



**HAL**  
open science

## Modeling Noisy Feedback in Decentralized Self-Configuring Networks

Samir M. Perlaza, Merouane Debbah

► **To cite this version:**

Samir M. Perlaza, Merouane Debbah. Modeling Noisy Feedback in Decentralized Self-Configuring Networks. Asilomar Conference on Signals, Systems, and Computers, Nov 2011, Pacific Grove, United States. 5p. hal-00651612

**HAL Id: hal-00651612**

**<https://centralesupelec.hal.science/hal-00651612>**

Submitted on 13 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modeling Noisy Feedback in Decentralized Self-Configuring Networks

S.M. Perlaza and M. Debbah

Alcatel Lucent Chair in Flexible Radio - SUPELEC. France  
 {samir.medina-perlaza, merouane.debbah}@supelec.fr

**Abstract**—This paper introduces a generalization of the notion of Nash equilibrium (NE), namely **quantal response equilibrium (QRE)**. In the QRE, radio devices choose their transmit/receive configuration taking into account that the estimation of their own performance contains a noise component. Here, it is shown that the notion of QRE neatly models decentralized self-configuring networks (DCSN) where feedback messages are impaired by quantization noise or decoding errors. The main contribution of the paper is twofold. First, we show that under the presence of noise in the estimation expected utility, the notion of NE no longer holds, as players cannot be considered rational. Second, we introduce a learning technique that converges to a QRE in a fully decentralized fashion. We present numerical results in the context of a channel selection problem in a parallel multiple access channel in order to illustrate our theoretical results.

## I. INTRODUCTION

Game theory is by now accepted as one of most suited theories for the study of decentralized self-configuring networks (DCSN) [1]. In particular, the concept of Nash equilibrium (NE) [2] has played a key role in the performance analysis of DCSN. The interest on NE relies mainly on two reasons. First, it is a network state at which the transmit configuration of each radio device is optimal with respect to the configurations adopted by all the other devices. Second, the NE can be the result of the interaction of several radio devices following particular behavioral rules, for instance, best response dynamics, fictitious play, etc, [3]. From this point of view, radio device or network designers can develop algorithms that allow DCSN to achieve NE in a fully decentralized way [1]. Nonetheless, the application of the notion of NE in DCSN is largely constrained by several real-implementation constraints. As a first constraint, we identify the fact that radio devices are assumed rational. That is, radio devices must be capable of reliably determining the performance they achieve by playing each of their actions (transmit configurations). This assumption clearly does not always hold, since often this procedure involves feedback messages that are subject to quantization and decoding errors. Thus, the assumption of rational radio devices simply fails. Some

references, where rationality is the main assumption, are for instance [4], [5] in the analysis of interference channels and [6], [7] in the analysis of the multiple access channels.

In this article, we introduce the idea of quantal response equilibrium (QRE) [8], [9], which can be considered as a generalization of the notion of NE. At the QRE, players choose their actions taking into account that the observed utility contains an unknown noise component. This implies that players are not always able to choose the action that maximizes their individual utility. Indeed, the noise component might lead them to choose any of their feasible actions. As a result, the assumption of rationality is no longer needed, which makes this notion of equilibrium better suited for DCSN. The notion of QRE can be specialized for certain particular types of noise. For instance, when the noise component added to the observed average utility follows a Gumble distribution, the QRE is reduced to the notion of logit equilibrium (LE) [10] already used in [11] in the context of spectrum sharing. Another special case, the probit equilibrium, appears when the noise is Gaussian distributed [10].

The main contribution of this article is a game theoretical framework based on the notion of QRE to study DCSN when radio devices are not fully rational due to the imperfect estimation of their individual performance. This comprehends mainly the case of noisy feedback. We also show that the QRE, as the NE, can be learnt by radio devices in a fully decentralized manner. We provide some behavioral rules than convergence to QRE.

The sequel of this paper unfolds as follows. In Sec. II, we present the general system model which is later described by a one-shot game. In Sec. III, we introduce the notion of QRE. In Sec. IV, we present a behavioral rule that allows players to achieve QRE in the general context of DCSN. In Sec. V, we present some numerical examples to illustrate our results in the context of a parallel multiple access channel. Finally, in Sec. VI, we state our conclusions.

