An MIMO Configuration Mode and MCS Level Selection Scheme by Fuzzy Q-Learning for HSPA^+ Systems

Wen-Ching Chung, *Member*, *IEEE*, Chung-Ju Chang, *Fellow*, *IEEE*, Kai-Ten Feng, *Member*, *IEEE*, and Ying-Yu Chen

Abstract—In this paper, we propose a fuzzy Q-learning-based MIMO configuration mode and MCS level (FQL-MOMS) selection scheme for high speed packet access evolution (HSPA^+) systems. The FQL-MOMS selection scheme intends to enhance the system throughput under the block error rate (BLER) requirement guarantee. It will determine an appropriate MIMO configuration mode and MCS (modulation and coding scheme) level for packet data transmission in HSPA^+ systems, under the situations that the channel status is varying and the channel quality indication (CQI) has report delay. The FQL-MOMS scheme considers not only the reported CQI and the last transmission result but also the BLER performance metric and the transmission efficiency. Moreover, it is effectively configured, where the fuzzy rules and the reinforcement signals for the Q-learning algorithm are sophisticatedly designed. Simulation results show that the proposed FQL-MOMS scheme increases the system throughput by up to 49.3 and 35.9 percent, compared to the conventional adaptive threshold selection (ATS) scheme [12] and the Q-HARQ scheme [14], respectively, under the BLER requirement fulfillment.

Index Terms—HSPA^+, MIMO, MCS, HARQ, BLER, fuzzy logic, and Q-learning.

1 Introduction

High speed packet access evolution (HSPA^+) has been introduced in Release 7 by the 3rd generation partnership project (3GPP) for universal mobile telecommunications system (UMTS) [1]. HSPA^+ adopts many effective techniques to enhance the performance of high speed downlink packet access (HSDPA) services proposed in Release 5, such as multiple-input multiple-output (MIMO), higher order modulation and coding, continuous packet connectivity, and so on [2].

When MIMO is enabled in HSPA^+, a larger peak data rate can be achieved through the spatial multiplexing (SM) when channel quality is good, or a higher link reliability can be provided through the spatial diversity (SD) when channel quality is bad [3], [4], [5]. 3GPP extended the double transmitter antenna array (D-TxAA) in Release 99 to be the standard of Release 7 [6], [7]. The use of multiple transmit antennas at the base station can provide diversity gain without additional receiver chains at the mobile terminal. When the channel quality is good, the D-TxAA transmits two data streams (transport blocks) simultaneously over the radio channel to increase the system throughput by using the same channelization codes. Each data stream is processed and coded separately. When the channel quality is bad, the D-TxAA transmits one data stream through two antennas to increase the probability of the successful decoding. Consequently, there are totally two MIMO configuration (transmission) modes. How to determine an appropriate MIMO configuration mode to increase the system throughput is an interesting and important issue.

At the data link layer of protocol for packet data transmission in HSPA^+ systems, the hybrid automatic repeat request (HARQ) is conventionally the error control scheme. The HARQ control scheme combines forward error correction (FEC) mechanism with the original ARQ scheme. Li and Zhao analyzed the ARQ scheme adopted in parallel multichannel communications for error control [8]. Multiple parallel channels are often created by using orthogonal frequency division multiplex (OFDM) technology or MIMO technology. Wang and Chang studied the performance analysis of stall avoidance schemes for HSDPA with parallel HARQ mechanisms and determined a proper number of processes for the parallel HARQ mechanisms [9]. On the other hand, there are traditionally three kinds of schemes to implement the HARQ scheme: chase combining (CC), incremental redundancy (IR), and partially IR schemes. The performance of these three schemes was studied in [10] and [11]. Since the IR scheme has higher error correction ability, we adopt the IR scheme to implement the HARQ scheme in this paper.

As for the modulation and coding scheme (MCS) level selection in the HARQ control scheme, the 3GPP has predefined a table to show the relationship between the MCS level and the required channel quality indication (CQI) to fulfill the block error rate (BLER) performance requirement, according to fixed signal-to-interference-plus-noise ratio (SINR) thresholds [1]. However, the relationship definition is by a fixed threshold selection (FTS) method

- *The authors are with the Department of Electrical Engineering, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan 300, ROC. E-mail: wenching.ece88g@nctu.edu.tw, cjchang@mail.nctu.edu.tw, ktfeng@mail.nctu.edu.tw, skocea.cm96g@g2.nctu.edu.tw.*

Manuscript received 12 Aug. 2010; revised 3 June 2011; accepted 10 June 2011; published online 22 June 2011.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2010-08-0379.
Digital Object Identifier no. 10.1109/TMC.2011.139.

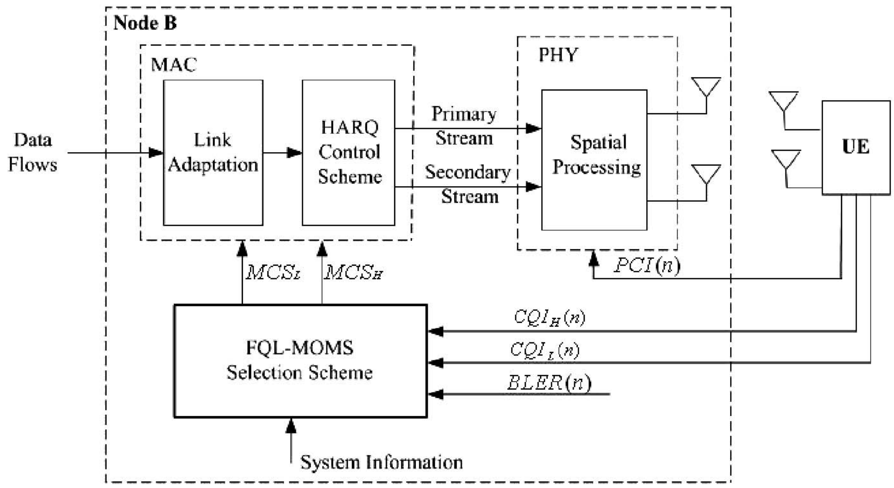

Fig. 1. The HSPA^+ system with the FQL-MOMS scheme.

which has some drawbacks since the CQI has report delay and measurement inaccuracy. Nakamura et al. proposed an adaptive threshold selection (ATS) scheme to adjust the SINR threshold for the selected MCS level by considering the last transmission result [12]. Also, Muller and Chen further adjusted the SINR threshold for each MCS level according to not only the last transmission result but also the CQI delay [13]. However, these two adaptive schemes still cannot cope with the rapidly varying channel status since the granularity of the SINR threshold adjustment according to the reported acknowledgment (ACK) or negative acknowledgment (NACK) is coarse or, say, inflexible. A new HARQ scheme by using Q-learning algorithm (Q-HARQ) was proposed in [14] to learn the policy of MCS decision. The Q-learning algorithm is a kind of unsupervised learning methods, which can learn optimal control rules through dynamic interactions with the environment [15]. By using the feedback reinforcement signal obtained from the HSDPA system, the Q-HARQ scheme proposed in [14] outperforms the conventional IR scheme and two other link adaptation IR (LA-IR) schemes proposed in [16] in the total system throughput. The reason is that the Q-learning algorithm is performed in a closed-loop iteration manner such that the optimal solution for the initial packet MCS can be found.

On the other hand, fuzzy logics has been well developed [17], [18]. Fuzzy logics emulates the way of human thinking to describe the behavior of systems which are too complicated to tackle mathematically. Also, it can deal with uncertainty and imprecision knowledge to reason a suitable decision. By combining the fuzzy logics with the Q-learning algorithm, the fuzzy Q-learning (FQL) algorithm [19], [20], [21], [22], [23], [24], [25] takes advantages of both Q-learning algorithm and fuzzy logics to efficiently learn optimal control rules for uncertainty and imprecision problems. Therefore, the FQL algorithm can use domain knowledge to accelerate the Q-learning algorithm to find the optimal solution of the MCS selection for packet transmission. A preliminary study on the selection of MCS level but without MIMO configuration mode for an HARQ system using the FQL algorithm has been given in [26], under the BLER requirement guarantee.

