Self-Optimization of Capacity and Coverage in LTE Networks Using a Fuzzy Reinforcement Learning Approach

R. Razavi[1], S. Klein[2] and H. Claussen[1]
[1] Bell Laboratories, Alcatel-Lucent, Blanchardtown Industrial Park, Dublin 15, Republic of Ireland
[2] Bell Laboratories, Alcatel-Lucent, Lorenzstr10, 70435 Stuttgart, Germany

Abstract—**This paper introduces a solution to enable self–optimization of coverage and capacity in LTE networks through base stations' downtilt angle adjustment. The proposed method is based on fuzzy reinforcement learning techniques and operates in a fully distributed and autonomous fashion without any need for a priori information or human interventions. The solution is shown to be capable of handling extremely noisy feedback information from mobile users as well as being responsive to the changes in the environment including self-healing properties. The simulation results confirm the convergence of the solution to the global optimal settings and that the proposed scheme provides up to 20% performance improvement when compared with an existing fuzzy logic based reinforcement learning approach.**

Keywords-component; Self-x Networks; LTE; Downtilt Adjustment; Reinforcment Learning; Fuzzy Logic

I. INTRODUCTION

A major challenge faced by mobile operators is to ensure that mobile services are of a high quality while reducing capital expenditures (CAPEX) and operational expenditures (OPEX) of the complex radio access network (RAN). Based on recent research [1], approximately 17% of wireless operator's CAPEX is being spent on engineering and installation services. By employing self-configuration and self-optimization mechanisms several human interventions from network operation and maintenance can be removed. These autonomous functions aim towards self-organising behaviours of network that increase the network performance and service quality by reacting in response to the dynamics of the network.

In the wireless domain, quick evolution of networks has led to parallel operation of different technologies and infrastructures (e.g. 2G, 3G and LTE). Ideally, not only simultaneous support of these entirely different infrastructures should be feasible with minimal additional operation cost, but also introduction of a new system should be straight forward. This is especially important during the early deployment phase of a new system when the efforts to setup and optimise the system are significant and traditionally lead to lengthy periods of instability before an optimised and stable network is achieved. It is therefore essential to have the necessary self-organising mechanisms available before commencing the deployment of a new system. Stemming from the rapid growth of mobile data services, operators now face the challenge of a drastic increase in data traffic demands and therefore are currently planning to upgrade and evolve their existing mobile networks. In respect to wireless technologies, LTE is known as the last step toward the 4th generation (4G) of radio technologies. High throughput, low delays, and low costs are the three basic requirements for 4G network development, which LTE is destined to realize.

A key phase of topology planning when a new technology is to be employed is the definition of antenna configuration and especially the antenna downtilt angle. By utilizing antenna downtilt, signal levels within a cell can be improved and interference radiation towards other cells is effectively reduced due to more precise aiming of the antenna radiation pattern. However, an aggressive downtilting might lead to coverage problems at cell border areas. According to [2], as a source of coverage improvement, beam-tilt optimisation can yield 10dB signal strength increases over large areas. Fundamentally, downtilting is possible both mechanically and electrically. Adjusting mechanical downtilt requires site visits which are time and cost inefficient and inflexible. Moreover, mechanical tilting is only effective in the forward direction and has no effect on side radiation and moves the rear lobe upward. Fortunately, advances in electrical downtiliting have enabled remote adjustment of radiation pattern for all azimuth angles by means of signal phasing alteration. Techniques such as the continuously adjustable electrical down-tilt (CAEDT) [3] enable intelligent algorithms to be employed for autonomous configuration and optimisation of the downtilt angles and remove or at least minimize the need for expensive drive testing and statistical analysis of network performance and fault reports.

