electronics

Article

Post-Cancellation-Based LLR Refining for MIMO Multiple ARQ Systems

Sangjoon Park

School of Electronic Engineering, Kyonggi University, Suwon 16227, Republic of Korea; sj.park@kgu.ac.kr

Abstract: In multiple-input multiple-output (MIMO) multiple automatic repeat request (ARQ) systems, multiple streams with independent hybrid ARQ (HARQ) processes can be simultaneously sent. Thus, the interference from other streams can affect future retransmissions of a packet as well as the current transmission, and proper management of interference at the receiver is required. Therefore, in this paper, a post-cancellation-based log-likelihood ratio (LLR) refining scheme is proposed for MIMO multiple ARQ (MMARQ) systems. In the proposed scheme, after the end of the entire conventional reception procedure for packet decoding, LLR refining is performed for the non-terminated packets that will be sent during the next transmission time interval. For LLR refining, the packet cancellation is performed to cancel only the successfully decoded packets. Thus, the LLRs of the non-terminated packets are refined without any error propagation, including the inter-transmission error propagation. Consequently, the proposed scheme can compensate for the interference problem in MMARQ systems and improve system performance. In order to utilize the error detection results of decoded codewords, the proposed scheme should be performed after the end of the entire reception procedure for packet decoding. Therefore, as the post-processing scheme, the proposed scheme can be employed to any existing LLR-level combining-based MMARQ receiver without changing the original procedure. Simulation results verify that the proposed scheme can significantly improve the error performance and throughput of MMARQ systems, especially for hard-decision interference cancellation-based receivers and high-order modulation. In addition, compared with the conventional reception procedure, the proposed scheme requires a smaller computational complexity in most of the simulated SNR region. Therefore, the proposed LLR refining scheme can be considered as an effective and practical post-processing scheme for an MMARQ receiver.

Keywords: MIMO; MMARQ; HARQ; LLR refining; post-processing; interference cancellation; error propagation

Citation: Park, S. Post-Cancellation-Based LLR Refining for MIMO Multiple ARQ Systems. *Electronics* **2024**, *13*, 200. https://doi.org/10.3390/electronics13010200

Academic Editors: Stefano Scanzio, Dalia Nashaat and Angie Reda Eldamak

Received: 6 December 2023
Revised: 27 December 2023
Accepted: 30 December 2023
Published: 2 January 2024

1. Introduction

The combined technique of forward error correction (FEC) and automatic repeat request (ARQ), hybrid ARQ (HARQ), is an essential technique in wireless communication systems [1–11]. When the wireless channel condition is poor, by jointly using FEC to improve the error correction capability without considering retransmissions, HARQ can reduce the number of retransmissions in comparison to ARQ. On the other hand, when the channel condition is fine, by using ARQ to prepare for possible decoding failures, HARQ can reduce the amount of redundancy in comparison to with FEC. Thus, HARQ-employed systems can improve both the error performance and system throughput over non-HARQ systems, making HARQ to be adopted into modern wireless communication standards [12–15].

Among the recent advances in the literature, in [2], the autonomous retransmission-based HARQ was investigated for multicasting scenarios in wireless sensor networks. In [3], a fast HARQ protocol omitting some feedback signaling and packet decoding was proposed for low-latency reliable communications. Secure transmission and achievable diversity order in HARQ-assisted non-orthogonal multiple-access networks were investigated in [4,5], respectively. Ref. [6] analyzed the outage performance and diversity gain of HARQ schemes in satellite-terrestrial transmissions. Ref. [7] developed a non-orthogonal HARQ mechanism that shares time slots for conducting retransmission in order to achieve reliability and guaranteed packet-level latency, and ref. [8] provided a contemporary survey of HARQ in wireless communications systems and standards. The joint use of variable-power allocation and HARQ for low-earth orbit satellite communication systems was investigated in [9], and the physical-layer security performance of HARQ systems with intelligent reflecting surface was studied in [10]. In [11], an adaptive HARQ scheme based on reinforcement learning for selecting the timing and frequency of retransmission was investigated.

When HARQ is employed to multiple-input multiple-output (MIMO) systems, various applications are possible by using the natures of HARQ retransmissions and MIMO transmission/reception. Thus, many studies have been conducted on MIMO systems with HARQ in recent years [16–30], e.g., the transmission efficiently analysis [16], utilization of the acknowledgment/non-acknowledgment (ACK/NACK) feedback bundling [17], modulation design considering a MIMO-coordinated multi-point scenario [18], optimization of the energy efficiency for a point-to-point massive MIMO system [19], adaptive modulation and coding design [20], integration for vehicle-to-vehicle communications [21], etc.

