Reinforcement-based data transmission in temporally-correlated fading channels: Partial CSIT scenario

Behrooz Makki*, Tommy Svensson*, Merouane Debbah†
*Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
{behrooz.makki, tommy.svensson}@chalmers.se
†Alcatel-Lucent Chair - SUPELEC, Gif-sur-Yvette, France
merouane.debbah@supelec.fr

***Abstract*—Reinforcement algorithms refer to the schemes where the results of the previous trials and a reward-punishment rule are used for parameter setting in the next steps. In this paper, we use the concept of reinforcement algorithms to develop different data transmission models in wireless networks. Considering temporally-correlated fading channels, the results are presented for the cases with partial channel state information at the transmitter (CSIT). As demonstrated, the implementation of reinforcement algorithms improves the performance of communication setups remarkably, with the same feedback load/complexity as in the state-of-the-art schemes.**

I. Introduction

In machine learning, reinforcement algorithms refer to the schemes where dynamic parameter adaptation is performed based on a reward-punishment strategy [1]. The previous trial(s) being successful, more aggressive parameter settings are risked. On the other hand, the parameters of the upcoming trials are designed more conservatively, if the previous gambling fails. Reinforcement learning differs from the standard *supervised* learning in that the correct input/output pairs are never presented, but a reward-punishment signal is used for parameter adaptation, and the goal is to maximize some notion of the cumulative reward. Due to its generality, the reinforcement algorithm is applied in different fields, such as game theory, control theory, simulation-based optimization and statistics, e.g., [2]–[5]. However, except some works in these last years, e.g., [4]–[10], the reinforcement concept has not been well studied in wireless communication.

In this paper, we elaborate on the performance of communication systems utilizing reinforcement algorithms. The problem is cast in form of optimizing the data transmission efficiency of wireless networks in the cases with partial channel state information at the transmitter (CSIT). The results are obtained for temporally-correlated fading channels, and the reward-punishment signal is used to dynamically adapt the data transmission rates/powers.

The partial CSIT systems are mainly based on two different channel state information (CSI) feedback models, namely, CSI quantization [11]–[15] and hybrid automatic repeat request (HARQ) [14]–[21][1]. With CSI quantization, the channel quality information is fed back before the codeword transmission. The HARQ methods, on the other hand, are based on reporting the decoding status of the previous messages. Here, to emphasize the generally of the reinforcement algorithms, we consider both the quantized CSI and the HARQ feedback models and the results are presented for different metrics. Specifically, considering the point-to-point communication setups, we address the following problems:

- **Problem 1:** *Power-limited throughput maximization in the presence of quantized CSI feedback.* Here, the reward-punishment signal is used for dynamic adaptation of the data transmission rates such that the throughput is maximized. The results are compared with the ones in the static quantization techniques [11], [14] which, with the same feedback load, show more than 6% throughput increment for a large range of fading correlations.
- **Problem 2:** *Outage-limited power minimization in the presence of HARQ feedback.* In this scenario, the transmitter uses the HARQ feedback signals to learn about the channel condition and update the data (re)transmission powers in a reinforcement-based fashion. As demonstrated, the proposed scheme improves the power efficiency of the HARQ protocols remarkably. For instance, consider a communication setup utilizing repetition time diversity (RTD) HARQ with codewords of rate 1 nats-per-channel-use (npcu) and outage probability 10^{-2}. Then, compared to uniform and the adaptive (non-reinforcement based) power allocation scheme of [21], the implementation of reinforcement scheme improves the power efficiency by 4 and 1 dB, respectively; The result is valid

Behrooz Makki and Tommy Svensson are supported in part by the Swedish Governmental Agency for Innovation Systems (VINNOVA) within the VINN Excellence Center Chase.

Merouane Debbah has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering).

978-1-4799-5863-4/14/$31.00

[1]Throughout the paper, we concentrate on the frequency-division duplexing (FDD) communication setups. However, the reinforcement algorithms are applicable in time-division duplexing (TDD) systems as well.

for a large range of fading correlations.

The remainder of the paper is organized as follows. In Section II, the system model is presented. Sections III and IV present the results for Problems 1 and 2, respectively. The conclusions are presented in Section V.