During the deployment of the 3G networks, there have been significant efforts to optimise the configuration of the base-stations antenna. The upgrade to the 4G would have been very swift and cost effective if such settings could be inherited from the more mature existing 3G networks. However, due to the fundamental differences between the technologies this is unfortunately not possible. For example, regarding the antenna downtiliting, a recent field measurement study [4] reveals that when LTE is overlaid into CDMA1x networks, independent optimization of the LTE downtiliting would result in over 26% throughput gains at the cell edges compared to the case when CDMA1x network settings are applied to LTE. That is mainly because LTE has a more robust link budget (around 5dB higher than CDMAx1) which can be traded to improve the throughput. On the other hand, WCDM supports macro diversity or soft(er) handover implying that some cell overlap is beneficial. However, in LTE where the same frequency band is used by all cells and no macro diversity is adopted, smaller cell overlap is desirable [5].

While most operators commonly seek to deploy their sites in a hexagonal layout, this is not always feasible due to many reasons like the terrain factors, site rental and zoning issues. Even in a strictly hexagonal site deployment the radio propagation characteristics of a cell might be completely different from another. In addition, there is the variation of geographical distribution of traffic load across the cells (due to formation of new hotspots for example) that can significantly impact on the overall performance [6]. These recommend that the downtilt adjustment should be performed in a distributed fashion on a cell-by-cell basis and that the solution is to be responsive to the environment changes. Compared to the network-wide tilt angle optimization, it is shown that cell-based scheme can provide a significant gain even in quite homogenous scenarios [7].

Many of the autonomous functionalities and automation mechanisms are based on intelligent algorithms such as the machine learning techniques. The objective of these algorithms is to enable the network to learn and use that experience for future actions. Reinforcement Learning (RL) [8] is a type of learning that involves an agent learning behaviour through interaction with its environment. The direct consequences of actions are given by the reinforcement signal, but the correct actions are not provided. However, major issues arise with RL, when the input/output space is continuous or when the environment is highly dimensional (curse of dimensionality). Besides, using fixed thresholds for girding and partitioning the input/output space into discrete states leads to sudden state transitions and abrupt actions that are naturally undesirable.

Fuzzy logic controllers, on the other hand, have recently found extensive applications in industry as they can cope well with inaccurate and approximate data and can mimic the human reasoning. By employing fuzzy sets through commonly overlapping fuzzy membership functions, not only a very good level of input abstraction, but also smooth state transitions can be achieved. However, the performance of fuzzy controllers relies on the availability of human expertise. Such expertise may not be available or optimal in all application scenarios implying the need for a self-learning algorithm. This suggests successful combination of fuzzy systems and RL techniques. While the earlier provides a high level of abstraction to cope with curse of dimensionality issue, the later is particularly needed for training/tuning the controller.

This paper introduces a solution to combine fuzzy logic and reinforcement learning in an effective fashion. While the concept of combining the fuzzy logic with the reinforcement learning already exists, we proposed an effective way to improve the efficiency of the learning process through appropriate distribution of the reinforcement signal and secondly to illustrate successful application of this scheme for autonomous coverage optimisation of LTE systems through downtilt adjustment. In addition, the paper introduces a simple yet effective framework to construct the reinforcement learning states from the real input parameters. In contrast to the common existing solutions where the environment is typically formulated as a finite-state Markov decision process (MDP), we do not assume a stationary environment and the scheme is therefore responsive to environment changes. This is especially important for distributed multi-agent algorithms and can additionally provide a self-healing capability which is vital for a majority of self-organising networks.

The rest of the paper is organised as follows: Section II includes a brief overview of the fuzzy logic and RL and describes the proposed approach. Section III describes the LTE simulation environment used, the associated metrics and parameters. Experimental results are presented and discussed in Section IV. Finally, Section V concludes the paper.

II. Fuzzy Logic Based Reinforcment Learning

Fuzzy Inference Systems (FIS) are popular computing frameworks based on the concepts of fuzzy set theory, which have been extensively applied with success in many fields like control, decision support, system identification etc. In contrast to the classical set theory where the membership of elements in a set is assessed in binary terms, fuzzy set theory permits the graded assessment of the membership of elements in a set. This is described with the aid of a membership function valued in the real unit interval [0, 1]. A membership function represents the degree of truth of a statement and allow for partial membership in a set. The process of mapping the input values into membership functions is usually referred to as 'fuzzification'. Membership functions may be combined in fuzzy "if…then" rules to make inferences, such as "if x is high and y is low, then z is normal" in which "high", "low" and "normal" are membership functions of the matching fuzzy subsets. Fuzzy rule sets usually have several antecedents that are combined using fuzzy operators, such as AND, OR, and NOT. Finally, 'defuzzification' is the process of producing a quantifiable result by interpreting the membership degrees of the fuzzy sets into real output values. Figure 1 shows the three main components of a FIS in a block diagram.