In this paper, we propose a fuzzy Q-learning-based MIMO configuration mode and MCS level (FQL-MOMS) selection scheme for HSPA^+ systems. The FQL-MOMS scheme intends to enhance the system throughput while guarantee the BLER performance requirement. In order to overcome the imprecise problem of CQI report delay, an effective configuration for the FQL-MOMS selection scheme is designed. The FQL-MOMS selection scheme adopts fuzzy logics to reason a suitable MIMO configuration (transmission) mode and an MCS level simultaneously for a new packet data transmission. Also, to accelerate the learning process, the fuzzy rule bases for SD and SM are all partitioned into three regions. Moreover, the FQL-MOMS selection scheme uses the Q-learning algorithm to learn optimal decision values for rules in its configuration. The fuzzy rule bases are updated by feedback reinforcement signals, which are designed according to not only the CQI and the transmission results but also the BLER performance metrics and the transmission efficiency. Because of the closed-loop iteration manner, the fuzzy Q-learning can learn an excellent relationship between the MIMO configuration mode combined with MCS level and the measured BLER with delayed CQI. Simulation results show that, when user mobility is at around 60 km/hr, the proposed FQL-MOMS selection scheme increases the system throughput by up to 49.3 percent compared to the conventional ATS scheme [12], and by up to 35.9 percent compared to the Q-HARQ scheme [14].

The remainder of the paper is organized as follows: The system model is introduced in Section 2. In Section 3, the design of the FQL-MOMS scheme is presented. It is followed by the performance analysis of the FQL-MOMS scheme in Section 4. Finally, conclusions are given in Section 5.

2 System Model

2.1 HSPA^+ System

In the HSPA^+ system shown in Fig. 1, both base station (Node B) and user equipment (UE) are assumed to be equipped with two antennas. Node B controls link adaptation and physical layer transmission and retransmission combining to fast react the variation of channel condition. The *link adaptation* performs a suitable MIMO configuration

TABLE 1
MCS Levels for the HSPA$^+$ System

MCS Level	Transport Block Size	Modulation Order	Coding Rate
1	4581	QPSK	1/3
2	6673	QPSK	1/2
3	8574	QPSK	2/3
4	10255	QPSK	3/4
5	12488	16QAM	1/3
6	14936	16QAM	1/2
7	17548	16QAM	2/3
8	20617	16QAM	3/4

mode and MCS level for data flow transmission according to the decision from the *fuzzy Q-learning-based MIMO configuration mode and MCS level selection scheme*, which will be designed in Section 3. The data flow is divided and coded into suitable data streams by the *HARQ control scheme*.

Data flows for UE are transmitted via high-speed downlink shared channel (HS-DSCH) in HSPA$^+$ [1], which supports multicode transmission and code multiplexing of different users. The coding scheme of HS-DSCH is by turbo code with lowest code rate $1/3$, and the code rate used in this study is considered as $\{1/3, 1/2, 2/3, 3/4\}$ [27]. The modulation order supported in the operation is QPSK and 16-QAM. Therefore, the HSPA$^+$ system totally has eight kinds of MCS levels [1], shown in Table 1. Each MCS level has different transport block size, which is the number of data bits transmitted in one data stream. Let MCS_k be the kth MCS level, $k = 1, \ldots, 8$. Note that the higher MCS level has higher system throughput but less reliability for data transmission.

The MIMO configuration has two modes of operation, based on the downlink channel condition reported in CQI by UE. Mode 1 is with SD transmission and mode 2 is with SM transmission. Let MCS_H (MCS_L) be the MCS level $H(L)$ assigned to the primary (secondary) data stream with the CQI_H (CQI_L), $1 \leq MCS_L \leq MCS_H \leq 8$. In the SD transmission mode, only primary stream carries data packet with proper MCS_H level to enhance the reliability of transmission. In the SM transmission mode, primary and secondary streams carry different data packets with proper MCS_H and MCS_L levels, respectively, to increase the system throughput. Notice that the primary (secondary) stream is the one with large (small) CQI. After the *spatial processing*, the data stream(s) will be transmitted to UE through the two antennas. A packet transmission time interval (TTI) is specified as 2 ms to achieve shorter round trip delay. Each data stream for UE is processed and coded separately per TTI. Node B sets precoding weights for each data stream according to the precoding control indication (PCI) reported by UE.

Control information for UE is delivered via high-speed shared control channel (HS-SCCH), which contains the channelization code set, modulation scheme, transport-block size, HARQ-related parameters, and CRC attachment. This control information is used to decode the data transported by the HS-DSCH and to perform the soft combining when the retransmission is carried out. Control information to Node B from UE is via high-speed dedicated physical control channel (HS-DPCCH), which consists of ACK or NACK messages, PCI, and CQI value. The ACK/NCAK message indicates the result of the HS-DSCH decoding. The CQI contains the channel quality indications of the primary and the secondary data streams, denoted by CQI_H and CQI_L, respectively, and $CQI_H \geq CQI_L$. The values of CQI_H and CQI_L are measured by UE through common pilot channel (CPICH) and are an integer ranging from 0 to 14 [1]. According to the reported CQIs, measured BLER, and some system information, the FQL-MOMS selection scheme designed in the Node B will choose a suitable MIMO configuration mode and MCS level for packet data transmission to enhance the system throughput and to satisfy the BLER requirement for the HSPA$^+$ system.

2.2 Channel Model

In this study, we consider the HSPA$^+$ system in an urban area with a terrestrial mobile radio channel. Three types of propagation factor are considered in the channel model: path loss, slow fading resulting from shadowing and scattering, and fast fading due to multipath effects. Let η be a normal-distributed random variable with zero mean and variance σ_L^2. The channel condition at time t, denoted by $F(t)$, can be expressed as

$$F(t) = \xi(d) \times 10^{\eta/10} \times \zeta(t), \tag{1}$$

where d is the distance between the Node B and the UE, $\zeta(t)$ is the short-term fading, and $\xi(d) \times 10^{\eta/10}$ is the long-term fading with path loss $\xi(d)$ and shadowing effect $10^{\eta/10}$. The short-term fading is caused by the multipath effects and is given by Stuber [29]

$$\zeta(t) = 2\sigma\sqrt{\frac{2}{L}}\sum_{m=1}^{M}\cos(2\pi f_D t\cos(2\pi m/L) + \theta_m)e^{j\phi_m}, \tag{2}$$

where $L = 4M + 2$ is the number of the signal path, $\phi_m = \pi m/(M+1), \theta_m = \phi_m + 2\pi s/(M+1), s = 0, 1, \ldots, M-1, \sigma$ is the radical of the average power signal, and f_D is the Doppler frequency.

When the position of the UE changes, the shadowing effect of the UE is different. Since the sampling rate in HARQ is very fast compared to the motion of the UE, the shadowing effects of two sampling points are highly correlated in practical. Let Δx be the user distance between two sampling points. According to [30], the correlation of shadow fading, denoted by $\rho(\Delta x)$, is modeled by a normalized autocorrelation function and is given by $\rho(\Delta x) = e^{-\frac{|\Delta x|}{d_{cor}}\ln 2}$, where d_{cor} is the decorrelation length.

3 FQL-MOMS Selection Scheme

The fuzzy Q-learning-based MIMO configuration (transmission) mode and MCS level (FQL-MOMS) selection scheme is here designed to determine a suitable MIMO configuration mode with MCS level for packet data transmission in HSPA$^+$ systems. As known, the selection decision for packet transmission should be based on current and past system states. Thus, this selection scheme can be modeled as a discrete-time Markov decision process (MDP). We here adopt the fuzzy Q-learning algorithm to solve this MDP problem.

Functional blocks of the FQL-MOMS selection scheme are depicted in Fig. 2. Since the MIMO transmission mode

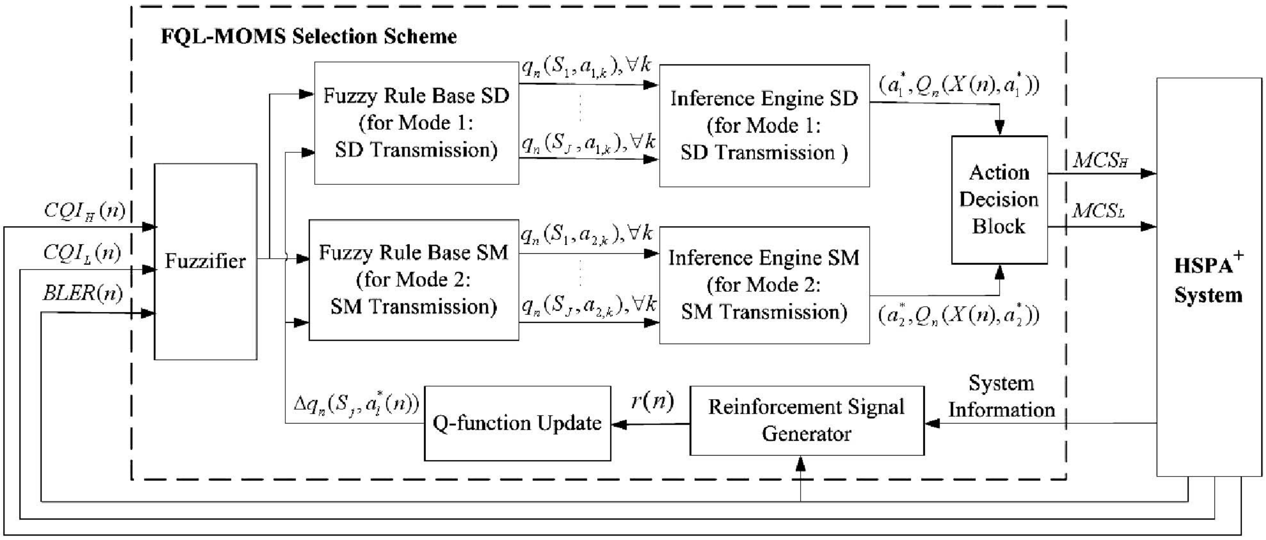

Fig. 2. The functional blocks of the FQL-MOMS selection scheme.

