

國立交通大學

電信工程學系

碩士論文

適用於 UMTS 高速下行封包擷取技術之 Q-Learning 式
混合自動重傳機制

Q-Learning-based Hybrid ARQ for High Speed
Downlink Packet Access in UMTS

研究生：張家源

指導教授：張仲儒 博士

中華民國 九十五年六月

適用於 UMTS 高速下行封包擷取技術之
Q-Learning 式混合自動重傳機制

Q-Learning-based Hybrid ARQ for High Speed
Downlink Packet Access in UMTS

研究生：張家源

Student: Chia-Yuan Chang

指導教授：張仲儒 教授

Advisor: Prof. Chung-Ju Chang

國立交通大學
電信工程學系
碩士論文



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
June 2006
Hsinchu, Taiwan

中華民國九十五年六月

適用於 UMTS 高速下行封包擷取技術之 Q-Learning 式混合自動重傳機制

研究生：張家源

指導教授：張仲儒

國立交通大學電信工程學系碩士班

摘要

為了在現有的 WCDMA universal mobile telecommunications system (UMTS) 下提供更高速、有效率、健全的下鏈路資料封包傳送，一種高速下行封包擷取技術 (high speed downlink packet access HSDPA) 被 3rd generation partnership project (3GPP) 所提出，且已經在 Release 5 被標準化。針對不同的連結，為了作到更好的適應性，HSDPA 除了動態的去調整使用不同的調變方式、不同的編碼速率 (AMC) 提供更多不同的傳輸速率；提供更大量的多碼 (multi-code) 使用運作；同時配合一種混合自動重傳機制 (H-ARQ)；以及大幅縮短成 2ms 的傳送時間間隔，希望去創造更有效率的資源分配。

在本篇論文中，我們提出了一個適用於 UMTS 高速下行封包擷取技術之 Q-learning 式混合自動重傳機制 (Q-HARQ)。我們先將整個 H-ARQ 的程序模擬成一種離散時間馬可夫決策過程 (Markov decision process MDP)，並將封包傳送所會付出的代價 (cost) 針對我們所希望去滿足的傳輸服務品質 (quality of service QoS) 的封包傳送錯誤率 (BLER) 作設計；再利用一種名為 Q-learning 的即時加強型學習演算法去估計每一次的傳送代價，不斷的去學習，針對每次封包的第一次傳送過程，來達到最佳且符合一定傳輸品質 (QoS) 傳送決策。

模擬結果顯示，我們所提出的方法可以在滿足我們所要求 BLER 下去選擇最佳的傳送決策。這意味著，針對封包的第一次傳送，我們能提供最有效率的傳輸方法以對抗變化劇烈的通道環境。另一方面，我們也證實了我們所提出的 Q-HARQ 機制在收斂時間，及處理運算時間均符合實際系統的要求，適用於現行的通訊系統。

Q-Learning-based Hybrid ARQ for High Speed Downlink Packet Access in UMTS

Student: Chia-Yuan Chang

Advisor: Chung-Ju Chang

Department of Communication Engineering
National Chiao Tung University

Abstract

WCDMA Release 5 has been standardized for universal mobile telecommunications system (UMTS) in the 3rd generation partnership project (3GPP), where high speed downlink packet access (HSDPA) is proposed to provide efficient, robust, and high-speed packet data services for UMTS. In HSDPA, the adaptive modulation and coding (AMC) technique and extensive multi-code operation are adopted for the link adaptation. Also, an advanced retransmission strategy based on hybrid automatic repeat request (H-ARQ) is proposed to upgrade the robustness against link adaptation errors. In this thesis, a Q-learning-based hybrid automatic repeat request (Q-HARQ) scheme for HSDPA in UMTS system is proposed to achieve efficient resource utilization. The Hybrid ARQ procedure is modeled as a discrete-time Markov decision process, where the transmission cost is defined in terms of the QoS parameters of transport block error rate for enhancing spectrum utilization subject to QoS constraint. The Q-learning reinforcement algorithm is employed to accurately estimate the transmission cost to perform the most suitable decision of modulation and coding scheme for the packet initial transmission while the requirement of transport block error rate is guaranteed.

Simulation results show that the QoS requirement of BLER for Q-HARQ is indeed fulfilled. In addition, the performance of the Q-HARQ can be improved under the specific QoS constraint of BLER. It is verified finally that the Q-HARQ scheme is feasible in the practical system.



誌謝

“呼”，兩年的碩士生活就這樣夾雜著許許多多的不同回憶而結束。能夠完成這篇碩士論文，要感謝很多人，感謝張仲儒教授對我的指導與教誨，引導我論文正確的研究方向；在我面對未知的未來，帶給我更多不同的經驗分享及應有的處世態度；同時在這兩年的碩士生涯裡，提供我一個完整、豐富、多元，令人依依不捨的實驗室環境；就像老師說的，“我不會忘記我們大家一起相處的 happy time，更會謹記跟老師一起做研究 hard time”。感謝芳慶學長總是在工作之餘，特地撥空指導我論文，討論、解答我研究過程所遇到的種種困難。感謝家慶、義昇、詠翰、文祥學長，總是適時針對我的論文給予我精闢的論點，及教導我如何從不同的觀點去剖析問題。再來，感謝無時無刻守護著實驗室的立峰學長，傳授我許多不同領域的知識技巧；總是細心的完美解決我們做論文時所遭遇到的總總阻礙。接著，感謝志明、朕逢、凱元、立忠、宗軒，在我還是輕澀懵懂的一年級研究生時，帶我快速的適應 701 這個大家庭，渡過豐富、精采的碩一生活。以及感謝在我寫論文的碩二生活，一起歡樂、打球的學弟妹，建安、建興、佳璇、佳泓、世宏、正昕、尚樺，當然還陪我一起成長，一起修課，一起上下工，一起打鬧的俊帆、琴雅、媛玉，很幸運這兩年有你們的陪伴。

最後，我要感謝我的爸爸、媽媽，謝謝你們給我這麼好的求學環境，讓我永遠沒有後顧之憂，盡情的走我自己想走的路；感謝我的姐姐、哥哥，一路伴我成長；還有我的好朋友們，因為你們，我的生活更多采多姿。當然，還有從大學以來一起互相扶持、關心、體貼我的女友，曉雯。

家源 謹誌

民國 九十五年八月

Contents

Mandarin Abstract	i
English Abstract	ii
Acknowledgements	iv
Contents	v
List of Figures	vii
List of Tables	ix
1 Introduction	1
2 System Model	7
2.1 HSDPA System Operation	7
2.2 The Q-HARQ for Packet Transmission in HS-DSCH	11
2.3 Propagation Model	12
3 Design of Q-HARQ	15
3.1 Markov Decision Process	15
3.2 State, Action, and Transmission Cost Function	16
3.3 Q-HARQ Scheme	21



3.4	Implementation Issue	25
4	Simulation Results and Discussion	28
4.1	System Environment and Simulation Parameters	28
4.2	Conventional Schemes	29
4.3	Performance Evaluation and Discussions	31
5	Concluding Remarks	42
	Vita	47



List of Figures

2.1	HSDPA protocol architecture	7
2.2	Release '99 and Release 5 HSDPA retransmission control in the network . . .	8
2.3	HS-SCCH and HS-DSCH timing relationship	10
2.4	The Q-HARQ for packet transmission in HS-DSCH	12
3.1	Block diagram of a learning system	16
3.2	The flow chart of the Q-HARQ procedure	19
3.3	Structure of the Q-learning-based H-ARQ scheme	24
4.1	The auto-correlation on shadow fading	30
4.2	The BLER with 0.1 BLER requirement for Q-HARQ	32
4.3	The BLER with 0.1 BLER requirement for Q-HARQ	32
4.4	The average number of transmission time with 0.1 BLER requirement for Q-HARQ	33
4.5	The average number of transmission time with 0.1 BLER requirement for Q-HARQ	34
4.6	The system throughput with 0.1 BLER requirement for Q-HARQ	35
4.7	The system throughput with 0.1 BLER requirement for Q-HARQ	35
4.8	The failure rate with 0.1 BLER requirement for Q-HARQ	36
4.9	The failure rate with 0.1 BLER requirement for Q-HARQ	37
4.10	The system throughput with 0.3 BLER requirement for Q-HARQ	39

4.11 The system throughput with 0.2 BLER requirement for Q-HARQ	39
4.12 The incremental change of Q-factors with 0.1 BLER requirement for Q-HARQ when the SIOR is equal to 15	40



List of Tables

4.1	Simulation parameters	29
4.2	Convergence Time	40
4.3	Processing Time	40



Chapter 1

Introduction

WCDMA Release 5 [1] has been standardized for universal mobile telecommunications system (UMTS) in the 3rd generation partnership project (3GPP), where high speed downlink packet access (HSDPA) is proposed to provide efficient, robust, and high-speed packet data services for UMTS. For HSDPA, a high speed downlink shared channel (HS-DSCH) is used among users with a fixed spreading factor. Differing from Release 99, fast power control and variable spreading factor are disabled, while adaptive modulation and coding (AMC) and extensive multi-code operation are adopted for the link adaptation in HS-DSCH [2]. Also, an advanced retransmission strategy based on hybrid automatic repeat request (H-ARQ) is proposed to upgrade the robustness against link adaptation errors. The retransmission blocks are controlled directly by the Node B, instead of the radio network controller (RNC), to accelerate the retransmission and reduce the transmission delay.

The largest difference between traditional ARQ and H-ARQ is that H-ARQ combines an extra forward error correction (FEC) mechanism with the original ARQ function to achieve more efficient channel usage and higher system throughput. There are two main kinds of schemes for implementing H-ARQ: chase combining (CC) and incremental redundancy (IR). In the chase combining scheme, the Node B will decide to retransmit the same packet according to feedback request of negative acknowledge (NACK) when a received packet cannot

be decoded successfully. The decoder combines these multiple copies of packets weighted by the received signal-to-noise ratio (SNR). In the IR scheme, the Node B executes each retransmission with different redundancy versions in accordance with the feedback channel quality indicator (CQI) when errors happen. The IR scheme needs more buffers at receiver and is more complicated than the chase combining scheme.

There is another classification for H-ARQ protocol, called type-I, type-II, and type-III H-ARQ. The conventional ARQ used in 3GPP Release 99 standard belongs to type-I H-ARQ, where the fixed coded packet is resent until the packet is decoded successfully and the previously received packet will be discarded at user equipment (UE). The type-II H-ARQ e.g. M-stage H-ARQ [3], is equivalent to the IR scheme. The type-III H-ARQ makes each retransmission packet self-decodable. The decoder at receiver combines these retransmitted packets weighted by the received channel quality while decoding the packet. Chase combining can be regarded as a kind of type-III H-ARQ with only one redundancy version. In another perspective, we also call the type-III H-ARQ partial IR scheme and the type-II H-ARQ full IR scheme.

In order to provide more flexible coding scheme for accommodating different channel conditions, concepts of rate-compatible punctured codes were researched by Hagenauer [4], and Mantha and Kschischang [5]. The rate-compatible punctured codes developed on various fundamental coding structures, such as the rate-compatible punctured convolution (RCPC) code and the rate-compatible punctured turbo (RCPT) code, are employed together with IR scheme, for example, the M-stage H-ARQ [3]. For a rate-compatible punctured code, the parity bits of a low-rate "mother" code, or say the parent code, are punctured to generate a family of higher rate codes which can be decoded by the same decoder. At the first

transmission, a high-rate error-detecting codeword called cyclic redundancy check (CRC) attachment packet is transmitted. If the higher rate code cannot be decoded successfully, the predetermined codeword with lower code rate from the acceptable set of punctured codewords is transmitted. By sharing a puncturing table with the transmitter, the receiver can simply insert erasures for all code symbols that have not yet been transmitted. Therefore, various code rates for a packet transmission can be decoded through the same one decoder.