## II. NOTATION AND GAME THEORETIC MODEL

### A. Notation

In the sequel of this paper, vectors and scalars are respectively denoted by boldface lower case symbols and lower case symbols. The expectation operator with respect to a random variable  $x$  is denoted by  $\mathbb{E}_x[\cdot]$ . The indicator function is denoted by  $\mathbb{1}_{\{condition\}}$  and it equals 1 (resp. 0) when *condition* is true (resp. false). The set of non-negative and strictly positive integers is denoted by  $\mathbb{N} = \{0, 1, 2, \dots\}$  and  $\mathbb{N}_1 = \mathbb{N} \setminus \{0\}$ , respectively. The space of real vectors of dimension  $n \in \mathbb{N}$  is denoted by  $\mathbb{R}^n$ . Let  $\mathcal{X}$  be a finite set. We denote by  $\Delta(\mathcal{X})$  the unit simplex over the elements of  $\mathcal{X}$ , that is, the set of all probability distributions over the elements of the set  $\mathcal{X}$ . The cardinality of  $\mathcal{X}$  is denoted by  $|\mathcal{X}| \in \mathbb{N}$  and the set of all subsets of  $\mathcal{X}$  including the set  $\mathcal{X}$  itself is denoted by  $2^{\mathcal{X}}$ . We denote by  $\{e_1^{(N)}, \dots, e_N^{(N)}\}$  the set of vectors of the canonical base spanning the space of the  $N$ -dimensional real vectors. Here,  $\forall n \in \{1, \dots, N\}$ ,  $e_n^{(N)} = (e_{n,1}^{(N)}, \dots, e_{n,N}^{(N)})$ , and  $\forall s \in \{1, \dots, N\} \setminus \{n\}$ ,  $e_{n,s}^{(N)} = 0$  and  $e_{n,n}^{(N)} = 1$ . Given a vector  $\mathbf{a} = (a_1, \dots, a_N)$  in a given space of dimension  $N \in \mathbb{N}$ , we denote by  $\mathbf{a}_{-n}$ , with  $n \in \{1, \dots, N\}$ , the vector  $\mathbf{a}_{-n} = (a_1, \dots, a_{n-1}, a_{n+1}, \dots, a_N)$  in the corresponding space of dimension  $N - 1$ . With a slight abuse of notation, we often write  $\mathbf{a} = (a_n, \mathbf{a}_{-n})$  to highlight the  $n$ -th component of the vector  $\mathbf{a}$ .

### B. Game Theoretical Model

Assume that the game in normal form  $\mathcal{G} = (\mathcal{K}, \{\mathcal{A}_k\}_{k \in \mathcal{K}}, \{u_k\}_{k \in \mathcal{K}})$  models a DSCN. The set  $\mathcal{K} = \{1, \dots, K\}$  contains the indexes of all transmit-receiver pairs. A transmitter-receiver pair  $k$ , with  $k \in \mathcal{K}$ , chooses its transmit/receive configuration out of a set of  $N_k$  possible choices. We denote the  $n$ -th choice by  $A_k^{(n)}$  and the set of all possible choices by  $\mathcal{A}_k = \{A_k^{(1)}, \dots, A_k^{(N_k)}\}$ . Here, a transmit/receive configuration describes, for instance, a particular power allocation policy, modulation-coding scheme, decoding order, scheduling policy or any combination of these parameters. In the following, we refer to the transmit-receiver pairs and their corresponding transmit/receive configurations as players and actions, respectively. An action profile of the game  $\mathcal{G}$  is a vector  $\mathbf{a} = (a_1, \dots, a_K) \in \mathcal{A}$ , where  $\mathcal{A} \triangleq \mathcal{A}_1 \times \dots \times \mathcal{A}_K$ . The utility function  $u_k : \mathcal{A} \rightarrow \mathbb{R}$ , for all  $k \in \mathcal{K}$ , measures the benefit (in the sense of Morgenstern - von Neumann [12]) player  $k$  obtains when it plays a specific action given the actions adopted by all the other players. Typical utility functions in DSCN are transmission rates, bit error rates, spectral efficiency, etc. See [13] and references therein. The game  $\mathcal{G}$  is played as follows. Player  $k$  chooses its action  $a_k$  following a probability distribution  $\pi_k =$