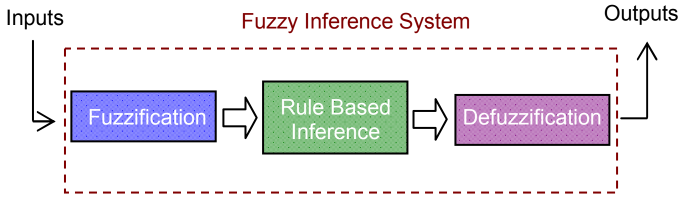

Figure 1. Components of a FIS system

Regarding the rule-based fuzzy systems, the main issue is to generate a rule-base which is adaptive and can be structured on line when no previous expert knowledge is available.

Reinforcement Learning (RL), on the other hand, is a sub-area of machine learning concerned with how an agent should take actions in an environment so as to maximize its long-term reward. Dynamic Programming, Monte Carlo methods and Temporal Difference (TD) schemes are considered as the main solution frameworks to solve RL problems. Amongst these, the TD-based methods are more appropriate for autonomous algorithms since they do not require any model of the environment and are naturally implemented in an online and fully incremental fashion [8]. TD methods learn their estimates in part on the basis of other estimates. Q-learning is a TD technique that works by learning an action-value function that gives the expected utility of taking a given action in a given state and following a fixed policy thereafter. The problem model consists of an agent, states S and a number of actions per state A. By performing an action a ($a \in A$), the agent can move from a state to another. Each state provides the agent a reward and the goal is to maximize the total reward. This is achieved by learning which action is optimal for each state. The algorithm therefore has a function which calculates the Quality of a state-action combination, Q(s, a). Each time the agent perform an action, it moves to a next state and is given a reward. Based on this, new quality values are calculated for each combination of a state and action. The core of the algorithm is a simple value iteration update:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha(s_t,a_t) \times \left[r_{t+1} + \gamma \times \underset{a}{Max} Q(s_{t+1},a) - Q(s_t,a_t) \right] \quad (1)$$

Where r_{t+1} is the reward given at time t+1and $\alpha_t(s,a)$ ($0<\alpha<1$) is the learning rates. The learning rate determines to what extent the newly acquired information will override the old information and the discount factor determines the importance of future rewards. (i.e. a reward received k time steps in the future is worth only γ^k times what it would be worth if it were received immediately). The exploitation versus exploration is another key consideration in RL systems. While the agent uses its best so-far known solution (highest Q-value) during the exploitation, exploration (defined by the exploration rate, ε) enables an agent to try other potentially suitable actions.

The main problem with RL is when the input state space or output action space is continuous or highly dimensional. Therefore, fuzzy logic can be used to provide the required level of abstraction both for state and action spaces. However, this requires a method to match the fuzzy logic to the reinforcement learning layer. In other words, there is a need for a mechanism for using the fuzzy logic input membership functions to form the reinforcement learning states and, if necessary, to identify the available actions based on the fuzzy output membership functions. For this purpose, we recommend that each input parameter is separately fuzzified. Then in the inference stage the fuzzified inputs should be combined together intersectionally using the AND fuzzy conjunction operator to form identical states. For example, a simple system consisting of 2 inputs (e.g. x and y) each being fuzzified through two membership functions (e.g. 'Low' and 'High') can create a maximum of 4 distinctive states. With this description, the case where x is 'Low' AND y is 'High' is treated as one state etc. This very simple construction of states also has the advantage that one and only one 'if...then' rule is sufficient to handle that state where the state representation appears in the antecedent part of the 'if...then' statement and the consequent is therefore the action selected for that state. Within this context, the task of the RL layer is find the best consequent for each fuzzy 'if...then' rule.