In particular, it is possible to simultaneously operate multiple HARQ processes using multiple inputs. MIMO multiple ARQ (MMARQ) systems simultaneously operate multiple HARQ processes in antenna domain, e.g., each transmit antenna sends a packet having an independent HARQ process from the other packets with the same time and frequency resources [23–30]. Unlike MIMO single ARQ (MSARQ) systems that operate a single HARQ process (e.g., a single packet) in the antenna domain, e.g., the modulated symbols of a packet are spatially multiplexed and sent together from all transmit antennas [23], MMARQ systems can benefit from a partial decoding success, i.e., a situation where some of the packets sent during a given transmission time interval (TTI) are successfully decoded. This provides the MMARQ systems a potential performance gain over the MSARQ systems [24,27].

Among the studies for MMARQ, in [24], joint and separate detection algorithms for both MMARQ and MSARQ systems were developed when the channel state information (CSI) is available only at the receiver. In [25], combining techniques for multi-user systems were exploited according to the availability of packet blanking and codeword cancellation. In [26], a combining technique adapting the structure of space–time block codes was investigated. The performance characteristics of MMARQ systems according to the combining scheme at the receiver were analyzed in [27], and a hybrid combining scheme based on the interference-to-noise reformulation was developed in [28]. In [29], a shared HARQ scheme using the piggyback technique was proposed for overloaded MIMO systems. In [30], a Kalman filtering-based combining scheme applicable for both MMARQ and MSARQ regardless of the employed HARQ retransmission strategy was investigated, which was extended to perform iterative detection and decoding for MSARQ systems in [22].

Because MMARQ systems are essentially based on spatial multiplexing, multiple streams (e.g., symbols or packets) with independent HARQ processes are simultaneously sent. This implies that the interference from other streams can affect future retransmissions of a packet as well as the current transmission in MMARQ systems. Therefore, proper management of interference at the receiver is required in order for improving the performance of MMARQ systems.

For the interference management at the receiver, interference cancellation (IC) can be employed in MMARQ systems. A typical hard-decision IC-based receiver, such as successive hard-decision IC (SHIC) [31,32] or iterative hard-decision IC (IHIC) [33,34], is known to achieve an improved error performance compared with the simple linear detection-based receiver without IC in spatially multiplexed MIMO systems without HARQ. However, by employing IC operations, there is a possibility that the error occurred in a previously detected stream can be propagated to the remaining streams to be detected later [35]. This fundamental problem in IC, known as the error propagation, becomes more significant in MIMO systems with HARQ as compared to the MIMO systems without HARQ. That is, once an incorrect IC operation occurs, the error can be propagated not only to the other streams in the current TTI but also to the streams in the future TTIs because of retransmissions, e.g., the log-likelihood ratios (LLRs) with the propagated errors obtained in the past TTIs are used for decoding in the future TTIs by combining LLRs. Thus, after the occurrence of an incorrect IC operation during a given TTI, the effect of the interference from other streams in MMARQ systems becomes more significant via error propagation, which can cause the performance of MMARQ systems in future TTIs to be severely degraded.

This error propagation in hard-decision IC can be compensated by using soft-decision IC [22,36], i.e., the IC operation is performed by the soft estimate of a stream based on the estimated reliability of the stream. However, the computational complexity of soft-decision IC is considerably larger than hard-decision IC. In addition, if hard-decision IC is used with a joint detection approach such as symbol-level combining [24,25,27], the inter-transmission error propagation can be eliminated by the LLR recalculation for all the remaining packets in every TTI. However, because of the LLR recalculation, this approach can also require large computational complexity, especially as the modulation order increases. Meanwhile, the linear detection-based receiver does not suffer from the error propagation, but it cannot manage the interference from the nature of MMARQ systems.

Therefore, in this paper, an LLR refining scheme is proposed for MMARQ systems. The proposed scheme is performed after the conventional reception procedure for packet decoding in MMARQ systems, i.e., after the end of the decoding of all the transmitted packets and decision of the ACK/NACK feedback. In the proposed LLR refining scheme, the packet cancellation is performed using the successfully decoded packets in the current TTI, and the LLR update of non-terminated packets are performed for LLR refining, where the non-terminated packets indicate the packets that are neither successfully decoded nor terminated in the current TTI and thereby will be retransmitted during the next TTI.