$(\pi_{k,A_k^{(1)}}, \dots, \pi_{k,A_k^{(N_k)}}) \in \Delta(\mathcal{A}_k)$ . Here,  $\forall k \in \mathcal{K}$  and  $\forall n_k \in \{1, \dots, N_k\}$ ,  $\pi_{k,A_k^{(n_k)}}$  represents the probability that player  $k$  plays action  $A_k^{(n_k)} \in \mathcal{A}_k$ , i.e.,

$$\pi_{k,A_k^{(n_k)}} = \Pr(a_k = A_k^{(n_k)}). \quad (1)$$

A vector  $\pi_k$  is often called mixed strategy of player  $k$ . The expected utility  $\bar{u}_k : \Delta(\mathcal{A}_1) \times \dots \times \Delta(\mathcal{A}_K) \rightarrow \mathbb{R}$  of player  $k$  is defined as follows,

$$\bar{u}_k(\pi_k, \pi_{-k}) = \sum_{\mathbf{a} \in \mathcal{A}} \left( \prod_{j=1}^K \pi_{j,a_j} \right) u_k(a_k, \mathbf{a}_{-k}). \quad (2)$$

The information available for player  $k$  is a vector,

$$\bar{\mathbf{u}}_k(\cdot, \pi_{-k}) = \left( \bar{u}_k(e_1^{(N_k)}, \pi_{-k}), \dots, \bar{u}_k(e_{N_k}^{(N_k)}, \pi_{-k}) \right), \quad (3)$$

where,  $\bar{u}_k(e_m^{(N_k)}, \pi_{-k})$  is the utility that player  $k$  hope to achieve by playing the action  $A_k^{(m)}$  when all the other players play the mixed strategies  $\pi_{-k}$ . However, the actual utility achieved by player  $k$  by playing action  $A_k^{(m)}$  is a random variable  $\tilde{u}_k(e_{k,m}, \pi_{-k})$  obtained as follows,

$$\tilde{u}_k(e_{k,m}, \pi_{-k}) = \bar{u}_k(e_{k,m}, \pi_{-k}) + \varepsilon_{k,A_k^{(m)}}, \quad (4)$$

where,  $\forall k \in \mathcal{K}$  and  $\forall m \in \{1, \dots, N_k\}$ ,  $\varepsilon_{k,A_k^{(m)}}$  is the realization of a random variable which represents an additive noise on the expected utility  $\bar{u}_k$ . Sources of such noise are the quantification of the control variables such as transmit power, or decoding errors on the feedback messages used to build the vector  $\bar{\mathbf{u}}_k(\cdot, \pi_{-k})$ . In this work, it is assumed that the additive error is independently and identically distributed (i.i.d) for all the players.

The behavioral assumption is that, player  $k$  must choose the action  $A_k^{(n_k)}$ , where  $n_k$  satisfies

$$n_k \in \arg \max_{m \in \{1, \dots, N_k\}} \tilde{u}_k(e_{k,m}, \pi_{-k}), \quad (5)$$

given that all the other players play the complementary mixed strategy profile  $\pi_{-k}$ . It is very important to note that such a maximization problem must be carried out without knowing the exact realization of the noise component. Note also that in the absence of the noise component, this behavioral rule corresponds to a fictitious play [14].

Under these assumptions, it becomes clear that players are unable to choose the action that maximizes the truth utility  $\bar{u}_k$ . Indeed, the noisy component can lead players to take any of their actions. In the following section, we introduce the game theoretical tools tackling such scenario.

### III. EQUILIBRIUM NOTIONS

In this section, we introduce the notion of QRE and a special case known as logit equilibrium.

