國 立 交 通 大 學

電信工程學系

碩 士 論 文

在多輸入多輸出高速下行封包擷取系統中
採用乏晰 Q-Learning 技術之混合自動重傳機制

HARQ by Fuzzy Q-learning for MIMO HSDPA System

研 究 生：陳盈仔

指導教授：張仲儒 博士

中 華 民 國 九 十 八 年 七 月

在多輸入多輸出高速下行封包擷取系統中
採用乏晰Q-Learning技術之混合自動重傳機制

HARQ by Fuzzy Q-learning for MIMO HSDPA System

研 究 生：陳盈伃　　　　Student：Ying-Yu Chen

指導教授：張仲儒 博士　　Advisor：Dr. Chung-Ju Chang

國 立 交 通 大 學

電 信 工 程 學 系

碩 士 論 文

A Thesis

Submitted to Department of Communication Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master of Science

in

Communication Engineering

July 2009

Hsinchu, Taiwan

中華民國九十八年七月

在多輸入多輸出高速下行封包擷取系統中
採用乏晰 Q-Learning 技術之混合自動重傳機制

研究生：陳盈伃　　　　　　　　　　　　指導教授：張仲儒 博士

國立交通大學電信工程學系碩士班

摘　　要

為了提供更高速的資料傳輸與更有效的資源利用, 第三代合作夥伴計畫（3rd generation partnership project, 3GPP）提出了多輸出多輸入高速下行封包擷取技術（Multiple input multiple output high speed downlink packet access, MIMO HSDPA）來提供更高速且安全的下鏈路資料封包傳送。在 3GPP 的規格裡面,有一個重要的服務品質要求：根據通道狀況決定原始傳送的 MCS 時，必須使封包的錯誤率小於 0.1。因此，我們提出在多輸出多輸入高速下行封包擷取技術之混和自動重傳機制下,採用乏晰 Q 學習演算法來解決這個問題（FQLM-HARQ）。乏晰 Q 學習法同時結合了乏晰邏輯運算與 Q 學習法的優點。在此,我們將混合自動重傳機制（hybrid automatic retransmission request, HARQ）程序模擬為離散時間馬可夫決策過程（Markov decision process, MDP）。根據 BLER 的表現,乏晰系統規則會設計成不同的部分，來實現 BLER 的服務品質需求。Q 學習演算法可以在不同的環境下,經由不斷的學習,選出最適當的 MCS 並且修正乏晰系統規則。在學習之後，我們期望多輸出多輸入高速下行封包擷取系統可以達到最高的資料輸出而又不違反 QoS 需求。

從模擬結果可知，所提出的 FQLM-HARQ 機制在通道訊息延遲的情況下，可以達到最大的輸出並且盡力維持 BLER 的要求。相較於其他系統，FQLM-HARQ 可以更適應通道的變化。

HARQ by Fuzzy Q-learning for MIMO HSDPA System

Student：Ying-Yu Chen　　　　　　　　Advisor：Dr. Chung-Ju Chang

Department of Communication Engineering
National Chiao Tung University

ABSTRACT

Multiple input multiple output high speed downlink packet access (MIMO HSDPA) system is proposed by 3rd generation partnership project (3GPP) to provide higher transmission data rate and more resource utilization. An important QoS requirement defined in spec is to choose a suitable MCS based on the channel quality indicator while maintaining the initial block error rate (BLER) smaller than 0.1. Therefore, we proposed a fuzzy Q-learning based MIMO HARQ (FQLM-HARQ) scheme for MIMO HSDPA system to solve this problem. The FQLM-HARQ scheme can take the advantage from both fuzzy logic and Q-learning. Here, the HARQ scheme is modeled as a Markov decision process (MDP). The fuzzy rule is designed to separate different parts according to the BLER performance and the Q-learning algorithm can learn the optimal MCS under different environment. After learning, we can expect the MIMO HSDPA system with higher throughput while not violating the BLER requirement.

From simulation results, the proposed FQLM-HARQ scheme can achieve higher system throughput and endeavor to maintain the BLER requirement with channel quality indicator delay consideration. Comparing to other traditional schemes, the FQLM-HARQ scheme can accommodate well in channel variation.

誌 謝

兩年的碩士生活與這篇論文的產生，我要感謝許多人。

謝謝指導教授張仲儒博士，帶領我走入研究的領域，傳授很多寶貴經驗和處事道理。還有感謝口試委員：鄭獻勳博士，陳震宇博士和召集人陳光禎博士，對於論文給予的意見。另外還要感謝文敬學長在論文上諸多的指教和建議，讓我的論文得以順利完成。感謝芳慶、志明、文祥、耀興、振宇學長在研究上的協助，還有謝謝尚樺、宗利、維謙、英奇、浩翔、邱胤、巧瑩學長姐們，讓我很快熟悉實驗室環境，渡過精彩的碩一生活。更要感謝一起做研究的志遠、和儁、欣毅，很開心這兩年有你們的陪伴，讓我學習了很多事情。謝謝耀庭、正忠、心瀅、蕊綺、苡仲學弟妹們，為實驗室增添歡樂氣息，以及即將進入實驗室的竣威學弟，祝你們未來研究順利。謝謝 sandy，給予很多的幫忙和關心。還有謝謝惟媜以及多位實驗室的學長姐們，豐富了我的研究生活，謝謝大家。

最後，感謝我的家人和朋友們，讓我更樂觀積極面對未來，謝謝你們。

盈伃 謹誌

民國 九十八年七月

Contents

List of Tables

List of Figures

Chapter 1
Introduction

High speed packet access evolution (HSPA+) has been proposed in 3rd generation partnership project (3GPP) release 7 of universal mobile telecommunications system (UMTS) [1]. In release 7, the downlink transmission is improved to supports the peak data rate up to 28.8 Mbps. In order to enhance the performance of high speed downlink packet access (HSDPA) proposed in release 5, release 7 adopted many useful mechanisms, such as multiple input multiple output (MIMO), higher order modulation (HOM), improved layer 2 support for high data rates, and continuous packet connectivity (CPC). Note that in release 7 MIMO and HOM can not be enabled simultaneously. In this thesis, we consider the HSDPA with MIMO enabled, since MIMO can not only achieve the highest peak data rate, but also increase link reliability. However, due to MIMO enabled in HSDPA system, it becomes more complexity to decide a suitable modulation and coding scheme for initial transmission in HARQ mechanism.

Release 7 is the evolution of release 5. Hence, release 7 inherits many favorable properties from release 5. A key characteristic of HSDPA is the use of shared-channel transmission, which can rapidly allocate a large fraction of downlink resources to a specific user through high-speed downlink shared channel (HS-DSCH). HS-DSCH inherits the capability of the common channel which transports data abruptly increased. In order to allow wideband code division multiple access (WCDMA) networks to achieve higher data rate and lower latency, there are several mechanisms used in HSDPA, such as fast scheduling, adaptive modulation and coding (AMC), and hybrid automatic repeat request (HARQ). The packet scheduler is shifted from radio network controller (RNC) to the Node B in HSDPA and the transmission time interval (TTI) is shortened from 10ms to 2ms. Hence, the packet scheduling decisions are faster than traditional WCDMA networks. Moreover, power control is carried out by AMC, which adjusts the transmission rate according to the channel quality. If the system can tolerate some jitter in the data rate, AMC can be more efficient.

Another important functionality of HSDPA is the HARQ mechanism which is used for error detections and for the retransmission of the error data. Comparing with the traditional automatic repeat request (ARQ), HARQ combines the forward error correction (FEC) bits with the existing error detection (ED) bits to increase the probability of the successful decoding. There are three different schemes to implement the HARQ technique. The first one is the chase combining (CC) scheme, that each retransmission data is identical to the original one. When the CC scheme is executed with maximal-ratio combining (MRC), the final received SINR is the total of each (re)transmission data [2]. The second one is the incremental redundancy (IR) scheme, where each retransmission data consists of new redundancy bits. The performance comparison of these two schemes was shown in [3] and [4]. The last one is the H-ARQ-type-Ⅲ scheme, which is a variety of the IR scheme. The difference between the H-ARQ-type-Ⅲ scheme and the IR scheme is that each retransmission data for the H-ARQ-type-Ⅲ scheme has a self-decodable ability. Other varieties of the HARQ mechanism are proposed to tolerate the fast-varying channel condition, such as asynchronous and adaptive hybrid ARQ (A^2IR HARQ) and enhancement hybrid ARQ (E-HARQ) [5]. In A^2IR HARQ, each (re)transmission rate can be changed according to the feedback information of the channel quality indicator (CQI) and the residual energy for the packet to have the more probability of the successful decoding. The (re)transmission rate is changed by the selection of different modulation order and coding rate schemes (MCSs).

The MCS selection for the initial transmission of each data stream will affect the retransmission times, block error rate and the system performance. Therefore, one important object of the HARQ mechanism is to find a suitable MCS level for first data transmission. In [1], the relationship between the CQI and MCS has decided in a predetermined table to achieve the block error rate (BLER) requirement. Nakamura et al. proposed an adaptive method to tune the SINR threshold for each MCS level based on the last transmission result [6]. In [7], Muller and Chen proposed a modified SINR threshold adaptation method for each MCS level not only based on the transmission result but also considered different CQI delay scenarios. Q-learning algorithm [8] is one kind of powerful reinforcement learning method. Since HARQ procedure can be modeled as a discrete-time Markov decision process (MDP), Chang et al. [9] adopted Q-learning algorithm on HARQ scheme, called Q-HARQ, to achieve more system efficiency and enhance the throughput. The Q-HARQ scheme uses the Q function and the reinforcement signal to learn the optimal solution for MCS selection. By assigning each MCS level a Q-value, the base station will choose the MCS level with minimum Q-value to reduce the cost for packet transmission.

In the recent years, MIMO have been widely used in wireless communication systems. It enhances the performance of the system, including the capacity and the coverage, as well as improved service provisioning. The MIMO scheme can increase the data rate through the spatial multiplexing (SM) [10] or provide the reliability of data transmissions through spatial diversity (SD) [11]. However, there is a trade-off relationship between these two through schemes [12]. At the base station (BS), the use of multiple transmit antennas is primarily of interest for the downlink since it provides an opportunity for diversity and beam-forming without the need of additional receiver chains at the mobile terminal. There are different approaches to realize the diversity by the multiple transmit antennas. For the open loop diversity, in general, there are the orthogonal transmit diversity and the transmit diversity via space-time coding (STC). Note that STC is a popular solution for diversity gain and coding gain, which can be easily combined with all kinds of multiple antenna systems. For the closed loop diversity, in general, there are the switched transmit diversity (STD) and the transmit adaptive array (TxAA) [13]. The closed-loop MIMO schemes have better performance than open-loop MIMO schemes in theory since the transmitter for the closed-loop MIMO schemes have the knowledge of the channel condition. The 3GPP extended the closed loop transmit diversity scheme, called double transmitter antenna array (D-TxAA) [14], [15], in release 99 to be the standard of release 7. With D-TxAA, up to two data streams (transport blocks) can be transmitted simultaneously over the radio channel to increase the throughput by using the same channelization codes. Here, each data stream is processed and coded separately. When the channel quality becomes worse, D-TxAA can transmit one data stream to increase the accuracy of transmitted data. Hence, to decide the number of transmitted data stream in the HSPA+ system becomes an important issue since it is highly related to the system performance. The detail of D-TxAA in release 7 can be found in [16] and [17].

The study of the MIMO technology with the adaptive HARQ mechanism has attracted considerable attention [18]-[19] since it can provide more throughput and reliability for packet data services. In [18], authors proposed the basis hopping method with the HARQ pre-combining scheme to provide dramatic gain for MIMO. In [19], authors proposed the TAS/STBC/HARQ scheme to enhance system performances, which adopted STC, transmit antenna selection (TAS), and receive diversity combining method.

Fuzzy Q-learning (FQL) method can be seen as an extension of Q-learning algorithm into fuzzy environments. Traditionally, the FQL method is used to model the motion and thinking way of human in the robot design [20], [21], [22]. Combining Q-learning scheme and fuzzy logic can help the system efficiently to adapt the environment and to choose the suitable actions.

In this thesis, we propose a fuzzy Q-learning based MIMO HARQ (FQLM-HARQ) scheme for HSDPA system in release 7 to increase the system throughput while guarantee the BLER requirement. It is very hard to use an explicit mathematics equation to represent the relationship between BLER requirement fulfillment and throughput maximization. Fuzzy logic is a very useful technique to solve the imprecise problem. Hence, the FQL scheme takes advantages from both fuzzy logic and Q-learning algorithm to get the best action-value function gradually for above imprecise problem. In our design, the FQLM-HARQ scheme can not only decide the number of data stream for transmission, but also choose the suitable MCS level for each transmission.