While using overlapping fuzzy membership functions resolves the issue of handling continuous inputs/output spaces, there is a new concern that is to be addressed. With overlapping membership functions more than one single membership function is activated during the fuzzification stage. Consequently different actions might be selected by different activated states and the final output of the fuzzy system is determined by the aggregation of all individual selected actions at the defuzzification stage. This means that not only more than one state is selected at each step, but also the final observed reinforcement is a result of mixture of different individual actions, making it difficult to relate the overall observed reinforcement signal to each contributing state-action pair. The firing strength of each rule is the degree to which the antecedent part of a fuzzy rule is satisfied. Since in our fuzzy setting each state is defined through only one fuzzy rule, the firing strength of each rule can be interpreted as a degree of truth of that state. We introduce the state strength of each state as the normalised firing strength of the state as:

$$\forall s \in S \qquad Ss(s) \equiv \frac{f_s}{\sum_{s'} f_{s'}} \qquad (2)$$

Where f_s represents the firing strength of the state s. Additionally, we can define a similar measure for the action space to determine the contribution of individual actions on the final collective action after the defuzzification stage. Generally speaking, since defuzzification is not a linear procedure, it is not an easy task to isolate and evaluate the impact of individual contributing actions on the overall collective result. However, the degree of truth of the state associated to each action can be considered as a fairly relevant measure. Besides, it is quite possible that identical actions are selected by a number of active states. The involvement of each individual action in deciding the final collective output would be in proportion to the sum of the firing strength of all states selecting that particular action. Therefore, we introduce a new metric as the action strength which is defined as

$$\forall a \in A \qquad As(a) \equiv \frac{\sum_{s} f_s \times Selected(s,a)}{\sum_{s'} f_{s'}} \qquad (3)$$

Where $Selected(s,a)$ is a binary metric that is defined to indicate whether or not action a is the selected action by state s:

$$Selected(s,a) = \begin{cases} 1 & if\ a\ is\ selected\ for\ state\ s \\ 0 & Otherwise \end{cases} \qquad (4)$$

For efficient credit distribution of the reinforcement signal, our proposal is to modulate the learning rate of the classical Q-learning (Equation 1) appropriately so that it firstly reflects the impact of both states and actions' strength as well as the solution being responsive to environment variations:

$$\forall s \in S, \forall a \in A \quad \alpha(s,a) = \alpha_{Max}(s,a) \times (Ss(s) \wedge As(a)) \wedge Selected(s,a) \quad (5)$$

Where Λ represent the conjunction operator and $\alpha_{Max}(s,a)$ is the maximum value of α for action *a* in state *s* and could be a fixed number for all action-sate pairs. The role of the *selected* parameter is to limit the update procedure only to the selected actions in activated states.

For the comparison purposes, we have identified a related state of the art framework referred to as Evolutionary Learning of Fuzzy Rules (ELF) [9] [10] [11] that combines the fuzzy logic and the reinforcement learning. ELF is different from the proposed scheme in various aspects: first of all, there is no concept of actions strength in the ELF's reinforcement distribution mechanism and the technique is merely based on the states' strength. Secondly, ELF is based on averaging all observations over time which makes it non-adaptive in respect to the environment's changes. The ELF reinforcement distribution and update process can be formulated as

$$Q(s_t,a_t) \leftarrow \frac{Sum(s,a)\times Q(s_t,a_t) + w(s_t,a_t)\times\left(r_{t+1} + \gamma\times \underset{a}{Max}Q(s_{t+1},a_t)\right)}{Sum(s,a) + w(s_t,a_t)} \quad (6)$$

Where $w(s,a)$ referees to the updating weight associated to the action *a* in state *s* and is defined by firing strength of the states and $Sum(s,a)$ represents the so-far sum of the weights. After updating the rule's ranking, the $Sum(s,a)$ value is updated. In ELF the issue of slow adaptation is addressed by introducing a parameter referred to as the '*EnoughTested*' which serves to define the length of the averaging window (i.e. the Sum value is capped when reaching this threshold). ELF recommends a value in range of 10-20 to be used for this parameter which unfortunately does not take into the account the requirements of the application scenario for which the algorithm is being used and the dynamics of the environment.