As for the AMC in HSDPA, its idea is to adapt the transmission to the fast varying channel quality. An adaptive scheduling algorithm was proposed in [6] for the AMC technique in HSDPA with multi-code transmission. From the viewpoint of link adaptation (LA), the joint consideration of the number of multi-codes and the AMC scheme can reach the high throughput under a requirement of the maximum tolerable frame error rate. A protocol which combines advantages of LA and IR was proposed in [7]. Unlike the fixed starting code rate in the conventional IR, two LA_IR protocols which adaptively choose the initial transmission code rate in accordance with the current channel condition to minimize the number of retransmission and maintain a desirable throughput. These LA_IR schemes usually achieve a higher effective throughput than the protocol [8] that only considered the LA. On the other hand, in order to achieve more efficient channel utilization, the method of that both the initial code rate and the mother code rate can be adaptively selected was proposed in [9].

Moreover, an adaptive incremental redundancy (AIR) algorithm was proposed in [10], where an information-theoretic model was developed to explain and predict the coding gains of the H-ARQ scheme. A bit-interleaved coded modulation (BICM) capacity, which is the function of the SNR and the modulation order, is adopted to calculate the accumulated conditional mutual information (ACMI) for a packet transmission. On the other hand, the

target data rate (bits per symbol) of the initial transmission is used for the basis of the outage event. While the SNR is known, the number of transmission symbols can just be chosen such that the accumulated information for this packet transmission will reach the target data rate. Through assigning adaptive code rate and modulation order for the retransmission packet, the channel resource can be efficiently utilized. However, it is a pity that it needs a modulation and coding schemes (MCS) table for the initial transmission in advance and it causes the system more complicated on the account of the BICM based on the information theory. Also, it is impractical because only single cell environment was considered.

Besides, a reliability-based H-ARQ was presented [11]. Taking the advantage of property of soft-input, soft-output (SISO) decoder, there were some methods further proposed [12, 13, 14, 15]. These methods determined the size of sequential retransmissions based on the reliability, which is defined as the magnitude of the log-likelihood ratios, of the received information bits. It was shown in [12] that the average value of the soft output is a good indicator of how many bits within the transmitted packet are in error. Employing the estimated mean of the soft output values, the variable retransmission size can be determined appropriately to target the bits of low reliability. The proposed reliability-based HARQ scheme [13] maximized the throughput subject to a maximum packet delay constraint. Utilizing the redefined reliability for a received packet, the successive retransmission size is adaptively chosen to achieve the target reliability given the average signal-to-noise ratio at the input to the decoder. In [14], the scheme not only minimized the retransmission size through aiming at those error bits but also tried to reduce the retransmission request packet according to the time correlation property in the output of the decoder for a convolution code. In order to avoid the large number of request packets, there are two operations, elimination and smoothing, performed to mark the window with errors in it. Instead of sending the total

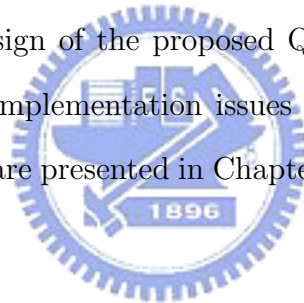
indices of unreliable bits in [12], only the first and last bit indices of each window need to be retransmitted. Another reliability-based H-ARQ in [15] used the source coding to achieve the small request size. In [16], the retransmission signal constellations were rearranged so that the reliability of each bit inputting to the turbo decoder would be averaged and improved to achieve an outstanding performance gain.

Except for the skills mentioned above, maximizing spectrum utilization while meeting QoS constraints suggests a constrained Markov decision process (MDP) [17]. This methodology has been successfully applied to solve many network control problems. However, they require an extremely large amount of state spaces to exactly model these problems. Consequently, the numerical computation is intractable due to the curse of dimensionality. Also, a priori knowledge of state transition probabilities is required. Alternatively, many researchers turned to adopt the reinforcement learning (RL) algorithm to solve the large state space problems [17, 18, 19, 20]. The most obvious advantage of RL algorithm is that it could approach an optimal solution from the on-line operation if the RL algorithm is converged.

In this thesis, we propose a Q-learning-based hybrid automatic repeat request (Q-HARQ) scheme for HSDPA in UMTS system to achieve efficient resource utilization. By improving the initial packet transmission under a QoS requirement of transport block error rate (BLER), Q-HARQ decreases the number of transmission time due to the good link adaptation for the initial transmission. Moreover, because the IR technique is combined together, the enhancement of the average system throughput will be expected. On the other hand, the Q-HARQ can efficiently combat the channel estimation errors through the converge property of Q-learning algorithm. The good performance would be maintained even the estimation error exists. Since WCDMA is an interference-limited system, interference profile is chosen

as the system state. The transport format for the initial packet transmission is the main action in the formulated H-ARQ process. An evaluation function is defined to appraise the cumulative discounted cost of the consecutive decisions for the Q-HARQ. Without knowing the state transition behavior, the evaluation function is calculated by a real-time RL technique known as Q-learning [21], [22]. After a decision is made, the consequent cost is used as an error signal feedback to the Q-HARQ control to adjust the state-action pairs. Thus, the learning procedure is performed in a closed-loop iteration manner which will help the value of evaluation function converge to the optimal solution point.

The remainder of this thesis is organized as follows. In Chapter 2, HSDPA system operation is outlined, and the system model, including the propagation model, is presented. The principal concept for the design of the proposed Q-HARQ is described in Chapter 3. Also, there are some significant implementation issues discussed in this Chapter. Finally, simulation results and discussion are presented in Chapter 4, followed by concluding remarks and future work in Chapter 5.



Chapter 2

System Model

2.1 HSDPA System Operation

Both the physical layer and the medium access control (MAC) specifications for UMTS HSDPA are defined by 3GPP. Fig. 2.1 shows the different protocol layers in HS-DSCH for HSDPA. For the HS-DSCH, an additional intelligence over the HSDPA MAC layer is installed in Node B. This key functionality of the new Node B MAC functionality (MAC-hs) is to handle the ARQ functionality and several scheduling procedures. The retransmission procedure can be controlled directly by Node B instead of serving RNC. Fig. 2.2 shows the difference of the retransmission control in the network between Release '99 and Release 5.

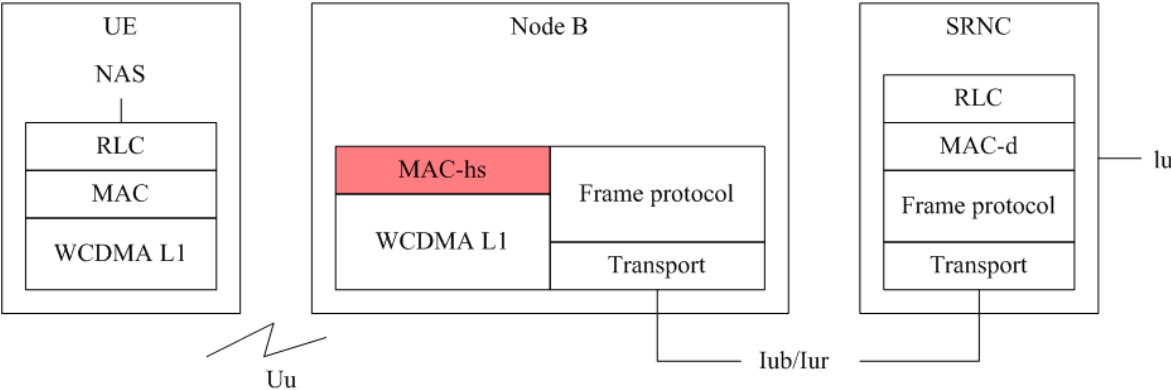


Figure 2.1: HSDPA protocol architecture

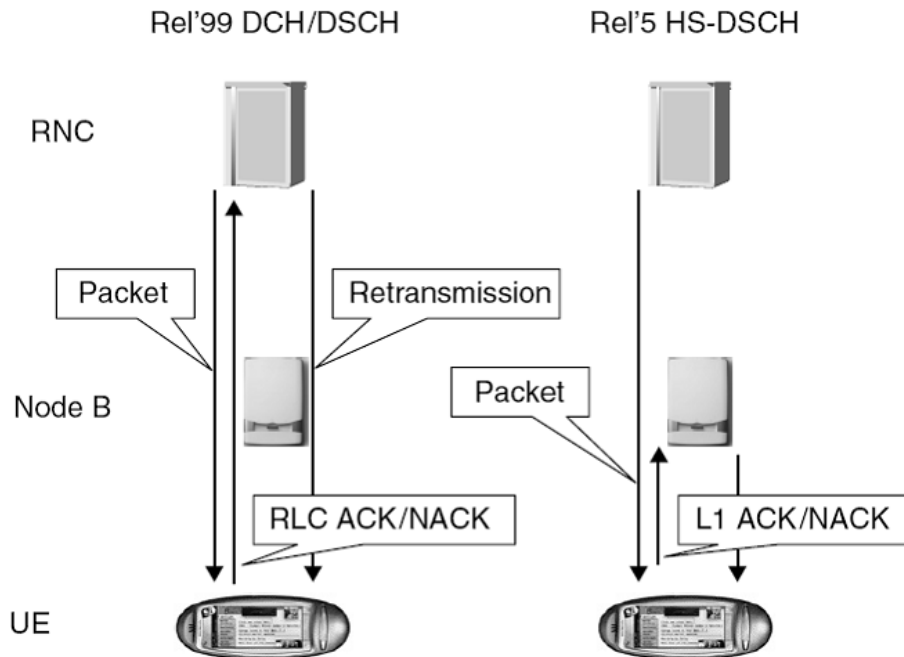


Figure 2.2: Release '99 and Release 5 HSDPA retransmission control in the network

The leading benefit is the faster retransmission and thus the shorter delay when retransmissions are needed. For the Iub interface between Node B and RNC in HSDPA, however, a flow control mechanism is required to control Node B buffers such that no data loss due to Node B buffer overflow. Even though the new MAC-hs is added at Node B, the serving RNC still reserve the inherent functionalities in Release '99 and Release 4.

In order to carry the feature for HSDPA, there are three new channel introduced in the physical layer specifications, as follows:

- *High-Speed Downlink Shared Channel (HS-DSCH)*

The HS-DSCH carries the user data in the downlink direction and code multiplexing is adopted such that it can be shared among scheduled users in a transmission time interval (TTI). In HSDPA, the TTI is defined to be 2ms, as well as three time slots, to achieve a short round trip delay for transmission. A higher order modulation scheme,

16-QAM, is added for improving the instantaneous peak data rate. The fixed spreading factor, 16, is used in HS-DSCH. Based on the spreading factor 16, the total number of channelization codes is 16. However, it is necessary to reserve an available code space for common channels, HS-SCCHs and the associated DCH, the maximum number of codes that can be allocated is 15. Also, the assigned number of multicode for users depends on the terminal capacity individually. A maximum of 5, 10 or 15 codes may be received for a specific terminal.

- *High-Speed Shared Control Channel (HS-SCCH)*

The high-speed shared control channel carries the important control information for HS-DSCH demodulation. The timing relationship between HS-DSCH and HS-SCCH is illustrated in Fig. 2.3. For a three slot HS-SCCH transmission block, it is divided into two parts which provide different signalling information, individually. The first slot, the first part, carries the time-critical message that is prepared to start the demodulation process in anticipative time to avoid packet buffering. For example, the modulation and coding information for the HS-DSCH data are included in the first part. The second part, the proceeding next two slots, involves less time-critical factor including the H-ARQ process information and the CRC check of the HS-SCCH. On the other hand, the additional terminal-specific masking is adopted for allowing the terminal to check whether the data and the control information is intended for it.