II. System model

Consider a communication setup where, at time slot t, the power-limited input message $x(t)$ multiplied by the fading coefficient $h(t)$ is summed with an independent and identically distributed (iid) complex Gaussian noise $z(t) \sim \mathcal{CN}(0,1)$ resulting in the output

$$y(t) = h(t)x(t) + z(t). \tag{1}$$

We study temporally-correlated Rayleigh block-fading conditions where the channel coefficients remain constant in a fading block, determined by the channel coherence time, and then change to other values according to the fading probability density function (pdf). Particularly, the channel changes in each codeword transmission period according to a first-order Gauss-Markov process

$$h(t+1) = \beta h(t) + \sqrt{1-\beta^2}\epsilon, \epsilon \sim \mathcal{CN}(0,1). \tag{2}$$

Here, β is the correlation factor of the fading realizations experienced in two successive codeword transmissions, with $\beta = 0$ (respectively, $\beta = 1$) representing the uncorrelated (respectively, fully-correlated) block-fading channel. This is a well-established model considered in the literature for different phenomena such as CSI imperfection, estimation error and channel/signals correlation [22]–[25]. In this way, defining the channel gain as $g(t) \doteq |h(t)|^2$, the joint and the marginal pdfs of the channel gains are found as

$$f_{g_{(t)},g_{(t+1)}}(x,y) = \frac{1}{1-\beta^2} e^{-\frac{x+y}{1-\beta^2}} B_0\left(\frac{2\beta\sqrt{xy}}{1-\beta^2}\right) \tag{3}$$

and

$$f_{g(t)}(x) = e^{-x}, x \geq 0, \tag{4}$$

respectively, where $B_0(.)$ is the zeroth-order modified Bessel function of the first kind [25].

In each block, the channel coefficient is assumed to be known by the receiver, which is an acceptable assumption in block-fading channels [11]–[21]. However, there is no instantaneous channel state information available at the transmitter except the reinforcement-based feedback signals. Moreover, all results are presented in natural logarithm basis, the throughput is presented in npcu and the arguments are restricted to Gaussian input distributions. Finally, we concentrate on the *continuous* data communication models where there is a large pool of information to be sent to the receiver, and a new codeword transmission starts as soon as the previous codeword transmission ends.

In Sections III and IV, we study Problems 1 and 2, respectively. Note that the considered problems are only examples and the reinforcement algorithms are applicable in various setups/problem formulations.

III. Power-limited throughput maximization via reinforcement-based CSI feedback

Considering the static (non-reinforcement based) CSI quantization scheme with N quantization regions, an encoding function

$$\begin{aligned} C(g(t)) = c_i, \text{ if } g(t) \in G_i = [g_{i-1}, g_i), i = 1, \ldots, N, \\ g_0 \doteq 0, g_N \doteq \infty, \end{aligned} \tag{5}$$

is applied at the receiver and the symbol c_i is fed back to the transmitter [11], [14]. Receiving c_i, the transmitter sends the data at rate r_i and power P.[2] If the instantaneous channel gain supports the data rate, i.e., $\log(1+g(t)P) \geq r_i$, the data is successfully decoded, otherwise outage occurs. In [11], [14], it has been proved that, to maximize the power-limited throughput, the optimal rate allocation rule of the static quantization schemes is given by $r_i = \log(1+\tilde{g}_i P)$ where

$$\tilde{g}_i = \begin{cases} \tilde{g}_1 \in [0, g_1), & \text{if } i = 1 \\ g_{i-1}, & \text{if } i \neq 1. \end{cases} \tag{6}$$

That is, to maximize the throughput, the channel gain is assumed to be its worst value within each quantization region, except the first one. In this way, using $r_i = \log(1+\tilde{g}_i P)$ and (6), the throughput of the static quantized CSI scheme is determined as

$$\begin{aligned} \eta^{\text{SQ}} &= E\{\text{Achievable rates}\} \\ &= \sum_{n=1}^{N} \Pr(g(t) \in G_n) \Pr(g(t) \geq r_n | g(t) \in G_n) r_n \\ &= \sum_{n=1}^{N} \log(1+\tilde{g}_n P)\left(F_{g(t)}(\tilde{g}^{n+1}) - F_{g(t)}(\tilde{g}_n)\right) \\ &= \sum_{n=1}^{N} \log(1+\tilde{g}_n P)(e^{-\tilde{g}_n} - e^{-\tilde{g}_{n+1}}), \end{aligned} \tag{7}$$