with the MCS level is tightly related with channel quality and the resulted performance metric BLER, at each nth TTI (episode), the FQL-MOMS selection scheme chooses three measures to be the input linguistic variables. They are $BLER(n)$, $CQI_H(n)$, and $CQI_L(n)$, where the $BLER(n)$ is defined as the number of packets needed retransmission over the total number of transmitted packets over an observation window at the nth TTI. The input state vector, denoted by $X(n)$, is $(BLER(n), CQI_H(n), CQI_L(n))$. There are two fuzzy rule bases designed in the FQL-MOMS selection scheme. *Fuzzy rule base SD* is for the SD transmission mode and *fuzzy rule base SM* is for the SM transmission mode. The Q-value for the state-action pair at the nth episode, denoted by q_n, can then be obtained form these two rule bases. The optimal action for each transmission mode is inferred from corresponding *inference engine*. By *action decision block*, a suitable MIMO transmission mode with an MCS level can be determined. The *reinforcement signal generator* will generate a reinforcement signal, denoted by $r(n)$, based on not only the results of transmission and retransmission but also the BLER performance metric and the transmission efficiency. This $r(n)$ will be used to finely update the Q-values in fuzzy rule base in the block of *Q-function update*. Detailed design of each functional blocks are given as follows:

3.1 Fuzzifier

The fuzzifier receives the system state vector $X(n) = (BLER(n), CQI_H(n), CQI_L(n))$ as input linguistic vector. The $BLER(n)$ is an essential performance metric for HSPA$^+$ service which has BLER requirement, denoted by $BLER^*$. This performance metric $BLER(n)$ would be in satisfaction, attention, or violation region with respect to the performance requirement $BLER^*$ of the HSPA$^+$ service. Therefore, the fuzzifier defines the fuzzy term set for $BLER(n)$ as $T(BLER(n)) = \{Satisfaction\,(S), Attention\,(A), Violation(V)\}$. The membership function of each term α in $BLER(n)$, denoted by $\mu_\alpha(BLER(n)), \alpha = S, A,$ or V, is given in Fig. 3. This membership function for the fuzzy term indicates that the intensity of the input variable belong to itself fuzzy labels, and is designed with preknowledge of the system. In Fig. 3, we set $\mathrm{A}_1 = 0.5 \times BLER^*(\mathrm{A}_4 = BLER^*)$ to indicate that the region of $(0, A_1)$ $((A_4, 1])$ where the BLER performance metric is resided is in the satisfaction (violation) situation. Also, we set $\mathrm{A}_2 = 0.7 \times BLER^*$ and $\mathrm{A}_3 = 0.9 \times BLER^*$ to represent whenever the BLER performance measure is in this region approaching to $BLER^*$, the BLER performance metric is in the attention situation.

The fuzzy term sets for $CQI_H(n)$ and $CQI_L(n)$ are defined as

$$T(CQI_H(n)) = T(CQI_L(n)) = \{Level\ 1\ (L1), Level\ 2\ (L2), Level\ 3\ (L3), Level\ 4\ (L4), Level\ 5\ (L5), Level\ 6\ (L6), Level\ 7\ (L7), Level\ 8\ (L8)\}.$$

Terms in $T(CQI_H(n))$ and $T(CQI_L(n))$ represent the ranges of the channel quality indication, and each term will correspond to one MCS level which can guarantee the BLER requirement $BLER^*$ during this CQI range. The membership functions of the term β in $CQI_H(n)$ and $CQI_L(n)$, denoted by $\mu_\beta(CQI_H(n))$ and $\mu_\beta(CQI_L(n))$, respectively, $\beta = L1, L2, \ldots,$ or $L8$, are designed to be the same and are shown in Fig. 4. In the figure, we set $\mathrm{B}_k, k = 1, \ldots, 8$, to be the required CQI to maintain $BLER^*$ for MCS_k.

Since $CQI_H(n)$ is greater than or equal to $CQI_L(n)$, there are totally $J = 108$ kinds of $X(n)$. Let S_j be the fuzzy linguistic terms of $X(n), 1 \leq j \leq J$. The intensity of $X(n)$ belonging to S_j, denote by $\mu_{j,n}$, is obtained by the membership functions of $BLER(n)$, $CQI_H(n)$, and $CQI_L(n)$ via a max-product operation [17], [18]. It is given by

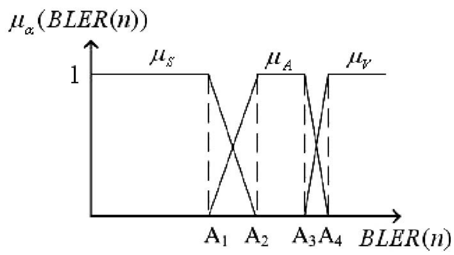

Fig. 3. The membership function of $\overline{BLER}(n)$.

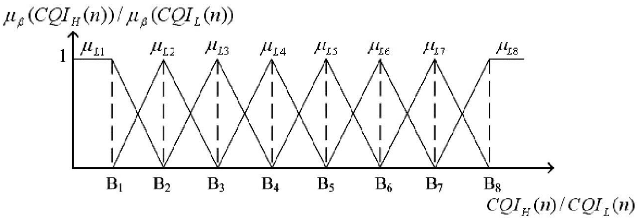

Fig. 4. The membership function of $CQI_H(n)$ or $CQI_L(n)$.

$$\mu_{j,n} = \mu_\alpha(BLER(n)) \times \mu_\beta(CQI_H(n)) \times \mu_\beta(CQI_L(n)). \quad (3)$$

3.2 Fuzzy Rule Base

Two fuzzy rule bases are constituted in the FQL-MOMS selection scheme. They are fuzzy rule base SD for the SD transmission mode and fuzzy rule base SM for the SM transmission mode, where each input vector has one fuzzy Q-learning rule. The fuzzy rule in the fuzzy rule base SD is designed as

$$\begin{aligned}&\text{Rule } j: \text{if } X(n) \text{is } S_j, \text{then } a_{1,k} \text{ with } q_n(S_j, a_{1,k}),\\&\quad \text{for } 1 \le j \le J, 1 \le k \le 8,\end{aligned} \quad (4)$$

where $a_{1,k}$ is the action for the SD transmission mode and $q_n(S_j, a_{1,k})$ is the Q-value for the state-action pair $(S_j, a_{1,k})$. If $a_{1,k}$ is selected, the FQL-MOMS selection scheme uses the kth MCS level for the primary data stream. Similarly, the fuzzy rule in the fuzzy rule base SM is designed as

$$\begin{aligned}&\text{Rule} j: \text{ if } X(n) \text{ is } S_j, \text{ then } a_{2,k} \text{ with } q_n(S_j, a_{2,k}),\\&\quad \text{for } 1 \le j \le J, 1 \le k \le 36,\end{aligned} \quad (5)$$

where $a_{2,k}$ is the action for the SM transmission mode. If $a_{2,k}$ is selected, the FQL-MOMS selection scheme assigns the MCS level $H(MCS_H)$ for the primary data stream with $CQI_H(n)$ and the MCS level $L(MCS_L)$ for the secondary data stream with $CQI_L(n)$. The mappings of MCS_H and MCS_L with respect to $a_{2,k}$ are given in Table 2, where $k = f(H, L) = H - 8 + 0.5L(17 - L)$.

In these two fuzzy rule bases, Q-value is learned via the reinforcement signal at each TTI. The reinforcement signal is designed to reward (punish) the selected action which can increase the system throughput (guarantee the BLER requirement) under the BLER performance requirement. Therefore, the Q-value can be seen as the preference value for each action under different system state.

