{-m}) = \sum_{\ell=0}^t (\delta_m)^\ell \cdot u_m(p_m; \mathbf{p}_{-m})$ is the discounted reward for taking joint action $(p_m; \mathbf{p}_{-m})$ by the user m in time step t ; $\delta_m \in [0, 1]$ is a “discount factor”; and the parameter T_m is a function which provides a randomness component to control exploration and exploitation of the actions (see [14] for more details).

After this update, each agent m chooses the best joint strategy $\tilde{\mathbf{p}}_m$, according to

$$\tilde{\mathbf{p}}_m = \arg \max_{\mathbf{p} \in \mathcal{P}} \pi_m^t(\mathbf{p}) \quad (11)$$

and finally updates its own *Q-value* according to the recursion

$$Q_m^{t+1}(\tilde{\mathbf{p}}_m) \leftarrow (1 - f_m^{t+1}) \cdot Q_m^t(\tilde{\mathbf{p}}_m) + f_m^{t+1} \cdot \left(u_m(\tilde{\mathbf{p}}_m) + \delta_m \cdot Q_m^t(\tilde{\mathbf{p}}_m) \cdot \prod_{i=1}^M \pi_i^t(\tilde{\mathbf{p}}_m) \right) \quad (12)$$

where f_m^t is the *learning rate*, which is a function of t [24].

Hu and Wellman in [23] proves that the *Q-value* updates (12) converge for all agents $m \in \mathcal{M}$ in the long-run with probability one. When this occurs, the profile $\sigma_m^* = [\sigma_m^*(p_m)]_{p_m \in \mathcal{P}_m}$, with elements

$$\sigma_m^*(p_m) = \sum_{\mathbf{p}_{-m} \in \mathcal{P}_{-m}} \pi_m^t(p_m; \mathbf{p}_{-m}) \quad (13)$$

for all $p_m \in \mathcal{P}_m$, coincides with a mixed-strategy Nash equilibrium of the game \mathcal{G} . For the reader’s convenience, the proposed algorithm is summarized in Table I. Note that, if there exists some agent $m \in \mathcal{M}$, such that its initial *Q-values* $Q_m^{t=0}(p_m; \mathbf{p}_{-m}) = u_m(p_m; \mathbf{p}_{-m}) = 0$ for all $\mathbf{p} \in \mathcal{P}$, then its strategy set becomes empty. In this case, the problem is declared to be unfeasible, as the network resources are not enough to accommodate all users given their QoS constraints and the channel realizations.

Initialization:

```

for every  $m \in \mathcal{M}$  do
  for every  $\mathbf{p} \in \mathcal{P} = \mathcal{P}_m \times \mathcal{P}_{-m}$  do
    set  $Q_m^0(\mathbf{p}) = u_m(\mathbf{p})$ ;
  end for
end for

```

Feasibility check:

```

if  $(\exists m \in \mathcal{M} \text{ s.t. } Q_m^0(\mathbf{p}) = 0 \forall \mathbf{p} \in \mathcal{P})$ 
then exit;
else set  $t = 0$  and a tolerance  $\varepsilon \ll 1$ ;

```

Loop:

```

repeat
  {updating probabilities}
  for every  $m \in \mathcal{M}$  do
    for every  $\mathbf{p} \in \mathcal{P} = \mathcal{P}_m \times \mathcal{P}_{-m}$  do
      update  $\pi_m^t(\mathbf{p})$  using (10);
    end for
    compute  $\tilde{\mathbf{p}}_m$  using (11);
  end for
  {updating Q-values}
  for every  $m \in \mathcal{M}$  do
    update  $Q_m^{t+1}(\tilde{\mathbf{p}}_m)$  using (12);
  end for
  update  $t = t + 1$ ;
until  $\max_{m \in \mathcal{M}} |Q_m^t(\tilde{\mathbf{p}}_m) - Q_m^{t-1}(\tilde{\mathbf{p}}_m)| \leq \varepsilon$ 

```

Output:

```

compute  $\sigma_m^*(p_m)$  using (13).

```

Table I: Q-learning-based algorithm.