where $E\{.\}$ denotes the expectation operator, $F_{g(t)}(.)$ is the cumulative distribution function (cdf) of the channel gain and the last equality is for Rayleigh-fading channels. Using (7), the power-limited throughput maximization problem of a static CSI quantization approach is formulated as

$$\max_{\forall \tilde{g}_i, i=1,\ldots,N} \sum_{n=1}^{N} \log(1+\tilde{g}_n P)(e^{-\tilde{g}_n} - e^{-\tilde{g}_{n+1}}) \tag{8}$$

and, as the problem is complex, the optimization parameters $\tilde{g}_i$'s are determined via iterative optimization algorithms, e.g., [11, Algorithm 1], [14, Algorithm 1]. Finally, setting $N = 1$ and $N \to \infty$, the throughput with no and perfect CSIT are respectively found as [26, Chapter 1.4.1]

$$\eta^{\text{No CSIT}} = \max_{\tilde{g}_1} \{e^{-\tilde{g}_1} \log(1+\tilde{g}_1 P)\} = \Lambda(P) e^{-\frac{e^{\Lambda(P)}-1}{P}} \tag{9}$$

[2] As Section III studies the effect of reinforcement algorithms on the rate adaptation, we consider a constant (peak) power P. It is straightforward to extend the results to cases with adaptive power allocation.

and

$$\eta^{\text{Perfect CSIT}} = \int_0^{\infty} e^{-g}\log(1+gP)\mathrm{d}g = e^{-\frac{1}{P}}\mathrm{Ei}(-\frac{1}{P}), \quad (10)$$

with $\Lambda(.)$ and $\mathrm{Ei}(.)$ representing the Lambert W function and the exponential integral function, respectively.

Compared to no-CSIT (open-loop) systems, the static quantizers increase the throughput considerably [11], [14]. However, as also demonstrated in (7), the channel temporal dependencies are not exploited for throughput increment/feedback load reduction. On the other hand, exploiting the temporal correlations has been previously shown to be crucial for practical implementation of many communication systems [12], [27][3].

To exploit the temporal dependencies of the channel, we propose a simple reinforcement-based algorithm as stated in Algorithm 1. In words, the algorithm is based on the following procedure. Start the data transmission with an initial transmission rate R and consider an adaptation coefficient δ. In each block, depending on whether the channel can support the data rate $R+\delta R$ or not, the receiver sends a reinforcement signal $\alpha = 1$ or $\alpha = 0$, respectively. Receiving the reward-punishment signal α, the transmitter updates its transmission rate correspondingly (For more details please see Algorithm 1). The throughput is achieved by averaging on the decodable rates over many codeword transmissions.

Algorithm 1 CSI-based data transmission by a reinforcement algorithm

Consider an initial transmission rate R and an updating coefficient δ. In each block, do the followings.

I. *Feedback report at the receiver*
Feed $\alpha = 1$ back, if $\log(1+g(t)P) > R+\delta R$.
Otherwise, send $\alpha = 0$.

II. *Rate adaptation at the transmitter*
$R \leftarrow R+\delta R$, if $\alpha = 1$
$R \leftarrow R-\delta R$, if $\alpha = 0$.
Send a codeword with rate R. The codeword is correctly decoded by the receiver if $\log(1+g(t)P) > R$. Otherwise, the codeword is dropped and an outage is declared.

III. Go to I.

As opposed to the static quantization scheme, there are only two optimization parameters in Algorithm 1 which can be determined by, e.g., exhaustive search. Also, in contrast to the static quantization scheme, the reinforcement-based scheme of Algorithm 1 *follows* the channel variations and dynamically updates the transmission rates. Finally, note that to represent N quantization regions $\log_2 N$ feedback bits per codeword is required in the static quantization scheme which, depending on the number of quantization regions, can be considerably high. However, the proposed algorithm is based on only 1 bit feedback per codeword.

[3]For instance, the amount of CSIT required for proper implementation of orthogonal frequency-division multiplexing (OFDM) and multiple-input-multiple-output (MIMO) broadcast channels is not practically affordable if temporal and frequency correlations are not exploited [12], [27].