III. SIMULATION ENVIRONMENT

To evaluate the efficiency of the proposed algorithm, a detailed LTE simulation package was developed and used through this study. This section introduces the simulation environment and the important metrics used through the experiments.

The LTE simulation case is based on a hexagonal 7-site (single layer) deployment with 3-sector sites and wraparound propagation. The latter means that the cell layout is folded like a torus in order to avoid boundary effects. The antenna is modelled according to [12] and consists of a horizontal and a vertical pattern each parameterized with a half power beamwidth, which is effectively an opening angle, and a backward attenuation. For modelling the pathloss, the extended Okumura Hata model [13] is used. The model defines the loss of transmission power based on the distance between the channel's end points and two more parameters commonly named as the A and B parameters. Two noise sources are considered in our simulations: the thermal and the receiver noise. Thermal noise depends on the bandwidth of the radio system whereas the receiver's noise is relevant to the quality of the receiver. Table 1 summarises the simulation parameters and their associated values used through the experiments (unless otherwise explicitly stated).

Table I. Simulation Parameters

Parameters	Values
Traffic Distribution	Uniform
No Hexagon Layers	1
Wrap Around	Enabled
Simulation time step	200s
Site-to-site distance	500m
No of Sectors	3
BS Antenna Azimut	0, 120, 240 degrees
BS Antenna Max Gain	15 dBi
BS Antenna Height	32 m
BS Transceiver Power	46 dBm
Channel Passloss A	128.1
Channel Passloss B	37.6
Shadow Fading Decorrelation Distance	50 m
Shadow Fading Standard Deviation	8 dB
Bandwidth	10 MHz
UE receiver noise	8 dB
User movement	Random Walk
User speed	3 km/h
UE's antenna gain	2 dBi
UE's antenna height	1.5 m

The overall system performance can be judged through many different metrics and KPIs whilst the average users' spectral efficiency is considered to be a very common and practical measure. The spectral efficiency refers to the information rate that can be transmitted over a given bandwidth in a specific communication system and can be therefore interpreted as a measure of how efficiently a limited frequency spectrum is utilized. This metric can be constructed based on the mobile users' feedback of the Signal to Interference-plus-Noise Ratio (SINR) values. Moreover, the users on the cell boundary (or edge) experience significantly poorer performance than the terminals in the cell interior. In acknowledgement of this reality, the specification for spectrum efficiency only requires that at least 95% of the terminals should be served at better than approximately 0.1 bps/Hz [2]. The system average, on the other hand, is expected to be 2-3bps/Hz per user. Therefore, it is useful to additionally account for the lower 5-percetile of the reported spectral efficiency distribution when the performance is judged. We defined a measure of fitness as

$$\mathrm{Fitness} = \mathrm{S_{Avrg}} + \mathrm{a} \times \mathrm{S_{edge}} \quad (7)$$

Where $\mathrm{S_{Avrg}}$ represents the mean of the users spectral efficiency in bps/Hz and $\mathrm{S_{edge}}$ represents the lower 5-percentile of the users' spectral efficiency distribution. *a* (>1) is a unitless coefficient that serves to magnify the contribution of the edge users capacity (set to 2 in herein). Moreover, with the fuzzy reinforcement learning optimisation algorithm, a reward function is to be defined. In this study, the immediate reward is calculated as a delta of the fitness function for each learning agent (i.e. $r=Fitness_{new}-Fitness_{old}$). The input is considered to be the current tilt angle of each sector which is fuzzified using five membership functions (Figure 2-a) and the systems' output is the adjustment to be applied to the tilt angle similarly represented via five membership functions (Figure 2-b) .