- *Uplink High-Speed Dedicated Physical Control Channel (HS-DPCCH)*

For the uplink HS-DPCCH, it is utilized to carry both the ACK/NACK information for the retransmission procedure in physical layer and the quality feedback information used in Node B scheduler to determine which terminal to transmit and at which data rate. It is also separated into two partition, the ACK/NACK information and the feedback channel quality indicator (CQI) information. The UE measures the downlink

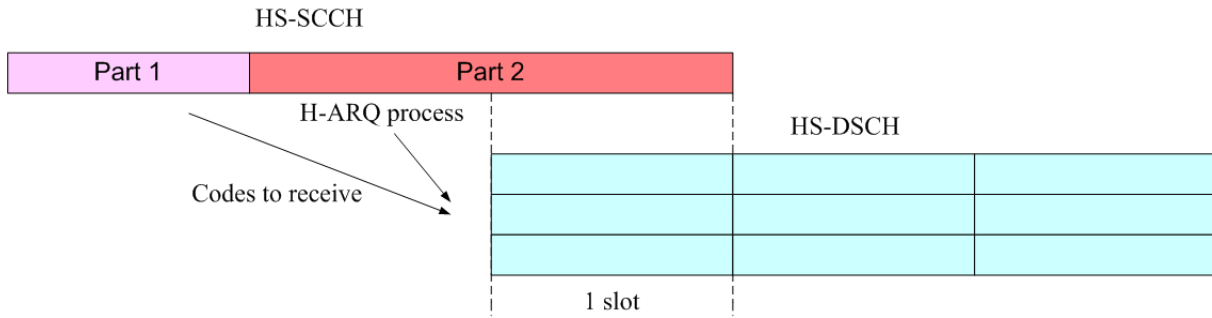


Figure 2.3: HS-SCCH and HS-DSCH timing relationship

channel quality and selects a proper CQI value corresponding to the combination of modulation, transport block size and number of parallel codes for HS-PDSCH. Also, the CQI value should be the one that the UE can expect a 0.1 block error probability under the measured channel quality. It is just a recommendation for the Node B scheduling.

At last, the detail HSDPA physical layer operation is encapsulated as the following steps:

- The scheduler in the Node B evaluates for different users what the channel conditions are, how much data is pending in the buffer for each user, for which users retransmissions are pending and how much time has elapsed since a particular user was last served and so forth.
- Once a terminal has been determined to be served in a particular TTI, the Node B identifies the necessary HS-DSCH parameters for example, how much codes are available or which kind of modulation order can be used and what are the terminal capability limitations.
- The Node B starts to transmit the HS-SCCH two slots ahead the corresponding HS-DSCH TTI to inform the terminal of the important messages. The HS-SCCH selection is free if there was no data for the terminal in the previous HS-DSCH frame.

- The terminal monitors the HS-SCCHs given by the network. While the terminal has decoded part 1 from an HS-SCCH intended for itself. This terminal will begin to decode the rest of the HS-SCCH and will buffer the necessary codes from the HS-DSCH.
- Until the HS-SCCH parameters has been decoded from part 2, the terminal can determine which H-ARQ process the data belongs to and whether it needs to be combined with data received previously in the soft buffer.
- Upon decoding the combined data, an ACK/NACK indicator will be sent in the uplink HS-DPCCH, depending on the outcome of the CRC check conducted on the HS-DSCH data.
- If the network continues to transmit data for the same terminal in consecutive TTIs, the terminal will stay on the same HS-SCCH used during the previous TTI.

2.2 The Q-HARQ for Packet Transmission in HS-DSCH

The Q-HARQ scheme in the HS-DSCH is shown in Fig. 2.4. The Q-HARQ scheme determines an optimal action which includes the effective code rate and the modulation order for the initial packet transmission according to a corresponding system state information. Also, the method of how to fragment arriving data packets is decided from Q-HARQ while the best suitable action is assigned. Before the encoded data packet is passed for interleaving, the two-stage rate matching is to match the physical layer transmission rate based on the determined action and to buffer the retransmission data packet for IR request. On the other hands, the CRC attachment and the interleaving technique are used for the error detecting and against the batch error. The turbo coding with a minimum 1/3 code rate is employed in the HS-DSCH.

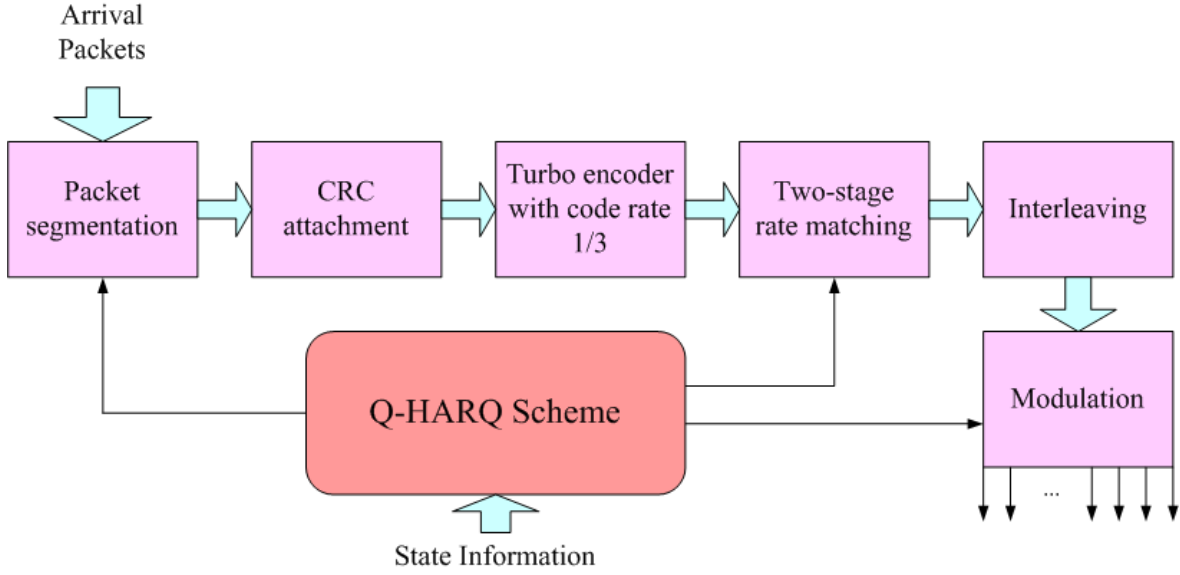


Figure 2.4: The Q-HARQ for packet transmission in HS-DSCH

2.3 Propagation Model

A terrestrial mobile radio channel for urban areas is considered in this thesis. For the terrestrial environments, the propagation effects are divided into three distinct types. These are path loss, slow variation about the mean due to shadowing and scattering, and the rapid variation in the signal due to multi-path effects. The model of channel fading for WCDMA cellular system is mainly determined by both the long-term fading and the short-term fading, which is presented by

$$F(t) = \xi(r) \times 10^{\eta/10} \times \zeta(t), \quad (2.1)$$

where $\xi(r) \times 10^{\eta/10}$ is the long-term fading including path loss and shadowing, r is the distance from the base station to the mobile station, and η is the normal-distributed random variable with zero mean and variance σ_L^2 . The short-term fading, $\zeta(t)$, caused by the multi-path transmission, is assumed to be the Jakes model [23], which is given by

$$\zeta(t) = 2\sigma \sqrt{\frac{2}{L}} \sum_{m=1}^M \cos(2\pi f_D t \cos(2\pi m/L) + \theta_m) e^{j\beta_m}, \quad (2.2)$$

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

$$\theta_m = \beta_m + 2\pi m s / (M + 1) \quad , s = 0, 1, 2, \dots M - 1. \quad (2.3)$$

Since the generated $\zeta(t)$ is mutually independent, this technique can produce up to M independent short-term fading. Therefore, we choose M equal to the number of total links in all cells of the system. Because it is reasonable to assume that the scattering geometry is time invariant within some small local area, we are able to assume further that parameters of the Jakes model are fixed in simulations.

On the other hand, the shadowing effect for a moving user will change with the user's location. For a practical system, however, the degradation degree of shadowing between two sampling time is small because of the limit of the mobility. In other words, the shadowing variation is correlated with the distance of two adjacent sampling point. In this paper, the normalized autocorrelation function $\rho(\Delta x)$ [24] is adopted for modeling the correlated shadow fading versus distance between two adjacent TTI. It can be described with sufficient accuracy by an exponential function as

$$\rho(\Delta x) = e^{-\frac{|\Delta x|}{d_{cor}} \ln 2}, \quad (2.4)$$

where d_{cor} is the decorrelation length.

Similarly, since the shadowing object is continuous in the real environment, the received signals from the same direction exist a high degree of correlation. At a specific time slot, the shadowing among the signals received from different base stations has a correlation, called cross-correlation on shadow fading. In order to describe the cross-correlation in the multi-cell system, the method [25] proposed by Viterbi is adopted in this thesis. The experienced

shadow effect for a user from different base stations links is divided into two components. One is the near field of the user that is common to all base stations, and the other that pertains solely to the receiving base station is independent from one base station to another.



Chapter 3

Design of Q-HARQ

3.1 Markov Decision Process

In the H-ARQ procedure, an objective what we want to reach is to find a suitable decision for an initial transmission of packets. According to the acquired information from the system environment, a proper policy can be obtained and, a expected QoS is guaranteed, simultaneously. Considering a learning system that interacts with its environment in the manner is illustrated in Fig. 3.1. The system operates in accordance with a finite, discrete-time Markov decision process that is characterized as follows:

- The environment evolves probabilistically occupying a finite set of discrete states. The state, however, does not contain past statistics, even though these statistics could be useful to the learning system. It means that the future of the process is completely summarized in the current state of the process, the Markov property.
- For each finite state, there is a finite set of possible actions that may be taken by the learning system.
- Every time the learning system takes an action, a certain cost is incurred.

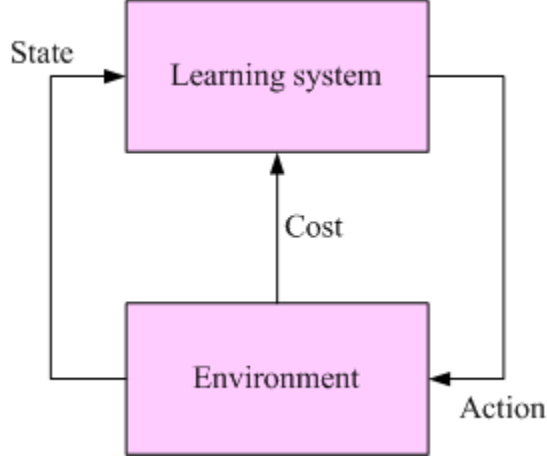


Figure 3.1: Block diagram of a learning system

- States are observed, actions are taken, and cost are incurred at discrete time.

Based on the characteristic of MDP, the estimated downlink channel quality can be seen as the system state for the H-ARQ procedure. The transport format for the packet initial transmission is regarded as the action of the system. The transmission cost is defined in terms of the QoS parameters for enhancing spectrum utilization subject to QoS constraint.

3.2 State, Action, and Transmission Cost Function

The H-ARQ procedure for HSDPA in a multimedia WCDMA system is modeled as a discrete-time MDP. We define the system state at the beginning of the initial transmission of the k -th packet for the H-ARQ process, denoted by x_k , as

$$x_k = (\widehat{SINR}_k), k = 0, 1, 2, \dots \quad (3.1)$$

where the \widehat{SINR}_k is the predicted downlink signal-to-interference-and-noise ratio (SINR) value performed at Node B for the k -th state. The \widehat{SINR}_k is classified into several degrees, and it is intuitively known that the transmitter should send a packet with more redundancy when \widehat{SINR}_k is smaller.