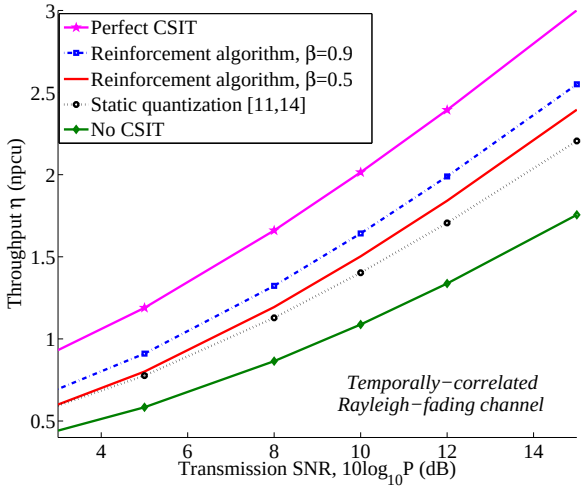

Figure 1. The throughput of different schemes vs the transmission SNR $10\log_{10} P$ dB, temporally-correlated Rayleigh-fading channel following (2).

As an example, Fig. 1 demonstrates the throughput of the reinforcement-based scheme in a Rayleigh-fading channel following (2). Also, the results are compared with the cases having perfect and no CSIT and when the static quantization (5) is implemented. The parameters R and δ of Algorithm 1 are found by exhaustive search such that the throughput is maximized in each signal-to-noise ratio (SNR). Also, the throughput of the static quantization scheme is obtained with $N = 2$ quantization regions which leads to 1 bit per codeword feedback, the same as in the reinforcement-based scheme. Moreover, Fig. 2 shows the relative throughput gain of the proposed scheme compared to the static CSI quantization approach, i.e., $\Delta = \frac{\eta-\eta^{\text{SQ}}}{\eta^{\text{SQ}}}\%$, where η is the throughput achieved via the data transmission approach of Algorithm 1. As demonstrated, the system throughput is remarkably increased by implementation of the reinforcement-based algorithm. For instance, with a correlation factor of $\beta \geq 0.5$ and transmission SNR of ≥ 8dB, the reinforcement-based scheme results in $\geq 6\%$ increase in the relative throughput. Also, the gain of the proposed scheme increases with the SNR.

Finally, to close the section, we should mention that, while the paper concentrates on a single-user setup, the reinforcement-based schemes are of particular interest when the number of base stations/users increases. There, the same approach as in Algorithm 1 can be implemented for user scheduling, where higher and higher data rates are considered for a user as long as it correctly decodes the message, otherwise the scheduler selects another user. Indeed, the gain of the reinforcement-based scheme, over the static quantization techniques, increases with the number of users, because to achieve the same throughput the reinforcement-based scheme requires less number of feedback bits compared to the cases with static quantization. This point becomes more interesting when we remember that, since the positive acknowledgement

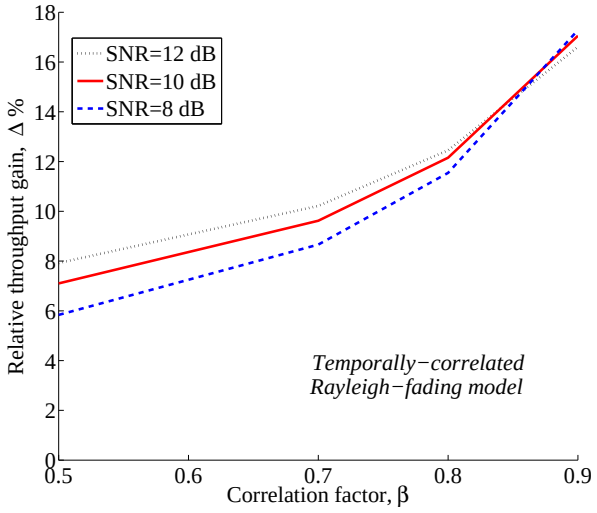

Figure 2. The relative throughput gain $\Delta = \frac{\eta - \eta^{\text{SQ}}}{\eta^{\text{SQ}}}\%$ for different correlation coefficients β, temporally-correlated Rayleigh-fading channel following (2).

is a standard provision of most practical link layers [16], the reinforcement signal feedback is not required in each slot. As a result, the reinforcement-based scheme requires even less than one-bit feedback per user/slot.

IV. Outage-limited power minimization via reinforcement-based HARQ

In contrast to the CSI-based schemes where the partial CSI is fed back before the codeword transmission, the HARQ-based schemes are based on reporting the message decoding status at the end of each codeword [14], [16].

In the following, we elaborate on the implementation of reinforcement algorithms in HARQ protocols. The results are presented for the RTD HARQ, also referred to as Type III HARQ, but the discussions are valid for the other HARQ protocols, such as the incremental redundancy [14], [20], as well.