#### A. Quantal Response Equilibrium

For doing so, we first focus on the determination of the mixed strategy  $\pi_k$  that player  $k$  must use to select its action  $a_k$  in order to maximize its actual utility function  $\tilde{u}_k$  given a vector  $\bar{\mathbf{u}}_k = (\bar{u}_k(e_{k,1}, \boldsymbol{\pi}_{-k}), \dots, \bar{u}_k(e_{k,N_k}, \boldsymbol{\pi}_{-k}))$ . Thus, consider the vector  $\boldsymbol{\xi}_k = (\xi_{k,A_k^{(1)}}, \dots, \xi_{k,A_k^{(N_k)}}) \in \mathbb{R}^{N_k}$  and let  $\xi_{k,A_k^{(n_k)}}$ , with  $n_k \in \{1, \dots, N_k\}$ , be distributed according to a joint distribution with density function  $f_k$ , where the marginal of  $f_k$  exists for each  $\xi_{k,A_k^{(n_k)}}$ , with  $n_k \in \{1, \dots, N_k\}$  and  $\mathbb{E}[\xi_{k,A_k^{(n_k)}}] = 0$ . When  $f_k$  satisfies these conditions,  $f_k$  is said to be *admissible* [9]. Now, assuming that the truly achieved utility follows the form of (4), the best response can be obtained as follows. Let  $\mathbf{E}_{k,A_k^{(n_k)}} : \mathbb{R}^{N_k} \rightarrow \mathbb{R}^{N_k}$  specify the set of errors that lead player  $k$  to choose action  $A_k^{(n_k)}$ , for all  $k \in \mathcal{K}$  and for all  $n_k \in \{1, \dots, N_k\}$ . Hence, given a vector  $\bar{\mathbf{u}}_k(\cdot, \boldsymbol{\pi}_{-k}) = (\bar{u}_k(e_{k,1}, \boldsymbol{\pi}_{-k}), \dots, \bar{u}_k(e_{k,N_k}, \boldsymbol{\pi}_{-k}))$ , it yields

$$\mathbf{E}_{k,A_k^{(n_k)}}(\bar{\mathbf{u}}_k(\cdot, \boldsymbol{\pi}_{-k})) = \left\{ \zeta_k \in \mathbb{R}^{N_k} : \forall m \in \{1, \dots, N_k\}, \bar{u}_k(e_{k,n_k}, \boldsymbol{\pi}_{-k}) + \zeta_{k,A_k^{(n_k)}} \geq \bar{u}_k(e_{k,m}, \boldsymbol{\pi}_{-k}) + \zeta_{k,A_k^{(m)}} \right\}.$$

Thus, the probability distribution over the set of actions  $\mathcal{A}_k$  which maximizes the chances that player  $k$  chooses the action which brings the highest (expected) utility based on the vector  $\bar{\mathbf{u}}_k(\cdot, \boldsymbol{\pi}_{-k})$ , can be determined by using the function  $\beta_k : \Delta(\mathcal{A}_1) \times \dots \times \Delta(\mathcal{A}_{k-1}) \times \Delta(\mathcal{A}_{k+1}) \times \dots \times \Delta(\mathcal{A}_K) \rightarrow \Delta(\mathcal{A}_k)$ , where, given an explicit form for the expected utility  $\bar{\mathbf{u}}_k$ , it follows that  $\beta_k(\boldsymbol{\pi}_{-k}) = (\beta_{k,A_k^{(1)}}(\boldsymbol{\pi}_{-k}), \dots, \beta_{k,A_k^{(N_k)}}(\boldsymbol{\pi}_{-k}))$  and  $\forall n \in \{1, \dots, N_k\}$ ,

$$\beta_{k,A_k^{(n)}}(\boldsymbol{\pi}_{-k}) = \int_{\zeta \in \mathbf{E}_{k,A_k^{(n_k)}}(\bar{\mathbf{u}}_k(\cdot, \boldsymbol{\pi}_{-k}))} f_k(\zeta) d\zeta. \quad (6)$$

Such probability distribution  $\beta_k(\boldsymbol{\pi}_{-k})$  is known as the *statistical reaction function* (SRF) or *quantal response function* (QRF) [9].

Finally, we define the QRE as follows.

*Definition 1 (Quantal Response Equilibrium):*

The strategy profile  $\boldsymbol{\pi}^* = (\boldsymbol{\pi}_1^*, \dots, \boldsymbol{\pi}_K^*) \in \Delta(\mathcal{A}_1) \times \dots \times \Delta(\mathcal{A}_K)$  is a *quantal response equilibrium* of the game  $\mathcal{G} = (\mathcal{K}, \{\mathcal{A}_k\}_{k \in \mathcal{K}}, \{u_k\}_{k \in \mathcal{K}})$ , if for all  $k \in \mathcal{K}$ ,

$$\boldsymbol{\pi}_k^* = \beta_k(\boldsymbol{\pi}_{-k}^*). \quad (7)$$

It is important to remark that Def. 1 implies a fixed point equation. For instance, let  $\zeta : \Delta(\mathcal{A}_1) \times \dots \times$