IV. NUMERICAL RESULTS

In this section, we show the performance of the proposed algorithm to control the power in a multiple-relay-aided communication network scenario, and compare it with that of well known power allocation schemes, namely: *i*) Nash bargaining solution (NBS) [25]; *ii*) max-min fairness solution [26]; and *iii*) max rate solution [27]. Throughout the simulations, unless otherwise specified, we make use of the following system parameters: $\Delta f = 10.938$ kHz, $\sigma_w^2 = 10$ nW, $p_c = 100$ mW [15], and $\bar{p}_m = 1$ W for all $m \in \mathcal{M}$. For simplicity, we set $\underline{R}_s = 1$ kb/s for all source nodes $s \in \mathcal{S}$. The distances of relays and source nodes from the base stations are assumed to be uniformly distributed between 10 and 100 m, the path loss exponent is set to $\zeta = 4$, and a 24-tap channel model [28] is used to reproduce the effects of shadowing and scattering. We also set $f_m^t = t^{-0.8}$, $\delta_m = 0.85$, and $T_m = 10^{-2} \cdot \bar{u}_m \cdot \exp\{-10^{-2} \cdot \bar{u}_m \cdot t\}$, with $\bar{u}_m = \max_{\mathbf{p} \in \mathcal{P}} u_m(\mathbf{p})$, that provide a good tradeoff between EE and SE, based on an exhaustive search [14], not reported here for the sake of brevity. To reduce the computational burden of the proposed algorithm, which is exponentially increasing with the number of power steps $K + 1$, we select the cases $K = 1$, corresponding to the situation in which the sum-rate is maximized and the computational load is the minimum one, at the cost of a reduced EE, and $K = 3$, which provides

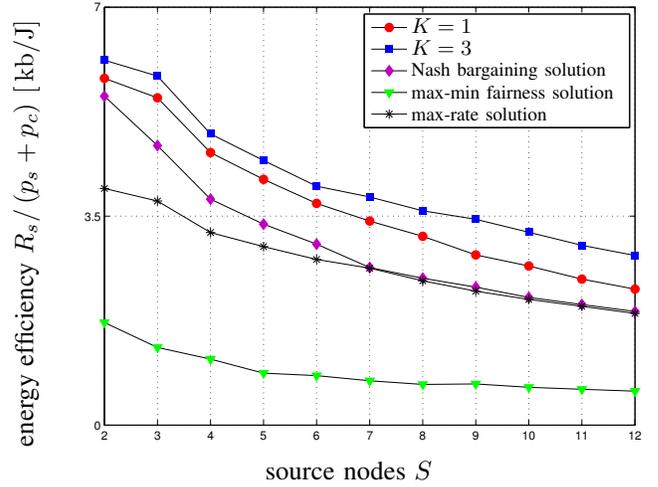


Fig. 2: Average source node's EE as a function of S ($N = 4$).

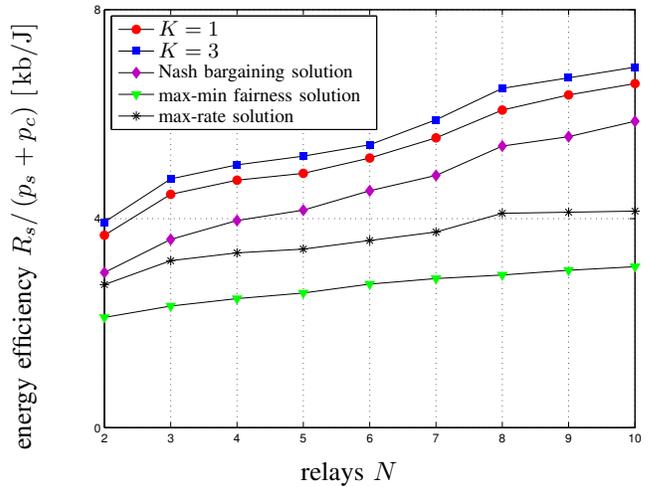


Fig. 3: Average source node's EE as a function of N ($S = 4$).

an interesting tradeoff between SE, EE, and computational complexity of the algorithm. Note that, using such values of K , the proposed algorithm converges after a few iteration steps (typically, less than 3) [14].

In Figs. 2-4, we will compare the performance of our proposed algorithm, using $K = 1$ (circles) and $K = 3$ (squares), with the following optimization techniques, formalized below, and depicted by diamonds, lower triangles, and asterisks, respectively:²

$$\text{NBS fairness: } \max_{p_m \in [0, \bar{p}_m]} \prod_{s \in \mathcal{S}} \frac{R_s - \underline{R}_s}{p_s + p_c} \quad (14a)$$

$$\text{max-min fairness: } \max_{p_m \in [0, \bar{p}_m]} \min_{s \in \mathcal{S}} \frac{R_s}{p_s + p_c} \quad (14b)$$

$$\text{max rate: } \max_{p_m \in [0, \bar{p}_m]} \sum_{s \in \mathcal{S}} R_s \quad (14c)$$

²Throughout the simulations, and unlike our proposed algorithm, the transmit power set here is assumed to be continuous in $[0, \bar{p}_m]$.