Consider the RTD HARQ with a maximum of M (re)transmission rounds, i.e., the data is retransmitted a maximum of $M-1$ times. Also, define a packet as the transmission of a codeword along with all its possible retransmissions. Using power-adaptive RTD HARQ, b information nats is encoded into a codeword of length L channel uses. Thus, the codeword rate is $R = \frac{b}{L}$ npcu. In the mth, $m = 1, \ldots, M$, (re)transmission round, the codeword is scaled to have power P_m. The codewords are (re)transmitted until the receiver correctly decodes the data or the maximum permitted retransmission rounds is reached. Also, in each round of a packet the receiver performs maximum ratio combining (MRC) of all received signals.

Considering the non-reinforcement based scheme and the continuous data communication model, the average power and the outage probability are obtained as follows (please see [21] as well). In the mth (re)transmission round, the transmission energy is LP_m. If the data transmission stops at the end of the mth (re)transmission round, the average power, i.e., the ratio of the total transmission energy and the total data transmission time, is $P_{(m)} = \frac{\sum_{n=1}^{m} LP_n}{mL} = \frac{1}{m}\sum_{n=1}^{m} P_n$. Thus, the average power, averaged over many packet transmissions, is obtained as

$$P^{\text{HARQ}} = \sum_{n=1}^{M} \left(\frac{1}{m}\sum_{n=1}^{m} P_n\right) \Pr(A_m), \tag{11}$$

where A_m is the event that the data transmission stops at the end of round m. Note that $\sum_{m=1}^{M} \Pr(A_m) = 1$, as a maximum of M (re)transmissions is considered.

As the same codeword is retransmitted, the equivalent data rate decreases to $\frac{b}{mL} = \frac{R}{m}$ at the end of the mth round. Also, the implementation of MRC increases the received SNR to $\sum_{n=1}^{m} g(n)P_n$ in round m. Thus, following the same procedure as in [20], [21], the data is successfully decoded at the end of the mth round (and not before) if $\log(1 + \sum_{n=1}^{m-1} g(n)P_n) < R \leq \log(1 + \sum_{n=1}^{m} g(n)P_n)$. This is based on the fact that, with an SNR x, the maximum achievable rate is

$$U_{(m)} = \frac{1}{m}\log(1+x),$$

if the same codeword is retransmitted m times.

In this way, the probability terms $\Pr(A_m)$ are obtained as

$$\begin{aligned}\Pr(A_m) &= \begin{cases} \Pr\Big(\log(1 + \sum_{n=1}^{m-1} g(n)P_n) < R \\ \qquad \leq \log(1 + \sum_{n=1}^{m} g(n)P_n)\Big), & m \neq M \\ 1 - \sum_{n=1}^{M-1} \Pr(A_n), & m = M \end{cases} \\ &= \begin{cases} \Pr\Big(\log(1 + \sum_{n=1}^{m-1} g(n)P_n) < R \\ \qquad \leq \log(1 + \sum_{n=1}^{m} g(n)P_n)\Big), & m \neq M \\ \Pr\Big(\log(1 + \sum_{n=1}^{M-1} g(n)P_n) < R\Big), & m = M, \end{cases} \end{aligned} \tag{12}$$

and, with the same arguments, the outage probability is found as [20], [21]

$$\Pr(\text{Outage}) = \Pr\left(\log(1 + \sum_{n=1}^{M} g(n)P_n) < R\right). \tag{13}$$

Therefore, with an initial rate R, (11)-(13) are used to rephrase the outage-limited power minimization problem as

$$\begin{aligned} \min_{P_m, m=1,\ldots,M} \quad & \sum_{n=1}^{M} \left(\frac{1}{m}\sum_{n=1}^{m} P_n\right) \Pr(A_m), \\ \text{s.t.} \quad & \Pr(\log(1 + \sum_{n=1}^{M} g(n)P_n) < R) = \epsilon, \end{aligned} \tag{14}$$

where ϵ denotes the outage probability constraint. Finally, as discussed in, e.g., [20], [21], there may be no closed-form solution for the optimal, in terms of (14), powers P_m and,

depending on the fading pdf and the number of retransmissions, we may need to find the optimal power allocation rules numerically.