$\Delta(\mathcal{A}_K) \rightarrow \Delta(\mathcal{A}_1) \times \dots \times \Delta(\mathcal{A}_K)$  be defined as follows,

$$\zeta(\boldsymbol{\pi}) = (\beta_1(\boldsymbol{\pi}_{-1}), \dots, \beta_K(\boldsymbol{\pi}_{-K})). \quad (8)$$

Then, if  $\boldsymbol{\pi}^*$  is a logit equilibrium it holds that  $\boldsymbol{\pi}^* = \zeta(\boldsymbol{\pi}^*)$ . This observation leads to the following result.

*Theorem 2 (Existence of the LE [9]):* The game  $\mathcal{G}$  has at least one QRE.

The proof of Theorem 2 (in [9]) relies on the fact that the mapping  $\zeta$  is a continuous mapping from a compact convex set into itself. Thus, by Brouwer's fixed point theorem, there exists at least one  $\boldsymbol{\pi}^*$  such that  $\boldsymbol{\pi}^* = \zeta(\boldsymbol{\pi}^*)$ . In the following, we deal with a particular case of QRE, known as logit equilibrium [10].

#### B. Logit Equilibrium

Assume now that the noise components  $\xi_{k,A_k^{(1)}}, \dots, \xi_{k,A_k^{(N_k)}}$  follow a Gumbel distribution (GD) with location and shape parameters  $\mu \in \mathbb{R}$  and  $\frac{1}{\gamma_k} \in \mathbb{R}_+$ , respectively. That is,

$$\Pr(\xi_{k,A_k^{(k)}} < z) = \exp(-\exp(\gamma_k(\mu - z))). \quad (9)$$

The relevance of the GD is that it models the maximum value of a set of independent and identically distributed random variables. From (6), it follows that the SRF can be written as follows,

$$\beta_{k,A_k^{(n_k)}}^{(\gamma_k)}(\boldsymbol{\pi}_{-k}) = \frac{\exp(\gamma_k \bar{u}_k(e_k^{(n_k)}, \boldsymbol{\pi}_{-k}))}{\sum_{m=1}^{N_k} \exp(\gamma_k \bar{u}_k(e_k^{(m)}, \boldsymbol{\pi}_{-k}))}, \quad (10)$$

where the index  $(\gamma_k)$  is used to differentiate the particular case from the general case in (6). In the next section, we provide some behavioral rules that allow achieving QRE in a fully decentralized fashion.

### IV. LEARNING QRE

In the following, we basically focus on the design of the behavioral rules that allow radio devices to achieve a logit equilibrium. Note that this assumes that the noise component in (4) follows a Gumbel distribution. Nonetheless, the same analysis hold for other assumption on the noise distribution as long as the density  $f_k$  aforementioned is admissible. Before we start, let the logit function  $\tilde{\beta} : \mathbb{R}^{N+1} \rightarrow \Delta(\mathcal{A}_k)$ , with parameter  $x \in \mathbb{R}^+$ , be defined as follows,

$$\tilde{\beta}(x, \mathbf{y}) = (\tilde{\beta}_1(x, \mathbf{y}), \dots, \tilde{\beta}_N(x, \mathbf{y})), \quad (11)$$

with  $\tilde{\beta}_s(x, \mathbf{y}) = \frac{\exp(x y_s)}{\sum_{m=1}^{N_K} \exp(x y_m)}$ ,  $\mathbf{y} = (y_1, \dots, y_N)$ ,  $s \in \{1, \dots, N\}$  and  $N \in \mathbb{N}$ . We relate the smoothed best

response in (10) with (11), as follows,

$$\beta_k^{(\gamma_k)}(\boldsymbol{\pi}_{-k}) = \tilde{\beta}(\gamma_k, \bar{\mathbf{u}}_k(\cdot, \boldsymbol{\pi}_{-k})). \quad (12)$$

Using these elements, our main result in this section is summarized in the following theorem.