The remainder of this thesis is organized as follows. The system model for the release 7 is described in Chapter 2. After that, the fuzzy Q-learning design for MIMO HARQ scheme is proposed in Chapter 3. It is followed by the performance analysis of the FQLM-HARQ scheme. Finally, conclusions are given in Chapter 5.

Chapter 2
System Model

The MIMO HSDPA system, the HARQ entity, the spatial processing, the channel model and the fuzzy Q-learning algorithm are given in this chapter.

2.1 MIMO HSDPA System

The protocol architecture of MIMO HSDPA is shown in Fig 2.1. At the serving radio network controller (SRNC), radio link control (RLC) handles the data of MAC-d which will be transmitted on HS-DSCH. At the Node B (the base station), a new MAC entity, called MAC-ehs, is used instead of MAC-hs to support the functionality of MIMO in the HSDPA system. The functionalities of MAC-ehs are included flow control, scheduling, priority handling, HARQ entity. Moreover, MAC-ehs can determine the number of data stream for MIMO scheme to be transmitted at the Node B.

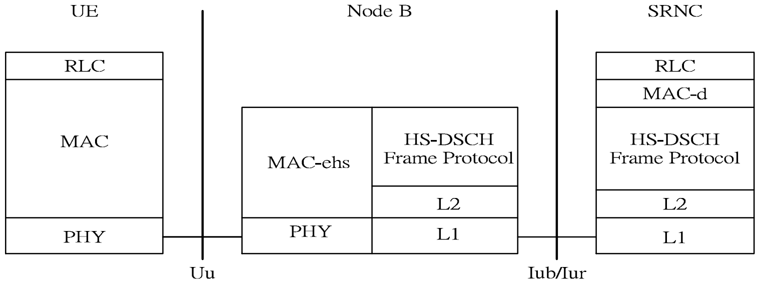

Figure 2.1: Protocol architecture of MIMO HSDPA

There are three special channels used in the physical layer specification of release 7 the same as that in release 5. They are HS-DSCH, high-speed shared control channel (HS-SCCH), and uplink high-speed dedicated physical control channel (HS-DPCCH). Details are given as follows:

(1) HS-DSCH: It carries the data to the user in the downlink direction. The code resource for HS-DSCH contains a set of channelization code, which spread factor (SF) is always fixed at 16. Each channelization code is known as for high-speed physical downlink shared channel (HS-PDSCH). Since common channels need a code space, the maximum number of codes which can be allocated to HS-DSCH is 15. The only coding scheme of HS-DSCH is turbo code with a minimum 1/3 coding rate.

(2) HS-SCCH: It carries the necessary physical layer control information to the user which includes transport format HARQ-related information and so on. This information is used to decode the data transmitted on HS-DSCH and to perform the soft combining when a retransmission is carried out. If there is no data transmitted on HS-DSCH, there is no need to send the data on HS-SCCH.

(3) HS-DPCCH: It carries the control information in the uplink direction, which consists of ACK/NACK messages, precoding control indication (PCI), and channel quality indication (CQI) value. The Node B needs the result of the HS-DSCH decoding and the instantaneous channel conditions for the purpose of scheduling and rate control. Note that, one HS-DPCCH is set up for each user with an HS-DSCH configured.

The MIMO HSDPA system with fuzzy Q-learning based MIMO HARQ (FQLM-HARQ) scheme is described in Fig 2.2. Details are given as follows:

(1) In this study, to maximize system throughput is our main concern. Hence, the scheduling in the MAC-ehs will choose the user with the best channel quality to be served. Note that, in our MIMO HSDPA system, only one user can be served per TTI.

(2) Once the served user has been selected in a particular TTI, the scheduler identifies the

necessary HS-DSCH parameters according to the output signal from our FQLM-HARQ scheme. These parameters includes the number of transmitted data stream, the kind of modulation, the size of transmitted block, and so on. The block of priority handling then segments the user data to a suitable size based on those HS-DSCH parameters.

(3) After the segmented packet(s) arrives the HARQ entity, the HARQ entity assigns each packet to a suitable HARQ process then generates new data stream for transmission. Note that, this data stream might included the retransmission data for previous packet.

(4) After the single or dual data stream is generated from the HARQ entity, it will pass through the spatial processing and then be transmitted through two physical antennas.

(5) After the terminal received the transmitted data, it checks which HARQ process the data belongs to and whether it needs to be combined with data received previously. The terminal then decodes the combined data. After that, the outcome of decoding and the channel information are sent on the HS-DPCCH to the Node B.

(6) After the Node B receives the feedback information from the terminal, the FQLM-HARQ scheme uses the channel information, block error rate, and the observed system information such as retransmission times, the packet information bits, and so on to adjust the fuzzy rule base and Q-learning.

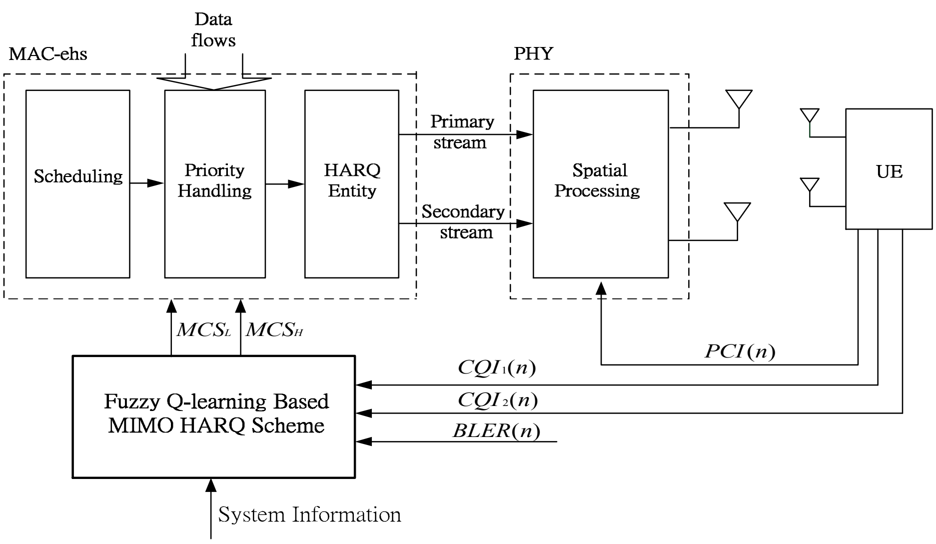

Figure 2.2: The MIMO HSDPA system

2.2 HARQ Entity

One HARQ entity handles the HARQ functionality for one user. HARQ entity is capable of supproting multiple HARQ process of HARQ protocols. HARQ entity determines a suitable HARQ process service the MAC-ehs protocol data unit (PDU). Each HARQ process has indepent acknowledgements and retransmission. There shall be one HARQ process for single stream transmission and two HARQ processes for dual streams transmission per TTI. In order to achieve continuous data transmission, a minimum of six HARQ processes needs to be configured in single stream transmission. Similarly, in dual stream transmission, a minimum of twelve HARQ processes needs to be configured to achieve continuous data transmission. However, HARQ processes on both streams run independently from one another will result in the signaling overhead, since each possible combination of HARQ processes should be addressed. To efficiently solve this problem, HARQ processes in release 7 are only signaled for the primary transport block with 4 bits. The HARQ process for the secondary transport block is directly derived from that for the primary transport block. Hence, there is a one-to-one mapping between the HARQ processes used for the primary transport block and the secondary transport block. In this study, the number of the HARQ process is set to 12 to achieve continuous data transmission. Fig 2.3 shows the mapping of the HARQ process between the primary stream and the second stream in our MIMO HSDPA system.

HARQ process identifier on primary	0	1	2	3	4	5
HARQ process identifier on secondary	6	7	8	9	10	11

Figure 2.3: The HARQ processes for primary and secondary stream

We adopt IR method to implement the HARQ scheme. In order to enhance the system throughput, we modify the traditional IR method to carry partially new data. That is, if the Node

B receives a NACK signal, the HARQ process combines the information and redundancy bits from new packet and the redundancy bits from the retransmission packet to transmit. Fig 2.4 shows an example of the bit allocation in the HARQ process. Let Z be the transport block size. There are Z bits for the new transmission at 1st TTI. After 5 TTIs, the Node B receives the ACK/NACK of this transmission. If a NACK indicator is received, the Node B sends A redundancy bits for the first packet and $Z-A$ bits as new data packet at 7th TTI. If the Node B receives a NACK indicator again after 5 TTIs, it sends A_1 redundancy bits for the first packet, B redundancy bits for the previous packets, and $Z-A_1-B$ bits for new data packet at 13th TTI. Note that, the transport block size can be changed per TTI depending on the channel condition. By increasing the redundancy bits for previous failure packet, it can improve the probability of successful decoding of previous failure packet. If the number of the retransmission for the same packet is more than two, this packet will be dropped.

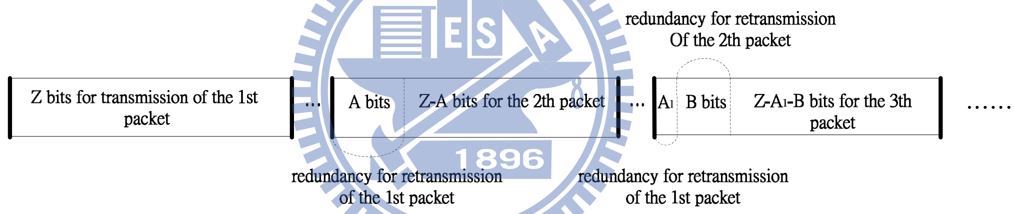

Figure 2.4: The bit allocation in the HARQ process

In release 7, the CQI table stores the information that which transport block size, modulation order, and coding rate are suitable for the given channel condition. The Node B scheduler determines exact (re)transmission format based on this CQI information for HARQ scheme. However, CQI has report delay or measurement inaccuracy. By changing the modulation order according to the transmission results, HARQ scheme can effectively reduce the number of retransmission and enhance the system throughput.

In our study, we call the combination of the transport block size, modulation order, and coding rate as the MCS level. There are 8 kinds of MCS level used for HARQ scheme and is given in Table 2.1. The lower MCS level stands for the more reliability of the transmission, but has the less throughput of the system. On the contrary, the higher MCS level stands for the less reliability of the transmission, but has higher throughput of the system.

Table 2.1: The 8 kinds of MCSs

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

2.3 Spatial Processing

The spatial processing for the MIMO HSDPA system as shown in Fig 2.5. The MIMO scheme used for MIMO HSDPA system is called double transmit adaptive arrays (D-TxAA), which is a multi-codeword scheme with rank adaptation and pre-coding. There are two transmit antennas at the Node B and two receive antennas at the UE. In release 7, the D-TxAA scheme is only applicable for the HS-DSCH. Each data stream is processed and coded separately per TTI. After spreading and scrambling, precoding weights w_1 and w_2 are applied to the primary data stream, and weights w_3 and w_4 are applied to the secondary data stream. Note that, if only single data stream is transmitted, it can get transmit diversity by exploiting both transmit antennas. For dual data streams transmission, pre-coding can aid the receiver to separate the two data streams.

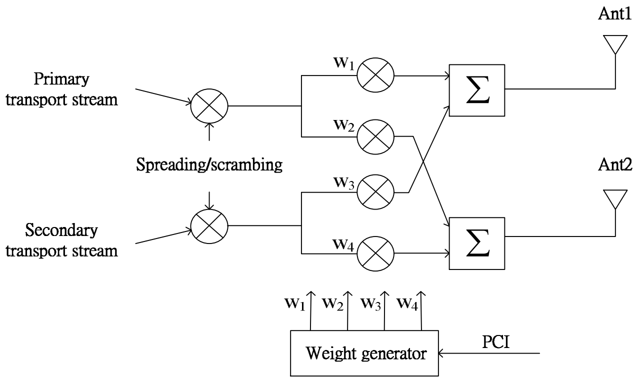

Figure 2.5: The downlink transmit structure for 2x2 MIMO approach

The pre-coding weights for primary data stream in release 7 are the same as those in closed-loop transmit diversity for release 99, $w_1 = 1/\sqrt{2}$, $w_2 \in \left\{\frac{1+j}{2}, \frac{1-j}{2}, \frac{-1+j}{2}, \frac{-1-j}{2}\right\}$. To reduce the overhead of PCI report, the terminal only report the preferred pre-coding weight for primary data stream. Weights w_3 and w_4 can be derived from weights w_1 and w_2, that is, $w_3 = w_1$ and $w_4 = -w_2$. Since w_1 is fixed, only w_2 is reported by PCI. Note that weights w_3 and w_4 are chosen to be orthogonal to weights w_1 and w_2. This can reduce the interference between the two data streams and lessen the burden on the receiver processing. According to this feedback PCI from UE, the Node B can determine the pre-coding weights for spatial processing.