Based on the system state x_k , the action for the initial transmission of the k -th packet, denoted by A_k , mainly contains two parts: the effective code rate $(ECR)_k$ and the modulation order M_k . The action A_k is expressed as

$$A_k = [(ECR)_k, M_k]. \quad (3.2)$$

The $(ECR)_k$ is defined as the ratio of the number of bits going into the turbo encoder to the number of bits going out the two-stage rate matching at the k -th packet initial transmission, and it is designed to be in five redundancy versions which are

$$\left\{ 1, \frac{3}{4}, \frac{2}{3}, \frac{1}{2}, \frac{1}{3} \right\}, \quad (3.3)$$

where $1/3$ is the specified lowest available code rate for the turbo encoder. The M_k is the option of QPSK or 16-QAM. After an A_k is decided for the initial transmission of the k -th packet, a suitable packet size can be subsequently obtained under the known fixed spreading factor. Then the fragment, as well as the packet segmentation shown in Fig. 2.4, is implemented for the exact packet block size in order to reach more efficient channel utilization. The action can be seen as a preprocess for the initial packet transmission. If the packet of the first transmission cannot be received successfully, the effect code rate for the packet retransmission will be increased by one redundancy version until the $1/3$ effective code rate is reached, and the same modulation scheme with the next state is assigned for retransmission to accommodate the initial transmission of the next initial packet transmission.

If the state-action pair (x_k, A_k) has been determined, an immediate transmission cost function is defined as the square of the normalized difference between the received SINR and the desired SINR for the k -th packet initial transmission. It is given by

$$C(x_k, A_k) = \left[\frac{SINR(x_k, A_k) - \widetilde{SINR}(A_k)}{\widetilde{SINR}(A_k)} \right]^2, \quad (3.4)$$

where $SINR(x_k, A_k)$ is the *received* SINR at the mobile station for the state x_k with action A_k and $\widetilde{SINR}(A_k)$ is the *desired* SINR with action A_k . The $\widetilde{SINR}(A_k)$ is the required SINR received at mobile station under a QoS requirement of the block error rate based on the current action A_k . As for the characteristic of Q-learning algorithm, the objective to minimize the cost function for the Q-HARQ is to make the $SINR(x_k, A_k)$ most close to the $\widetilde{SINR}(A_k)$. In other words, there are not too many unnecessary packet redundancy occupying the channel resource, and it does not happen that the inadequate packet redundancy cannot accommodate channel conditions, either. The most appropriate packet size depending on the decided action would be transmitted for a specific initial packet transmission state such that the received SINR value will optimally approach to the desired SINR value while satisfying the maximum block error rate bound. Therefore, the most efficient resource allocation for the initial packet transmission can be achieved. The average system throughput must be improved and the expected QoS of block error rate can be guaranteed. On the other hand, the enhancement of the average system throughput implies that the average packet delay, the number of transmission time, can be reduced while a fixed amount of packets needed to be sent in the long-term viewpoint. From another viewpoint, usually, packets transmissions with a little larger BLER can reach much higher throughput than that with smaller BLER. Bartering for data rates with transmission times achieves a good trade-off between the system throughput and the transmission delay. Based on the $\widetilde{SINR}(A_k)$ with the 0.1 BLER requirement in HSDPA Release 5 standard, the best throughput can be achieved while the system BLER is approach to 0.1. In summary, maximizing the initial packet transmission utilization benefits the average system throughput, and the average packet delay also can be effectively diminished.

Fig. 3.2 illustrates the entire H-ARQ process associated with the Q-learning algorithm

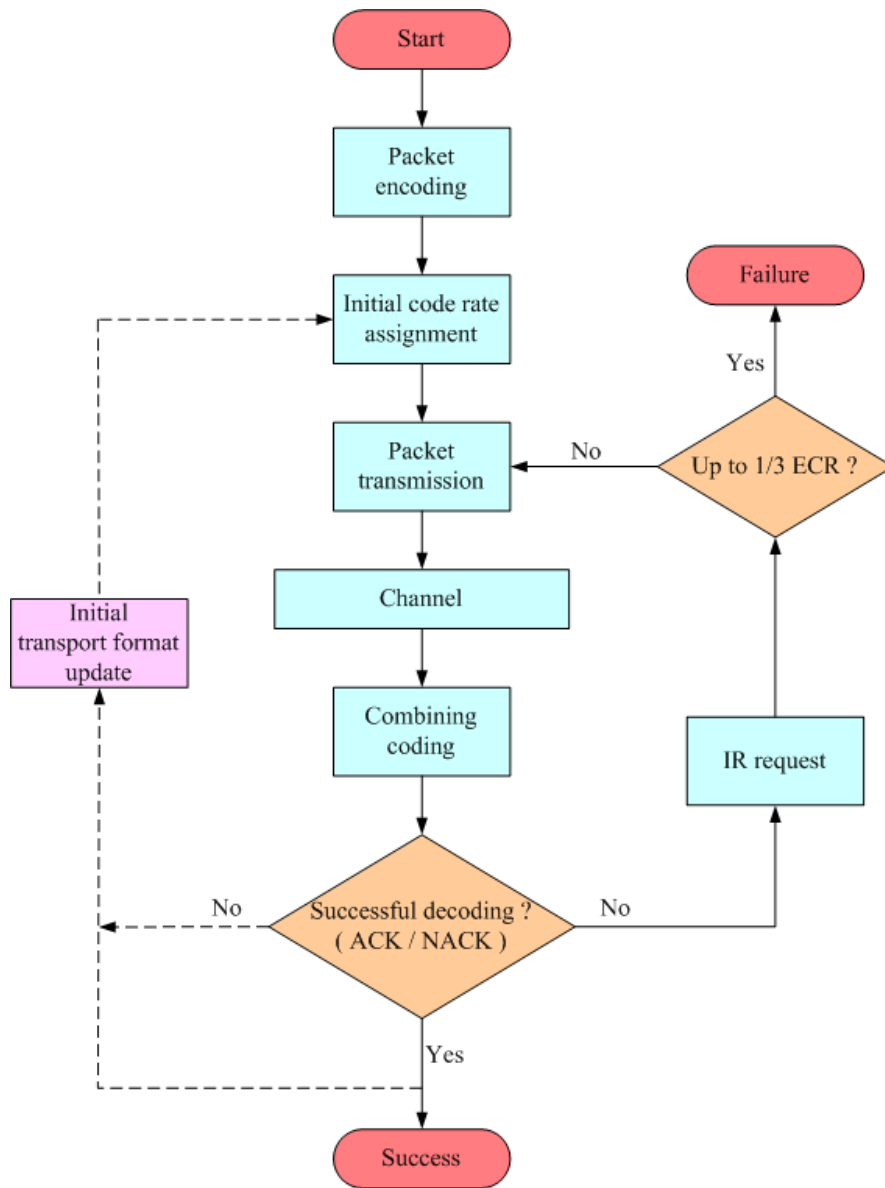


Figure 3.2: The flow chart of the Q-HARQ procedure

on the packet initial transport format selection. For the Q-HARQ scheme, whether the initial transmitted packet is successfully decoded or not, the state and cost information for updating the initial transport format based on the Q-learning algorithm will be feeded back to Node B. The portion with dotted lines represents these control plane processes. Especially speaking, once the retransmission packet still can not be successfully decoded while the effective code rate is up to 1/3, the 'failure' event happens. It means that the H-ARQ process will be restarted to transmit this failure packet again.

We further define an evaluation function, denoted by $Q(x, A)$, namely Q-function or Q-factor or Q-value, as the expected total discounted cost counting from the initial state-action pair (x, A) over an infinite time. It is given by

$$Q(x, A) = E \left\{ \sum_{k=0}^{\infty} \gamma^k C(x_k, A_k) | x_0 = x, A_0 = A \right\}, \quad (3.5)$$

where $E\{\cdot\}$ is the expectation operator and $0 \leq \gamma < 1$ is a discounted factor. The Q-HARQ scheme is to determine an optimal action, denoted by A^* , which corresponds to the minimum Q-function with respect to the current state. The minimization of Q-function implies the maximization of the system throughput and the fulfillment of QoS requirements.

Let $P_{xy}(A)$ be the transition probability from state x to the next state y on account of action A . Then $Q(x, A)$ can be expressed as

$$\begin{aligned} Q(x, A) &= E \{ C(x_0, A_0) | x_0 = x, A_0 = A \} + E \left\{ \sum_{k=1}^{\infty} \gamma^k C(x_k, A_k) | x_0 = x, A_0 = A \right\} \\ &= E \{ C(x, A) \} + \gamma \sum_y P_{xy}(A) \times E \left\{ \sum_{k=1}^{\infty} \gamma^{k-1} C(x_k, A_k) | x_1 = y, A_1 = B \right\} \\ &= E \{ C(x, A) \} + \gamma \sum_y P_{xy}(A) Q(y, B). \end{aligned} \quad (3.6)$$

From Eq. (3.6), it implies that the Q-function of the current state-action pair can be represented in terms of the expected immediate cost of the current state-action pair and the

Q -function of the next state-action pairs.

3.3 Q-HARQ Scheme

Based on the principle of Bellman's optimality [26], the optimal action A^* can be obtained with two-step optimality operation. The first step is to find a local minimum for the $Q(x, A)$, denoted by $Q^*(x, A)$. The intermediate evaluation function for each possible next state-action pair (y, B) is minimized while the optimal action B is performed with respect to every next state y . Thus the $Q^*(x, A)$ can be obtained by

$$Q^*(x, A) = E \{C(x, A)\} + \gamma \sum_y P_{xy}(A) \left\{ \min_B [Q^*(y, B)] \right\} \quad \text{for all } (x, A) \quad (3.7)$$

Then the optimal action A^* with respect to the current state x has to be determined for the next step such that $Q^*(x, A)$ is minimized. The minimum evaluation function for the state-action pair (x, A^*) can be expressed as

$$Q^*(x, A^*) = \min_A [Q^*(x, A)]. \quad (3.8)$$

However, it is difficult to get the $E \{C(x, A)\}$ and $P_{xy}(A)$ for solving the Eq. (3.7). Thus, we adopt a real-time reinforcement learning algorithm, named Q-learning algorithm [21], [22], to find the optimal resource allocation without a priori knowledge of $E \{C(x, A)\}$ and $P_{xy}(A)$. In order to get the optimal $Q^*(x, A)$, the Q-learning algorithm computes the Q value in a recursive method by using available information $(x, A, y, C(x, A))$, where x (y) is the current (next) state; A and $C(x, A)$ are the action for current state and its immediate cost of the state action pair, respectively.

A value iteration algorithm [22] was formulated in terms of the Q-factors to solve this

linear system of equations. Thus, for one iteration, it is given as

$$Q(x, A) = E \{C(x, A)\} + \gamma \sum_y P_{xy}(A) \left\{ \min_B [Q(y, B)] \right\} \quad \text{for all } (x, A) \quad (3.9)$$

The small step-size version of this iteration is described by

$$Q(x, A) = (1 - \eta)Q(x, A) + \eta \{E \{C(x, A)\} + \gamma \sum_y P_{xy}(A) \left\{ \min_B [Q(y, B)] \right\}\} \quad \text{for all } (x, A) \quad (3.10)$$

where η is a small leaning rate parameter that lies in the range $0 < \eta < 1$. As it stands, an iteration of the value iteration algorithm described in Eq. (3.10) requires knowledge of the transition probabilities. The need for this prior knowledge may be eliminated by formulating a stochastic version of Eq. (3.10). Specifically, the averaging performed in an iteration of Eq. (3.10) over all possible states is replaced by a single sample, thereby resulting in the following update for the Q-factor:

$$Q_{n+1}(x, A) = (1 - \eta_n(x, A))Q_n(x, A) + \eta_n(x, A) \left[C(x, A) + \gamma \min_B [Q(y, B)] \right] \quad \text{for all } (x, A) = (x_n, A_n) \quad (3.11)$$

where j is the successor state, and $\eta_n(x, A)$ is the learning rate parameter at time step n for the state-action pair (x, A) . For all other admissible state-action pairs, however, the Q-factors remain unchanged as shown by

$$Q_{n+1}(x, A) = Q_n(x, A) \quad \text{for all } (x, A) \neq (x_n, A_n) \quad (3.12)$$

Equations (3.11) to (3.12) constitute one iteration of the Q-learning algorithm. Moreover, for the state-action pair (x_n, A_n) , the Eq. (3.11) can be rewritten in the equivalent form

$$Q_{n+1}(x_n, A_n) = Q_n(x_n, A_n) + \eta_n(x_n, A_n) \left[C(x_n, A_n) + \gamma \min_B [Q_n(y_n, B)] - Q_n(x_n, A_n) \right] \quad (3.13)$$