*Theorem 3: Let the following algorithm be defined, for all  $k \in \mathcal{K}$  and for all  $n_k \in \{1, \dots, N_k\}$  and  $n \in \mathbb{N}_1$ , as follows,*

$$\begin{aligned} \bar{u}_{k,A_k^{(n_k)}}(n) &= \bar{u}_{k,A_k^{(n_k)}}(n-1) + \\ &\quad \alpha_k(n) \mathbb{1}_{\{a_k(n)=A_k^{(n_k)}\}} (\tilde{u}_k(n) \\ &\quad - \bar{u}_{k,A_k^{(n_k)}}(n-1)), \\ \pi_{k,A_k^{(n_k)}}(n) &= \pi_{k,A_k^{(n_k)}}(n-1) + \\ &\quad \lambda_k(n) (\tilde{\beta}_{n_k}(\gamma_k(n), \bar{\mathbf{u}}_k(n)) \\ &\quad - \pi_{k,A_k^{(n_k)}}(n-1)), \\ \gamma_k(n) &= \gamma_k(n-1) + \nu_k(n) \Psi_k(n), \end{aligned}$$

where,  $(\gamma_k(0), \bar{\mathbf{u}}_k(0), \boldsymbol{\pi}_k(0)) \in \mathbb{R}_+ \times \mathbb{R}_+^{N_k} \times \Delta(\mathcal{A}_k)$  is an arbitrary initialization of player  $k$ ,  $\Psi_k: \mathbb{R} \rightarrow \mathbb{R}$  is a positive continuous non-decreasing function and for all  $(j, k) \in \mathcal{K}^2$ ,

$$\begin{aligned} (i) \quad &\lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_k(t) = +\infty, \quad \lim_{T \rightarrow \infty} \sum_{t=1}^T \alpha_k(t)^2 < +\infty \\ (ii) \quad &\lim_{T \rightarrow \infty} \sum_{t=1}^T \lambda_k(t) = +\infty, \quad \lim_{T \rightarrow \infty} \sum_{t=1}^T \lambda_k(t)^2 < +\infty \\ (iii) \quad &\lim_{T \rightarrow \infty} \sum_{t=1}^T \nu_k(t) = +\infty, \quad \lim_{T \rightarrow \infty} \sum_{t=1}^T \nu_k(t)^2 < +\infty \\ (iv) \quad &\lim_{n \rightarrow \infty} \frac{\lambda_j(n)}{\alpha_k(n)} = 0, \quad \lim_{n \rightarrow \infty} \frac{\nu_j(n)}{\lambda_k(n)} = 0, \end{aligned} \quad (13)$$

and either,

$$\begin{aligned} (a) \quad &\forall k \in \mathcal{K}, \quad \lambda_k = \lambda \text{ or} \\ (b) \quad &\forall k \in \mathcal{K} \setminus \{K\}, \quad \lim_{n \rightarrow \infty} \frac{\lambda_k(n)}{\lambda_{k+1}(n)} = 0. \end{aligned} \quad (14)$$

Then, if the algorithm (13) converges, it holds that,

$$\lim_{n \rightarrow \infty} \boldsymbol{\pi}_k(n) = \boldsymbol{\pi}_k^*, \quad (15)$$

$$\lim_{n \rightarrow \infty} \bar{u}_{k,A_k^{(n_k)}}(n) = \bar{u}_k(\mathbf{e}_{k,A_k^{(n_k)}}, \boldsymbol{\pi}_{-k}^*), \quad (16)$$

where  $\boldsymbol{\pi}^* = (\boldsymbol{\pi}_1^*, \dots, \boldsymbol{\pi}_N^*) \in \Delta(\mathcal{A}_1) \times \dots \times \Delta(\mathcal{A}_K)$  is a (unique) NE.

The reasoning behind the proof of Theorem 3 uses elements of stochastic approximations [15] and in particular, the notion of multiple time scales as introduced in [16] and [17]. A formal proof of Theorem 3 will be provided in a future publication. Here, we limit our analysis to some numerical examples to justify the validity of our results due to the space constraints.

Note that Theorem 3 states that the learning process converges, if it does, to a NE instead of a QRE. This effect is due to the fact that the parameter  $\gamma_k$  depends on time and it is forced to grow to infinity, which implies that the SRF asymptotically approaches a uniform distribution over the actions which maximize  $\bar{u}_k$  given  $\boldsymbol{\pi}_{-k}$ . Thus, one can state that if the noise component can be faded out in time, a NE selection process arises naturally since the NE to which the system converges is determined by the initial state of the network. This implies using additive noise to solve the problem of equilibrium selection.