The main characteristics of the proposed scheme can be summarized as below:

- [Performance] The packet cancellation in the proposed scheme is performed only with the successfully decoded packets. Thus, even if the LLRs of the non-terminated packets are contaminated by incorrect IC during the prior reception procedure for decoding, the refined LLRs without any error propagation (i.e., inter-packet and inter-transmission error propagation) can be stored in the buffer for decoding during future TTIs. In this way, the proposed LLR refining scheme can compensate for the interference problems in MMARQ systems. This enables the proposed scheme to improve both the error performance and throughput of MMARQ systems, as verified in numerical simulations.
- [Practicality] The proposed LLR refining scheme is a post-processing scheme for the receiver. That is, the proposed scheme is performed after the end of the decoding for all transmitted packets in a given TTI. Therefore, the proposed scheme can be employed with any conventional LLR-level combining-based receivers for MMARQ systems as the post-processing scheme, e.g., linear detection or hard-decision IC-based LLR-level combining receivers. In addition, the proposed scheme can be utilized regardless of the HARQ retransmission strategy, e.g., Chase combining or incremental redundancy. Thus, the proposed scheme is suitable for practical MMARQ systems.
- [Complexity] Although employing the proposed scheme to the conventional receiver requires additional complexity, the required computational complexity is smaller than that of the detection and LLR calculation of the conventional reception procedure for decoding. In addition, no computational operations are performed for the proposed scheme when there is no possibility of error propagation or there does not exist any packet that benefits from LLR refining. Consequently, the proposed scheme can provide LLR refining with a small amount of extra complexity for the conventional MMARQ receiver, as verified in numerical simulations.

It is worthwhile to mention that the concept of the post-processing in this paper is different from those in the existing literature for MIMO systems [37–39]. The existing concepts refer to a certain procedure performed relatively later during the whole transmission and/or reception process, i.e., requiring the design of a new process. On the other hand, the proposed scheme is employed to the receiver after the end of the conventional reception procedure, i.e., without modification of the existing procedure. To the best of the author's knowledge, there have been no studies on similar post-processing approaches for the LLR refining and/or improvement in HARQ employed systems.

This paper is organized as follows. Section 2 describes the MMARQ system model considered in this paper. Section 3 describes the proposed LLR refining scheme, including the computational complexity analysis in detail, and Section 4 shows various numerical simulation results to verify the effectiveness of the proposed scheme. Finally, Section 5 makes conclusions.

Throughout this paper, the following notation is used. Vectors and matrices are denoted by lowercase and uppercase boldface letters, respectively. The superscripts T, H, and -1 represent the transpose, conjugate-and-transpose, and inverse operations, respectively. $\mathbf{I}_j$ is the $j \times j$ identity matrix. $[\mathbf{A}]_{i,j}$ denotes the element of the matrix $\mathbf{A}$ at the ith row and jth column. In addition, $[\mathbf{A}]_{i,:}$ and $[\mathbf{A}]_{:,i}$ denote the ith row and column of the matrix $\mathbf{A}$, respectively. Finally, $\mathrm{diag}(\mathbf{A})$ is a vector where the diagonal elements of the matrix $\mathbf{A}$ and $|\mathcal{B}|$ are the number of elements in the set $\mathcal{B}$.

2. MMARQ System Model

In Figure 1, the overall MMARQ system model is illustrated. We consider a multi-packet transmission system with N transmit and M receive antennas, where N packets are simultaneously sent during a given TTI. Each packet is sent from a different transmit antenna and has an independent HARQ process. For each packet, K data bits are encoded to the mother codeword of a length J, i.e., a code rate of K/J, where the encoding procedure includes the error detection encoding such as cyclic redundancy check (CRC) codes and error correction encoding such as low-density parity-check (LDPC) codes [40]. Then, using a Q-ary constellation $\mathcal{S}$, the mother codeword is modulated to the symbol sequence of a length L, i.e., the modulation order Q satisfies $\log_2 Q = J/L$. The elements in $\mathcal{S}$ satisfy $\sum_{s \in \mathcal{S}} s = 0$ and $\sum_{s \in \mathcal{S}} s^2 = Q$, e.g., quadrature amplitude modulation (QAM).

