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

Chung-Ju Chang*, Chia-Yuan Chang†, and Fang-Ching Ren‡
*Department of Communication Engineering
National Chiao Tung University, Hsinchu, Taiwan 300
E-mail:cjchang@cc.nctu.edu.tw
† Chi Mei Communication Systems, Inc.
‡ Information and Communications Research Laboratories
Industrial Technology Research Institute, Hsinchu, Taiwan 300
E-mail:frank_ren@itri.org.tw

***Abstract*-In this paper, a Q-learning-based hybrid automatic repeat request (Q-HARQ) scheme is proposed to achieve efficient resource utilization for high speed downlink packet access (HSDPA) in universal mobile telecommunications system (UMTS). The Hybrid ARQ procedure is modeled as a discrete-time Markov decision process (MDP), where the transmission cost is defined in terms of the signal-to-interference-and-noise (SINR) which is based on the desired (quality-of-service) QoS parameters of transport block error rate (BLER) 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 block error rate for Q-HARQ is nearly met around a reasonable value indeed. In addition, the system throughput 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 because of the short processing and convergence time.**

I. INTRODUCTION

High speed downlink packet access (HSDPA) was proposed in the 3rd generation partnership project (3GPP) Release 5 to provide efficient, robust, and high-speed packet data services for universal mobile telecommunications system (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 [1]. Also, advanced hybrid automatic repeat request (H-ARQ) retransmission is adopted to upgrade the robustness against link adaptation errors. Instead of the radio network controller (RNC), Node B directly controls the retransmission procedure to reduce the transmission delay.

The major difference between ARQ and H-ARQ is that H-ARQ combines an 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). By chase combining, the transmitter will retransmit the same packet format according to feedback of negative acknowledge (NACK) when the previous packet cannot be decoded successfully at receiver side. While receiving the retransmitted packet, the decoder combines these multiple copies of packets weighted by the received signal-to-noise ratio (SNR). By IR scheme, the transmitter sends different redundancy versions for each retransmission 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.

As for the AMC in HSDPA, the idea is to adapt the transmission to the fast varying channel quality. With a proper combination of the modulation order and channel coding rate, an adaptive selection is made per transmission time intervel (TTI) a definite set of modulation and coding schemes (MCS) such that an enhanced spectral efficiency can be achieved in good channel conditions. The selection criterion of the MCS should be made such that the transmission block error rate (BLER) is reasonable. An adaptive scheduling algorithm was proposed in [2] 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. Furthermore, a strategy which combines advantages of LA and IR was proposed in [3]. Unlike the fixed starting code rate in the conventional IR, two LA_IR protocols adaptively choose the initial transmission code rate in accordance with the current channel condition to minimize efficiently the number of retransmission. These LA_IR schemes usually achieve a higher effective throughput than that only considered LA scheme. 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 [4].

Moreover, an adaptive incremental redundancy (AIR) algorithm was proposed in [5], where an information-theoretic model was developed to predict the coding gains of the H-ARQ scheme. An accumulated conditional mutual information (ACMI) for a packet transmission is found by using a bit-interleaved coded modulation (BICM) capacity, which

1550-2252/$25.00 ©2007 IEEE

is determined by the SNR and the modulation order. 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 is obtained 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 this AIR scheme just uses an modulation and coding schemes (MCS) table for the initial transmission in advance. It is not sufficiently enough to adapt channel conditions for the initial packet transmission. Also, it causes the system more complicated on the account of the BICM while calculating the accumulated information bit by bit.

In this paper, we propose a Q-learning-based hybrid automatic repeat request (Q-HARQ) scheme for HSDPA in WCDMA systems to achieve efficient resource utilization. Since WCDMA is an interference-limited system, channel condition is chosen as the system state in the formulated H-ARQ process. The selection of modulation and coding schemes for initial packet transmissions is a critical action while experiencing various system state. On the other hand, an evaluation function is defined to appraise the cumulative discounted cost of the consecutive decisions for H-ARQ, where the cost function is presented in terms of the expected BLER. The evaluation function is calculated by a real-time reinforcement learning (RL) technique known as Q-learning [6], [7] because of the difficulty for obtaining the state transition behavior. After a decision is made, the consequent cost is used as a feedback parameter to the Q-HARQ controller to adjust adaptively the evaluation Q-function which is utilized as a basis for determining the exact action. Thus, the learning procedure is performed in a closed-loop iteration manner such that the optimal solution for the initial packet MCS can be found. By improving the initial packet transmission under a QoS requirement of transport BLER, the suitable transport decisions for initial packet transmission can be optimized while the desired QoS requirements is meet. Also, it is expected that the number of transmission time can be efficiently diminished due to the good link adaptation for the initial transmission.