2.4 Channel Model

In this thesis, we consider a terrestrial radio channel for urban areas just as that given in [23]. Three types of propagation factor are included in the channel model. They are path loss, slow variation resulting from shadowing and scattering, and the rapid variation due to the multi-path effects. Denote $F(t)$ the channel condition function at time t for WCDMA cellular system. $F(t)$ is modeled by long-term fading and short-term fading, and can be given by:

$$F(t) = \xi(d) \times 10^{\eta/10} \times \zeta(t). \tag{2-1}$$

Here, $\xi(d) \times 10^{\eta/10}$ is the long-term fading including path loss and shadowing, d is the distance from the antenna of the base station to the antenna of the terminal, and η is a normal-distributed random variable with zero mean and variance σ_L^2. $\zeta(t)$ is the short-term fading factor caused by the multi-path effect, and can be recalled from the Jakes model [24],

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

Here, σ is the radical of the average power signal, f_D is the Doppler frequency, $L = 4M + 2$ is the number of the signal path, $\beta_m = \pi m/(M+1)$, and

$$\theta_m = \beta_m + 2\pi ms/(M+1) \qquad \text{for } s = 0,1,2,\ldots,M-1. \tag{2-3}$$

Since the summation components of $\zeta(t)$ are mutually independent to each other, this model can produce up to M independent short-term fading. Therefore, we choose M to be equal to the number of the total links in all cells of the system. It is reasonable to suppose that the scattering geometry is time invariant within some small local area. We then can assume that parameters of the Jakes model are fixed in simulations.

The shadowing effect of a moving user would be different when the position of the user changes. However, compared to the motion of the user, the sampling frequency in HARQ is very short. This caused that, for a practical system, the degradation degree between two sampling time is small. Hence, the shadowing effects of these sampling points can be expected to be highly correlated. The correlation function is then defined by the distance between two adjacent sampling points. Let Δx be the position difference between two adjacent TTI. The correlated shading fading can then be obtained from the normalized autocorrelation function $\rho(\Delta x)$ [25]. Here,

$$\rho(\Delta x) = e^{-\frac{|\Delta x|}{l_{dcor}}\ln 2}, \tag{2-4}$$

with l_{dcor} being the decorrelation length.

2.5 Fuzzy Q-learning Algorithm

Fig 2.6 is a general learning system which is consists of five elements. The learner will select the optimal action according to the knowledge of the state by interacting with the environment. After applying the action, the environment will give some reward feedback to the learner. These rewards can help justifying the action decision policy for the better system performance.

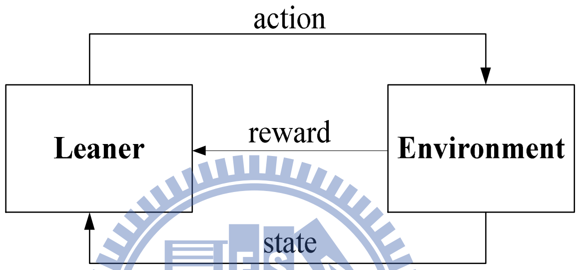

Figure 2.6: Block diagram of a learning system

The expectation of accumulated rewards is called Q function and it will be affected by the selected action at each time. Eq. (2-5) is the Q-function which depends on the system state denoted by x and the action denoted by a, respectively:

$$Q\left(x_0, a_0\right) = E\left\{\sum_{n=0}^{\infty} \gamma^n r(n) \middle| x(0) = x_0, a(0) = a_0\right\}, \tag{2-5}$$

where γ is the discount factor, $r(n)$ is the reinforcement signal (reward), $E\{\cdot\}$ is the expectation operation and n is the episode index. The Q-value is the output value of Q-function. Because the Q-value is the output value of the accumulation rewards in the future from now on, it will be affect by the selected action a_0 under state x_0 at current decision episode $n = 0$. Therefore, the reinforcement learning will choose the optimal action which can maximize the accumulation of rewards, denoted by a^* as Eq. (2-6):

$$a^* = \arg\max_a Q(x_0, a). \tag{2-6}$$

Since the system state in the future, $x(n)$ for $n > 0$, the expectation value $r(n)$ are usually unavailable in the real world. Hence, Watkins et al [18] proposed a recursive method, called Q-learning algorithm. The Q-learning algorithm can solve above problems and obtain the optimal Q function at next stage. It is given by [18] as:

$$Q_{n+1}(x,a) = \begin{cases} Q_n(x,a) + \eta\big[r(n) + \\ \qquad \gamma \max_b [Q_n(x(n+1), b)] - Q_n(x(n), a(n))\big], & \text{for } (x,a) = (x(n), a(n)), \\ Q_n(x,a), & \text{for } (x,a) \neq (x(n), a(n)). \end{cases} \tag{2-7}$$

where $Q_n(x,a)$ is the optimal Q function at episode n, and η is the learning rate, $\eta \in [0,1]$. In order to get the new optimal Q function $Q_{n+1}(x,a)$ at next stage, it is assume that the system state $x(n+1)$ is available at next stage and the feedback reward $r(n)$ caused by the last action is used to update $Q_n(x,a)$. Only the state-action pair (x,a) occurs in the previous episode can have information to correct its Q-value. After updating Q-value at each episode, we can get a more accurate Q-function approximation and use Eq. (2-7) to decide the optimal action.

The fuzzy Q-learning (FQL) algorithm can be regarded as the Q-learning algorithm which combined with fuzzy logic. Eq. (2-8) is the general form of fuzzy if-then rules which shows the characteristic of this combination scheme as:

$$\text{Rule } j: \text{if } X(n) \text{ is } S_j, \text{then } a_k \text{ with } q_n(S_j, a_k),\ 1 \le k \le K. \tag{2-8}$$

where $X(n) = [x_1(n), \ldots, x_H(n)]$ is the vector of input linguistic variables, H is the number of input variables, $q_n(S_j, a_k)$ is the Q-value for the state-action pair (S_j, a_k) at episode n, $\mathrm{S} = \{S_j, j = 1, \ldots, J\}$ is the set of system state, and $\mathrm{A} = \{a_k, k = 1, \ldots, K\}$ is the set of action candidate. Since each S_j containing a rule, there are J rules with different Q-values for each pair (S_j, a_k), $k = 1, \ldots, K$. The select-max strategy is adopted for each rule to choose the suitable action $a_j^*(n)$

$$a_j^*(n) = \arg\max_{a_k \in A} q_n(S_j, a_k) \ . \tag{2-9}$$

Because $X(n)$ belongs to each S_j with different intensity which depends on the combination of the membership function of each input linguistic variable, it can infer to J consequences from these J rules separately. Then these consequences $a_j^*(n), j = 1,...,J$ will be gathered to infer the optimal action, denoted by $a^*(n)$

$$a^*(n) = \frac{\sum_{j=1}^{J} \mu_{j,n} \times a_j^*(n)}{\sum_{j=1}^{J} \mu_{j,n}}, \tag{2-10}$$

where $\mu_{j,n}$ is the intensity of each rule j at episode n.

After applying the optimal action, the reward caused by the action is feedback to update the Q function which infers to Eq. (2-7):

$$q_{n+1}(S_j, a^*(n)) = q_n(S_j, a^*(n)) + \eta \times \Delta q_n(S_j, a^*(n)), \quad \text{for } 1 \le j \le J, \tag{2-11}$$

and

$$\Delta q_n(S_j, a^*(n)) = \frac{\mu_{j,n}}{\sum_{i=1}^{J} \mu_{i,n}} \times \left\{ r(n) + \gamma \times Q_n^*\left(X(n+1), a(n+1)\right) - Q_n\left(X(n), a^*(n)\right) \right\}. \tag{2-12}$$

$Q_n\left(X(n), a^*(n)\right)$ is the Q-value for state-action pair $(X(n), a^*(n))$ and is the weighted summation of the J Q-values $q_n(S_j, a_j^*(n)), j = 1,...,J$ by using the rule intensity $\mu_{j,n}$:

$$Q_n\left(X(n), a^*(n)\right) = \frac{\sum_{j=1}^{J} \left[\mu_{j,n} \times q_n\left(S_j, a_j^*(n)\right) \right]}{\sum_{j=1}^{J} \mu_{j,n}}. \tag{2-13}$$

$Q_n^*(X(n+1), a(n+1))$ is the transient optimal Q function, denoted by Q^* at next stage. Since the next stage Q-values $q_{n+1}(S_j, a_k), j = 1,\ldots,J,\ k = 1,\ldots,K$ are not available, $Q_n^*(X(n+1), a(n+1))$ will be calculated by $q_n(S_j, a_j^*), j = 1,\ldots,J$ which is defined as:

$$Q_n^*\left(X(n+1), a(n+1)\right) = \frac{\sum_{j=1}^{J}\left[\mu_{j,n+1} \times q_n\left(S_j, a_j^*(n)\right)\right]}{\sum_{j=1}^{J}\mu_{j,n+1}}. \tag{2-14}$$

Chapter 3
Fuzzy Q-learning based MIMO HARQ Scheme

In this chapter, the design of the fuzzy Q-learning-based MIMO HARQ (FQLM-HARQ) scheme is proposed to choose the MIMO transmission mode and the MCS level for each initial transmission. Since the decision in the HARQ scheme is based on current and past system state, this process can be modeled as a Discrete-time Markov decision process (MDP). Applying the fuzzy Q-learning-based algorithm, the FQLM-HARQ scheme can solve this MDP problem.

The overall structure of the FQLM-HARQ scheme is given in Fig 3.1. The measure parameters $BLER(n)$, $CQI_1(n)$, and $CQI_2(n)$ are translated into fuzzy rule base indicator $\beta(n)$ and input linguistic variables $X(n)$ by the block of operation. $\beta(n)$ indicates which fuzzy rule base $X(n)$ is suitable for. There are two fuzzy rule bases in FQLM-HARQ scheme. One is fuzzy rule base A (for mode 1), and the other is fuzzy rule base B (for mode 2). Note that, there are two transmission modes. They are spatial diversity (SD) and spatial multiplexing (SM). Here, mode 1 denotes SD transmission is used and mode 2 denotes SM transmission is used. After the calculation of the fuzzy rule base, we can get Q-value q_n for each system state. According to these Q-values, the inference engine can infer a optimal action for each mode. The action decision then decides the MCS level for each data stream. When the packet transmission is finished, the reinforcement signal generator will create a reinforcement signal $r(n)$ to the block of Q-function update based on system information. The Q-values in fuzzy rule base will be updated by Q-function update. The detailed design is given as follows.

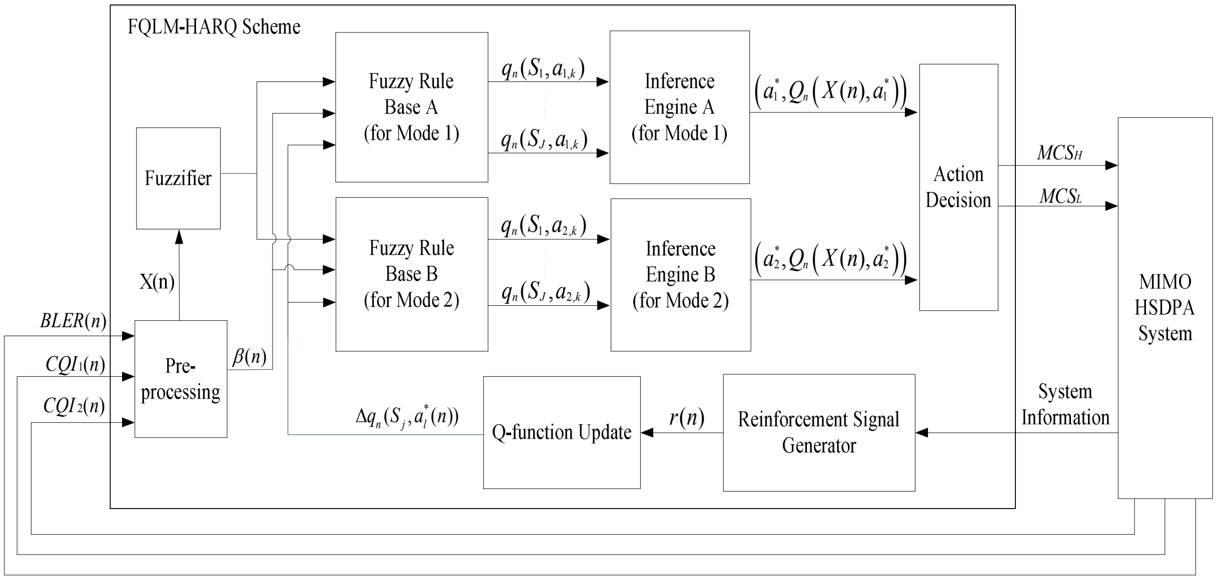

Figure 3.1: The FQLM-HARQ scheme

3.1 Pre-Processing

In this block, input data is pre-processing to reduce the amount of calculation in the FQLM-HARQ scheme. Let $CQI_1(n)$ and $CQI_2(n)$ be the channel quality indicator for the primary and secondary data streams, respectively. Note that, the values of $CQI_1(n)$ and $CQI_2(n)$ are measured by user equipment (UE) through common pilot channel (CPICH), and are integer values between 0 and 14. In order to reduce the number of fuzzy rule, we define $CQI_H(n)$ and $CQI_L(n)$ are the maximum and the minimum values, respectively, among $CQI_1(n)$ and $CQI_2(n)$. That is,

$$CQI_H(n) = \max\left\{CQI_1(n), CQI_2(n)\right\}, \tag{3-1}$$

and

$$CQI_L(n) = \min\{CQI_1(n), CQI_2(n)\}, \quad (3\text{-}2)$$

Assume CQI_{th} is the threshold of channel quality indicator for the selection of transmission mode. If $CQI_L(n)$ is less than CQI_{th}, the MIMO HSDPA system will use SD transmission. On the contrary, if $CQI_L(n)$ is greater than CQI_{th}, the MIMO HSDPA system will use SM transmission. However, since the CQI has report delay or measurement inaccuracies, there should be a fuzzy margin for the selection of MIMO transmission mode. Let $\beta(n)$ be the fuzzy rule base indicator and β_{th} be the threshold for the selection of fuzzy rule base. We then have the following three rules:

(1) if $CQI_L(n) - CQI_{th} < -\beta_{th}$, only fuzzy rule base A enabled,

(2) if $CQI_L(n) - CQI_{th} > \beta_{th}$, only fuzzy rule base B enabled,

(3) if $-\beta_{th} \le CQI_L(n) - CQI_{th} \le \beta_{th}$, both fuzzy rule A and fuzzy rule B enabled.

Let $\beta(n) = 0$ follow the (1) rule. This indicates the channel quality is really bad. In order to enhance the reliability, the single data stream is adopted for SD transmission and only the fuzzy rule A is enabled. Let $\beta(n) = 1$ follow the (2) rule. It implies channel quality is very good. Hence, only the fuzzy rule B is enabled and dual data streams are used for SM transmission to increase the system throughput. Let $\beta(n) = 2$ follow the (3) rule. It means that the channel quality is in the indistinct region such that we can't exactly decide which transmission mode is suitable. Hence, both fuzzy rule bases A and B are enabled. In this case, the actual transmission mode will be judged by the action decision.

3.2 Fuzzifier

Let $X(n) = \{BLER(n), CQI_H(n), CQI_L(n)\}$ be the input linguistic variables in the FQLM-HARQ scheme. $BLER(n)$ is the block error rate indicator, which is defined as the number of the packets with retransmissions over the total transmission packets at episode n.

The fuzzy term set for $BLER(n)$ is defined as $T\left(BLER(n)\right)=\left\{\text{Low, Middle, High}\right\}$ $=\left\{\text{L,M,H}\right\}$. Here, "Low" means $BLER(n)$ is in the safe region of maintaining the QoS requirement; "Middle" means $BLER(n)$ is in the normal region of achieving the QoS requirement; "High" means $BLER(n)$ is in the dangerous region with the probability of violating the QoS requirement. The fuzzy term sets for $CQI_H(n)$ and $CQI_L(n)$ are denoted by $T\left(CQI_H(n)\right)$ and $T\left(CQI_L(n)\right)$, respectively. They are discriminated eight levels which shown as $T\left(CQI_H(n)\right)=T\left(CQI_L(n)\right)=\left\{\text{Level 1, Level 2, Level 3, Level 4, Level 5, Level 6, Level 7, Level 8}\right\}=\left\{\text{L1, L2, L3, L4, L5, L6, L7, L8}\right\}$. Each term in $T\left(CQI_H(n)\right)$ and $T\left(CQI_L(n)\right)$ stands for the judgment of channel quality indication which implies the QoS requirement of block error rate while using the corresponding MCS level during this term.

The membership function for fuzzy terms can indicate the intensity of the input variable belong to itself fuzzy label, and is designed with pre-knowledge of the system. Before designing the membership function for fuzzy terms in $T\left(BLER(n)\right)$, $T\left(CQI_H(n)\right)$ and $T\left(CQI_L(n)\right)$, we first define a triangle function $f(\cdot)$ and a trapezoid shape function $g(\cdot)$.

A triangle function $f(\cdot)$ is defined in Fig. 3.2, and can be expressed as:

$$f\left(y;y_0,y_1,y_2\right)=\begin{cases}\dfrac{y-y_0}{y_1-y_0}, & \text{for } y_0\le y\le y_1,\\ \dfrac{y_2-y}{y_2-y_1}, & \text{for } y_1\le y\le y_2,\\ 0, & \text{otherwise,}\end{cases} \tag{3-3}$$

where y_0, y_1, and y_2 in $f(\cdot)$ is the left edge, center, and right edge, respectively, of the triangular function.

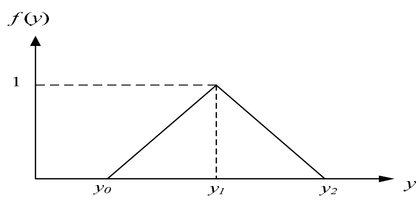

Figure 3.2: A triangle function $f(\cdot)$

A basic shape of a trapezoid shape function $g(\cdot)$ is shown in Fig. 3.3, and can be defined as:

$$g(x;x_1,x_2,x_3,x_4)=\begin{cases}\dfrac{x-x_1}{x_2-x_1}, & \text{for } x_1\le x\le x_2,\\ 1, & \text{for } x_2\le x\le x_3,\\ \dfrac{x_4-x}{x_4-x_3}, & \text{for } x_3\le x\le x_4,\\ 0, & \text{otherwise,}\end{cases} \tag{3-4}$$

where x_1 ,x_2, x_3, and x_4 in $g(\cdot)$ are the left edge, center-right, center-left and right edge, respectively, of the trapezoid.

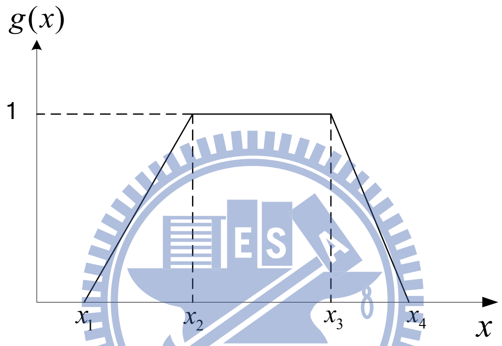

Figure 3.3: A trapezoid shape function $g(\cdot)$

The membership function of $BLER(n)$, denoted by $\mu(BLER(n))$, composed of the membership functions of the terms, "L", "M", and "H" of $T(BLER(n))$, denoted by $\mu_{\mathrm{L}}(BLER(n))$, $\mu_{\mathrm{M}}(BLER(n))$, and $\mu_{\mathrm{H}}(BLER(n))$, respectively. Here,

$$\mu_{\mathrm{L}}(BLER(n))=g(BLER(n);\ -\infty,\ 0,\ \mathrm{A}_1,\ \mathrm{A}_2), \tag{3-5}$$

$$\mu_{\mathrm{M}}(BLER(n))=g(BLER(n);\ \mathrm{A}_1,\ \mathrm{A}_2,\ \mathrm{A}_3,\ \mathrm{A}_4), \tag{3-6}$$

$$\mu_{\mathrm{H}}(BLER(n))=g(BLER(n);\ \mathrm{A}_3,\ \mathrm{A}_4,\ \mathrm{A}_5,\ \infty), \tag{3-7}$$

In our design, we set $\mathrm{A}_1=0.5\times BLER^*$, $\mathrm{A}_2=0.7\times BLER^*$, $\mathrm{A}_3=BLER^*$ and $\mathrm{A}_4=1.2\times BLER^*$, where $BLER^*$ denotes the BLER requirement. We then have the membership function of $BLER(n)$ as shown in Fig. 3.4.

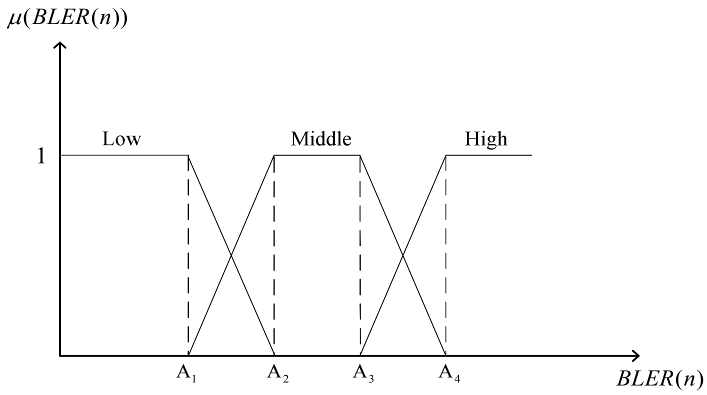

Figure 3.4: The membership function of $BLER(n)$

The membership functions of $CQI_H(n)$ and $CQI_L(n)$, denoted by $\mu\left(CQI_H(n)\right)$ and $\mu\left(CQI_L(n)\right)$, are shown in Fig 3.5. Here, $\mathrm{B}i$, for $i=1,\ldots,8$ is set to the required value of channel quality indication to maintain $BLER^*$ while using the i-th MCS level. Since $CQI_H(n)$ and $CQI_L(n)$ have the same fuzzy set. We use $CQI(n)$ instead of $CQI_H(n)$ and $CQI_L(n)$ to conveniently describe the membership functions of $CQI_H(n)$ and $CQI_L(n)$. $\mu\left(CQI(n)\right)$ is composed of the membership functions of the term $\mathrm{L}1,\ldots,\mathrm{L}8$ of $CQI(n)$, denoted by $\mu_{\mathrm{L}1}\left(CQI(n)\right),\ldots,\mu_{\mathrm{L}8}\left(CQI(n)\right)$, respectively.

$$\mu_{\mathrm{L}1}\left(CQI(n)\right)=g\left(CQI(n);\ \text{-}\infty,\ 0,\ \mathrm{B}1,\ \mathrm{B}2\right). \tag{3-8}$$

$$\mu_{\mathrm{L}2}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}1,\ \mathrm{B}2,\ \mathrm{B}3\right). \tag{3-9}$$

$$\mu_{\mathrm{L}3}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}2,\ \mathrm{B}3,\ \mathrm{B}4\right). \tag{3-10}$$

$$\mu_{\mathrm{L}4}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}3,\ \mathrm{B}4,\ \mathrm{B}5\right). \tag{3-11}$$

$$\mu_{\mathrm{L}5}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}4,\ \mathrm{B}5,\ \mathrm{B}6\right). \tag{3-12}$$

$$\mu_{\mathrm{L}6}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}5,\ \mathrm{B}6,\ \mathrm{B}7\right). \tag{3-13}$$

$$\mu_{\mathrm{L}7}\left(CQI(n)\right)=f\left(CQI(n);\ \mathrm{B}6,\ \mathrm{B}7,\ \mathrm{B}8\right). \tag{3-14}$$

$$\mu_{\mathrm{L}8}\left(CQI(n)\right)=g\left(CQI(n);\ \mathrm{B}7,\ \mathrm{B}8,\ 14,\ \infty\right). \tag{3-15}$$

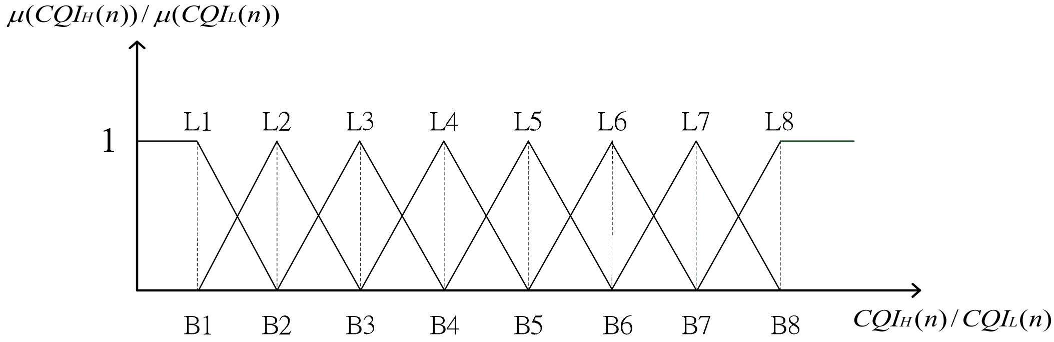

Figure 3.5: The membership function of $CQI_H(n)$ and $CQI_L(n)$

3.3 Fuzzy Rule Base

The fuzzy rule base is consisted of if-then rules. In the FQLM-HARQ scheme, we have two fuzzy rule bases. One is fuzzy rule base A for mode 1, and the other is fuzzy rule base B for mode 2. Since the value of $CQI_H(n)$ is greater than or equal to the value of $CQI_L(n)$, there are 108 kinds of combination of $T\left(BLER(n)\right), T\left(CQI_H(n)\right)$ and $T\left(CQI_L(n)\right)$. Therefore, the state in our system is $S_j, j = 1,\ldots,108$. Assume each state has one fuzzy Q-learning rule. We then have the following two rule bases.