Treating the expression inside the square brackets on the right-hand side of Eq. (3.13) as the error signal involved in updating the current Q-factor $Q_n(x_n, A_n)$, the target Q-factor at time step n may be identified as

$$Q_n^{target}(x_n, A_n) = C(x_n, A_n) + \gamma \min_B [Q_n(y_n, B)] \quad (3.14)$$

where $y_n = x_{n+1}$ is the successor state. It should be noted that the successor state j_n plays a critical role when determining the target Q-factor. Based on the definition of the target Q-factor, the Q-learning algorithm can be reformulated as

$$Q_{n+1}(x, A) = Q_n(x, A) + \Delta Q_n(x, A) \quad (3.15)$$

The incremental change in the current Q-factor is represented as

$$\Delta Q_n(x, A) = \begin{cases} \eta_n (Q_n^{target}(x, A) - Q_n(x, A)) & , \text{ for } (x, A) = (x_n, A(n)) \\ 0 & , \text{ otherwise} \end{cases} \quad (3.16)$$

As the description in Eq. ([?]), the optimal action A_n at current state x_n is the unique action at the state for which the Q-factor at time step n is minimum. Therefore, once the optimal Q-values are available, an optimal policy $A^*(x)$ at state x can be determined with relatively little computation as

$$A^*(x) = \arg \min_A [Q^*(x, A)] \quad (3.17)$$

Fig. 3.3 shows the structure of the Q-learning-based H-ARQ scheme. When there are packets for transmission at system state x , the *Q-function computation* block computes the value of $Q(x, A)$ for every possible action A . The *transport format selection* block then determines the optimal packet initial transmission format A^* among all the current Q values of all possible actions. Afterwards, the immediate cost $C(x, A^*)$ can be observed and the value of $Q(x, A)$ is adjusted based on the Q-learning rule and is updated every time when the corresponding state-action pair happens. The Q-learning rule is formulated as

$$Q(x, A) = \begin{cases} Q(x, A) + \eta \Delta Q(x, A) & , \text{ if } A = A^* \\ Q(x, A) & , \text{ otherwise} \end{cases} \quad (3.18)$$

Predicted Channel Quality

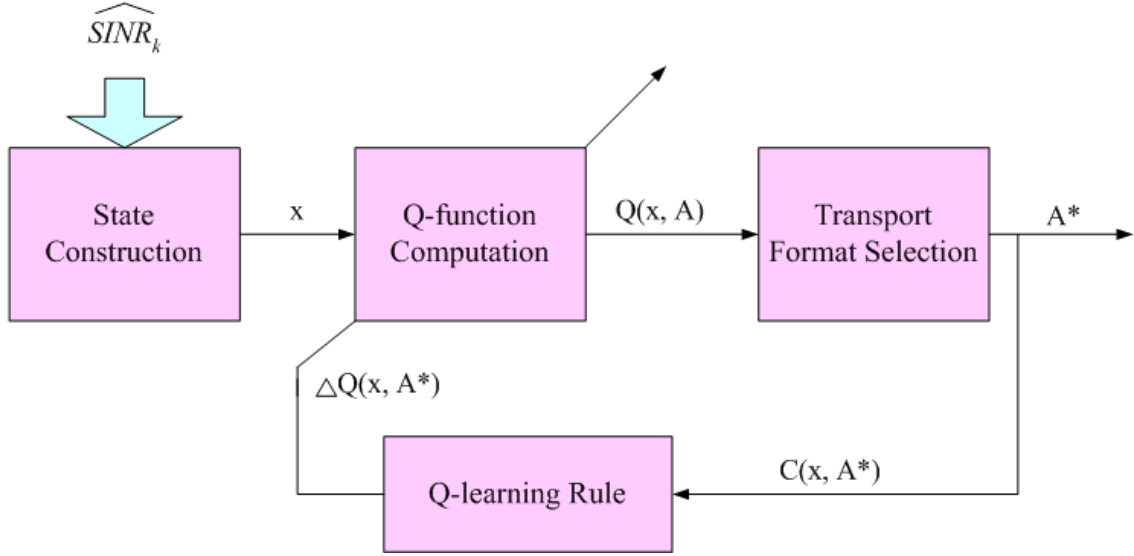


Figure 3.3: Structure of the Q-learning-based H-ARQ scheme

and

$$\Delta Q(x, A^*) = \left\{ C(x, A^*) + \gamma \min_B [Q(y, B)] \right\} - Q(x, A^*) \quad (3.19)$$

From the Eq. (3.18), only the Q value for the selected state-action pair is updated while others are kept unchanged. In other words, only one state-action pair is chosen for evaluation in each learning epoch. On the other hand, the operation of $\min_B [Q(y, B)]$ is executed by comparing the Q value of all the possible action candidates for state y in Eq. (3.19) and then choosing the desired action B which has the minimum Q value.

For the Q-learning algorithm, the convergence theorem had been proven by Watkins and Dayan in [21]. The theorem is here restated as follows: *if the value of each admissible pair is visited infinitely often and the learning rate is decreased to zero in a suitable way, then the value of $Q(x, A)$ in Eq. (3.18) will converge to $Q^*(x, A)$ with probability 1.*

3.4 Implementation Issue

Before the Q-HARQ is performed for the on line operation, a proper assignment of initial values is necessary. In this paper, however, the initialization for Q values is assigned arbitrarily. For channel states, we adopt directly the received SINR measurement of this state at mobile station as the predicted downlink channel state of the next state without using any channel predictor at Node B. It is reasonable because the channel condition between two transmission time interval is correlated.

For the designed cost function, the desired SINR, $S\widetilde{I}N\widetilde{R}(A_k)$, for k -th state is a significant topic needed to be discussed. First, in order to find a relation between BLER and defined system actions, the channel profile received at the mobile terminal is assumed to be an additive white Gaussian noise (AWGN) channel. Even though the assumption of AWGN channel may be not practical in real world environment, it indeed provides a meaningful speculation to approximate the realistic channel. The Gray encoding is used for signal mapping and we assume that errors only happen within neighbor transmitting signals in the signal constellation. It means that there is only one bit error for a symbol error such that the symbol error rate is equal to $\log_2 M$ times the bit error rate, where M is the modulation order. Also, we suppose that each symbol energy is the same as the average symbol energy. Therefore, two equations which describe the relation of bit error rate, denoted by P_e , and E_b/N_0 for different modulation order are derived. It is given by

$$P_e = \begin{cases} \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) & , \text{ for QPSK} \\ \frac{3}{4}p - \frac{9}{16}p^2 & , \text{ for 16-QAM} \end{cases} \quad (3.20)$$

where $p = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{2E_b}{5N_0}}\right)$ and $\text{erfc}(\cdot)$ is a complementary error function described as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \quad (3.21)$$

Moreover, Since $\frac{E_b}{N_0} = SINR \cdot PG$, where PG is the processing gain and is equal to 16 in HSDPA, the Eq. (3.20) can be rewritten as

$$P_e = \begin{cases} \frac{1}{2} \text{erfc} \left(\sqrt{SINR \times 16} \right) & , \text{ for QPSK} \\ \frac{3}{4}p - \frac{9}{16}p^2 & , \text{ for 16-QAM} \end{cases} \quad (3.22)$$

where $p = \frac{1}{2} \text{erfc} \left(\sqrt{SINR \times \frac{32}{5}} \right)$. For simplicity, we assume that the transport block will be regarded as an unsuccessful transmission while any one bit error of the block happens. Over the statements mentioned above, the desired SINR value, \widetilde{SINR} , without using any channel coding technique can be easily acquired based on the distinct modulation order from Eq. (3.22) while the requirement of the block error rate is determined. Second, in order to model the effect of the error correcting code, we make an assumption that the relation between the reciprocal of ECR and the received SINR gain is linear. In other words, the performance for a code rate $\frac{1}{3}$ is equal to the uncoded performance with three times received SINR. Based on the Eq. (3.22) and the error correcting code model, the $\widetilde{SINR}(A_k)$ for k -th system state can be found under satisfying a QoS requirement of block error rate while the effective code rate and modulation are selected. Finally, the computation of the complementary error function in Eq. (3.21) is directly found by looking up the complementary error function table.

In summary, the procedure of Q-HARQ is implemented iteratively as the following four steps.

Step 1: [*State-Action Construction*]

Construct the state $x_k = (\widehat{SINR}_k)$ of the k -th packet initial transmission and find a set of all possible actions for state x_k , denoted by $A(x)$, when the k -th packet transmission is requested for a terminal.

Step 2: [*Q-Value Computation*]

Compute the respective $Q(x, A)$ values for the set of state-action pairs $(x, A) | A \in A(x)$.

Step 3: [*Transport Format Selection*]

Determine the optimal action A^* such that the value of $Q(x, A^*)$ is minimum, i.e.

$$Q(x, A^*) = \min_{A \in A(x)} [Q(x, A)].$$

Step 4: [*Q-Value Update*]

Update the Q values by Eq. (3.18) while the next state y and the immediate cost $C(x, A^*)$ is obtained. Go to **Step 1**.



Chapter 4

Simulation Results and Discussion

4.1 System Environment and Simulation Parameters

In the simulation, a hexagonal grid cell structure with two-tier multi-cell system is considered. Including the own cell base station, there are 19 sites in the multi-cell system. For a HSDPA data transmission link, we assume that up to 80% of total transmission power is allocated to HS-DSCH, HS-SCCH, and the associated DCH. The total HSDPA transmission power, however, will depend on the number of data channels in use. The residual power allocated for other service, like voice transmissions on DCHs, in the own cell will interfere the HSDPA service with a little amount. For the transmitted power from base stations of other adjacent cells, it is always regarded as interference to the HSDPA link in the own cell. On the other hand, both the auto-correlation in time for the same link and the cross-correlation for different links on shadow fading are considered to approximate the real fading environment. In Fig. 4.1, it is evidently shown that the shadow effect for a specific link is correlated in time instead of the conventional model with log-normal distribution.

Each packet is transmitted per TTI length. The propagation plus the processing delay of ACK/NACK and CQI is assumed to be 6ms. This implies that after a channel measurement

Table 4.1: Simulation parameters

Parameter	Assumption
Cellular layout	Hexagonal grid, 19 sites, 2000 m cell radius
Path loss model ($\xi(r)$)	$128.1 + 37.6 \log_{10}(r)$ r is the base station separation in kilometers
Decorrelation length (d_{cor})	20 m
σ_L	8.0
Mobility assignment	0, 20, 40, 60 km/hr, random distribution
Carrier frequency	2.0 GHz
Channel bandwidth	5.0 MHz
Chip-rate	3.84 Mcps
Spreading factor	16
Thermal noise density	-174 dBm/Hz
TTI length	2 ms
Number of UE in one cell	4, random distribution
Number of multi-codes	12
Discounted factor (γ)	0.1
Scheduling algorithm	Proportional fair algorithm
BS total Tx power	Up to 44 dBm
Power for HSDPA data transmission	Maximum of 80% of total maximum available transmission power

is made at UE, it requires two additional TTIs before it can be used at the Node B. In order to identify the performance more easily, we also assume that the users always have information bits to be transmitted. That is, users are always in a saturation mode. As for the other detail simulation parameters, it is listed on Table 4.1.