In order to accelerate the learning process, we have divided each fuzzy rule base into satisfaction, attention, and violation regions according to the fuzzy terms of $BLER(n)$. The candidate set for action selection in each region is decided by the fuzzy terms of $CQI_H(n)$ and $CQI_L(n)$, since each fuzzy term for CQIs indicates one kind of MCSs to guarantee the BLER requirement $BLER^*$ during this CQI region. If $BLER(n)$ is in the $Satisfaction$ region, it means that the BLER performance is much smaller than the $BLER^*$, and a more aggressive action should be selected so as to increase the system throughput. Since the action with larger MCS level can transmit more information bits, the actions with the required CQI larger than the reported CQI are considered as action candidates. If $BLER(n)$ is in the $Attention$ region, it means that the BLER performance is close to the $BLER^*$, and the BLER performance can be kept to still remain in this region. In order to keep the BLER performance unchanged, only the actions with the required CQI at around the reported CQI are considered as action candidates. If $BLER(n)$ is in the $Violation$ region, it means that the BLER requirement is violated, and a more conservative action should be selected so as to comply the BLER requirement of the HSPA service. In order to recover the BLER performance from the violation region, only the actions which the required CQI is less than the reported CQI are considered as action candidates. In summary, the fuzzy rules bases SD and SM are given by:

Fuzzy Rule Base SD:

- If $BLER(n)$ is in satisfaction region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{1,k}$ with $q_n((S, Lb, Lc), a_{1,k}), \lceil \frac{b+c}{2} \rceil \le k \le 8$.
- If $BLER(n)$ is in attention region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{1,k}$ with $q_n((A, Lb, Lc), a_{1,k}), k = \lfloor \frac{b+c}{2} \rfloor - 1, \ldots, b+1$.
- If $BLER(n)$ is in violation region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{1,k}$ with $q_n((V, Lb, Lc), a_{1,k}), k = c-1, \ldots, b-1$.

Fuzzy Rule Base SM:

- If $BLER(n)$ is in satisfaction region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{2,k}$ with $q_n((S, Lb, Lc), a_{2,k}), k = f(H, L), H > b, L > c, H \ge L$.

TABLE 2
The Mapping between the MCS Levels and Action $a_{2,k}$

$MCS_H \diagdown MCS_L$	1	2	3	4	5	6	7	8
8	$a_{2,8}$	$a_{2,15}$	$a_{2,21}$	$a_{2,26}$	$a_{2,30}$	$a_{2,33}$	$a_{2,35}$	$a_{2,36}$
7	$a_{2,7}$	$a_{2,14}$	$a_{2,20}$	$a_{2,25}$	$a_{2,29}$	$a_{2,32}$	$a_{2,34}$	
6	$a_{2,6}$	$a_{2,13}$	$a_{2,19}$	$a_{2,24}$	$a_{2,28}$	$a_{2,31}$		
5	$a_{2,5}$	$a_{2,12}$	$a_{2,18}$	$a_{2,23}$	$a_{2,27}$			
4	$a_{2,4}$	$a_{2,11}$	$a_{2,17}$	$a_{2,22}$				
3	$a_{2,3}$	$a_{2,10}$	$a_{2,16}$					
2	$a_{2,2}$	$a_{2,9}$						
1	$a_{2,1}$							

- If $BLER(n)$ is in attention region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{2,k}$ with $q_n((A, Lb, Lc), a_{2,k}), k = f(H, L), b-1 \leq H \leq b+1, c-1 \leq L \leq c+1, H \geq L$.
- If $BLER(n)$ is in violation region, $CQI_H(n)$ is Level b, and $CQI_L(n)$ is Level c, then $a_{2,k}$ with $q_n((V, Lb, Lc), a_{2,k}), k = f(H, L), H < b, L < c, H \geq L$.

3.3 Inference Engine and Action Decision

By using a select-max strategy in the exploration/exploitation policy (EEP) [23], [25], a most suitable action for rule j in fuzzy rule base SD, denoted by $a^*_{1,j}(n)$, can be obtained by

$$a^*_{1,j}(n) = \arg\max_k q_n(S_j, a_{1,k}). \tag{6}$$

The $a^*_{1,j}(n)$ in (6) is also called the greedy action. In order to explore the set of possible actions and acquire experience through the reinforcement signals, the action of each rule is selected according to the select-max strategy of EEP [23], [25]. In the exploitation policy, the greedy action is selected. In the exploration policy, one of the nongreedy actions is selected to produce a larger total reward in the long iterations. Then, the optimal action for the SD transmission mode, denoted by $a^*_1(n)$, can be inferred by

$$a^*_1(n) = \frac{\sum_{j=1}^{J} \mu_{j,n} \times a^*_{1,j}(n)}{\sum_{j=1}^{J} \mu_{j,n}}, \tag{7}$$

and the Q-value for $a^*_1(n)$ can then be obtained by

$$Q_n(X(n), a^*_1(n)) = \frac{\sum_{j=1}^{J} [\mu_{j,n} \times q_n(S_j, a^*_{1,j}(n))]}{\sum_{j=1}^{J} \mu_{j,n}}. \tag{8}$$

Using the same strategy in (6), a most suitable action for rule j in fuzzy rule base SM, denoted by $a^*_{2,j}(n)$, can be obtained. Similarly, taking $a^*_{2,j}(n)$ instead of $a^*_{1,j}(n)$ into (7) and (8), we can get the optimal action for the SM transmission mode, denoted by $a^*_2(n)$, and the Q-value for $a^*_2(n)$, denoted by $Q_n(X(n), a^*_2(n))$. Note that, since the output value of the FQL-MOMS selection scheme should be an integer, the optimal actions $a^*_1(n)$ and $a^*_2(n)$ are rounded off to be integer values.

The proper transmission mode at every episode n, denoted by $\bar{z}$, can then be decided in the *action decision block* by

$$\bar{z} = \arg\max_{z=1,2} Q_n(X(n), a^*_z(n)). \tag{9}$$

If $\bar{z} = 1$, the FQL-MOMS selection scheme selects the SD transmission mode and the outputs $(MCS_H, MCS_L) = (a^*_1(n), 0)$ for single primary data stream. If $\bar{z} = 2$, it chooses the SM transmission mode and outputs $(MCS_H, MCS_L) = (MCS_{H^*}, MCS_{L^*})$ based on $a^*_2(n) = f(H^*, L^*)$ for (primary, secondary) data streams given in Table 2.

3.4 Reinforcement Signal Generator

The reinforcement signal generator in the FQL-MOMS selection scheme will generate a reinforcement signal to effectively update the Q-value for each state-action pair, at every TTI. The reinforcement signals generated for fuzzy rules in the same region of the fuzzy rule base are designed to be in the same form. Also, in order to let Q-values in SD and SM fuzzy rule bases have the similar degree of reward and punishment, the reinforcement signals in these two rule bases have the similar form. The ratio of the largest reward over the severest punishment for the reinforcement signal is set around at 0.1 since the $BLER^* = 0.1$. In the following, the reinforcement signals is sophisticatedly designed so as to achieve a great granularity and large dynamic range to intelligently tune the Q-value in the fuzzy rule bases SD and SM.

3.4.1 Reinforcement Signal for SD Fuzzy Rule Base

If $BLER(n)$ is in the satisfaction region, a more aggressive action to increase the system throughput should be encouraged. As noted, the $BLER(n)$ can be getting better only when the initial packet is successfully received without retransmission. Therefore, in order to act in a more aggressive manner, a reward feedback reinforcement signal is given if a packet is successfully received no matter whether the retransmission is occurred or not. The reinforcement signal is designed, according to the results of both initial transmission and retransmissions, by

$$r(n) = \begin{cases} \dfrac{R_k}{R_{d,k} + R^*}, & \text{if packet is successfully received,} \\ -(\delta + \tau \times BLER(n)), & \text{if packet is dropped,} \end{cases} \tag{10}$$

where R_k is the number of information bits of the transmitted packet with the selected action $a_{1,k}$, $R_{d,k}$ is the summation of redundant bits in the initial transmission and the retransmission, R^* is the maximum information bits that the system can support, used as a normalized factor, δ is a bias constant for punishment if the packet is dropped, and τ is the weight constant for $BLER(n)$. From (10), we can see that if the selected action can deliver more information bits with fewer redundant bits (more efficient transmission), a larger reward feedback will be generated. Retransmission will decrease the reward since the value of $R_{d,k}$ increases. However, if a packet is dropped after three times of transmissions, a severe punishment is given and a negative reinforcement signal is designed as in (10). Here, the δ will be set to let the punishment be greater than the reward, for example, $\delta = 5$. Also, the smaller $BLER(n)$ should get the less punishment for packet dropped since it can tolerate more unsuccessful transmission. The τ is set to let the effect of $BLER(n)$ be observable and comparable with respect to the reward, for example, $\tau = 40$.