The remainder of this paper is organized as follows. In section 2, an HSDPA system model to investigate the proposed Q-HARQ method is presented. The principal concept for the design of the proposed Q-HARQ is described in section 3. Also, there are some significant implementation issues discussed in this section. Simulation results and discussion are presented in section 4, followed by concluding remarks in section 5.

II. SYSTEM MODEL

For the high-speed downlink packet access, an HS-DSCH carries the user data in the downlink direction with fixed spreading factor 16 such that it can be shared among scheduled users in transmission time interval (TTI) fashion. The TTI is specified as 2 ms to achieve a shorter round trip delay for better link adaptation. Except for QPSK, a higher order modulation scheme, 16-QAM, is adopted for supporting higher instantaneous peak data rate. Additionally, a high-speed shared control channel (HS-SCCH) carries layer 1 control information for HS-DSCH demodulation. Node B transmits the HS-SCCH two slots ahead the corresponding HS-DSCH TTI to inform the terminal of which H-ARQ process the data belongs to and whether it needs to be combined with data received previously in its buffer. Upon decoding the combined data, an ACK/NACK indicator will be sent in the uplink high-speed dedicated physical control channel (HS-DPCCH), depending on the outcome of the CRC check conducted on the HS-DSCH data. In addition, the channel quality indicator (CQI) information is periodically feedback to recommend which transport format is sufficient to satisfy for meeting the 0.1 BLER through the uplink HS-DPCCH. To evaluate impact of the fast varying channel, a terrestrial mobile radio channel for urban areas is considered in this paper. The propagation effect is divided into three parts, path loss, slow variation about the mean due to shadowing and the rapid variation in the signal due to multi-path effects [8]. Furthermore, the shadowing effect variation is correlated with the distance of two adjacent sampling point. In this paper, the normalized autocorrelation function [9] is adopted for modeling the correlated shadow fading versus distance between two adjacent TTI. 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. To describe the cross-correlation in the multi-cell system, the method [10] proposed by Viterbi is adopted in this paper. 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.

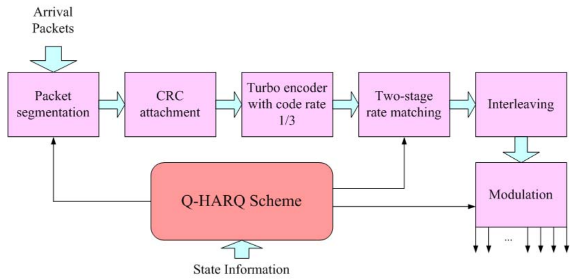

Fig. 1. The Q-HARQ for packet transmission in HS-DSCH

The proposed Q-HARQ scheme working behavior is shown in Fig. 1 to determine 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 date 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 generate different sizes of redundancy. 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.

III. DESIGN OF Q-HARQ

The major work in the H-ARQ procedure is to determine a suitable modulation and coding scheme for an initial transmission of packets. For all kinds of channel state, there is a finite set of possible actions that may be taken by the Node B. Every time the system takes an action, a certain cost is incurred and used as an response signal feedback to the system to adaptively select an action in a specific system state. States are observed, actions are taken, and cost are incurred at discrete time. Therefore, the problem for the decision control of packet transmissions in H-ARQ can be seen as a discrete-time MDP.