4.2 Conventional Schemes

In this thesis, we propose the Q-HARQ scheme to adjust adaptively the packet initial transmission rate based on the Q-learning algorithm over a span of training. Besides the IR scheme, there are also two schemes proposed in [7] for dealing with the initial packet code rate of H-ARQ. These conventional schemes are described as follows :

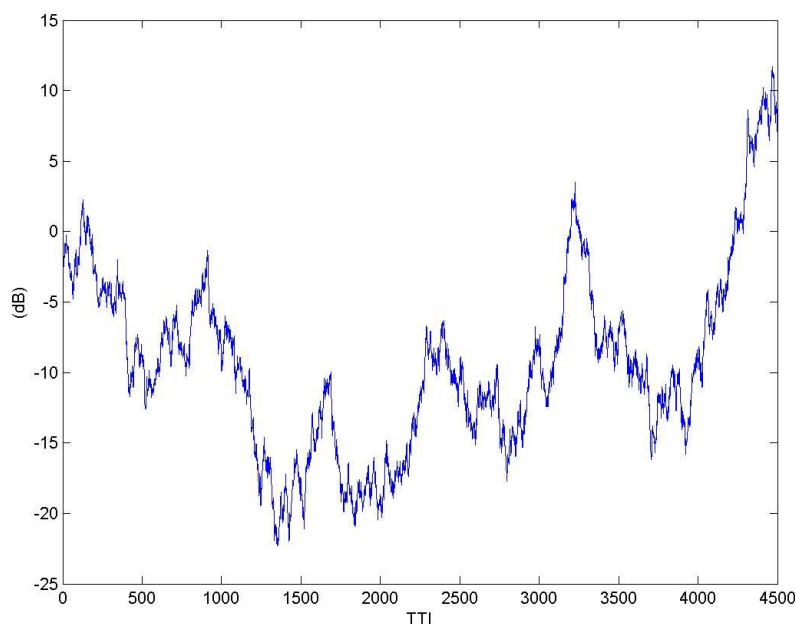


Figure 4.1: The auto-correlation on shadow fading

- IR : The initial code rate is always set as the largest effective code rate, and for the retransmission procedure, the redundancy is incrementally transmitted by one version until the decoding is successful.
- LA_IR_1 : The initial code rate is decided by retreating one redundancy version of the most recent effective code rate which achieves a successful packet decoding. For the retransmission procedure, the redundancy is incrementally transmitted by one version until the decoding is successful.
- LA_IR_2 : The initial code rate is decided by keeping the code rate as the last effective code rate which achieves a successful packet decoding. If there is no retransmission request for two consecutive times with this initial code rate, then retreat two redundancy of this effective code rate. For the retransmission procedure, the redundancy is incrementally transmitted by one version until the decoding is successful.

4.3 Performance Evaluation and Discussions

As mentioned above, we assumed that the interference for the user of HSDPA services from other service links in the own cell is existing with a little effect. Since the traffic which are not carried on the HS-DSCH does not be schedule with HSDPA traffic, we suppose that it interferes the HSDPA user with a fixed quantity. From another point of view, the increase of interference implies that the power allocated to HSDPA users decrease. That is, the load of real time service of this system raises. In this thesis, we define a signal to other services' interference ratio, SIOR, which describes the transmitting power for HSDPA services to other services signal in the own cell.

Fig. 4.2 and Fig. 4.3 show the block error rate of four schemes versus SIOR while the Q-HARQ is employed based on the 0.1 BLER requirement. For the Q-HARQ, it can be found that the BLER is exactly falling around the 0.1 which is conformed to the Release 5 specification. However, the BLER of those comparative schemes violate seriously the specification of BLER. Even the best of them under good channel condition just can not reach 0.27 BLER with 16-QAM and can not reach 0.17 BLER with QPSK. For this reason, these three schemes can not be used in the real system except the proposed Q-HARQ.

Fig. 4.4 and Fig. 4.5 show the average number of transmission time versus SIOR for a successful packet transmission, while the Q-HARQ is employed based on the 0.1 BLER requirement, under different modulation orders, 16-QAM and QPSK, respectively. It is shown that the Q-HARQ needs less number of transmission time than the comparative schemes for transmitting a packet successfully. It always maintains below 1.2 times for a successful packet transmission. On the other hand, it can be found that the conventional schemes need much number of transmission time for a successful packet transmission while

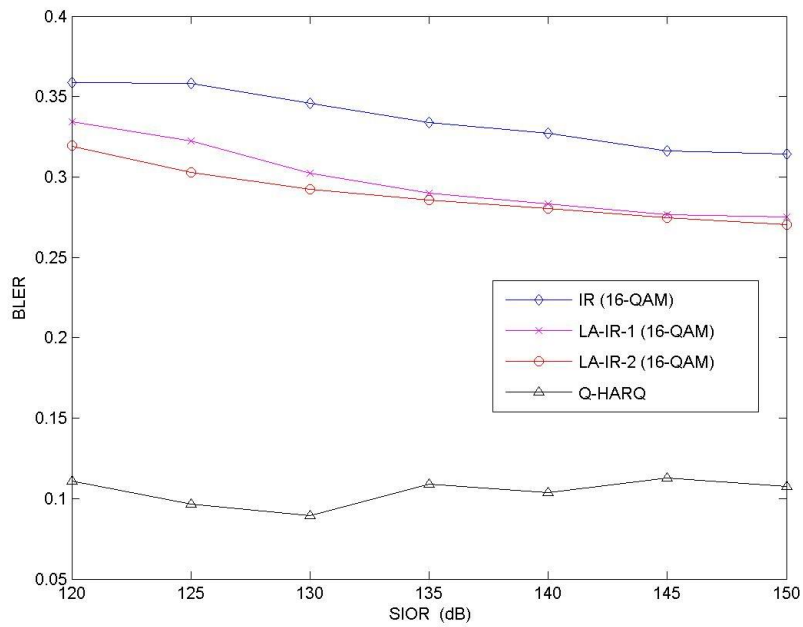


Figure 4.2: The BLER with 0.1 BLER requirement for Q-HARQ

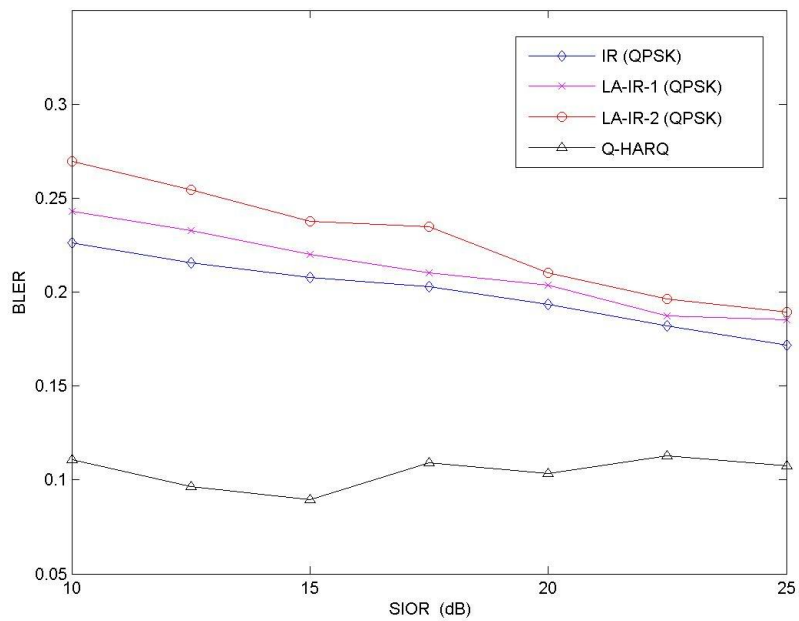


Figure 4.3: The BLER with 0.1 BLER requirement for Q-HARQ

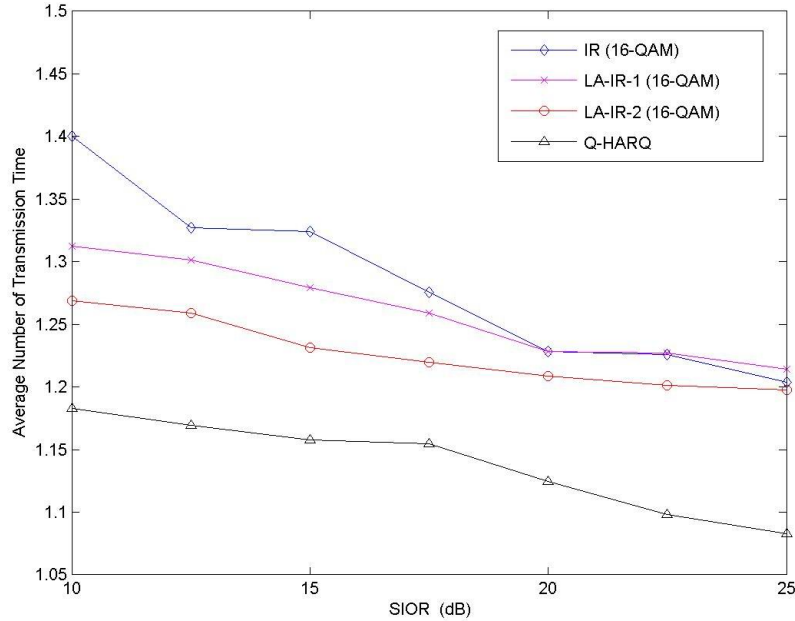


Figure 4.4: The average number of transmission time with 0.1 BLER requirement for Q-HARQ

the SIOR is low. However, the Q-HARQ can select the most suitable transport format to adapt the instantaneous channel condition more efficiently under satisfying the specified BLER requirement. In other words, the average number of transmission time for a successful packet transmission is less sensitive to the channel quality than the IR schemes. It stands for the intelligence of the Q-HARQ scheme. In Fig. 4.5, it is institute that symbols carried on QPSK will be more reliable than that carried on 16-QAM. Thus, the comparative schemes with QPSK demand less number of transmission time than those with 16-QAM. Oppositely, it also can be seen that schemes with 16-QAM will perform higher throughput than those with QPSK in Fig. 4.6 and 4.7. Especially speaking, among these comparative schemes, it can be found that the IR scheme take more risk for utilizing channel resource than LA-IR-1 and much more than LA-IR-2. Therefore, the LA-IR-2 outperforms the other two schemes when using 16-QAM and the IR scheme performs the best when using QPSK among these

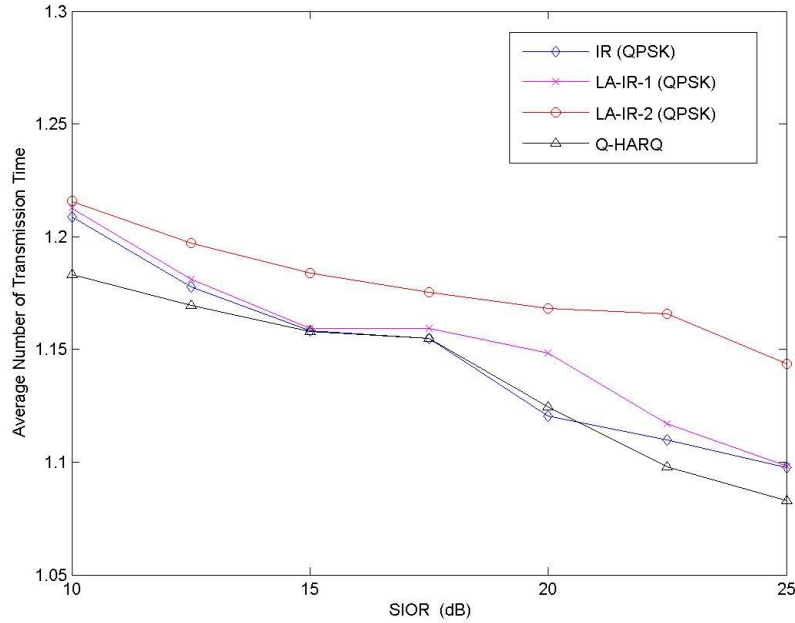


Figure 4.5: The average number of transmission time with 0.1 BLER requirement for Q-HARQ

three schemes for considering the number of transmission time.