If $BLER(n)$ is in the attention region, an action should be determined to keep the BLER performance unchanged. Therefore, the reward in this region should be smaller than that in the satisfaction region. The reinforcement signal is then designed as

$$r(n) = \begin{cases} \dfrac{1}{(1+N_{r,k})} \dfrac{R_k}{R_{d,k}+R^*}, & \text{if packet is successfully received,} \\ -(\delta + \tau \times BLER(n)), & \text{if packet is dropped,} \end{cases} \tag{11}$$

where $N_{r,k}$ is the number of the retransmission for the transmitted packet with the selected action $a_{1,k}$, and $0 \leq N_{r,k} \leq 2$. If the packet can be received without retransmission, the selected action can get the same reward as that in the satisfaction region. If the retransmission occurs, the reward is divided by $(1+N_{r,k})$ to decrease the reward of the selected action since the BLER performance is still within the $BLER^*$. However, if the packet is dropped, a similar severe punishment is given. Since the $BLER(n)$ in this attention region is greater than that in the satisfaction region, the punishment in the attention region is usually larger than that in the satisfaction region.

If $BLER(n)$ is in the violation region, a conservative action should be determined to recover the BLER requirement. Therefore, a reward reinforcement signal is given only if the packet is successfully received without retransmission. The reinforcement signal is given by

$$r(n) = \begin{cases} \dfrac{R_k}{R_{d,k}+R^*}, & \text{if packet is successfully received without retransmission,} \\ -\dfrac{(1+N_{r,k})}{3}\dfrac{R_k}{R_{d,k}+R^*}, & \text{if packet is successfully received with retransmission,} \\ -(\delta+\tau\times BLER(n)), & \text{if packet is dropped.} \end{cases} \tag{12}$$

If the packet is successfully received without retransmission, the reward for the selected action is the same form as that in the satisfaction region since the $BLER(n)$ can be reduced. However, if the packet is successfully received but with retransmission, the selected action will still attain a slight punishment since the BLER performance has been violated and is becoming deteriorate. This punishment increases as the number of retransmission increases. If the packet is dropped, a severe punishment is put on the selected action. Notice that the punishment in this region is the severest since the $BLER(n)$ in this region is the largest.

3.4.2 Reinforcement Signal for SM Fuzzy Rule Base

Similarly, if $BLER(n)$ is in the satisfaction region, the reinforcement signal is designed to increase the system throughput by

$$r(n) = \begin{cases} \dfrac{R_k}{R^*+R_{d,k}}, & \text{if both packets are successfully received,} \\ -\dfrac{N_{d,k}}{2}\times(\delta+\tau\times BLER(n)), & \text{if one or two packets are dropped,} \end{cases} \tag{13}$$

where $N_{d,k}$ is the number of the dropped transmitted packets with action $a_{2,k}$, $N_{d,k}=0,1,$ or 2. Notice that, in the SM transmission mode, R_k is the summation of information bits in the two transmitted packets with action $a_{2,k}$, and the value of R_k would be greater than $(R^*+R_{d,k})$ if the action $a_{2,k}$ adopts higher MCS levels for the two data streams. The selected action $a_{2,k}$ would get a higher reward than that in fuzzy rule base SD. This is a necessity since the SM transmission mode can enhance the system throughput more than the SD transmission mode. If both packets are dropped, a severe punishment is set to $-(\delta+\tau\times BLER(n))$. However, if one of two packets is dropped, the punishment is set to be the half of that for two packets dropped since there is still one packet can be successfully received.

If $BLER(n)$ is in the attention region, the reinforcement signal is designed to keep the BLER performance unchanged by

$$r(n) = \begin{cases} \dfrac{1}{(1+N_{r,k})}\dfrac{R_k}{R_{d,k}+R^*}, & \text{if both packets are successfully received,} \\ -\dfrac{N_{d,k}}{2}\times(\delta+\tau\times BLER(n)), & \text{if one or two packets are dropped.} \end{cases} \tag{14}$$

Notice that in the SM transmission mode, $0 \le N_{r,k} \le 4$. As $N_{r,k}$ increases, the reward for the selected action $a_{2,k}$ is decreased. The reason of this is the same as that given in the fuzzy rule base SD.

If $BLER(n)$ is in the violation region, the reinforcement signal is designed to recover the BLER requirement by

$$r(n) = \begin{cases} \dfrac{R_k}{R_{d,k}+R^*}, & \text{if both packets are successfully received without retransmission,} \\ -\dfrac{(1+N_{r,k})}{5}\dfrac{R_k}{R_{d,k}+R^*}, & \text{if both packets are successfully received but with retransmission,} \\ -(\delta+\tau\times BLER(n)), & \text{if one or two packets are dropped.} \end{cases} \tag{15}$$

Similar to that designed for fuzzy rule base SD, only when two transmitted packets are successfully decoded without retransmission, the selected action $a_{2,k}$ can get a reward since the BLER performance is improved. If there is any retransmission in the two packets transmission, a slight punishment is put on the selected action $a_{2,k}$ since the BLER performance has been out of BLER requirement $BLER^*$. This slight punishment is divided by 5 since the maximum value of $N_{r,k}$ is 4. If any packet dropped occurs, the selected action obtains a severe punishment since the BLER performance has been violated the requirement and cannot be allowed to deteriorate.

3.5 Q-Function Update

Applying the fuzzy Q-learning algorithm in [20], [22], the Q-values in the fuzzy rule base for transmission mode $\bar{z}$ in the next episode is updated by the Bellman equation [22], [28] given below

$$\begin{aligned} q_{n+1}(S_j, a_{\bar{z}}^*(n)) &= q_n(S_j, a_{\bar{z}}^*(n)) + \epsilon \times \frac{\mu_{j,n}}{\Sigma_{i=1}^{J}\mu_{i,n}} \\ &\times \Delta q_n(S_j, a_{\bar{z}}^*(n)), \text{ for } 1 \le j \le J, \end{aligned} \tag{16}$$

where ϵ is a learning rate, $0 \le \epsilon < 1$, and the $\Delta q_n(S_j, a_{\bar{z}}^*(n))$ is the temporal difference of the current episode and the next episode Q-values. By the ordinary gradient descent method [21], the $\Delta q_n(S_j, a_{\bar{z}}^*(n))$ can be obtained by

$$\begin{aligned} \Delta q_n(S_j, a_{\bar{z}}^*(n)) &= r(n) + \gamma \times Q_n^*(X(n+1), a_{\bar{z}}(n+1)) \\ &- Q_n(X(n), a_{\bar{z}}^*(n)), \end{aligned} \tag{17}$$

where γ is a discount factor, $0 < \gamma < 1$, and $Q_n(X(n), a_{\bar{z}}^*(n))$ is the Q-value for the state-action pair $(X(n), a_{\bar{z}}^*(n))$. By using the rule intensity $\mu_{j,n}$ of $X(n)$, we can get the $Q_n(X(n), a^*_{\bar{z}}(n))$ by

$$Q_n(X(n), a^*_{\bar{z}}(n)) = \frac{\Sigma_{j=1}^{J}[\mu_{j,n} \times q_n(S_j, a^*_{\bar{z},j}(n))]}{\Sigma_{j=1}^{J}\mu_{j,n}}. \quad (18)$$

Also, the $Q^*_n(X(n+1), a_{\bar{z}}(n+1))$ in (17) is the optimal Q-value at the next episode. Since the next episode Q-value for each state-action pair is unavailable, we use $q_n(S_j, a_{\bar{z},k})$ instead of $q_{n+1}(S_j, a_{\bar{z},k})$ to get $Q^*_n(X(n+1), a_{\bar{z}}(n+1))$ by the *center of area* (COA) method [17]

$$Q^*_n(X(n+1), a_{\bar{z}}(n+1)) = \frac{\Sigma_{j=1}^{J}[\mu_{j,n+1} \times q_n(S_j, a^*_{\bar{z},j}(n))]}{\Sigma_{j=1}^{J}\mu_{j,n+1}}. \quad (19)$$

Based on the principle of Bellman's optimality [28], as n is sufficiently large, $a^*_{\bar{z}}(n)$ will be an optimal action for the transmission mode $\bar{z}$ at which the maximum of Q-value is attained. Therefore, the fuzzy Q-learning algorithm can make an optimal decision for this MDP problem.

Note that, the learning rate ϵ controls the convergence of Q-value. As it is larger, the system learns faster but tends to more oscillation. The discount factor γ makes the Q-value at the next episode less valuable. As it is larger, the system becomes more foresight but requires a longer training time.