First, 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

$$x_k = (\widehat{SINR_k}), k = 0, 1, 2, \ldots \tag{1}$$

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

$$A_k = [(ECR)_k, M_k]. \tag{2}$$

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

$$\left\{1, \frac{3}{4}, \frac{2}{3}, \frac{1}{2}, \frac{1}{3}\right\}, \tag{3}$$

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. 1, 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

$$C(x_k, A_k) = \left[\frac{SINR(x_k, A_k) - \widetilde{SINR}(A_k)}{\widetilde{SINR}(A_k)}\right]^2, \tag{4}$$

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. Therefore, the objective to minimize the cost function for the H-ARQ is to make the $SINR(x_k, A_k)$ most close to the $\widetilde{SINR}(A_k)$.

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

$$Q(x, A) = E\left\{\sum_{k=0}^{\infty} \gamma^k C(x_k, A_k) | x_0 = x, A_0 = A\right\}, \tag{5}$$

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

$$\begin{aligned} Q(x, A) &= E\left\{C(x_0, A_0) | x_0 = x, A_0 = A\right\} + \\ &\quad E\left\{\sum_{k=1}^{\infty} \gamma^k C(x_k, A_k) | x_0 = x, A_0 = A\right\} \\ &= E\left\{C(x, A)\right\} + \gamma \sum_y P_{xy}(A) \times \\ &\quad E\left\{\sum_{k=1}^{\infty} \gamma^{k-1} C(x_k, A_k) | x_1 = y, A_1 = B\right\} \\ &= E\left\{C(x, A)\right\} + \gamma \sum_y P_{xy}(A) Q(y, B). \end{aligned} \tag{6}$$

From Eq. (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. Based on the principle of Bellman's optimality [11], 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 respective to every next state y. Thus the $Q^*(x, A)$ can be obtained by

$$\begin{aligned} Q^*(x, A) &= E\left\{C(x, A)\right\} + \\ &\quad \gamma \sum_y P_{xy}(A) \left\{\min_B \left[Q^*(y, B)\right]\right\} \\ &\quad \text{for all } (x, A) \end{aligned} \tag{7}$$

Then the optimal action A^* with respective 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

$$Q^*(x, A^*) = \min_A \left[Q^*(x, A)\right]. \tag{8}$$

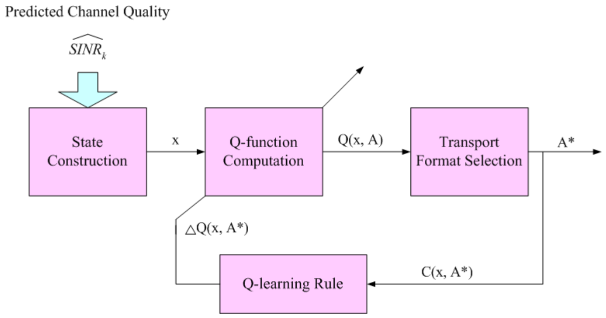

Fig. 2. Structure of the Q-learning-based H-ARQ scheme

However, it is difficult to get the $E\{C(x,A)\}$ and $P_{xy}(A)$ for solving the Eq. (7). Thus, we adopt a real-time reinforcement learning algorithm, named Q-learning algorithm [6], [7], 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.

Fig. 2 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

$$Q(x,A)=\left\{\begin{array}{ll} Q(x,A)+\eta\Delta Q(x,A) & \text{, if } A=A^* \\ Q(x,A) & \text{, otherwise} \end{array}\right. , \quad (9)$$

and

$$\Delta Q(x,A^*)=\Big\{C(x,A^*)+\gamma\min_B\left[Q(y,B)\right]\Big\}-Q(x,A^*) \quad (10)$$

From the Eq. (9), 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 $\min_B\left[Q(y,B)\right]$ is executed by comparing the Q value of all the possible action candidates for state y in Eq. (10) 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 [6]. 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. (9) will converge to $Q^*(x,A)$ with probability 1.*

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

Step 4: [*Q-Value Update*]
Update the Q values by Eq. (9) while the next state y and the immediate cost $C(x,A^*)$ is obtained. Go to **Step 1**.