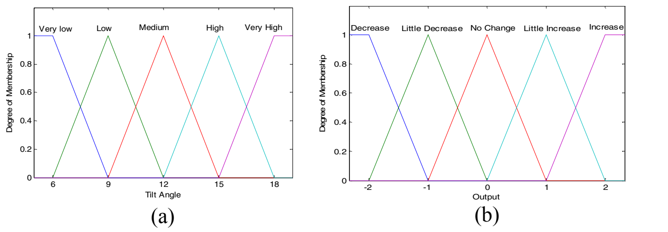

Figure 2. a) Input and b) output membership functions used through the experiments

IV. Results and Disscussions

This section presents the quantified performance of the proposed fuzzy reinforcement learning algorithm when employed for tilt angle optimisation in a LTE system. Not only that the results have been compared with the ELF scheme, but also the global optimal setting for the given scenario was found heuristically (through a global search) to confirm the convergence of the proposed solution to the optimal settings. Figure 3 compares the performance of the proposed and ELF scheme when the initial tilt angles are deliberately set randomly in range of 6 to 10 degrees and the initial Q-tables were also initialised with random generated values in range of 0 to 1. This implies a worst case scenario where no prior knowledge is available to the learning agents but also the initial tilt angles are far from their optimal range. Note that, here, the x-axis represents the total number of elapsed time steps for the whole system consisting of 21 agents (7 sites each with 3 sectors). In the discrete event simulator of this study, only a single agent is selected during each time step and the selection follows a uniform random distribution that reflects the distributed and asynchronous nature of the algorithm.

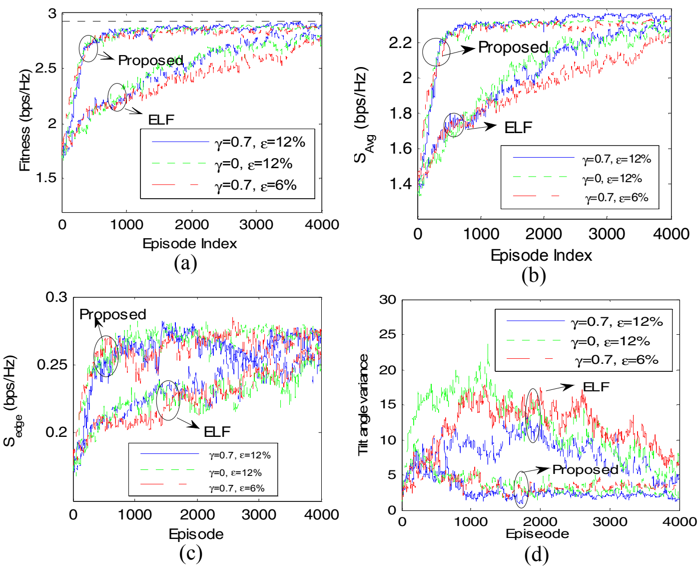

Figure 3. Performance of the proposed and the ELF scheme a) overall fitness function b) Average capacity c) Edge capacity d) variance of tilt angles over 4000 episodes

In Figure 3-a, the top dashed line represents the global optimal fitness value and as it can be seen the performance of the proposed scheme is very close to the optimal settings. The experiments were performed for a different set of learning parameters. A comparison between the performance of a myopic (γ=0) and a non-myopic system, where the γ was set to 0.7, suggests a strong similarity between these two approaches for the both proposed and the ELF scheme. The reason behind this observation is associated to the relatively simple structure of the environment and its response to the parameter under control (i.e. the downtilt angle herein). More specifically speaking, there are no local maxima points in the reward function and the immediate reward from the environment is well aligned with the global optimisation goal.

The exploration rate (ε) is also another important parameter. As the graphs show, the impact of the exploration rate is not very significant for the proposed algorithm but has a noticeable impact on the ELF scheme. Despite using a limiting window, the fact that ELF employs averaging operation over a number of observations makes it slow in learning and therefore, further reduction of the exploration rate can intensify this effect and impact on its performance. Since the placement of the sites follows a hexagonal arrangement with uniform users' distribution, the optimal tilt angle of all sectors should converge to a single value, implying that the tilt angle variance should decreases as the optimisation proceed. This measure can serve as an indication of the convergence of the algorithm as well. Figure 3-d compares the performance of proposed versus the ELF algorithm in this regard.