If on the contrary, the noise component can not be altered, as it models external variables, then it implies that  $\Psi_k(n) = 0$ , for all  $n \in \mathbb{N}_1$ . In this case, the system converges to a logit equilibrium with parameters  $\gamma_1, \dots, \gamma_K$ , which is, in deed the result stated in [17]. In the following, we provide some numerical results to illustrate our theoretical results.

## V. NUMERICAL RESULTS

Consider a parallel multiple access channel with a set  $\mathcal{K} = \{1, \dots, K\}$  of transmitters. Each transmitter sends private information to the receiver by using a set  $\mathcal{S} \triangleq \{1, \dots, S\}$  of orthogonal channels. Here, the orthogonality is assumed in the frequency domain. All channels are available to all the transmitters and thus, communications are subject to mutual interference. The receiver implements single user decoding. The aim of each radio device is to select a single channel and a power level such that its individual (Shannon) transmission rate is maximized. Note that under this formulation transmitting at the maximum power is a dominant action, and thus, at the (Nash) equilibrium those actions should not be used.

An action or power allocation (PA) vector for transmitter  $k \in \mathcal{K}$  is any vector

$$\mathbf{p}_k(n) = (p_{k,1}(n), \dots, p_{k,S}(n)) \in \mathcal{A}_k,$$

where,  $\forall k \in \mathcal{K}$ ,

$$\mathcal{A}_k = \left\{ \mathbf{p}_k^{(s)} = p_{k,\max} \mathbf{e}_s : \forall s \in \mathcal{S}, \mathbf{e}_s = (e_{s,1}, \dots, e_{s,S}) \forall r \in \mathcal{S} \setminus s, e_{s,r} = 0, \text{ and } e_{s,s} \in [0, 1] \right\}.$$

Thus, the utility of radio device  $k$  can be written as follows,

$$u_k(\mathbf{p}_k(n), \mathbf{p}_{-k}(n)) = \sum_{s \in \mathcal{S}} \log_2(1 + \gamma_{k,s}(n)) \text{ [bps/Hz]},$$

where the bandwidth per channel is normalized to 1Hz. and  $\gamma_{k,s}(n)$  is the signal-to-interference plus noise ratio (SINR) seen by player  $k$  over its channel  $s$  at time  $n$ .

In Fig. 1, we show the average sum Shannon rates (network spectral efficiency) as a function of the number

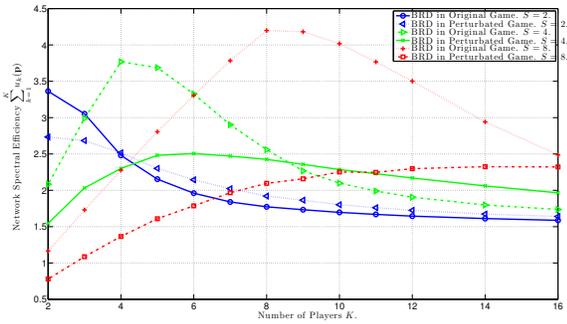


Fig. 1. Average sum Shannon rates (network spectral efficiency) as a function of the number of players  $K$  for different number of channels  $S$ . Average SNR  $\frac{p_{k,\max}}{\sigma_k^2} = 10$  dBs,  $K = S = 2$ , with  $Q = 4$  power levels per channel.

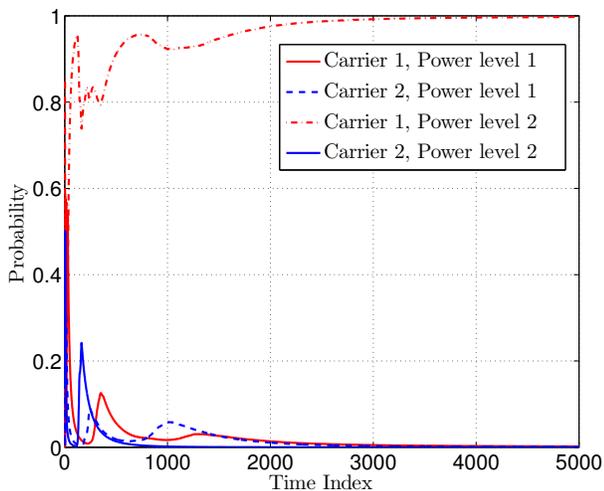


Fig. 2. Strategy of player  $k = 1$  as a function of the iterations.  $K = 2$ ,  $S = 2$ ,  $Q = 2$  power levels and average SNR  $\frac{p_{k,\max}}{\sigma_k^2} = 10$  dB.