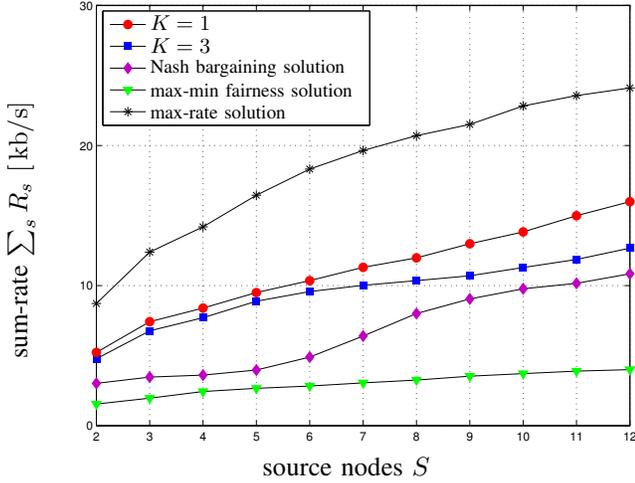


Fig. 4: Average sum-rate as a function of S ($N = 4$).

Figs. 2 and 3 report the average EE as functions of the number of source nodes S , and the number of relays N , respectively. As expected, the EE is decreasing with S and increasing with N . When N is fixed, increasing S increases the multiple access interference (MAI), thus reducing the EE. On the contrary, increasing N while S is constant increases the sum-rate in the long run, without additional power expenditure at the source side. Note that the case $K = 3$ outperforms the case $K = 1$. However, even in the extreme case $K = 1$ (i.e., each node selects either zero power or its maximum one), the proposed algorithm outperforms the well-known solutions (14). Similar conclusions can be drawn for the average SE, reported in Fig. 4, where the average sum-rate is reported as a function of S . Obviously, by definition the maximum sum-rate is given (14c), and confirmed by our numerical results. However, the performance gap paid by the proposed algorithm is still acceptable, especially when compared with the performance gap achieved by NBS and max-min approaches.

To measure the improvement achieved by including the relay powers into the resource allocation problem, which, to the best of our knowledge, is not present in the available literature, we now compare the proposed algorithm, where $\mathcal{M} = \mathcal{S} \cup \mathcal{N}$, with the same scheme, but with $\mathcal{M} = \mathcal{S}$, while each relay nodes $n \in \mathcal{N}$ adopts the same power level $p_n = \bar{p}_n = 1$ W at all steps of the algorithm. For the sake of completeness, we also report the comparison between the source-plus-relay versus source-only power control schemes for the NBS, max-min fairness, and max-rate solutions. Fig. 5 reports the percentage of improvement in terms of average EE at the mixed-strategy Nash equilibrium of the game, where u_s^* and \tilde{u}_s^* represent the source's average utilities achieved by the proposed algorithm using $\mathcal{M} = \mathcal{S} \cup \mathcal{N}$ and $\mathcal{M} = \mathcal{S}$, respectively, as a function of N , using $S = 4$. As can be seen, apart from the max-min fairness case,³ allowing the relays to

³Here, we fall into a Braess-type paradox, already observed in other different contexts (e.g., [29], [30]).

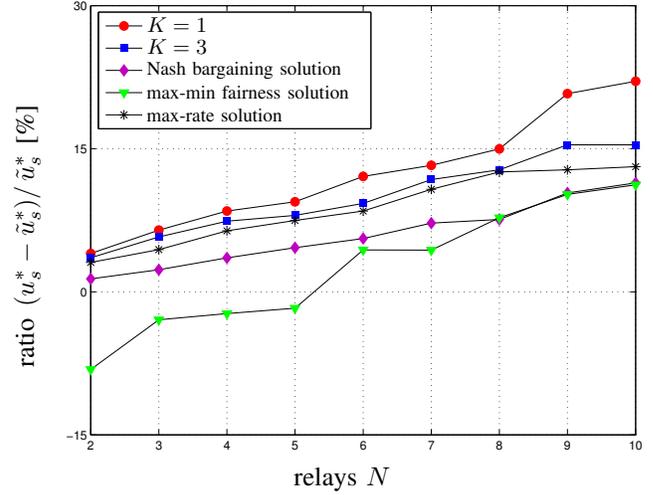


Fig. 5: Percentage of EE improvement as a function of N compared to the case $p_n = 1$ W for all $n \in \mathcal{N}$ ($S = 4$).