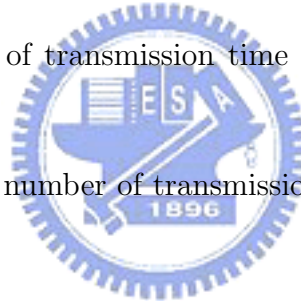


Fig. 4.6 and Fig. 4.7 illustrate the total system throughput versus SIOR, while the Q-HARQ is employed based on the 0.1 BLER requirement, under different modulation orders, 16-QAM and QPSK, respectively. For the conventional schemes, it can be found that the less number of transmission time for a successful packet transmission implies the higher system throughput. It is reasonable that the less number of transmission time for a successful packet transmission, the more channel resource can be fully utilized. Nevertheless, the Q-HARQ can not outperform the comparative schemes on system throughput, *while the Q-HARQ is employed based on the 0.1 BLER requirement.*

In chapter 3, the failure event has been defined as that moment, when the packet with

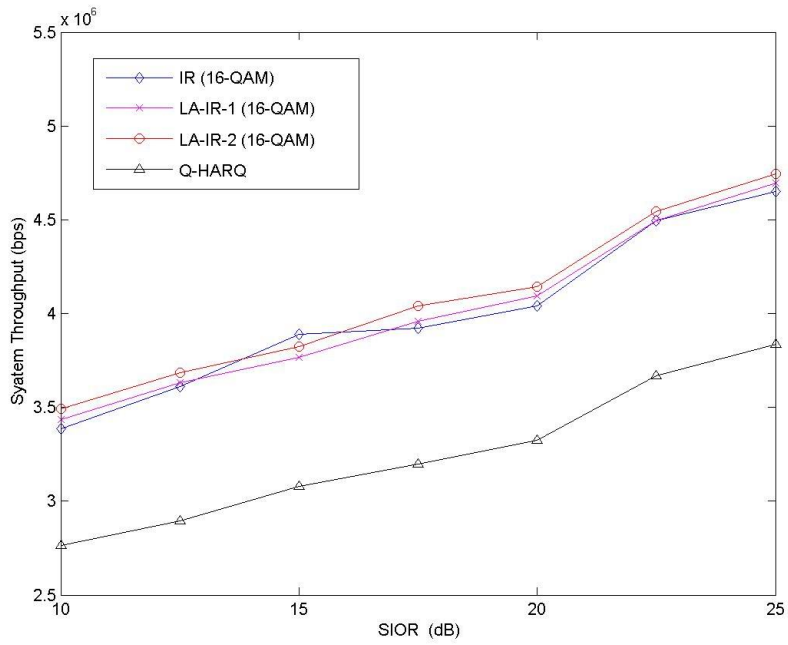


Figure 4.6: The system throughput with 0.1 BLER requirement for Q-HARQ

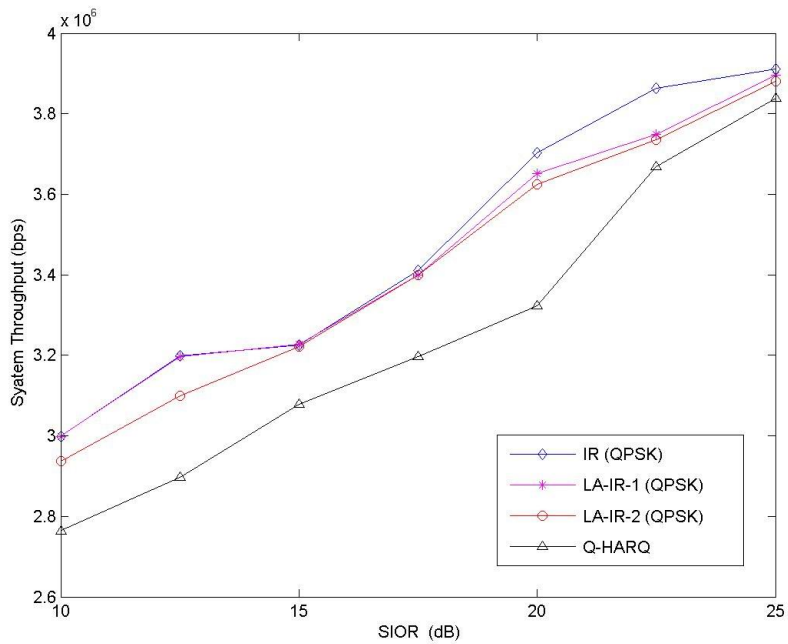
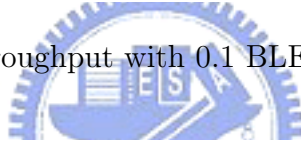


Figure 4.7: The system throughput with 0.1 BLER requirement for Q-HARQ

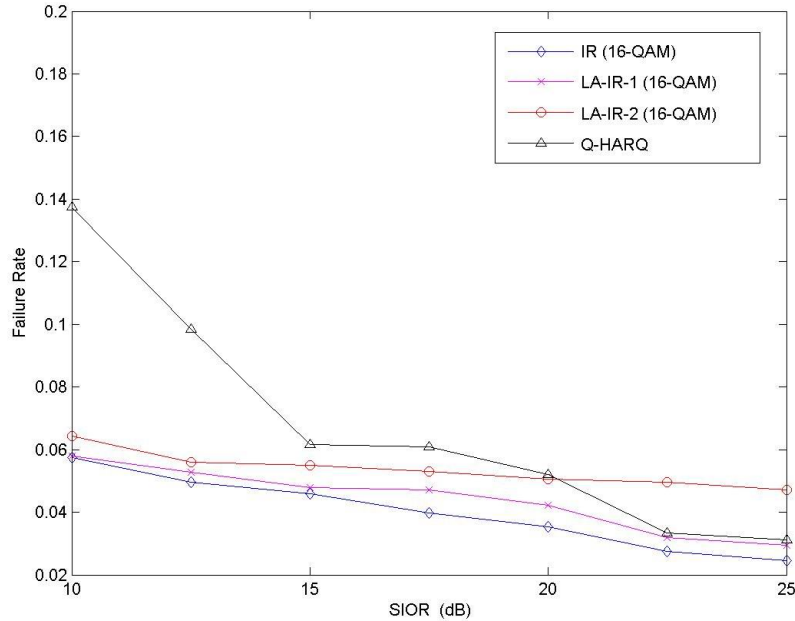


Figure 4.8: The failure rate with 0.1 BLER requirement for Q-HARQ

ECR up to $1/3$, still can not be decoded successfully. Here, we further define a *failure rate* as a ratio of *the number of failure events to the summation of the number of failure events and the number of successful transmitted packets*. Fig. 4.8 and Fig. 4.9 depict the failure rate versus SIOR, while the Q-HARQ is employed based on the 0.1 BLER requirement, under different modulation orders, 16-QAM and QPSK, respectively. From these two figures, It can be concluded that the failure rate of Q-HARQ is higher when the SIOR is low than that when the SIOR is high. This is because the Q-HARQ will select higher ECR as the initial packet transmission under the fixed 0.1 BLER requirement while the channel quality is worse, and also the packet error probability at bad channel condition is high. On the other hands, the failure rate of LA-IR-2 scheme is higher than LA-IR-1 and much higher than the IR scheme. The reason is that LA-IR-2 scheme adopts conservative manner to make transport decisions among these three schemes. As the simulation results of the system throughput, the performance of the failure rate of the Q-HARQ is worse than the conventional schemes,

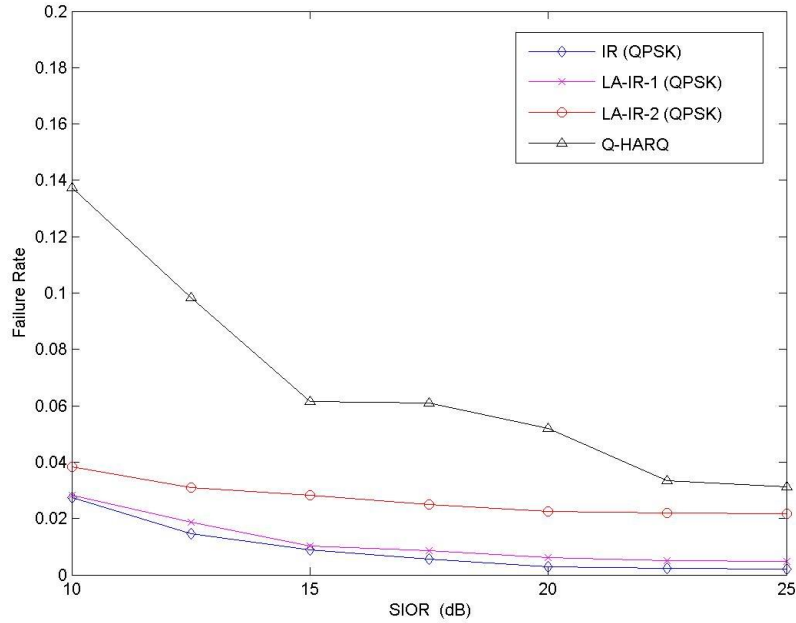


Figure 4.9: The failure rate with 0.1 BLER requirement for Q-HARQ

while the Q-HARQ is employed based on the 0.1 BLER requirement.



Over all the analysis of performance above, it can be concluded that why the system throughput of the Q-HARQ is the worst one while the number of transmission time of it is the best. This is because the circumstance exactly exists that sacrificing the requirement BLER brings much better benefit on system throughput in the communication system. For example, it is possible that the transmission with 0.3 BLER requirement will take a great risk such that 150 information bits can be sent for each transmission, and the transmission with 0.1 BLER requirement will send only 100 information bits for each transmission in a conservative manner. In a round of ten packet transmissions, the transmission with 0.3 BLER requirement will get higher throughput than the transmission with 0.1 BLER requirement. In addition, because of the tighter demand of BLER than conventional schemes, the failure rate of the Q-HARQ will be much large. However, the better performance of the number

of transmission time for the Q-HARQ can identify that the Q-HARQ always can select the most appropriate action even they are under different requirement of BLER. Additionally, it can also be seen that the system throughput of the Q-HARQ will approach much more the results of these conventional schemes with QPSK than with 16-QAM since the BLER with QPSK shown in Fig. 4.3 is closer to 0.1 than with 16-QAM.

In order both to reserve the design spirit of comparative schemes and to verify the proposed Q-HARQ scheme, the requirement of BLER for the Q-HARQ is reset to 0.3 for comparing with the conventional schemes with 16-QAM. In Fig. 4.10, it can be shown that the Q-HARQ with 0.3 BLER requirement enhances apparently the system throughput. The system throughput of the Q-HARQ with 0.2 BLER requirement, similarly, in Fig. 4.11 can be promoted substantially. It implies that the design of the Q-HARQ scheme outperforms the conventional schemes while the criterion is nearly the same.

For a practical communication system, the warm up time is necessary, and it is not appropriate for spending too long for warming up. In Table 4.2, the convergence time under different SIOR is represented, where a cycle is a period of 45 seconds. We claim that the Q-learning algorithm is converging when the average incremental change of Q-factor over 45 seconds is smaller than a specific value after a period of 4500 seconds for training all kinds of system states. The average convergence time of these seven channel conditions is about 7302 seconds which is equal to two hours and 1.7 minutes. It means that the Q-HARQ just needs less than one hour to achieve steady after one hour and 15 minutes training period. As the simulation results shown, the warm up time for the Q-HARQ is feasible for the real system. In addition, Fig. 4.12 illustrates the status of the average incremental change of Q-factors versus cycle with 0.1 BLER requirement when SIOR is equal to 15.