of players  $K$ , with  $S \in \{2, 4, 8\}$ . Therein, we assume that players are able to know the exact realization of the noise component and thus, they can play the action which maximizes the truth utility  $\tilde{u}_k$ . Note how the existence of the noise component makes the players to take different actions and thus, achieve a different performance in terms of  $\bar{u}_k$ .

In Fig. 2, we show the convergence of the strategy of player 1. Note that since  $\gamma_k$  is time dependent, and thus, the noise component can be faded out in time, the dynamics converges to a NE after a short transient. Indeed, the transmitters end up transmitting at the maximum power using different channels, which correspond to a NE.

## VI. CONCLUSIONS

In this paper, we have presented a mathematical framework to study the effect of bounded rationality in

wireless communications. The notion of quantal response equilibrium has been introduced in order to model the fact that players are not able to properly estimate their own utility and thus, they are subject to choosing other actions than the optimal one. Here, we have also presented a behavioral rule which allows the networks to achieve LE or NE in a fully distributed manner, relying only in the observation of the utility function.

## REFERENCES

- [1] S. Lasaulce and H. Tembine, *Game Theory and Learning in Wireless Networks: Fundamentals and Applications*. Elsevier Academic Press, Oct. 2011.
- [2] J. F. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, 1950.
- [3] L. Rose, S. M. Perlaza, S. Lasaulce, and M. Debbah, "Learning equilibria with partial information in wireless networks," *IEEE Communications Magazine, special issue Game Theory in Wireless Communications*, Sep. 2011.
- [4] E. G. Larsson and E. Jorswieck, "Competition Versus Cooperation on the MISO Interference Channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, 2008.
- [5] R. A. Berry and D. N. C. Tse, "Shannon meets nash on the interference channel," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2821–2836, May 2011.
- [6] F. Meshkati, M. Chiang, H. V. H.V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier cdma systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 6, pp. 1115–1129, 2006.
- [7] E.-V. Belmega, S. Lasaulce, and M. Debbah, "Power allocation games for mimo multiple access channels with coordination," *IEEE Trans. on Wireless Communications*, vol. 8, no. 6, pp. 3182–3192, June 2009.
- [8] R. D. McKelvey and T. R. Palfrey, "Quantal response equilibria for extensive form games," *Experimental Economics*, vol. 1, no. 1, pp. 9–41, June 1998.
- [9] R. D. McKelvey and T. R. Palfrey, "Quantal response equilibria for normal form games," *Games and Economic Behavior*, vol. 10, no. 1, pp. 6–38, 1995.
- [10] S. P. Anderson, J. K. Goeree, and C. A. Holt, "The logit equilibrium: A perspective on intuitive behavioral anomalies," University of Virginia, Department of Economics, Virginia Economics Online Papers, Nov. 1999.
- [11] S. M. Perlaza, H. Tembine, S. Lasaulce, and M. Debbah, "Learning to use the spectrum in self-configuring heterogeneous networks," in *4rd ICST/ACM International Workshop on Game Theory in Communication Networks*, May 2011.
- [12] J. V. Neumann and O. Morgenstern, "Theory of games and economic behavior," *Princeton University Press*, 1944.
- [13] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine, Special issue on Game Theory for Signal Processing*, vol. 26, no. 5, pp. 41–52, Sep. 2009.
- [14] G. W. Brown, "Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation*, vol. 13, no. 1, pp. 374–376, 1951.
- [15] V. Borkar, "Stochastic approximation: A dynamical system viewpoint," *Cambridge University Press*, 2008.
- [16] —, "Stochastic approximation with two timescales," *Systems Control Lett.*, vol. 29, pp. 291–294, 1997.
- [17] S. D. Leslie and E. J. Collins, "Convergent multiple-timescales reinforcement learning algorithms in normal form games," *Ann. Appl. Probab.*, vol. 13, no. 4, pp. 1231–1251, 2003.