To get an insight to the core of the reinforcement learning process, Figure 4-a illustrates the tilt angle and the strength of all states starting from the state 1 on top to the state 5 at the bottom for one of the sites. In this particular example, the initial Q-table was deliberately set completely against the optimal policy to examine the response of the learning agent. Figure 4-b shows the final response of the optimiser that is the change to the current tilt angle in degrees after defuzzification. Considering that in this scenario the optimal tilt angle is around 15°, this results confirms that despites a complete inappropriate initialisation of the Q-table (and fuzzy 'if...then' rules) the agent can rapidly policy learn and converge toward optimal settings.

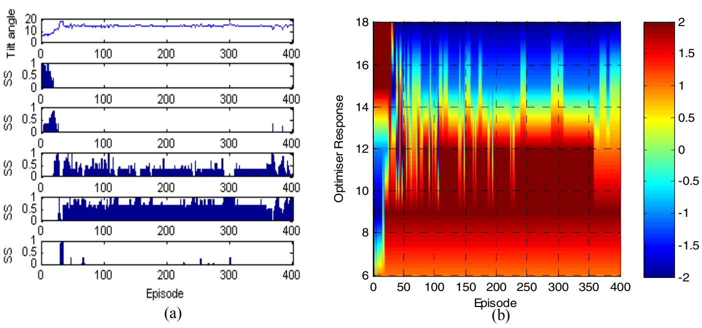

Figure 4. Tilt angle and the strength of different states from state 1 (second top) to 5 (the bottom) b) Optimisers' response

To show the self healing properties of the proposed method, the central site and all its associated sectors were disabled during the operation of the algorithm. Figure 5-a shows the response of the proposed and the ELF algorithm when the central site is disabled in the 20th episodes after the network was operating close to its optimal region. As the graph shows, while both algorithms are successful in recovering the loss of the central site, the recovery time is improved when using the proposed algorithm. Such recovery occurs when the neighbouring sectors extending their transmission range by reducing their tilt angle and cover the area where was formerly served by the failed site. In fact with an appropriate Q-table update procedure and reinforcement distribution mechanism, the inherent exploration feature of the RL enables the algorithm to be responsive to the environment's changes by allowing different solutions to be tried out. Figure 5-b shows the response of one of the neighbouring sites when reducing the tilt angle after the failure occurs.

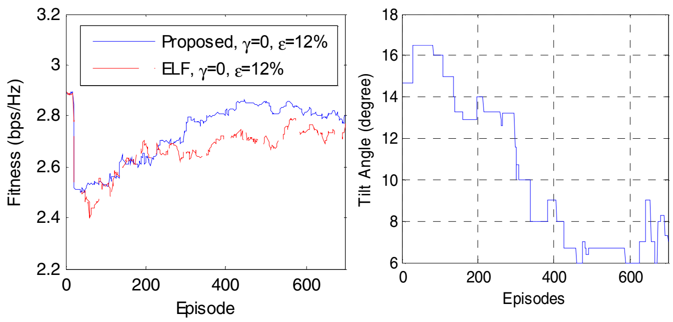

Figure 5. a) Fault recovery functionality in the proposed and ELF scheme b) Response of a neighbouring sector to the fault

It is very critical for an algorithm to be able to operate reliably under various circumstances. Therefore, the robustness of the algorithm against the environment's noise is of special importance. A major benefit of fuzzy logic systems is their ability to handle imprecise and abstract information which is important for many self-organizing networks. For tilt optimisation, there is a high probability of inaccurate feedback reports from the mobile users. This can be due to many factors including the imprecise measurement of the reported performance metrics. Additionally, one can easily trade off between the accuracy of the reported feedbacks and speed of the overall algorithm by shortening the learning time steps (episodes' length). Fortunately, the experiments confirm the ability of the proposed scheme to cope with extremely noisy environment. In a set of experiments, zero-mean and uniformly distributed random noise was applied to the most critical element of the learning process that is the reported reward value from the environment (mobile users herein). As illustrated in Figure 6, the proposed algorithm performs reasonably even under exceedingly noisy environment where the Signal-to-Noise-Ratio of the reward signal is set to 8dB. Interestingly, due to the averaging behaviour, the ELF algorithm also performs well in a noisy environment.