3.3.1 Fuzzy Rule Base A

In the fuzzy rule base A, the Q-value for the state action pair $q_n(S_j, a_{1,k})$ can be got through the fuzzy if-then rule as:

$$\text{Rule } j: \text{if } X(n) \text{ is } S_j, \text{ then } a_{1,k} \text{ with } q_n(S_j, a_{1,k}), \text{ for } k = 1,\ldots,8 \qquad (3\text{-}16)$$

where $a_{1,k}$ is the action for mode 1 transmission, which means the FQLM-HARQ scheme chooses the k-th MCS level for the single data stream. The design concept of MIMO-HARQ scheme is to discriminate the prefer actions $a_{1,k}$ according to $BLER(n)$. If $BLER(n)$

performance is better, then the choice of the action is more aggressive. On the other hand, if $BLER(n)$ performance is worse, then the choice of the action is more conservative. While $CQI_H(n)$ and $CQI_L(n)$ are help selecting the prefer actions for $BLER(n)$, they can accelerate the learning procedure. In the following, we will divide the fuzzy rules into three parts based on the 3 terms, "Low", "Middle" and "High" of $BLER(n)$.

Green Part:

If $BLER(n)$ *is Low,* $CQI_H(n)$ *is Level m, and* $CQI_L(n)$ *is Level n, then* $a_{1,k}$ *with* $q_n((BLER(n)\ is\ Low,\ CQI_H(n)\ is\ Level\ m,\ CQI_L(n)\ is\ Level\ n), a_{1,k})$,

$$k \geq \left\lceil \frac{m+n}{2} \right\rceil$$

Because of $1 \leq \mathrm{m} \leq 8$, $1 \leq \mathrm{n} \leq 8$, and $\mathrm{m} \geq \mathrm{n}$, there are 36 fuzzy rules in Green Part. Here, $BLER(n)$ is in the safe region and the main purpose is to maximize the throughput of the system. Therefore, we only consider $a_{1,k}$ with $k \geq \left\lceil \frac{\mathrm{m}+\mathrm{n}}{2} \right\rceil$ while k up to 8 for more aggressive way.

Yellow Part:

If $BLER(n)$ *is Middle,* $CQI_H(n)$ *is Level m, and* $CQI_L(n)$ *is Level n, then* $a_{1,k}$ *with* $q_n((BLER(n)\ is\ Middle,\ CQI_H(n)\ is\ Level\ m,\ CQI_L(n)\ is\ Level\ n), a_{1,k})$,

$$k = \left\lfloor \frac{m+n}{2} \right\rfloor - 1, \ldots, m+1$$

There are 36 fuzzy rules in Yellow Part. Here, $BLER(n)$ is in the normal region and we choose $a_{1,k}$ with $k = \left\lfloor \frac{\mathrm{m}+\mathrm{n}}{2} \right\rfloor - 1, \ldots, \mathrm{m}+1$ under the feedback $CQI_H(n)$ and $CQI_L(n)$. If the value of $\left\lfloor \frac{\mathrm{m}+\mathrm{n}}{2} \right\rfloor$ is equal to 1, we will choose k from 1. While the value of m is equal to 8, we will choose k up to 8. By this selection region, we want to balance the QoS requirement for $BLER^*$ and the throughput of the system.

Red Part:

If $BLER(n)$ *is High,* $CQI_H(n)$ *is Level m, and* $CQI_L(n)$ *is Level n, then* $a_{1,k}$ *with* $q_n((BLER(n)\ is\ High,\ CQI_H(n)\ is\ Level\ m,\ CQI_L(n)\ is\ Level\ n), a_{1,k})$,

$k = n-1, \ldots, \left\lceil \frac{m+n}{2} \right\rceil + 1$

There are also 36 fuzzy rules in Red Part. Here, $BLER(n)$ is in the dangerous region, the action decision in this part should be more conservative. Therefore, we take $a_{1,k}$ with $\mathrm{n}-1, \ldots, \left\lceil \frac{\mathrm{m}+\mathrm{n}}{2} \right\rceil + 1$ into account. If the value of $\left\lfloor \frac{\mathrm{m}+\mathrm{n}}{2} \right\rfloor$ is equal to 8, we will choose k up to 8. While the value of n is equal to 1, we will choose k from 1.

3.3.2 Fuzzy Rule Base B

In the fuzzy rule base B, the Q-value $q_n(S_j, a_{2,k})$ for the state-action pair $(S_j, a_{2,k})$ can be got through the fuzzy if-then rule as:

$$\text{Rule } j: \text{if } X(n) \text{ is } S_j, \text{ then } a_{2,k} \text{ with } q_n(S_j, a_{2,k}), \text{ for } k = 1, \ldots, 36 \tag{3-17}$$

where $a_{2,k}$ is the action for mode 2 transmission, which means that the FQLM-HARQ scheme chooses the p-th MCS level, denoted by MCS_p^*, for the data stream with best channel quality and the q-th MCS level, denoted by MCS_p^*, for the data stream with worst channel quality, where $k = p - \frac{q^2}{2} + \frac{17q}{2} - 8$. Fig 3.6 shows the relationship between the action $a_{2,k}$ and MCS levels. Here, MCS_H is the MCS level for the data stream with best channel quality and MCS_L is the MCS level for the data stream with worst channel quality.

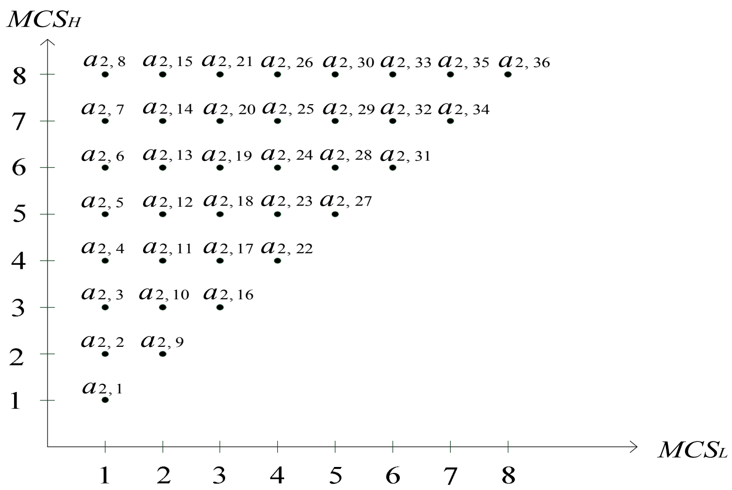

Figure 3.6: The relationship between MCS levels and action $a_{2,k}$

The design concept is also to discriminate the prefer actions $a_{2,k}$ according to $BLER(n)$ for each state which is the same form as fuzzy rule base A. In the following, we will divide the fuzzy rules into three parts based on the 3 terms, "Low", "Middle" and "High" of $BLER(n)$.

Green Part :

If $BLER(n)$ *is Low,* $CQI_H(n)$ *is Level m, and* $CQI_L(n)$ *is Level n, then* $a_{2,k}$

with $q_n((BLER(n) \text{ is Low}, CQI_H(n) \text{ is Level } m, CQI_L(n) \text{ is Level } n), a_{2,k})$,

$$k = p - \frac{q^2}{2} + \frac{17q}{2} - 8,\ p > m,\ q > n,\ p \geq q$$

Because of $1 \leq m \leq 8$, $1 \leq n \leq 8$, and $m \geq n$, there are 36 fuzzy rules in Green Part. Here, $BLER(n)$ is in the safe region and we also take the action with more aggressive for mode 2 to increase the system throughput. Therefore, we only consider $a_{2,k}$ with $k = p - \frac{q^2}{2} + \frac{17q}{2} - 8,\ p > \text{m},\ q > \text{n},\ p \geq q$ and the limit of p and q are up to 8.

Yellow Part:

If $BLER(n)$ *is Middle,* $CQI_H(n)$ *is Level m, and* $CQI_L(n)$ *is Level n, then*

$a_{2,k}$ *with* $q_n((BLER(n) \text{ is Middle}, CQI_H(n) \text{ is Level } m, CQI_L(n) \text{ is Level } n), a_{2,k})$,

$$k = p - \frac{q^2}{2} + \frac{17q}{2} - 8,\ m-1 \le p \le m+1,\ n-1 \le q \le n+1,\ p \ge q$$

We take $a_{2,k}$ with $k = p - \frac{q^2}{2} + \frac{17q}{2} - 8$, m-1 $\le p \le$ m+1, n-1 $\le q \le$ n+1, $p \ge q$ into account, because of $BLER(n)$ is in the normal region. Therefore, $a_{2,k}$ is chosen as contained the most information while the block error rate is near $BLER^*$ under the feedback $CQI_H(n)$ and $CQI_L(n)$. There are also 36 rules in this part. Note that, while the value of m is equal to 1, the selection of p is from 1. While the value of m is equal to 8, the selection of p is up to 8. There are the same way for the selection of q according to the value of n.

Red Part:

If $BLER(n)$ is High, $CQI_H(n)$ is Level m, and $CQI_L(n)$ is Level n, then

$a_{2,k}$ *with* $q_n((BLER(n)\ is\ High,\ CQI_H(n)\ is\ Level\ m,\ CQI_L(n)\ is\ Level\ n), a_{2,k}),$

$$k = p - \frac{q^2}{2} + \frac{17q}{2} - 8,\ p < m,\ q < n,\ p \ge q$$

Because of $BLER(n)$ is in the dangerous region, $a_{2,k}$ will select the conservative policy with $k = p - \frac{q^2}{2} + \frac{17q}{2} - 8$, $p <$ m, $q <$ n, $p \ge q$ and the value of p and q count from 1. While The goal in this region is to achieve the QoS requirement $BLER^*$. As the same, there are 36 rules in this part.

3.4 Inference Engine

There are two inference engines for each fuzzy rule base. One is inference engine A for model 1, and the other is inference engine B for mode 2. Details are given as follows.

3.4.1 Inference Engine A

By using the select-max strategy, a suitable action for each rule in fuzzy rule base A can be obtained by:

$$a_{1,j}^{*}(n) = \arg \max_{a_{1,k} \in \mathrm{A}_1} q_n(S_j, a_{1,k}) \ , \tag{3-18}$$

where $\mathrm{A}_1 = \left\{ a_{1,k}, \text{ for } k = 1,...,8 \right\}$ is the candidate action for mode 1 transmission at episode n. The optimal action for mode 1 transmission, denoted by $a_1^*(n)$, can be derived from Eq. (3-19) as:

$$a_1^*(n) = \frac{\sum_{j=1}^{J} \mu_{j,n} \times a_{1,j}^*(n)}{\sum_{j=1}^{J} \mu_{j,n}}, \tag{3-19}$$

The Q-value for $a_1^*(n)$ can be obtained by:

$$Q_n\left(X(n), a_1^*(n)\right) = \frac{\sum_{j=1}^{J} \left[\mu_{j,n} \times q_n\left(S_j, a_{1,j}^*(n)\right) \right]}{\sum_{j=1}^{J} \mu_{j,n}}. \tag{3-20}$$

3.4.2 Inference Engine B

By using the fuzzy select-max strategy, a suitable action for each rule in fuzzy rule base B can be got by:

$$a_{2,j}^{*}(n) = \arg \max_{a_{2,k} \in \mathrm{A}_2} q_n(S_j, a_{2,k}), \tag{3-21}$$

where $\mathrm{A}_2 = \left\{ a_{2,k}, \text{ for } k = 1,...,36 \right\}$ is the set of candidate action for mode 2 transmission. The optimal action for mode 2 transmission, denoted by $a_2^*(n)$, can then be obtained by:

$$a_2^*(n)=\frac{\sum_{j=1}^{J}\mu_{j,n}\times a_{2,j}^*(n)}{\sum_{j=1}^{J}\mu_{j,n}}. \tag{3-22}$$

The Q-value for $a_2^*(n)$ can be calculated by:

$$Q_n\left(X(n),a_2^*(n)\right)=\frac{\sum_{j=1}^{J_2}\left[\mu_{j,n}\times q_n\left(S_j,a_{2,j}^*(n)\right)\right]}{\sum_{j=1}^{J_2}\mu_{j,n}}. \tag{3-23}$$

3.5 Action Decision

In the block, the final action at episode n, is denoted by:

$$l^*=\arg\max_{l\in\{1,2\}} Q_n\left(X(n),a_l^*(n)\right) \tag{3-24}$$

If $l^*=1$, the mode 1 transmission is selected. Hence, only one data stream is transmitted. The output of this block is set to $MCS_H=a_1^*(n)$ and $MCS_L=0$. Based on this output signal, the MIMO HARQ system uses $a_1^*(n)$ to transmit the data stream with largest delay time in one of two HARQ processes. Note that, for each episode, there are two HARQ processes can be used. If $l^*=2$, the mode 2 transmission is selected. Hence, dual data stream is transmitted. The output of this block is set according to the action $a_2^*(n)=(MCS_p^*,MCS_q^*)$. That is, $MCS_H=MCS_p^*$ and $MCS_L=MCS_q^*$. Based on this output signal, the MIMO HARQ system sets MCS_p^* for the data stream with best channel condition and MCS_q^* for the data stream with worst channel condition.