IV. SIMULATION RESULTS AND DISCUSSION

In the simulation, a hexagonal grid multi-cell system is considered. where each observed cell is assumed to suffer two-tier neighbor cell interference. For a HSDPA data transmission link, we assume that 80% of total transmission power is allocated to HS-DSCH, HAS-SCCH, and the associated DCH. The residual power is completely allocated for other services on DCHs. The other service links may interference with the HSDPA service links each other in home cell because orthogonality between HSDPA and other services links will be degraded after passing through multipath fading channel. We define a signal to other services' interference ratio, SIOR, to quantify the imperfect orthogonality. in home cell. To consider neighbor cell interference, both the auto-correlation in time for a link and the cross-correlation for different links on shadow fading are considered to simulate a real fading environment.

Each packet is transmitted per TTI length. The propagation delay plus the processing delay of ACK/NACK and CQI is assumed to be 6ms. This implies that after a channel measurement is made at UE, it can be used at the Node B two TTI later. 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. The other simulation parameters are listed on Table I. The IR scheme and two advanced IR schemes in [3] for dealing with the initial packet code rate of H-ARQ are simulated to check whether the proposed Q-HARQ scheme is more efficient.

Fig. 3 shows the block error rate of four schemes versus SIOR. Since the proposed Q-HARQ is employed based on the 0.1 BLER requirement recommended BLER for HSDPA [12], it keeps BLER around 0.1. However, those comparative schemes with QPSK and 16-QAM modulation have much higher BLER. It means that these comparative schemes can not be used in the real system to meet the recommended BLER for SIOR within 6 dB (without orthogonality) to 20 dB (near perfect orthogonality).

Fig. 4 shows the average number of transmission time versus SIOR for a successful packet transmission It is shown that the Q-HARQ needs less number of transmission time than the comparative schemes for transmitting a packet successfully.

TABLE I

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
The standard deviation of shadow fading	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

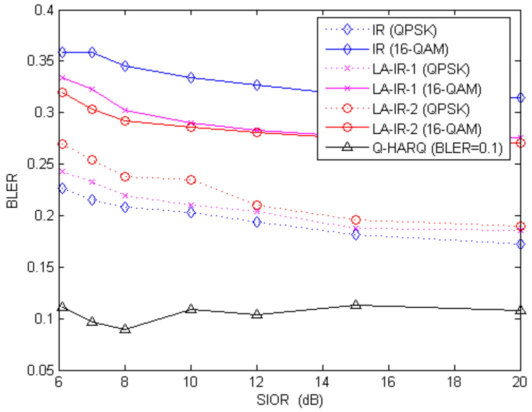

Fig. 3. The BLER versus SIOR for Q-HARQ (with 0.1 BLER requirement) and comparative schemes

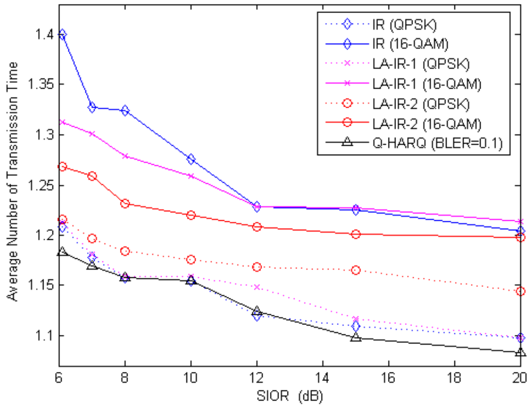

Fig. 4. The average number of transmission time versus SIOR for Q-HARQ (with 0.1 BLER requirement) and comparative schemed

TABLE II

CONVERGENCE TIME

SIOR	6.02	7.0	8.0	10.0	12.0
Convergence Time (sec)	4500	4500	4500	8370	8010
SIOR	15.0	20.0			
Convergence Time (sec)	8820	12420			

It is because that 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 addition, the comparative schemes with QPSK are more reliable to have smaller BLER and demand less number of transmission time than those with 16-QAM. To have a fair comparison under similar BLER level, only the results with QPSK modulation for the comparative schemes are shown in the following paper.