4 Simulation Results

4.1 Simulation Environment

In the simulations, we consider the HSPA^+ system supported in a hexagonal grid multicell CDMA system, where each cell suffers interference from two-tier neighbor cells. The spread factor (SF) of the CDMA system is fixed at 16, where one spreading code is reserved to common channels and 15 spreading codes are assigned to HS-DSCH. The channel model consists of the long-term fading and the short-term fading as given in (1) and (2), respectively. The channel condition is fixed within a TTI. The delay of either ACK/NACK or CQI report is set to be 10 ms, $5 \times$ TTI. Also, assume that the maximum power allocated to HS-DSCH is up to 80 percent of the total transmission power of the Node B, and the residual power is allocated to other control and service channels. Users always have packets waiting for transmission in the buffer. Since the variation of channel condition is very fast, the FQL-MOMS scheme would become myopic and we set $\gamma = 0.1$. Also, the FQL-MOMS selection scheme has to make decision for the HSPA^+ system per TTI (2 ms) interval. Thus, it needs fast learning speed and we set $\epsilon = 0.9$. Other parameters of the HSPA^+ in CDMA mobile system and the FQL-MOMS selection scheme are given in Table 3.

TABLE 3
Parameters for the HSPA^+ System and the FQL-MOMS Selection Scheme

Parameters	Assumption
Cellular layout	Hexagonal grid, 19 sites, 1000 m cell radius
Path loss model ($\xi(d)$)	$128.1 + 37.6\log_{10}(d)$, where d is in kilometers
Decorrelation length (d_{cor})	30 m
σ_L	8.0
Carrier frequency	2.0 GHz
Channel bandwidth	5.0 MHz
Chip-rate	3.84 Mcps
Spreading factor	16
Thermal noise density	-174 dBm/Hz
BS total Tx power	Up to 44 dBm
Signal path (L)	18
HARQ scheme	Full IR
BLER requirement	0.1

We compare the FQL-MOMS selection scheme with two conventional MCS selection schemes. One is the ATS scheme proposed in [12]. The ATS scheme adaptively sets the SINR threshold for each MCS level, where the SINR thresholds which are close to the SINR of the last transmission will be adjusted after each transmission. The increasing step of the threshold adjustment for the failed (NAK) transmission is set to be $\frac{BLER^*}{1-BLER^*}$, while the decreasing step for the successful (ACK) transmission is set to be $0.1 \times \frac{BLER^*}{1-BLER^*}$. The other is the Q-learning-based HARQ (Q-HARQ) scheme studied in [14]. The Q-HARQ scheme uses only the Q-learning algorithm to learn an optimal policy to choose the MCS level. The reinforcement signal of the Q-HARQ scheme is defined as

$$\left(\frac{SINR(x_k, a_k) - \widetilde{SINR}(a_k)}{\widetilde{SINR}(a_k)}\right)^2,$$

where $SINR(x_k, a_k)$ is the received SINR at UE for the state x_k with action a_k and $\widetilde{SINR}(a_k)$ is the required SINR with action a_k. It is a transmission cost function. The action of which SINR is closest to the required SINR has the smallest transmission cost value. After learning, the Q-HARQ scheme selects the MCS with minimum Q-value. In the meantime, the selection of the MIMO configuration mode is assumed for both the ATS and the Q-HARQ schemes. It is according to CQI_L. If the CQI_L is larger (less) than CQI_{th}, the SM (SD) transmission mode is selected since the channel condition is good (bad). Here, we set $CQI_{th} = 7$, the middle value of CQI and also the threshold for the switching between QPSK and 16-QAM, which is a suitable threshold to judge the quality of the channel.

In the following figures, we define an HSDPA service power ratio (HSPR) as the transmission power allocated to HS-DSCH over the total transmission power of the Node B. If HSPR is too small, there is no enough power to transmit the data. On the contrary, if HSPR is too large, the interference from other cells will seriously affect the data transmission.

4.2 Performance Evaluation

Fig. 5 shows the system throughput versus the user mobility in the HSPA^+ system for the FQL-MOMS selection scheme, the ATS scheme [12], and the Q-HARQ scheme [14], with HSPR equal to 80 percent. It can be found that the FQL-MOMS selection scheme enhances the system throughput of the HSPA^+ system by an amount of 49.3 percent more than the ATS scheme and by 35.9 percent more than the Q-HARQ scheme. Reasons for this are as follows: The decision by the FQL-MOMS selection scheme is intelligently inferred from fuzzy rules, the rules are adaptively refined

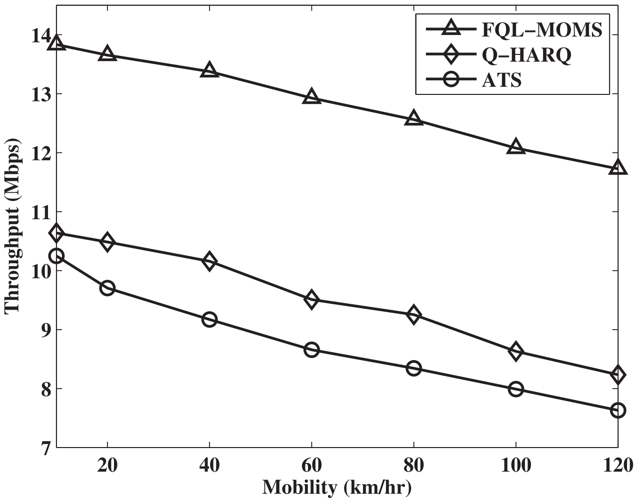

Fig. 5. The system throughput versus the user mobility.

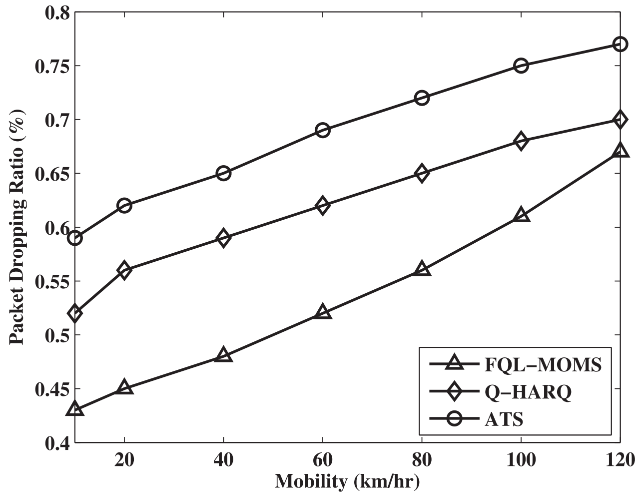

Fig. 6. The packet dropping ratio versus the user mobility.

by reinforcement signals in the Q-learning algorithm, and the reinforcement signals are effectively designed. The FQL-MOMS selection scheme considers not only the CQI and the last transmission results but also the BLER performance metric and the transmission efficiency. Also, based on the BLER performance metric, the FQL-MOMS selection scheme makes an aggressive (conservative) decision if the BLER performance falls in the satisfaction (violation) region. The granularity and the dynamic range of the adjustment for the decision of the MIMO transmission mode and MCS level are more flexible and sophisticated to cope with the varying channel environment. Consequently, the excellent balance between the system throughput enhancement and BLER requirement guarantee can be intelligently achieved. On the other hand, the ATS scheme determines the MIMO configuration mode and MCS level by an adjustable SINR. However, the adjustment of SINR threshold is less flexible, meaning that either the granularity of the adjustment is fixed or the dynamic range of the adjustment is small. The Q-HARQ scheme uses only the Q-learning algorithm to learn the optimal policy of MCS level selection but the designed reinforcement signal is a little coarse since it only considers the received SINR and the required SINR. Therefore, the Q-HARQ scheme can attain the system throughput higher than the ATS scheme but lower than the FQL-MOMS scheme.

Fig. 6 shows the packet dropping ratio versus the user mobility at HSPR equal to 80 percent. In the HSPA^{+} system, if a packet decoding is failed after three times of transmissions, this packet will be dropped. It can be observed that the FQL-MOMS (ATS) scheme has the smallest (largest) packet dropping ratio and the Q-HARQ scheme is in the between. Reasons for this are similar to as those given in Fig. 5, denoting that a larger system throughput would usually result in a smaller packet dropping ratio. Since the FQL-MOMS scheme infers a suitable MIMO transmission mode and MCS level by fuzzy rules, it achieves a lowest packet dropping ratio.