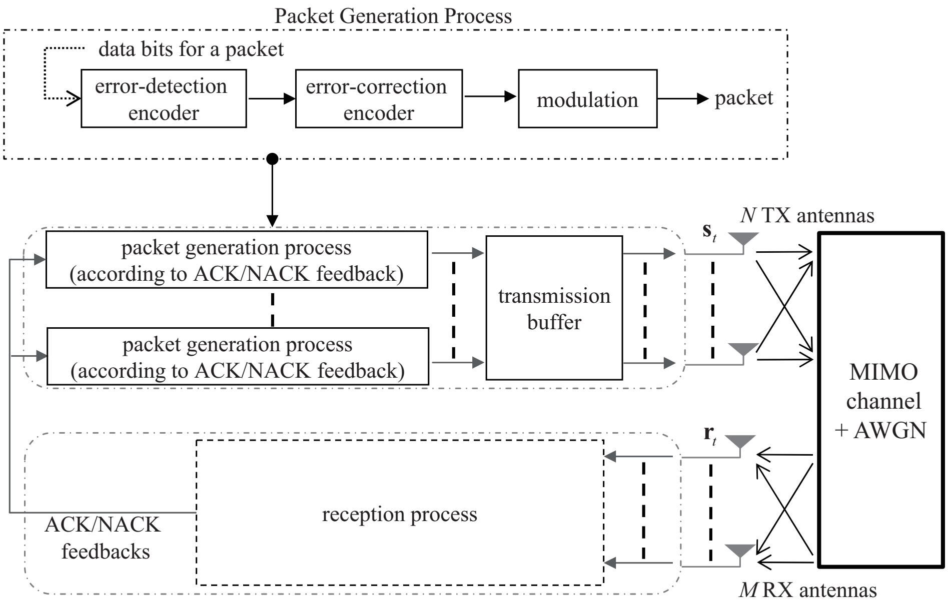

Figure 1. MMARQ system model block diagram.

For simplicity, Chase combining [1] is considered to be the HARQ retransmission strategy. That is, the identical symbol sequence generated from the mother codeword is repeatedly transmitted from the first to the Rth HARQ rounds of a packet, where R is the maximum HARQ round, i.e., the maximum number of transmissions for a packet. Let $\mathbf{s}_{l,t} = [s_{l,t}(1), \cdots, s_{l,t}(N)]^T$ be the $l(1 \leq l \leq L)$th $N \times 1$ transmit signal vector for the tth TTI, where $s_{l,t}(n)$ for $1 \leq n \leq N$ is the lth symbol of the packet sent from the nth transmit antenna during the tth TTI. Without loss of generality, the index l is omitted throughout the remainder of this paper. Therefore, $\mathbf{s}_t = [s_t(1), \cdots, s_t(N)]^T$ denotes the $N \times 1$ transmit signal vector for the tth TTI, where $s_t(n)$ for $1 \leq n \leq N$ is the symbol of the packet sent from the nth transmit antenna during the tth TTI.

Based on the above definitions and modeling, the receive signal vector for the tth TTI can be written as

$$\mathbf{r}_t = \mathbf{H}_t \mathbf{s}_t + \mathbf{w}_t, \tag{1}$$

where $\mathbf{r}_t$ is the $M \times 1$ receive signal vector, $\mathbf{w}_t$ is the $M \times 1$ additive white Gaussian noise (AWGN) vector with zero mean and variance σ^2, and $\mathbf{H}_t$ is the $M \times N$ channel matrix.

Let $r_{t,n}$ denote the HARQ round of the packet sent from the nth transmit antenna during the tth TTI. If $r_{t,n} = R$, i.e., the maximum HARQ round, the packet is terminated regardless of the error detection result and ACK/NACK feedback. In this case, a new packet will be sent from the nth transmit antenna during the next $(t+1)$th TTI, i.e., $r_{t+1,n} = 1$. Further, if the packet from the nth transmit antenna is successfully decoded (i.e., no errors are detected in the error detection operation at the receiver) during the tth TTI, a new packet will be sent from the nth transmit antenna during the next $(t+1)$th TTI as well, i.e., $r_{t+1,n} = 1$, regardless of $r_{t,n}$. Meanwhile, if $r_{t,n} < R$ and the packet sent from the nth transmit antenna fails at decoding (i.e., errors are detected in the error detection operation at the receiver) during the tth TTI, the same packet will be retransmitted during the next $(t+1)$th TTI, i.e., $r_{t+1,n} = r_{t,n} + 1$.