3.6 Reinforcement Signal Generator

We design the reinforcement signal for each rule to update the Q-value of each action and accomplish the Q-learning operation. There are two reinforcement signal for two kinds of transmission mode. One is the reinforcement signal for mode 1 and the other is the reinforcement signal for mode 2. The details are as following:

3.6.1 Reinforcement Signal for Mode 1

The reinforcement signal are designed according to the three part in fuzzy rule base A. Rules in the same parts will have the same reinforcement signal.

Green Part: For $BLER(n)$ *is Low*

$$r(n)=\begin{cases}\dfrac{B_{infor.}}{B_{infor.}+B_{redun.}}, & \text{if the packet is correctly received,}\\ -(5+\alpha_1\times BLER(n)), & \text{if the packet is dropped,}\end{cases} \tag{3-25}$$

$B_{infor.}$ represents the number of information bits in the packet of the single data stream and $B_{redun.}$ represents the required redundancy bits of successful transmission which included the initial transmission and the retransmission in this packet. It can expect the higher successful transmitted data rate of the packet will get the larger reward feedback. On the contrary, if the packet is dropped after three failed decoding, we give it a punishment as $-(5+\alpha_1\times BLER(n))$. It will get the more punishment when the larger block error rate at this episode n. This is because the larger average block rate, the dropped packet will increase the more load for the system. We will update the Q-value after each packet transmission is completed.

Yellow Part: For $BLER(n)$ *is Middle*

$$r(n)=\begin{cases}\dfrac{1}{(1+RT_1(n))}\dfrac{B_{infor.}}{B_{infor.}+B_{redun.}}, & \text{if the packet is correctly received,}\\ -(5+\alpha_2\times BLER(n)), & \text{if the packet is dropped,}\end{cases} \tag{3-26}$$

When the packet is correctly received, we can see that the degree of reward feedback is inverse proportional to $(1+RT_1(n))$. Here, $RT_1(n)$ is the retransmission time of the packet for mode 1 when the packet transmission in single data stream is completed at episode n. It means the more retransmission times, the less reward feedback by expecting the more conservative policy than Green Part. If the packet is dropped, it will get $-(5+\alpha_2\times BLER(n))$ for punishment. We give it the more punishment than Green Part because of the worse $BLER(n)$ performance.

Red Part: For $BLER(n)$ *is High*

$$r(n)=\begin{cases}\dfrac{B_{infor.}}{B_{infor.}+B_{redun.}}, & \text{if the packet is correctly received without retransmission,}\\ -(5+\alpha_3\times\min\{BLER(n),1.5\times BLER^*\}), & \text{if the packet is dropped,}\\ -\dfrac{(1+RT_1(n))}{3}\dfrac{B_{infor.}}{B_{infor.}+B_{redun.}}, & \text{if the packet is correctly received with retransmission,}\end{cases} \tag{3-27}$$

If the packet is correctly received at initial transmission, the reward feedback is design as $B_{infor.}/(B_{infor.}+B_{redun.})$ which is the same as previous part. However, if the packet is correctly received with retransmission, we give it a punishment according to the successful transmitted data rate of the packet while proportional to normalized $(1+RT_1(n))$ by dividing 3. This is because the purpose for this part is to maintain the QoS requirement for $BLER^*$. If the packet is dropped, we will give it the severest punishment $-(5+\alpha_3\times BLER(n))$ with the limit of $BLER(n)$ no more than $1.5\times BLER^*$.

3.6.2 Reinforcement Signal for Mode 2

The designed reinforcement signal are also according to the three part in fuzzy rule base B.

Rules in the same parts will have the same reinforcement signal.

Green Part: For $BLER(n)$ *is Low*

$$r(n)=\begin{cases}\dfrac{B_{infor.total}}{B_{infor.total}+B_{redun.total}}, & \text{if both packets is correctly received,}\\ -(5+\alpha_1\times BLER(n))(\dfrac{N_{dropped}}{2}), & \text{if the dropped packet happened,}\end{cases} \tag{3-28}$$

$B_{infor.total}$ represents the total number of information bits of the packets in dual data streams and $B_{redun.total}$ represents the total required redundancy bits which included the initial transmission and the retransmission of these packets. The higher data rate for the successful transmission of the packets, the larger reward feedback will get. $N_{dropped}$ represents the number of the dropped packets in dual data streams. As long as the dropped packet occurs, it will get the punishment as $-(5+\alpha_1\times BLER(n))(N_{dropped}/2)$ which is proportional to the number of the dropped packets. We also update the Q-value after the transmission streams is completed.

Yellow Part: For $BLER(n)$ *is Middle*

$$r(n)=\begin{cases}\dfrac{1}{(1+\overline{RT}_2(n))}\dfrac{B_{infor.total}}{B_{infor.total}+B_{redun.total}}, & \text{if both packets correctly received,}\\ -(5+\alpha_2\times BLER(n))(\dfrac{N_{dropped}}{2}), & \text{if the dropped packet happed,}\end{cases} \tag{3-29}$$

$\overline{RT}_2(n)$ is the average retransmission time of the packets for mode 2 when the packets transmission in dual data streams are completed at episode n. We can see that the more average retransmission time, the less reward feedback. Here, the punishment is the same form as Green Part if the dropped packet happened. However, we will give it the more punishment than Green Part, while it has the worse $BLER(n)$ performance.

Red Part: For $BLER(n)$ *is High*

$$r(n)=\begin{cases}\dfrac{B_{infor.total}}{B_{infor.total}+B_{redun.total}}, & \text{if both packets correctly received without retransmission,}\\ -(5+\alpha_3\times\min\{BLER(n),1.5\times BLER^*\}), & \text{if the dropped packet happened,}\\ -\dfrac{(1+\overline{RT}_2(n))}{3}\dfrac{B_{infor.total}}{B_{infor.total}+B_{redun.total}}, & \text{if both packets correctly received with retransmission,}\end{cases} \tag{3-30}$$

If both packets are successful received at initial transmission, the reward feedback is according the achievable data rate $B_{infor.total}/(B_{infor.total}+B_{redun.total})$. If both packets correctly received with retransmissions, we will give it a punishment which is proportional to normalized $(1+\overline{RT}_2(n))$ by dividing 3. However, in this part, just one packet is dropped through three failed decoding and we will give it the severe punishment according to $-\left(5+\alpha_3\times BLER(n)\right)$ with the limit of $BLER(n)$ no more than $1.5\times BLER^*$.

3.7 Q-function Update

According to the feedback reinforcement signal, the purpose of Q-function updating operation is to get new $q_{n+1}(S_j,a_l^*(n)), j=1,...,J$ for fuzzy rule base A (for mode 1) if $l^*=1$ in Eq. (3-24) or for fuzzy rule base B (for mode 2) if $l^*=2$ in Eq. (3-24). In the following, we describe how Q-function updates for these two transmission mode, respectively.

When the FQLM-HARQ scheme decides using the mode 1 transmission, we use the function of Eq. (3-31) to update the Q-function and get $q_{n+1}(S_j,a_1^*(n)), j=1,...,J$.

$$q_{n+1}(S_j,a_1^*(n))=q_n(S_j,a_1^*(n))+\eta\times\Delta q_n(S_j,a_1^*(n)),\ \text{ for } 1\le j\le J, \tag{3-31}$$

and

$$\Delta q_n(S_j,a_1^*(n))= \frac{\mu_{j,n}}{\sum_{i=1}^{J}\mu_{i,n}}\times\left\{r(n)+\gamma\times Q_n^*\left(X(n+1),a(n+1)\right)-Q_n\left(X(n),a_1^*(n)\right)\right\}. \tag{3-32}$$

The Q-value for state-action pair $(X(n),a_1^*(n))$, denoted by $Q_n\left(X(n),a_1^*(n)\right)$, is the weighted summation of the J Q-values $q_n(S_j,a_{1,j}^*(n)), j=1,...,J$ by using the rule intensity $\mu_{j,n}$ of $X(n)$, that is:

$$Q_n\left(X(n), a_1^*(n)\right) = \frac{\sum_{j=1}^{J}\left[\mu_{j,n} \times q_n\left(S_j, a_{1,j}^*(n)\right)\right]}{\sum_{j=1}^{J}\mu_{j,n}}. \tag{3-33}$$

$Q_n^*(X(n+1), a(n+1))$ is the transient optimal Q function at next episode. Since the next episode Q-values $q_{n+1}(S_j, a_{1,j,}^*), j=1,\ldots,J,$ is not available, it will be calculated by $q_n(S_j, a_{1,j}^*), j=1,\ldots,J$, defined as Eq. (3-34):

$$Q_n^*\left(X(n+1), a(n+1)\right) = \frac{\sum_{j=1}^{J}\left[\mu_{j,n+1} \times q_n\left(S_j, a_{1,j}^*(n)\right)\right]}{\sum_{j=1}^{J}\mu_{j,n+1}}. \tag{3-34}$$

When the FQLM-HARQ scheme decides using the mode 2 transmission, we use the function of Eq. (3-35) to update the Q-function and get $q_{n+1}(S_j, a_2^*(n)), j=1,\ldots,J$.

$$q_{n+1}(S_j, a_2^*(n)) = q_n(S_j, a_2^*(n)) + \eta \times \Delta q_n(S_j, a_2^*(n)), \text{ for } 1 \le j \le J, \tag{3-35}$$

and

$$\Delta q_n(S_j, a_2^*(n)) = \frac{\mu_{j,n}}{\sum_{i=1}^{J}\mu_{i,n}} \times \left\{r(n) + \gamma \times Q_n^*\left(X(n+1), a(n+1)\right) - Q_n\left(X(n), a_2^*(n)\right)\right\}. \tag{3-36}$$

The Q-value for state-action pair $(X(n), a_2^*(n))$, denoted by $Q_n\left(X(n), a_2^*(n)\right)$, is the weighted summation of the J Q-values $q_n(S_j, a_{2,j}^*(n)), j=1,\ldots,J$ by using the rule intensity $\mu_{j,n}$ of $X(n)$, that is:

$$Q_n\left(X(n), a_2^*(n)\right) = \frac{\sum_{j=1}^{J}\left[\mu_{j,n} \times q_n\left(S_j, a_{2,j}^*(n)\right)\right]}{\sum_{j=1}^{J}\mu_{j,n}}. \tag{3-37}$$

$Q_n^*(X(n+1), a(n+1))$ is the transient optimal Q-value at next stage. Since the next stage Q-values $q_{n+1}(S_j, a_{2,j,}^*), j=1,\ldots,J,$ is not available, it will be calculated by

$q_n(S_j, a_{2,j}^*), j = 1, \ldots, J$, defined as Eq. (3-38).

$$Q_n^*\left(X(n+1), a(n+1)\right) = \frac{\sum_{j=1}^{J}\left[\mu_{j,n+1} \times q_n\left(S_j, a_{2,j}^*(n)\right)\right]}{\sum_{j=1}^{J}\mu_{j,n+1}}. \tag{3-38}$$

After the Q-function update, we can get more accurate Q-value for fuzzy rule base to reflect the performance in time and achieve the expected goal.

Chapter 4
Simulation Results and Discussions

4.1 System Environment and Parameters

We consider a hexagonal grid cell structure in our simulation. There are 19 base stations (BS) in the multi-cell system. We assume that the HS-DSCH is allocated at maximum up to 80% of the total power of a BS. Hence, we define the HSDPA service power ratio (HSPR) to represent the ratio of transmission power on HS-DSCH to the total transmission power each antenna at BS for the user. The residual power will be used other service and control channels within the same cell. Here, HSPR controls the amount of the transmission power and the interference from self cell. The channel condition is assumed to be constant within a TTI and is described in section 2.4. We assume the user always has data to be transmitted for simplifying the simulation complexity and reach the higher data rate.