Fig. 7 shows the BLER performance versus the user mobility at HSPR equal to 80 percent. The Q-HARQ scheme has the best BLER performance, while the FQL-MOMS selection scheme attains the BLER performance larger than the Q-HARQ scheme by an amount of 0.01, and all the three schemes can guarantee the BLER requirement whatever the user mobility is. The phenomena are explained below. The FQL-MOMS selection scheme intends to enhance the system throughput while guarantee the BLER performance via the learning process for fuzzy rule bases. As the $BLER(n)$ is over the BLER requirement, the FQL-MOMS scheme adopts a conservative action to recover the BLER performance. As the $BLER(n)$ is far lower than the BLER requirement, the FQL-MOMS scheme adopts an aggressive action to enhance the system throughput and may increase the $BLER(n)$. Therefore, the FQL-MOMS selection scheme can achieve an excellent balance between the system throughput enhancement and BLER requirement guarantee. As it can be found, the FQL-MOMS attains the highest system throughput in Fig. 5, and in the meantime it maintains the BLER performance at *just below* the BLER requirement in this figure. On the other hand, the Q-HARQ scheme is too sensitive to the fast varying channel condition. It selects the action with minimum Q-value, where the Q-value is adjusted only based on the received SINR and the required SINR. The selected action which SINR is closer to the received SINR

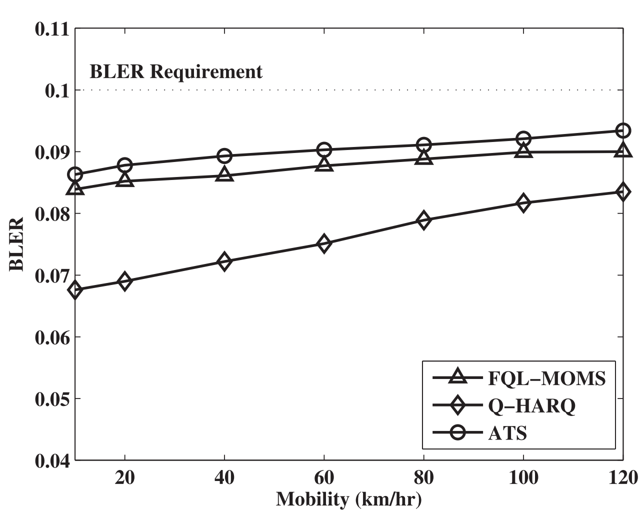

Fig. 7. The BLER performance versus the user mobility.

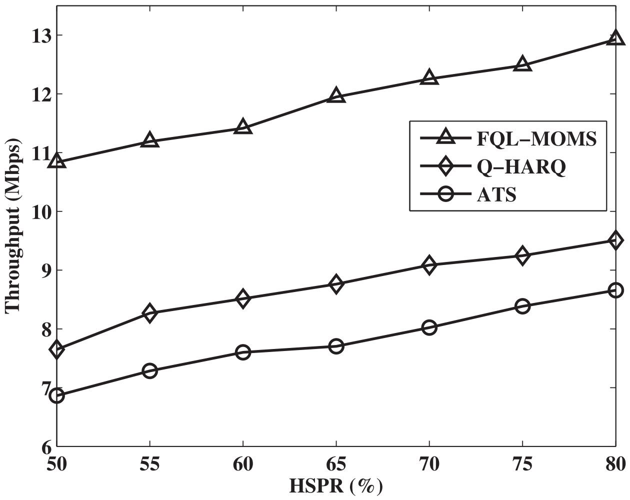

Fig. 8. The system throughput versus HSPR.

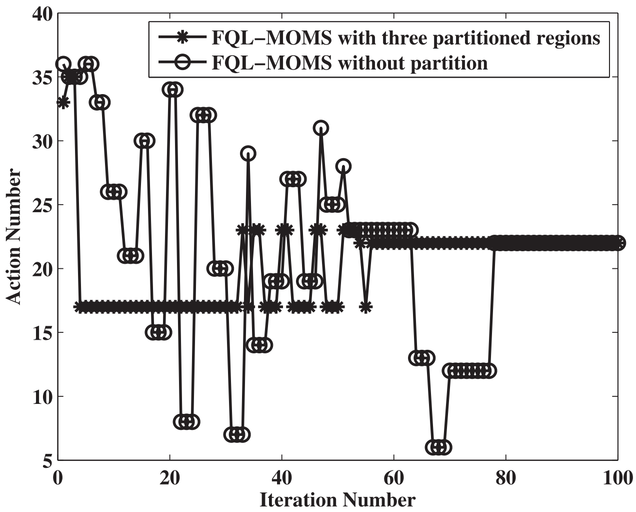

Fig. 9. The learning process of the FQL-MOMS selection scheme.

has a smaller reinforcement signal. This makes that the Q-HARQ scheme selects a conservative action and has the best BLER performance. The ATS scheme can rapidly adjust SINR thresholds based on the current ACK/NACK information, where the ratio of the decreasing step for successful transmission over the increasing step for failed transmission is set to 0.1. The decision of selection is too rough and thus makes that the ATS scheme has the largest BLER performance.

Fig. 8 presents the system throughput versus HSPR at the user mobility equal to 60 km/hr. It can be found that the FQL-MOMS (ATS) scheme still has the largest (smallest) system throughput among the three schemes no matter what the HSPR is. Reasons for this are the same as those given in Fig. 5. Also, as the HSPR increases, since the Node B has enough power to deliver the data and the UE has better received SINR, the system throughput increases. Performance metrics such as the packet dropping ratio and BLER versus HSPR are also measured in the simulations. The figure showing the packet dropping ratio (BLER) versus HSPR in contrast to Fig. 6 (Fig. 7) also has the same phenomena as Fig. 8 with respect to Fig. 5, which is not shown here.

Fig. 9 shows a worst case learning process of the FQL-MOMS selection scheme in the 1,000 simulations, where SINRs related with a desired action $a_{2,22}$ are randomly generated and input to the FQL-MOMS scheme with three partitioned regions and the FQL-MOMS scheme without partition. It is observed that the FQL-MOMS scheme with three partitioned regions first stays at a temporary action $a_{2,17}$ after four iterations, then it oscillates at either $a_{2,17}$ or $a_{2,22}$ and finally achieves the desired action after 56 iterations, while the FQL-MOMS scheme without partition oscillates quite a lot and finally reaches the desired action after 78 iterations. The FQL-MOMS scheme with three partitioned regions accelerates the learning process by up to 28 percent compared to the FQL-MOMS scheme without partition. Also, averagely speaking in the 1,000 simulations, the former can increase the learning speed by up to 61 percent than the latter. The phenomenon is explained below. In the FQL-MOMS scheme with three partitioned regions, only actions in a region are enabled for each update learning. On the other hand, in the FQL-MOMS scheme without partition, all actions are enabled. Although the former might attain a local optimal action for the time being, if the initial BLER performance metric results in an improper region of actions in the fuzzy rule base, it will fastly go to the desired action. It is because the purpose of fuzzy rule bases is to keep the BLER performance in the attention region where the desired action is enabled in each fuzzy rule.

5 Conclusions

In this paper, we propose a fuzzy Q-learning-based MIMO configuration mode and MCS level selection scheme for HSPA^+ systems, intending to efficiently utilize the system radio resource. The FQL-MOMS selection scheme combines the fuzzy logics and Q-learning algorithm to intelligently choose an MIMO configuration mode and MCS level for packet data transmission of HSPA^+ systems to enhance the system throughput while guarantee the BLER requirement. It is designed to have a clear and effective configuration which contains the SD fuzzy rule base and the SM fuzzy rule base. Also, in order to accelerate the learning process, each rule base is partitioned into satisfaction, attention, and violation regions according to the BLER performance. The fuzzy rules with Q-values in the three regions of the SD and SM fuzzy rule bases are carefully designed by the domain knowledge to maximize the system throughput and fulfill the BLER requirement. Moreover, the Q-learning algorithm is used to update the Q-values of fuzzy rules in both rule bases to adapt to the variation of channel condition, where the reinforcement signals consider BLER performance metric, transmission efficiency, and transmission results of both transmission and retransmission to reward or punish the selected action. By using the closed-loop iteration manner, the FQL-MOMS selection scheme can determine an appropriate MIMO configuration mode and MCS level for packet data transmission in HSPA^+ systems. It can indeed achieve an excellent balance between the system throughput maximization and BLER requirement fulfillment. Simulation results show that the FQL-MOMS selection scheme can achieve higher system throughput than conventional ATS and Q-HARQ schemes, under the BLER requirement guarantee.

Acknowledgments

The authors would like to give thanks to the editor and the four anonymous reviewers for their kind comments to improve the presentation of the paper. This work was supported by the National Science Council (NSC), Taiwan, under contract number NSC 97-2221-E-009-098-MY3, and the Ministry of Education, Taiwan, under the Aiming for Top University plan.