The drawback of power allocation based on (14) is that the channel quality information gathered in the previous packet transmissions is not exploited for parameter adaptation in the next packet. That is, the power terms of a packet are not affected by the message decoding status of the previous packet transmissions. To tackle this problem, we propose a reinforcement-based algorithm as illustrated in Algorithm 2.

In the algorithm, the data transmission starts with some initial power. Then, in each time slot, depending on whether the message is correctly decoded or not, the transmission power decreases or increases, respectively. In this way, the feedback signal makes it possible to *learn* about the channel quality and update the power based on all previous message decoding status. The initial power P_{initial} and the adaptation coefficients $d_m, d'_m, m = 1, \ldots, M, d_m \in (0,1)$, of the algorithm are determined by exhaustive search such that the average transmission power, averaged over many packet transmissions, is minimized for an outage probability constraint. Finally, note that to implement the reinforcement algorithm we changed the feedback model of the quantized CSI scheme in Section III. However, Algorithm 2 uses the same acknowledgement/negative acknowledgement (ACK/NACK) signal as in the standard HARQ to perform parameter adaptation.

As an example, setting $R = 1$ npcu and $\beta = 0.9$, Fig. 3 demonstrates the outage-limited average power of the RTD protocol with a maximum of $M = 2$ (re)transmissions. Also, the results are compared with the cases utilizing uniform power allocation, i.e., $P_m = P_n, \forall m, n$, and when the power terms are optimized based on (14). To solve (14), we have used the same iterative optimization algorithm as in [21, Algorithm 1]. As it can be seen, remarkable power efficiency gain is achieved by the reinforcement algorithm. For instance, with an outage probability $\epsilon = 10^{-2}$, the implementation of the reinforcement-based algorithm reduces the average power, compared to the uniform power allocation and the power allocation scheme of (14), by 4 and 1 dB, respectively. Also, the effect of reinforcement algorithm increases as the outage probability constraint becomes harder, i.e., ϵ decreases. Finally, although not demonstrated in the figure, (almost) the same average power reduction is observed for the cases with $\beta \geq 0.2$.

V. Conclusion

This paper studied the data transmission efficiency of the communication systems utilizing reinforcement algorithms. Considering temporally-correlated fading channels, the reinforcement feedback signals were used for parameter adaptation in the cases with partial CSIT. As illustrated, the reinforcement algorithms lead to remarkable performance improvement, compared to the state-of-the-art schemes, with the same feedback load. Specially, considerable throughput and power efficiency increment is achieved with 1 bit per codeword feedback, if the reinforcement algorithms are utilized.

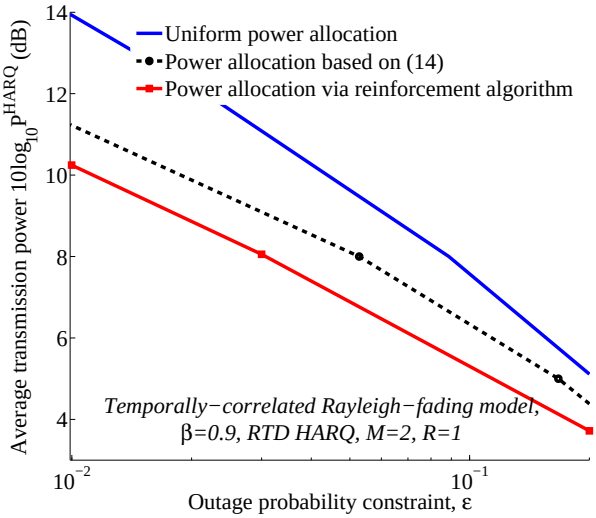

Figure 3. Outage-limited average power for different power allocation schemes, RTD HARQ, $M = 2$. Correlated Rayleigh fading channel model (2), $\beta = 0.9$. For the reinforcement-based scheme, Algorithm 2 is used where the constants $P_{\text{initial}}, d_m, d'_m, \forall m$, are optimized, in terms of average power, for every given outage probability.

Algorithm 2 HARQ-based data transmission by a reinforcement algorithm

I. For a given initial transmission rate R, set the initial transmission power to $\breve{P} = P_{\text{initial}}$ and consider the adaptation coefficients $d_m, d'_m, m = 1, \ldots, M, d_m \in (0,1)$.