In the simulation, we assume the CQI delay is set to be 10ms. The system performance considers the different HSPR with fixed user mobility at mean user mobility 60 km/hr and the different user mobility with fixed power allocation at 60%. The detailed simulation environment parameters are shown in Table 4.1. In order to evaluate the performance, we will discuss about the system throughput, BLER and dropping rate that comparing to other schemes.

Table 4.1: Simulation parameters

Parameter	Assumption
Cellular layout	Hexagonal grid, 19 sites, 1000m cell radius
Path loss model $(\xi(r))$	$128.1 + 37.6\log_{10}(r)$ r is the base station separation in kilometers
Decorrelation length (l_{cor})	30m
σ_L	8.0
Mobility assignment	0 km/hr to 120 km/hr
Carrier frequency	2.0 GHz
Channel bandwidth	5.0 MHz
Chip-rate	3.84 Mcps
Spreading factor	16
Thermal noise density	-174 dBm/Hz
Forgetting factor (γ)	0.1
Learning rate (η)	0.9
TTI length	2ms
CQI_{th}	4
β_{th}	2
α_1	100
α_2	125
α_3	150
$BLER^*$	0.1
Power for HSDPA data transmission	Maximum of 80% of total maximum available transmission power
ACK/NACK delay	10ms
HARQ	IR

4.2 Conventional Schemes

In the simulation, we will compare the proposed FQLM-HARQ scheme with two conventional schemes, which are described in the following:

(1) Fixed threshold selection (FTS):

FTS sets the SINR threshold for each MCS based on the pre-known BLER performance [1], [26]. The SINR threshold is the required SINR that the MCS has BLER equal to the requirement 0.1. This scheme will choose the MCS whose corresponding threshold is just under and closest to the measured SINR at each TTI.

(2) Q-learning based HARQ (QL-HARQ) [9]:

QL-HARQ uses the Q-learning algorithm to learn an optimal policy for each initial transmission. The reinforcement signal is designed to be the normalized difference square of the last received SINR and the required SINR of the last MCS decision. After learning, QL-HARQ will choose an optimal MCS to meet the BLER requirement.

In the following section, we will show the simulation results and discuss about them.

4.3 Performance Evaluation and Discussions

Fig 4.1 is the BLER versus HSPR with fixed user mobility at 60 km/hr. The motion incurs the Doppler Effect and the channel variance with CQI delay. Hence, the actual channel condition will be different from the channel information used for determination. It can be seen in Fig 4.1 that the proposed FQLM-HARQ satisfies the BLER requirement with HSPR more than 60%, the next is QL-HARQ with HSPR more than 70%. However, the FTS violates the BLER requirement even with HSPR up to 80%. The FQLM-HARQ has better learning way by considering the situation in different part of $BLER(n)$ and then adjusts the selection of MCS level to do its possible maintaining the BLER requirement. As the more BS transmission power is allocated for HS-DSCH service, the more MCS level can adaptively select in Red Part to reduce the effect of the channel variance. The MCS level selection of FTS just depends on the current channel condition while it is regardless of the channel variation. This results the BLER performance violating the BLER requirement. Although QL-HARQ also adjusts the MCS level based on the last transmission decision, it does not take the information of transmission results into account. Therefore, it is not flexible enough to accommodate to the channel variation like FQLM-HARQ.

Fig 4.2 is the dropping rate versus HSPR with fixed user mobility at 60 km/hr. In the simulation, when the total transmission times including retransmissions of the same transmission block is more than three, this block will be dropped. It can see that FQLM-HARQ has the lowest dropping rate and FTS has the highest dropping rate despite the HSPR. This has the relation between initial BLER shown in Fig 4.1 and the MCS level selection with the conservative or aggressive way. The smaller BLER performance can result in the lower dropping rate. In FQLM-HARQ, once the dropping block occurs, the design of the reinforcement signal will give the most punishment at each part according the value of $BLER(n)$. After updating the Q-function, we can expect conservative MCS level selection. This design can limit the dropping rate. The QL-HARQ just considers the difference between the received SINR and the required SINR at the last transmission. Therefore, the MCS level will be more aggressive than the FQLM-HARQ and results the more dropping rate.

Fig 4.3 is the system throughput versus HSPR with fixed user mobility at 60 km/hr. It can see that when HSPR increases, the throughputs of the three schemes increase, absolutely. With Fig 4.1, Fig 4.2 and Fig 4.3, FQLM-HARQ can select the optimal MCS level for the largest throughput than the other two schemes while endeavoring to maintain the BLER requirement and result the least dropping rate simultaneously.

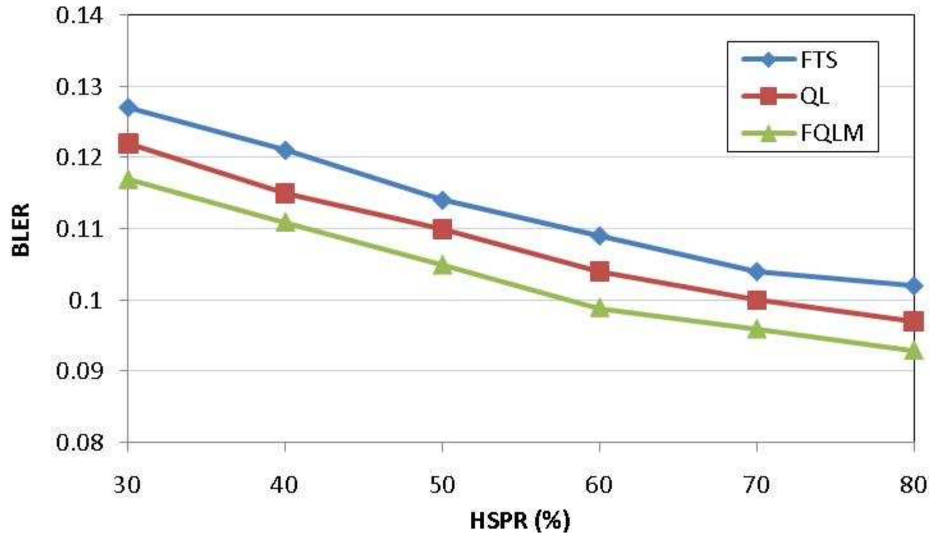

Figure 4.1: The BLER versus HSPR with fixed user mobility at 60 km/hr.

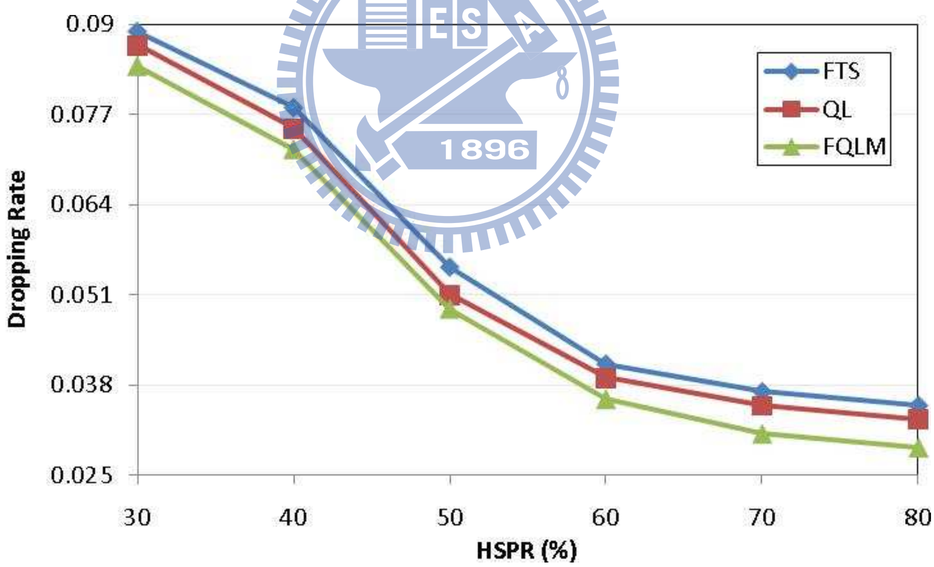

Figure 4.2: The dropping rate versus HSPR with fixed user mobility at 60 km/hr.

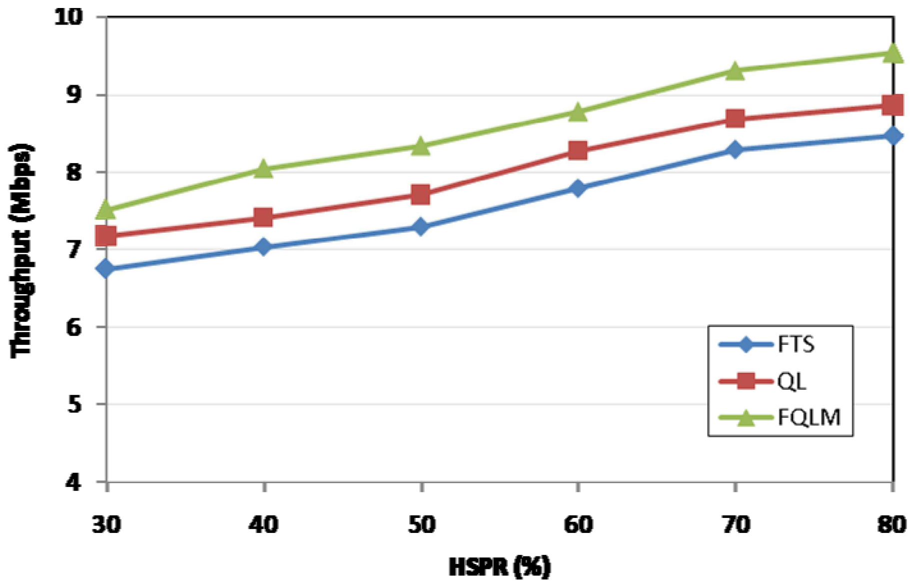

Figure 4.3: The system throughput versus HSPR with fixed user mobility at 60 km/hr.

In the following, we will discuss about the different user mobility with fixed HSPR. Fig 4.4 shows the BLER versus HSPR for FTS with different user mobility. There are two characters in this figure. As the user mobility increasing, the BLER will increase with the same HSPR. As the HSPR increasing, the difference of BLER between the different mobility will increase. When HSPR is more than 70%, the BLER of the same user mobility will not increase. This is due to the power achieve the saturation despite the distance between the user and the BS. On the other hand, the total number of MCS levels used in our system is 8 and the maximum MCS level will not reflect the channel quality. This limits the BLER performance. Therefore, we consider the effect of different user mobility with fixed HSPR at 60%.

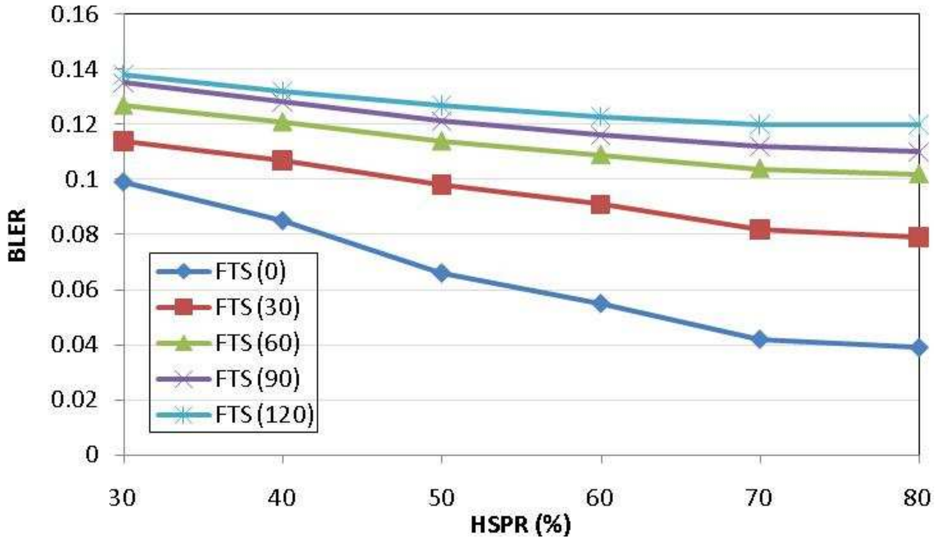

Figure 4.4: The BLER versus HSPR for FTS with different user mobility.