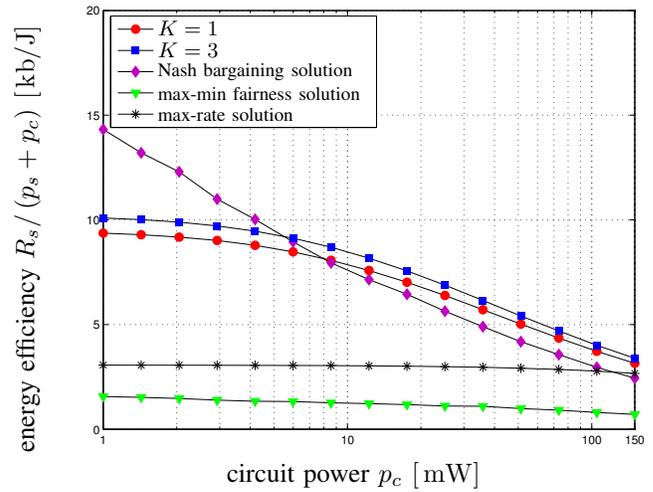


Fig. 6: Average source node's EE as a function of p_c ($S = 6$, $N = 4$).

regulate their transmit powers jointly with the source nodes is always beneficial in terms of both average EE and average sum-rate. This is particularly apparent for the case $K = 3$ when S increases, but the same behavior can also be observed for all other curves. More importantly, the proposed source-plus-relay approach outperforms not only the source-only case, as the improvement is always positive, but also both versions (i.e., source-plus-relay and source-only) at all other schemes.

Finally, let us investigate the impact of the circuit power p_c on the performance of the resource allocation techniques introduced above. As mentioned in Sect. II, p_c depends upon characteristics of electrical architecture of the nodes. Using some results available in the literature (e.g., [16], [31]), it is known that the EE increases as p_c decreases. To verify it, Fig. 6 reports the average EE as a function of p_c for a network consisting of $S = 6$ source nodes and $N = 4$ relay nodes.

As can be seen, p_c has a weak impact on max-min fairness and max-rate solutions, whose performance in terms of EE is always lower than that achieved by the technique illustrated in Sect. III. On the contrary, the NBS outperforms the proposed algorithm for low values of p_c (in this particular scenario, when $p_c < 7$ mW). However, note that such small values are not yet available in practical systems [32], and typical values of p_c are in the order of 100 mW [15], [32], wherein the proposed algorithm shows significantly higher performance.

V. CONCLUSION

In this work, we considered a reliable wireless communication network consisting of multiple sources, multiple parallel relays, and one destination, in which each active node is modeled as a player in a non-cooperative finite game, whose interest is formulated as follows: each source node is a self-interested player which aims at maximizing its own energy efficiency; and each relay node is an altruistic player which aims at maximizing the network sum-rate with constraints on the minimum rates of all sources. The mixed-strategy Nash equilibrium points of the game are then computed using a Q -learning-based algorithm, which leads the players to compute a probability distribution among all available pure strategies.

Numerical results show that our proposed technique, that uses a low-complexity algorithm, outperforms the Nash bargaining solution and the max-min fairness approach in terms of both energy efficiency and network sum-rate, and significantly outperforms the max-rate solution in terms of energy efficiency, which is our major performance metric, while paying a tolerable performance gap in terms of network sum-rate. Further work is needed to assess the feasibility of the problem given a particular network realization, and to extend the formulation of the problem to a multicarrier system, so that the additional degrees of freedom may improve the both the energy and the spectral efficiency of the network.