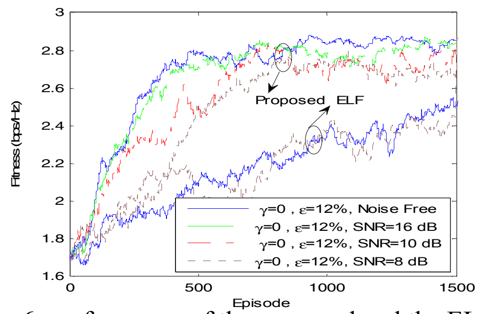

Figure 6. performance of the proposed and the ELF scheme in a noisy learning environment

V. CONCLUSION AND REMARKS

Presented is a fuzzy logic based RL method for enabling self-organizing features for the next generation communication networks. In this study, the coverage and capacity optimization of LTE networks through downtilt angle adjustment is considered. Subsequently, a new reinforcement distribution mechanism was proposed to address multiple simultaneous states' selection when fuzzy logic and RL are combined. As a by product of this, a simple yet effective RL state construction algorithm is introduced. The overall solution is fully distributed with no signaling overhead between LTE base-stations and requires no a *priori* information or human interventions. Through the experiments, it was shown that not only the solution is capable of self-configuring and self-optimizing the LTE base-stations' tilt angle, but it also provides self-healing properties and can operate acceptably under exceedingly noisy environments with inaccurate input information. Compared to an existing Fuzzy logic RL framework, up to 20% performance gain was observed.

REFERENCES

[1] M.Celentano "Carrier Capital Expenditures" IEEE Comms Mag, Jun 2008.
[2] S. Smith, "System and Method for Determining Optimal Broadcast Area of an Antenna", US Patent No. 7006038, 28 Feb 2006
[3] J.Niemela, etc "Sensitivity of optimum downtilt angle for geographical traffic load distribution in WCDMA" 62nd IEEE Vehicular Technology Conference, Pp. 1202- 1206, Dallas fall 2005.
[4] D.Barker and K. Radousky "The Path LTE Overlay Optimization", Quintel technical white paper, January 2010.
[5] F. Gunnarsson et al., "Downtilted Base Station Antennas – A Simulation Model Proposal and Impact on HSPA and LTE Performance", IEEE Vehicular Technology Conference, Calgary, Fall 2008.
[6] A. Wacker, etc "Automated and remotely optimization of antenna subsystem based on radio network performance" IEEE Symposium on Wireless Personal Multimedia Communications, pp. 752 - 765, Hawaii, 2002.
[7] J.-S. Wu etc "Hot-Spot Traffic Relief with a Tilted Antenna in CDMA Cellular Networks", IEEE Trans. Veh. Tech., Vol. 47, pp. 1-9, 1998
[8] R. Sutton "Reinforcement Learning: An Introduction" MIT Press, 1998.
[9]A. Bonarini etc "Evolutionary learning of fuzzy rules: Competition and cooperation," Fuzzy Modeling: Paradigms and Practice, W. Pedrycz, Ed. Norwell, MA: Kluwer, 1996, pp. 265-284.
[10] A. Bonarini, "Reinforcement distribution to fuzzy classifiers: A methodology to extend crisp algorithms," in IEEE International Conference Evolutionary Computation, pp. 51–56 Los Alamitos, CA: 1998,
[11] A. Bonarini, "Learning fuzzy classifier systems," in Learning Classifier System: New Directions and Concepts, Springer-Verlag, pp. 83–106, 2000.
[12] 3rd Generation Partnership Project, "Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)," 3GPP TR25 814, Sept. 2006.
[13] 3rd Generation Partnership Project, "RF System Scenarios (Release 1999)", 3GPP TR 25.942 V3.3.0 June 2002
[14] "Ultra Mobile Broadband Technology Overview and Competitive Advantages", Qualcomm Technical White Paper, 2008.