Fig. 5 illustrates the total system throughput versus SIOR, while the Q-HARQ is employed based on the 0.1, 0.2, and 0.3 BLER requirement, and the comparative schemes adopt QPSK. While the proposed Q-HARQ scheme is employed on 0.1 BLER requirement, it has lower system throughput than the comparative schemes who have BLER in range of 0.17 to 0.27. This is because that each transmission in H-ARQ can be variable code rate. For a given channel condition, an aggressive code rate would induce higher BLER and larger number of re-transmissions but gain higher throughput in one successful transmission. The Q-HARQ with 0.1 BLER performs a less aggressive action so that it lose some throughput. While let Q-HARQ employed on 0.2 and 0.3 BLER requirements for a fair comparison with other schemes, the results show that the proposed Q-HARQ scheme has higher system throughput than the comparative schemed. It means that the Q-HARQ can perform more accurately aggressive code rate to achieve higher system throughput but keep BLER requirement.

By using Q-leaning manner, the warm up time is necessary to learn the accurate state-action mapping, and it is not appropriate for spending too long for warming up. In Table II, the convergence time under different SIOR is presented. To evaluate the convergence, we set a cycle to be 45 seconds and check whether the Q-learning algorithm is convergent cycle by cycle. After a minimum training time of 100 cycles, the Q-learning algorithm is regarded convergent when the average incremental change of Q-factor over a cycle is smaller than a specific value for all kinds of system states. The simulation results show that the warm up time increases as SIOR increases. It is because that a larger SIOR implies higher orthogonality within the observed cell and the interference variation is dominated by neighboring cell. The behavior of neighbor cells is more complex to learn so it take more time to make Q-HARQ convergent. The results show that the Q-HARQ has a feasible warm-up time for the real system.

Besides the warm up time, we also evaluate the processing time of the Q-HARQ scheme to verify whether the method is practical. The overall processing time, including the state construction time and the searching time for action assignment, takes about 1.39 ms by the testing platform equipped with a 3.0 GHz processor and 1.0 GB RAM. The processing time for the Q-HARQ scheme is smaller than a transmission time

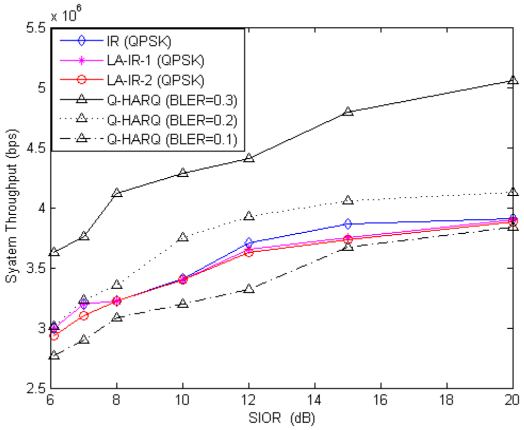

Fig. 5. The system throughput versus SIOR for Q-HARQ (with 0.1, 0.2, and 0.3 BLER requirements) and the comparative schemed (with QPSK)

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

V. CONCLUDING REMARKS

In this paper, 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. Then we solve the discrete-time Markov decision process to enhance spectrum utilization subject to QoS constraint. The Q-learning reinforcement algorithm is adopted to accurately estimate the transmission cost for optimal decision while the requirement of transport block error rate is guaranteed. By self-tuning capability of Q-learning algorithm, the optimal actions in terms of of coding rate and modulation order for the initial packet transmission is obtained.

Simulation results show that the Q-HARQ can work with the BLER around the QoS requirement of 0.1. 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 BLER. Furthermore, the Q-HARQ can effectively perform link adaptation on account of the channel prediction errors by reinforcement learning process. Finally, the analysis results of the convergence time and the processing time show that the Q-HARQ is feasible to use in real HSDPA systems.

REFERENCES

[1] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*, 3rd ed. JohnWiley & Sons, 2002.
[2] 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.
[3] 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.
[4] 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.
[5] 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.
[6] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8.
[7] S. Haykin, *Neural Networks*, 2nd ed. Prentice Hall, 1999.
[8] L. Gordon, *Principle of Mobile Communication*. Kluwer Academic, 1958.
[9] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *IEE Electron. Lett.*, vol. 27, pp. 2145–2146, Nov. 1991.
[10] 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.
[11] R. Bellman, *Dynamic Programming*. Princeton, 1957.
[12] *3rd Generation mobile system Release 5 specifications*, 3GPP Std. TS 25.214, 2005.