References

[1] "UMTS Physical Layer Procedures (FDD)," technical report 3GPP TR 25.214, Third Generation Partnership Project, May 2008.
[2] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband.* Elsevier, 2007.
[3] G.J. Foschini, "Layered Space-Time Architecture for Wireless Communication in Fading Environment when Using Multiple Antennas," *Bell Labs Technical J.,* vol. 1, no. 2, pp. 41-59, 1996.
[4] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. Information Theory,* vol. 45, no. 5, pp. 1456-1467, July 1999.
[5] L. Zheng and D.N.C Tse, "Diversity and Multiplexing: A Fundamental Trade Off in Multiple-Antenna Channels," *IEEE Trans. Information Theory,* vol. 49, no. 5, pp. 1073-1096, May 2003.
[6] M. Wrulich, S. Eder, I. Viering, and M. Rupp, "Efficient Link-to-System Level Model for MIMO HSDPA," *Proc. IEEE GlobeCom,* pp. 1-6, Dec. 2008.
[7] C. Mehlführer, S. Caban, M. Wrulich, and M. Rupp, "Joint Throughput Optimized CQI and Precoding Weight Calculation for MIMO HSDPA," *Proc. Signals, Systems and Computers Conf.,* pp. 1320-1325, Oct. 2008.
[8] J. Li and Y.Q. Zhao, "Resequencing Analysis of Stop-and-Wait ARQ for Parallel Multichannel Communications," *IEEE/ACM Trans. Networking,* vol. 17, no. 3, pp. 817-830, June 2009.
[9] L.C. Wang and C.W. Chang, "Gap Processing Time Analysis of Stall Avoidance Schemes for High-Speed Downlink Packet Access with Parallel HARQ Mechanism," *IEEE Trans. Mobile Computing,* vol. 5, no. 11, pp. 1591-1605, Nov. 2006.
[10] T. Cheng, "Coding Performance of Hybrid ARQ Schemes," *IEEE Trans. Comm.,* vol. 54, no. 6, pp. 1017-1029, June 2006.
[11] S. Chen, J. Du, M. Peng, and W. Wang, "Performance Analysis and Improvement of HARQ Techniques in TDD-HSDPA/SA System," *Proc. Sixth Int'l Conf. ITS-Telecomm.,* pp. 523-526, 2006.
[12] M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive Control of Link Adaptation for High Speed Downlink Packet Access in WCDMA," *Wireless Personal Multimedia Comm.,* vol. 2, pp. 382-386, 2002.
[13] A. Muller and T. Chen, "Improving HSDPA Link Adaptation by Considering the Age of Channel Quality Feedback Information," *Proc. IEEE Vehicular Technology Conf.,* pp. 1643-1647, Sept. 2005.
[14] C.J. Chang, C.Y. Chang, and F.C. Ren, "Q-Learning-Based Hybrid ARQ for High Speed Downlink Packet Access in UMTS," *Proc. IEEE Vehicular Technology Conf.,* pp. 2610-2615, Apr. 2007.
[15] C.J.C.H. Watkins and P. Dayan, "Technical Note: Q-Learning," *Machine Learning,* vol. 8, no. 3, pp. 279-292, 1992.
[16] L. Zhao, J.W. Mark, and T.C. Yoon, "A Combined Link Adaptation and Incremental Redundancy Protocol for Enhanced Data Transmission," *Proc. IEEE GlobeCom,* vol. 2, pp. 1277-1281, Nov. 2001.
[17] C.T. Lin and C.S.G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems.* Prentice Hall, 1996.
[18] J. Ye, X. Shen, and J.W. Mark, "Call Admission Control in Wideband CDMA Cellular Networks by Using Fuzzy Logic," *IEEE Trans. Mobile Computing,* vol. 4, no. 2, pp. 129-141, Mar./Apr. 2005.
[19] H.R. Berenji, "Fuzzy Q-Learning: A New Approach for Fuzzy Dynamic Programming," *Proc. IEEE Int'l Conf. Fuzzy Systems,* pp. 486-491, 1994.
[20] T. Horiuchi, A. Fujino, O. Katai, and T. Sawaragi, "Fuzzy Interpolation-Based Q-Learning with Profit Sharing Plan Scheme," *Proc. IEEE Int'l Conf. Fuzzy Systems,* pp. 1707-1712, 1997.
[21] P.Y. Glorennec and J. Jouffe, "Fuzzy Q-Learning," *Proc. IEEE Int'l Conf. Fuzzy Systems,* pp. 659-662, 1997.
[22] L. Jouffe, "Fuzzy Inference System Learning by Reinforcement Methods," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev.,* vol. 28, no. 3, pp. 338-355, Aug. 1998.
[23] C.F. Juang and C.M. Lu, "Ant Colony Optimization Incorporated with Fuzzy Q-Learning for Reinforcement Fuzzy Control," *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 39, no. 3, pp. 597-608, May 2009.
[24] C.F. Juang, "Combination of Online Clustering and Q-Value Based GA for Reinforcement Fuzzy System Design," *IEEE Trans. Fuzzy Systems,* vol. 13, no. 3, pp. 289-302, June 2005.
[25] Y.H. Chen, C.J. Chang, and C.Y. Huang, "Fuzzy Q-Learning Admission Control for WCDMA/WLAN Heterogeneous Networks with Multimedia Traffic," *IEEE Trans. Mobile Computing,* vol. 8, no. 11, pp. 1469-1479, Nov. 2009.
[26] C.Y. Huang, W.C. Chung, C.J. Chang, and F.C. Ren, "Fuzzy Q-Learning-Based Hybrid ARQ for High Speed Downlink Packet Access," *Proc. IEEE Vehicular Technology Conf.,* pp. 1-4, Sept. 2009.
[27] "Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4)," technical report 3GPP TR 25.848, Third Generation Partnership Project, Mar. 2001.
[28] R. Bellman, *Dynamic Programming.* Princeton Univ., 1957.
[29] G.L. Stuber, *Principle of Mobile Communication.* Kluwer Academic Publisher, 2001.
[30] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," *IEE Electronics Letters,* vol. 27, no. 23, pp. 2145-2146, Nov. 1991.

Wen-Ching Chung received the BE and PhD degrees in electrical control engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1999 and 2006, respectively. In January 2008, he joined the Department of Electrical Engineering of National Chiao Tung University in Taiwan as a postdoctor. His research interests are in the areas of radio resources management for wireless communication networks, link adaptation for broadband networks, and intelligent control systems. He is a member of the IEEE.

Chung-Ju Chang received the BE and ME degrees in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1972 and 1976, respectively, and the PhD degree in electrical engineering from National Taiwan University in 1985. From 1976 to 1988, he was with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Communications, Taiwan, as a design engineer, supervisor, project manager, and then division director. He also acted as a science and technical advisor for the Minister of the Ministry of Communications from 1987 to 1989. In 1988, he joined the Faculty of the Department of Communication Engineering, College of Electrical Engineering and Computer Science, National Chiao Tung University, as an associate professor. He has been a professor since 1993 and a chair professor since 2009. He was the director of the Institute of Communication Engineering from August 1993 to July 1995, chairman of the Department of Communication Engineering from August 1999 to July 2001, and dean of the Research and Development Office from August 2002 to July 2004. Also, he was an advisor for the Ministry of Education to promote the education of communication science and technologies for colleges and universities in Taiwan during 1995-1999. He is acting as a committee member of the Telecommunication Deliberate Body, Taiwan. Moreover, he once served as editor for *IEEE Communications Magazine* and associate editor for the *IEEE Transactions on Vehicular Technology*. His research interests include performance evaluation, radio resources management for wireless communication networks, and traffic control for broadband networks. He is a member of the Chinese Institute of Engineers (CIE) and the Chinese Institute of Electrical Engineers (CIEE). He is a fellow of the IEEE.

Kai-Ten Feng received the BS degree from National Taiwan University, Taipei, in 1992, the MS degree from the University of Michigan, Ann Arbor, in 1996, and the PhD degree from the University of California, Berkeley, in 2000. Since August 2007, he has been with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan, as an associate professor. He was an assistant professor with the same department between February 2003 and July 2007. He was with OnStar Corp., a subsidiary of General Motors Corporation, as an in-vehicle development manager/senior technologist between 2000 and 2003, working on the design of future Telematics platforms and in-vehicle networks. His current research interests include cooperative and cognitive networks, mobile ad hoc and sensor networks, embedded system design, wireless location technologies, and Intelligent Transportation Systems (ITSs). He received the Best Paper Award from the IEEE Vehicular Technology Conference in Spring 2006, which ranked his paper first among the 615 accepted papers. He was also the recipient of the Outstanding Young Electrical Engineer Award in 2007 from the Chinese Institute of Electrical Engineering (CIEE). He has served on the technical program committees of VTC, ICC, and APWCS. He is a member of the IEEE.

Ying-Yu Chen received the BE degree in electrical engineering from National Chung Hsing University in 2007 and the ME degree in electrical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2009. Her research interests include radio resource management and link adaptation for wireless communication networks.