Fig 4.5 is the BLER versus different user mobility with fixed HSPR at 60%. When the user is motionless, FQLM-HARQ has the largest BLER than the other two schemes while not violating the BLER requirement. Because the channel quality is very similar between two adjacent TTI, the design of FQLM-HARQ will try to transmit the more aggressive MCS level that the channel condition can tolerate without dropping. When the user mobility is from 0 km/hr to 30 km/hr, the BLER of three schemes rush up owing to the path loss effect. The increasing range of FQLM-HARQ is the least and this means it can learning much well than the other two schemes. After 30 km/hr, the BLER of three schemes will increase due to the increasing channel variation.

Fig 4.6 is the dropping rate versus different user mobility with fixed HSPR at 60%. The increasing mobility will induce the larger dropping rate for three schemes. As the same reason, the higher channel variation will results unexpected channel condition at next TTI. Therefore, even the MCS level determination from learning will have higher probability to be dropped.

Fig 4.7 is the system throughput versus different user mobility with fixed HSPR at 60%. As the mobility during 0 km/hr to 30 km/hr, the purpose of MCS level selection in this region is to balance the BLER and throughput. When the mobility is more than 30 km/hr, the FQLM-HARQ is trying to reach the QoS requirement and will select the more conservative MCS level. After 90 km/hr, the throughput of three schemes almost not increase due to the learning speed can't catch up the channel variation.

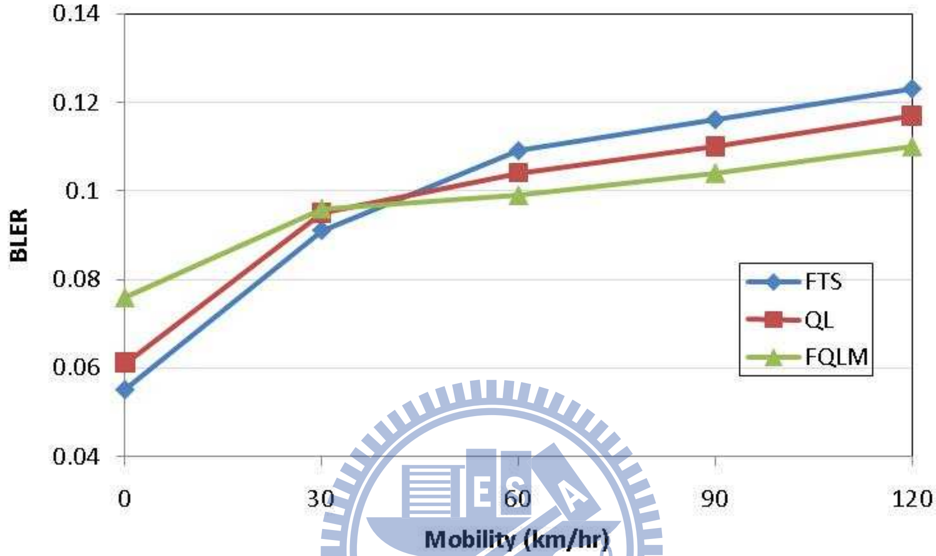

Figure 4.5: The BLER versus different user mobility with fixed HSPR at 60%.

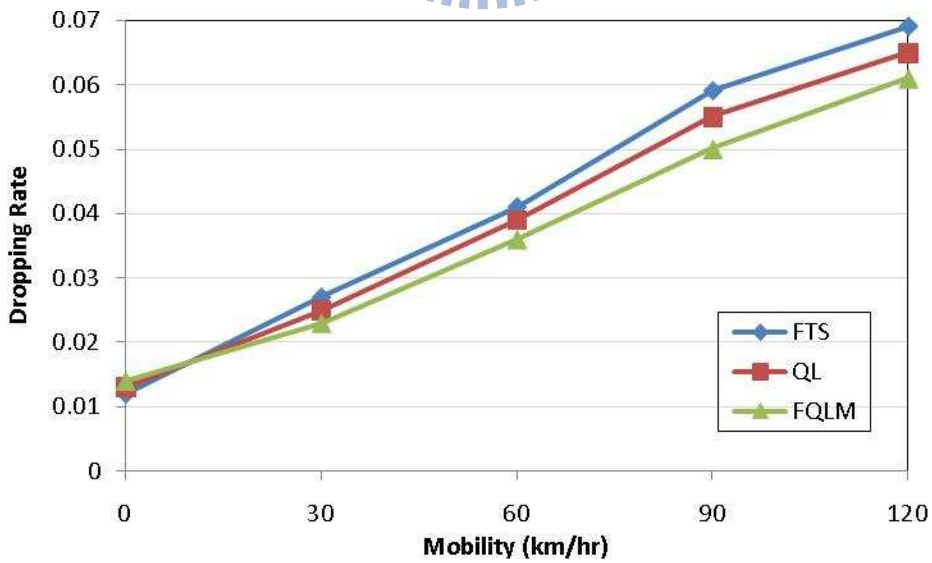

Figure 4.6: The dropping rate versus different user mobility with fixed HSPR at 60%.

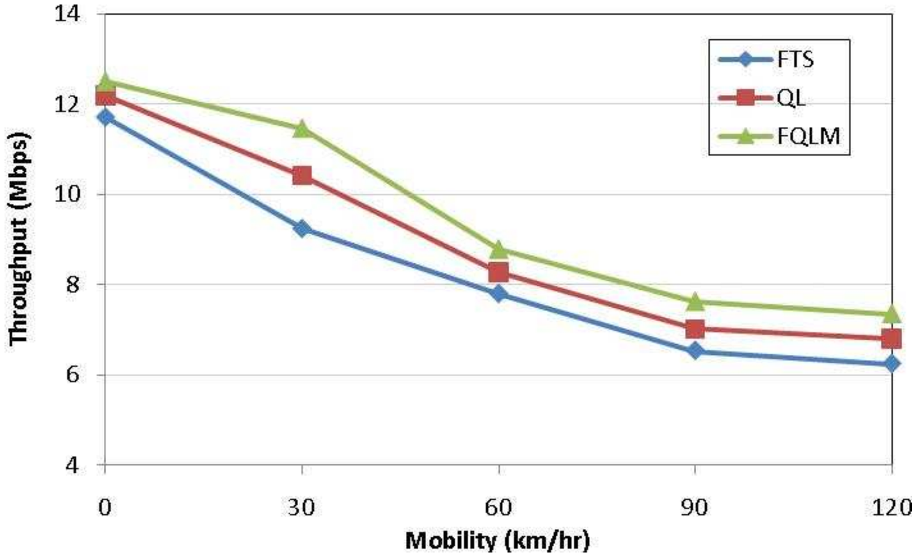

Figure 4.7: The system throughput versus different user mobility with fixed HSPR at 60%.

Chapter 5
Conclusion

Fuzzy Q-learning based MIMO HARQ scheme (FQLM-HARQ) is proposed to achieve efficient resource utilization in MIMO HSDPA system. The hybrid ARQ can be modeled as a Discrete-time Markov decision process (MDP). We apply fuzzy Q-learning algorithm to adjust the MIMO transmission mode and MCS level selection of initial transmission each TTI. The fuzzy rule is designed based on the channel quality indicator and BLER performance. The reinforcement signal is designed not only according to BLER performance but also considering the past transmission results. These can much help to accommodate the channel variation and the channel quality delay. By self-learning step by step, FQLM-HARQ can expect to maximize the system throughput while not violating the BLER requirement by choosing the optimal MCS level for each transmission.

From simulation results, with fixed user mobility, FQLM can have better performance than the other two schemes in different HSPR. Because of the BLER performance will affect the fuzzy rule base and then considering the different MCS level selection. When BLER is higher, the MCS level selection policy will become more conservative to satisfy the BLER requirement. On the other hand, when BLER is lower, the MCS level selection policy will become more aggressive to increase the system throughput. As the same reason, while in different user mobility with fixed HSPR, FQLM-HARQ also has more ability to resist the imprecise channel quality. Finally, the analysis shows that FQLM-HARQ scheme can achieve our designed target.

Bibliography

[1] 3GPP TS 25.214, "UMTS Physical layer procedures (FDD)" 3rd Generation Partnership Project, Tech. Rep., May 2008.

[2] D. G. Brennan, "Linear diversity combining techniques," *Proc. IRE*, vol. 91, no. 2, pp. 331-356, Feb. 2003.

[3] S. Parkvall and E. D. Hlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," *Proc. IEEE VTC-Spring* 2001, pp. 1829-1833.

[4] T. Cheng, "Coding performance of hybrid ARQ schemes," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1017-1029, Jun. 2006.

[5] M. Peng and W. Wang, "Advanced HARQ and scheduler schemes in TDD-CDMA HSDPA systems," *Proc. the 10th Asia-Pacific Conf. Commun. and 5th Int. Symposium on Multi-Dimensional Mobile Commun.*, vol. 1, pp. 67-70, Aug. 2004.

[6]. M. Nakamura, Y. Awad, and S. Vadgama, "Adaptive control of link adaptation for high speed downlink packet access in W-CDMA," *Wireless Personal Multimedia Commun.*, vol. 2, pp. 382-386, 2002.

[7] A. Muller and T. Chen, "Improving HSDPA link adaptation by considering the age of channel quality feedback information," *Proc. IEEE VTC-Fall* 2005, pp. 1643-1647.

[8] C. J. C. H. Watkins and P. Dayan, "Q-learning," Machine Learning, vol. 8, pp. 279-292, 1992.

[9] C. J. Chang, C. Y. Chang, and F. C. Ren, "Q-learning-based hybrid ARQ for high speed downlink packet access in UMTS," *Proc. of VTC-Spring* 2007, pp.2610-2615.

[10] G. J. Foschini, "Layered space-time architecture for wireless communication in fading

environment when using multiple antennas," *Bell Labs Syst. Tech. J.*, vol. 1, no. 2, pp. 41-59, Autumn 1996.

[11] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456-1467, Jul. 1999.

[12] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental trade off in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[13] R. T. Derryberry, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman, "Transmit diversity in 3G CDMA systems," *IEEE Commun. Magazine*, pp. 68-75, Apr. 2002.

[14] J. Peisa, S. Wager, M. Sågfors, J. Torsner, B. Göransson, T. Fulghum, C. Cozzo, and S. Grant, "High speed packet access evolution – concept and technologies," *Proc. IEEE 65th VTC-Spring* 2007, pp. 819-824.

[15] H. Holma, A. Toskala, K. Ranta-aho, and J. Pirskanen, "High-speed packet access evolution in 3GPP release 7," *IEEE Commun. Magazine*, pp. 29-35, Dec. 2007.

[16] T. Nihtil¨ and V. Haikola, "HSDPA MIMO system performance in macro cell network," *IEEE Sarnoff Symposium*, 2008.

[17] S. Semenov, "Modification of the D-TxAA scheme for fading channel," *Proc. of the Fourth Advanced Int. Conf. on Telecommun.*, 2008.

[18] E. N. Onggosanusi, A. G. Dabak, Y. Hui, and G. Jeong, "Hybrid ARQ transmission and combining for MIMO systems," *Proc. IEEE Int. Conf. Commun.*, vol. 5, pp. 3205-3209, May 2003.

[19] F. Wang, T. Peng, and W. Wang, "Multiple-input multiple-output system antenna subset selection with HARQ," *Proc. IEEE 18th Int. Symposium on PIMRC*, Sep. 2007.

[20] P. Y. Glorennec and L. Jouffe, "Fuzzy Q-Learning," *Proc. IEEE Int. Conf. Commun. on Fuzzy System*, vol. 2, pp. 659-662, 1997.

[21] Y. Maeda, “Fuzzy adaptive Q-Learning method with dynamic learning parameter,” *Proc. of IFSA World Congress and 20th NAFIPS International Conf.,* vol. 5, pp.2778-2780, 2001.

[22] M. Fei and S. L. Ho, “Progress in on-line adaptive, learning and evolutionary strategies for Fuzzy logic control,” *IEEE Int. Conf. on Power Electonics and Drive Systems,* vol. 2, pp.1108-1113, 1999.

[23] C. J. Chang, C. Y. Chang, and F. C. Ren, “Q-learning-based hybrid ARQ for high speed downlink packet access in UMTS,” *Proc. of VTC-Spring* 2007, pp. 2610-2615.

[24] L. Gordon, “Principle of Mobile Communication”, Kluwer Academic publishers, 2001.

[25] M. Gudmundson, “Correlation model for shadow fading in mobile radio systems,” *IEEE Electron. Lett*., vol. 27, pp. 2145-2146, Nov. 1991.

[26] B. Josef and F. Ivan, **“**CQI reporting for MIMO transmission in a wireless communication system” U.S. WO 6374 July 17, 2008.

Vita

Ying-Yu Chen was born in 1983 in Chunghwa, Taiwan. She received the B.E. degree in electrical and engineering from National Chung Hsing University, Taichung, Taiwan, in 2007 and M.E. degree in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2009, respectively. Her research interests include radio resource management and wireless communication systems.