REFERENCES

- [1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
- [2] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.
- [3] Z. Sahinoglu and P. Orlik, "Regenerator versus simple-relay with optimum transmit power control for error propagation," *IEEE Commun. Letters*, vol. 7, no. 9, pp. 416–418, Sept. 2003.
- [4] I. Hammerström, M. Kuhn, and A. Wittneben, "Impact of relay gain allocation on the performance of cooperative diversity networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 3, Los Angeles, CA, Sept. 2004, pp. 1815–1819.
- [5] D. Wang, X. Wang, and X. Cai, "Optimal power control for multi-user relay networks over fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 199–207, 2011.
- [6] J. Paredes and A. Gershman, "Relay network beamforming and power control using maximization of mutual information," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4356–4365, Dec. 2011.
- [7] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.
- [8] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 26, no. 7, pp. 1226–1237, Sept. 2008.
- [9] S. Ren and M. van der Schaar, "Pricing and distributed power control in wireless relay networks," *IEEE Trans. Signal Processing*, vol. 59, no. 6, pp. 2913–2926, June 2011.
- [10] Q. Cao, H. V. Zhao, and Y. Jing, "Power allocation and pricing in multiuser relay networks using Stackelberg and bargaining games," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3177–3190, Sept. 2012.
- [11] H. Khayatian, R. Saadat, and J. Abouei, "Coalition-based approaches for joint power control and relay selection in cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 835–842, Feb. 2013.
- [12] I. Stupia, L. Vandendorpe, L. Sanguinetti, and G. Bacci, "Distributed energy-efficient power optimization for relay-aided heterogeneous networks," in *Intl. Workshop Wireless Networks: Communication, Cooperation and Competition*, Hammamet, Tunisia, May 2014, submitted.
- [13] C. J. C. H. Watkins and P. Dayan, "Q-Learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.
- [14] F. Shams, G. Bacci, and M. Luise, "Game-theoretic power control for multiple-relay cooperative networks," *IEEE Trans. Wireless Commun.*, 2014, in preparation.
- [15] G. Miao, N. Himayat, and G. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.
- [16] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.
- [17] J. F. Nash, "Equilibrium points in n -person games," *Proc. National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, Jan. 1950.
- [18] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a Nash equilibrium," in *Proc. ACM Symp. Theory of Computing*, Seattle, WA, May 2006.
- [19] R. Porter, E. Nudelman, and Y. Shoham, "Simple search methods for finding a Nash equilibrium," in *Proc. Nat. Conf. Artificial Intelligence (AAAI)*, San Jose, CA, July 2004.
- [20] R. B. Dillts and T. A. Epstein, *Dynamic Learning*. Capitola, CA: MeTa Publications, 1995.
- [21] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
- [22] S. Singh and D. Bertsekas, "Reinforcement learning for dynamic channel allocation in cellular telephone systems," *Advances in Neural Information Processing Systems*, vol. 9, pp. 974–980, 1997.
- [23] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Machine Learning Research*, vol. 4, pp. 1039–1069, 2003.
- [24] E. Even-Dar and Y. Mansour, "Learning rates for Q-learning," *J. Machine Learning Research*, vol. 5, pp. 1–25, Dec. 2004.
- [25] T. Sanguanpuak and R. M. A. P. Rajatheva, "Power bargaining for amplify and forward relay channel," in *Proc. Intl. Conf. Commun. and Networking in China (ChinaCOM)*, Xi'an, China, Aug. 2009.
- [26] K. Phan, T. Le-Ngoc, S. Vorobyov, and C. Tellambura, "Power allocation in wireless multi-user relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2535–2545, May 2009.
- [27] J. Zhang, F. Roemer, M. Haardt, A. Khabbazi-basmenj, and S. Vorobyov, "Sum rate maximization for multi-pair two-way relaying with single-antenna amplify and forward relays," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.
- [28] ITU Radiocommunication Sector (ITU-R), "Guidelines for evaluation of radio transmission technology for IMT-2000," *Recommendation ITU-R M.1225*, 1997.
- [29] E. Altman, V. Kamble, and H. Kameda, "A Braess type paradox in power control over interference channels," in *Proc. Int. Symp. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Berlin, Germany, Apr. 2008.
- [30] G. Bacci, L. Sanguinetti, M. Luise, and H. V. Poor, "Energy-efficient power control for contention-based synchronization in OFDMA systems with discrete powers and limited feedback," *EURASIP J. Wireless Communications and Networking*, vol. 2013:192, no. 1, July 2013.
- [31] G. Bacci, E. V. Belmega, P. Mertikopoulos, and L. Sanguinetti, "Energy-aware competitive link adaptation in small-cell networks," in *Proc. Intl. Workshop Resource Allocation in Wireless Networks*, Hammamet, Tunisia, May 2014.
- [32] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?" in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014.