II. Start a new packet transmission with power $\breve{P}$ and do the following procedure

1) For $m < M$,
If the codeword is correctly decoded, set $\breve{P} \leftarrow (1 - d_m)\breve{P}$, $m \leftarrow 1$ and go to II.
If the codeword is not decoded, set $\breve{P} \leftarrow (1 + d'_m)\breve{P}$, $m \leftarrow m + 1$ and retransmit the codeword.

2) For $m = M$,
If the codeword is correctly decoded, set $\breve{P} \leftarrow (1 - d_M)\breve{P}$, $m \leftarrow 1$ and go to II.
If the codeword is not decoded, declare an outage, set $\breve{P} \leftarrow (1 + d'_M)\breve{P}$, $m \leftarrow 1$ and go to II.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 1998.

[2] C. Ye, N. H. C. Yung, and D. Wang, "A fuzzy controller with supervised learning assisted reinforcement learning algorithm for obstacle avoidance," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 33, no. 1, pp. 17–27, Feb. 2003.

[3] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 38, no. 2, pp. 156–172, March 2008.

[4] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1904–1919, May 2009.

[5] M. A. Khan, H. Tembine, and A. V. Vasilakos, "Game dynamics and cost of learning in heterogeneous 4G networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 198–213, Jan. 2012.

[6] C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan, "Non-cooperative power control for wireless ad hoc networks with repeated games," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1101–1112, Aug. 2007.
[7] D. V. Djonin and V. Krishnamurthy, "MIMO transmission control in fading channels-A constrained Markov decision process formulation with monotone randomized policies," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5069–5083, Oct. 2007.
[8] D. Djonin and V. Krishnamurthy, "Q-learning algorithms for constrained Markov decision processes with randomized monotone policies: Application to MIMO transmission control," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2170–2181, May 2007.
[9] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, July 2013.
[10] M. Simsek, M. Bennis, and A. Czylwik, "Dynamic inter-cell interference coordination in HetNets: A reinforcement learning approach," in *GLOBECOM*, Dec. 2012, pp. 5446–5450.
[11] T. T. Kim and M. Skoglund, "On the expected rate of slowly fading channels with quantized side information," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 820–829, April 2007.
[12] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
[13] S. Ekbatani, F. Etemadi, and H. Jafarkhani, "Outage behavior of slow fading channels with power control using partial and erroneous CSIT," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6097–6102, Dec. 2010.
[14] B. Makki and T. Eriksson, "On hybrid ARQ and quantized CSI feedback schemes in quasi-static fading channels," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 986–997, April 2012.
[15] ——, "Feedback subsampling in temporally-correlated slowly-fading channels using quantized CSI," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2282–2294, June 2013.
[16] C. E. Koksal and P. Schniter, "Robust rate-adaptive wireless communication using ACK/NAK-feedback," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1752–1765, April 2012.
[17] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1129–1141, April 2010.
[18] ——, "Coding versus ARQ in fading channels: How reliable should the PHY be?" in *GLOBECOM*, Nov. 2009, pp. 1–6.
[19] B. Makki, A. Graell i Amat, and T. Eriksson, "On noisy ARQ in block-fading channels," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 731–746, Feb. 2014.
[20] D. Tuninetti, "On the benefits of partial channel state information for repetition protocols in block fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5036–5053, Aug. 2011.
[21] B. Makki, A. Graell i Amat, and T. Eriksson, "Green communication via power-optimized HARQ protocols," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 161–177, Jan. 2014.
[22] H. A. Suraweera, P. J. Smith, and M. Shafi, "Capacity limits and performance analysis of cognitive radio with imperfect channel knowledge," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1811–1822, May 2010.
[23] K. S. Ahn and R. W. Heath, "Performance analysis of maximum ratio combining with imperfect channel estimation in the presence of cochannel interferences," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1080–1085, March 2009.
[24] K. Huang, R. Heath, and J. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1959–1975, May 2009.
[25] C. Tellambura and A. D. S. Jayalath, "Generation of bivariate Rayleigh and Nakagami-M fading envelopes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 170–172, May 2000.
[26] B. Makki, "Data transmission in the presence of limited channel state information feedback," Ph.D. dissertation, Dept. Sig. and Sys., Chalmers Uni. Tech., Nov. 2013.
[27] T. Eriksson and T. Ottosson, "Compression of feedback for adaptive transmission and scheduling," *Proc. IEEE*, vol. 95, no. 12, pp. 2314–2321, Dec. 2007.