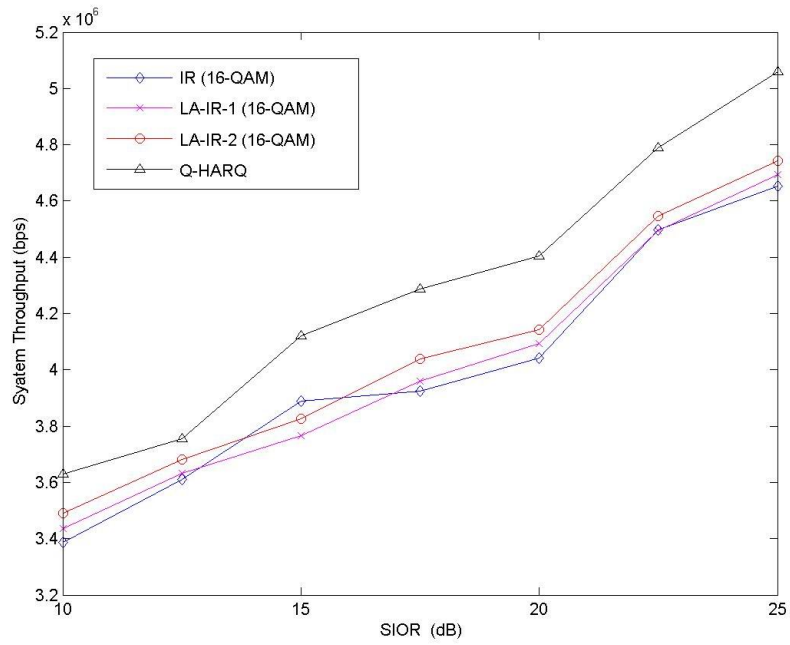


Figure 4.10: The system throughput with 0.3 BLER requirement for Q-HARQ

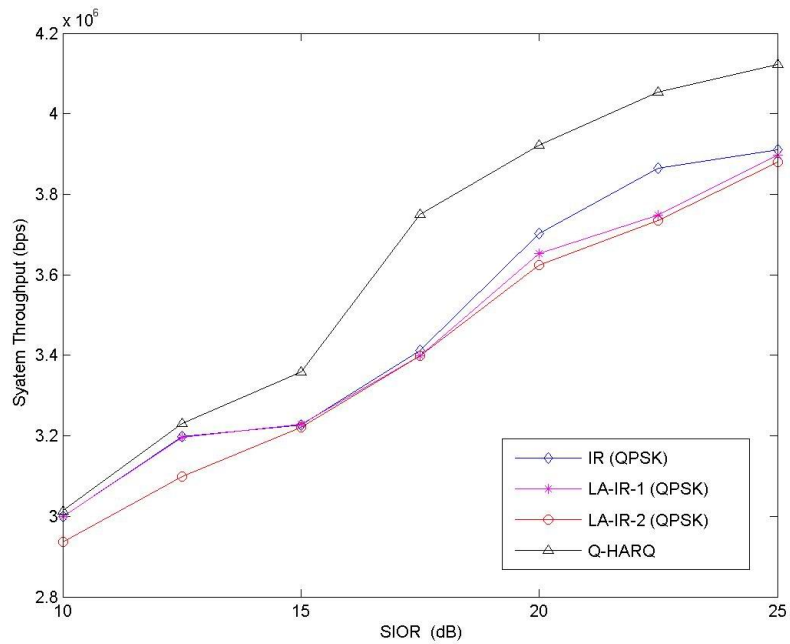


Figure 4.11: The system throughput with 0.2 BLER requirement for Q-HARQ

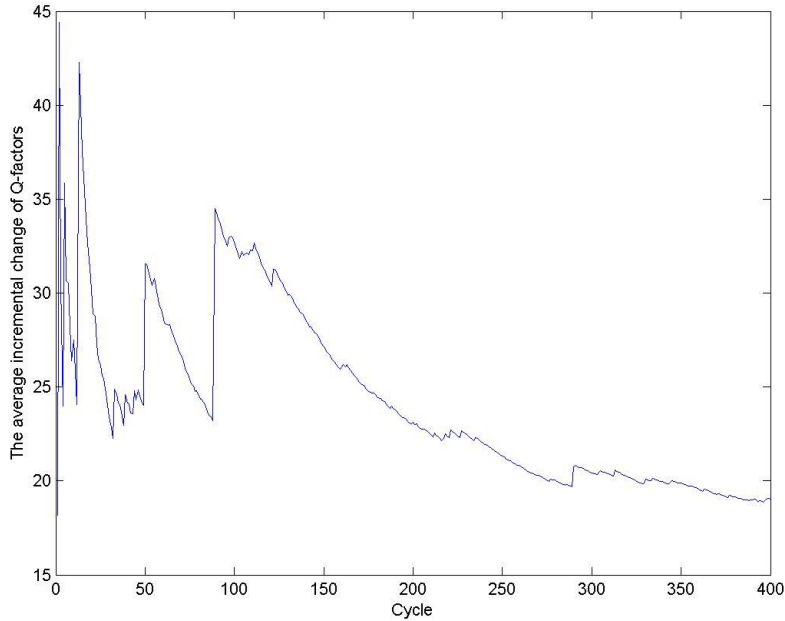


Figure 4.12: The incremental change of Q-factors with 0.1 BLER requirement for Q-HARQ when the SIOR is equal to 15

Table 4.2: Convergence Time

SIOR	120	125	130	135	140	145	150
Convergence Time (sec)	4500	4500	4500	8370	8010	8820	12420

Besides the warm up time, the processing time of the Q-HARQ scheme is another critical issue to verify whether the method is practical or not. From Table 4.3, there are five random trials for calculating the processing time including the state construction time and the searching time for action assignment. The equipment of the testing platform is provided with a 3.0 GHz processor and 1.0 GB RAM. It can be concluded that the processing time for

Table 4.3: Processing Time

Trial	1	2	3	4	5
Processing Time (ms)	0.001520	0.001440	0.001360	0.001340	0.001280

the Q-HARQ scheme is extremely trifling for a transmission time interval of 2 ms. Hence, both considering the practicality and the performance, the Q-HARQ is a suitable method which can be implemented for the HSDPA system.



Chapter 5

Concluding Remarks

In this thesis, a Q-learning-based hybrid automatic repeat request (Q-HARQ) scheme for HSDPA in UMTS system is proposed to achieve efficient resource utilization. The Hybrid ARQ procedure is modeled as a discrete-time Markov decision process, where the transmission cost is defined in terms of the QoS parameters of transport block error rate for enhancing spectrum utilization subject to QoS constraint. The Q-learning reinforcement algorithm is adopted to accurately estimate the transmission cost to perform the most suitable decision while the requirement of transport block error rate is guaranteed. By means of the self-tuning capability of Q-learning algorithm, the optimal actions of coding rate and modulation order for the initial packet transmission is obtained after a period of convergence time under distinct channel states.

Simulation results show that the Q-HARQ is the only scheme which satisfies indeed the QoS requirement of 0.1 BLER specified in 3GPP Release 5 when consider the benchmark schemes. Also, the Q-HARQ scheme can improve the total system throughput for HSDPA in UMTS over the conventional IR, LA-IR-1, and LA-IR-2 schemes under the same constraint of the QoS requirement of transport block error rate. Under satisfying the specific 0.1 transport block error rate criterion, the number of transmission time for a successful packet transmission can be maintained below 1.2 times. By reducing the delay due to the retransmission

procedures, the channel resource can be utilized much efficiently. On the other hand, it can be found that the Q-HARQ can avoid the bad link adaptation on account of the channel prediction errors. Based on the property of the Q-learning algorithm, the prediction errors can also be count in the reinforcement training process. Finally, the analysis results of the convergence time and the processing time confirm exactly that the Q-HARQ is feasible for employing in the real system.



Bibliography

- [1] *3rd Generation mobile system Release 5 specifications*, 3GPP Std. TS 21.103, 2001.
- [2] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*, 3rd ed. JohnWiley & Sons, 2002.
- [3] E. F. C. LaBerge and J. M. Morris, “Expressions for the mean transfer delay of generalized M-stage hybrid ARQ protocols,” *IEEE Trans. Commun.*, vol. 52, pp. 999–1009, June 2004.
- [4] J. Hagenauer, “Rate compatible punctured convolutional codes (RCPC) and their applications,” *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1998.
- [5] R. Mantha and F. R. Kschischang, “A capacity-approaching hybrid ARQ scheme using turbo codes,” in *Proc. IEEE Global Telecommunications Conference*, vol. 5, Dec. 1999, pp. 2341–2345.
- [6] R. Kwan, P. Chong, and M. Rinne, “Analysis of the adaptive modulation and coding algorithm with the multicode transmission,” in *Proc. IEEE 56th Vehicular Technology Conference*, vol. 4, Sept. 2002, pp. 24–28.
- [7] L. Zhao, J. W. Mark, and T. C. Yoon, “A combined link adaptation and incremental redundancy protocol for enhanced data transmission,” in *Proc. IEEE Global Telecommunications Conference*, vol. 2, Nov. 2001, pp. 1277–1281.

- [8] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, pp. 653–663, May 1991.
- [9] Y. Xu and P. Zhang, "Adaptive incremental redundancy scheme on high-speed wireless communication system," in *Proc. IEEE PIMRC*, vol. 1, Sept. 2003, pp. 294–296.
- [10] J. F. Cheng, Y. P. E. Wang, and S. Parkvall, "Adaptive incremental redundancy," in *Proc. IEEE 58th Vehicular Technology Conference*, vol. 2, Oct. 2003, pp. 6–9.
- [11] J. M. Shea, "Reliability-based hybrid ARQ," *IEE Electron. Lett.*, vol. 38, pp. 644–645, June 2002.
- [12] H. Kim and J. Shea, "New turbo-ARQ techniques based on estimated reliabilities," in *Proc. IEEE WCNC*, vol. 2, Mar. 2003, pp. 16–20.
- [13] V. Tripathi and E. Visotsky, "Reliability-based type II hybrid ARQ schemes," in *Proc. IEEE ICC*, vol. 4, May 2003, pp. 2899–2903.
- [14] A. Roongta and J. Shea, "Reliability-based hybrid ARQ using convolutional codes," in *Proc. IEEE ICC*, vol. 4, May 2003, pp. 2889–2893.
- [15] —, "Reliability-based hybrid ARQ and rate-compatible punctured convolutional (RCPC) codes," in *Proc. IEEE WCNC*, vol. 4, Mar. 2004, pp. 2105–2109.
- [16] H. Chen and P. Fan, "An adaptive coded modulation scheme associated with improved HARQ," in *Proc. IEEE PIMRC*, vol. 2, Sept. 2003, pp. 1292–1296.
- [17] B. Makarevitch, "Application of reinforcement learning to admission control in CDMA network," in *Proc. IEEE PIMRC*, vol. 2, Sept. 2000, pp. 1353–1357.

- [18] J. Nie and S. Haykin, "A Q-learning-based dynamic channel assignment technique for mobile communication systems," *IEEE Trans. Vehicular Technology*, vol. 48, pp. 1676–1687, Sept. 1999.
- [19] H. Tong and T. X. Brown, "Adaptive admission call admission control under quality of service constraints: a reinforcement learning solution," *IEEE J. Select. Areas. Commun.*, vol. 18, pp. 209–221, Feb. 2000.
- [20] P. Marbach, O. Mihatsch, and J. N. Tsisiklis, "Call admission control and routing in integrated services networks using neuro-dynamic programming," *IEEE J. Select. Areas. Commun.*, vol. 18, pp. 197–208, Feb. 2000.
- [21] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8.
- [22] S. Haykin, *Neural Networks*, 2nd ed. Prentice Hall, 1999.
- [23] L. Gordon, *Principle of Mobile Communication*. Kluwer Academic, 1958.
- [24] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *IEE Electron. Lett.*, vol. 27, pp. 2145–2146, Nov. 1991.
- [25] A. J. Viterbi, A. M. Viterbi, K. S. Gilhousen, and E. Zehavi, "Soft handoff extends cdma cell coverage and increases reverse link capacity," *IEEE J. Select. Areas. Commun.*, vol. 12, pp. 1281–1288, Oct. 1994.
- [26] R. Bellman, *Dynamic Programming*. Princeton, 1957.

Vita

Chia-Yuan Chang was born on 1982 in Taichung, Taiwan. He received the B.E. degree in electrical engineering from Nation Cheng-Kung University, Tainan, Taiwan, in 2004, and the M.E. degree in the department of communication engineering, college of electrical and computer engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2006. His research interests include radio resource management and wireless communication systems.