3. Proposed Post-Cancellation-Based LLR Refining Scheme

3.1. Conventional MMARQ Receiver

In this subsection, we describe the conventional LLR-level combining-based MMARQ receiver for the system model given in Section 2, where the LLR combining is used as the packet combining strategy for retransmission. For simplicity, it is assumed that the full CSI is available at the receiver so that the channel estimation operation is omitted during the reception procedure for packet decoding.

The conventional LLR-level combining-based MMARQ receiver can include LLR calculation and recalculation; LLR combining, decoding, error detection, and ACK/NACK generation; and buffering operations [27,28,30]. During the LLR calculation and recalculation operation, the LLRs of the coded bits of each packet are calculated from the symbols estimated by using (1), where linear zero-forcing (LZF) or linear minimum mean-squared-error (LMMSE) detection with or without IC operation can be used for the symbol estimation. Then, during the LLR combining operation, the LLRs of a packet calculated from the previous operation are combined with the LLRs of the packet stored in the buffer, which were calculated from the previous TTIs (i.e., previous HARQ rounds of the packet). These combined LLRs are used as input for decoding, where the output of the decoder can be utilized for the LLR calculation and recalculation operation according to the implementation, e.g., iterative processing such as IHIC. After the decoding of all the transmitted packets, error detection is performed to determine the decoding success of each packet, and the error detection results are used for ACK/NACK feedback generation. Finally, the LLRs of the packets terminated at the current TTI (i.e., successfully decoded or reaching the maximum HARQ round R) are cleared from the buffer, and the calculated LLRs of the non-terminated packets that will be retransmitted during the next TTI (i.e., detected in errors with the HARQ round smaller than R) are stored in the buffer.

The detailed reception procedure of an LLR-level combining-based MMARQ receiver can be different according to the implemented method. We consider the linear detection-based receiver, coded SHIC (CSHIC)-based receiver, and coded IHIC (CIHIC)-based receiver. For CSHIC and CIHIC, the hard-decision IC operation is performed using the decoded codeword(s) of other packet(s).

The reception procedure of the linear detection-based receiver is illustrated in Figure 2. The LLRs of the coded bits of the packets for the current TTI are initially calculated, and the calculated LLRs are combined with the LLRs stored in the buffer. The decoding operation is performed using the combined LLRs, and the error detection and corresponding ACK/NACK feedback generation are performed after the decoding operation. Further, the initially calculated LLRs of the non-terminated packets (i.e., the packets that will be retransmitted during the next TTI) are stored in the buffer.

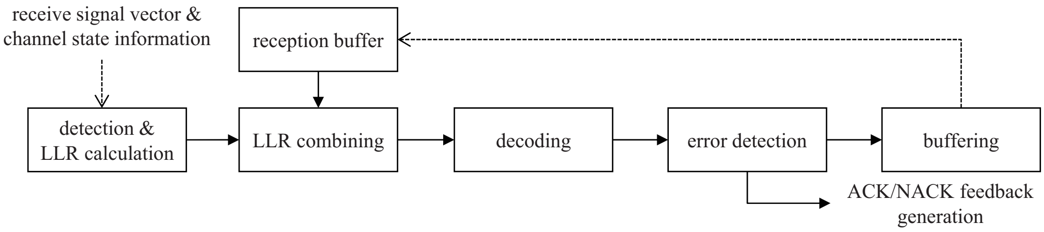

Figure 2. The reception procedure of the linear detection-based receiver.

Next, the reception procedures of the CSHIC- and CIHIC-based receivers are illustrated in Figures 3 and 4, respectively. Unlike linear detection, by using IC, the decoding result of other packets is utilized for the LLR calculation of CSHIC and CIHIC. In CSHIC, according to the pre-determined order of each packet, packets are sequentially detected and decoded, while IC of previously decoded packets is performed prior to the detection and LLR calculation. Meanwhile, in CIHIC, packets are iteratively detected and decoded, while IC of the other packets is performed prior to the detection and LLR calculation. In this way, for CSHIC and CIHIC, IC using the decoded codeword(s) of packet(s) can be performed prior to the LLR calculation. Then, the calculated LLRs are combined with the LLRs stored in the buffer, and the decoding operation is performed using the combined LLRs. After the end of the decoding for all packets, e.g., successive decoding of all packets in CSHIC and reaching the maximum number of outer iterations in CIHIC, the error detection, ACK/NACK feedback generation, and buffering operations are performed as in the linear detection-based receivers.

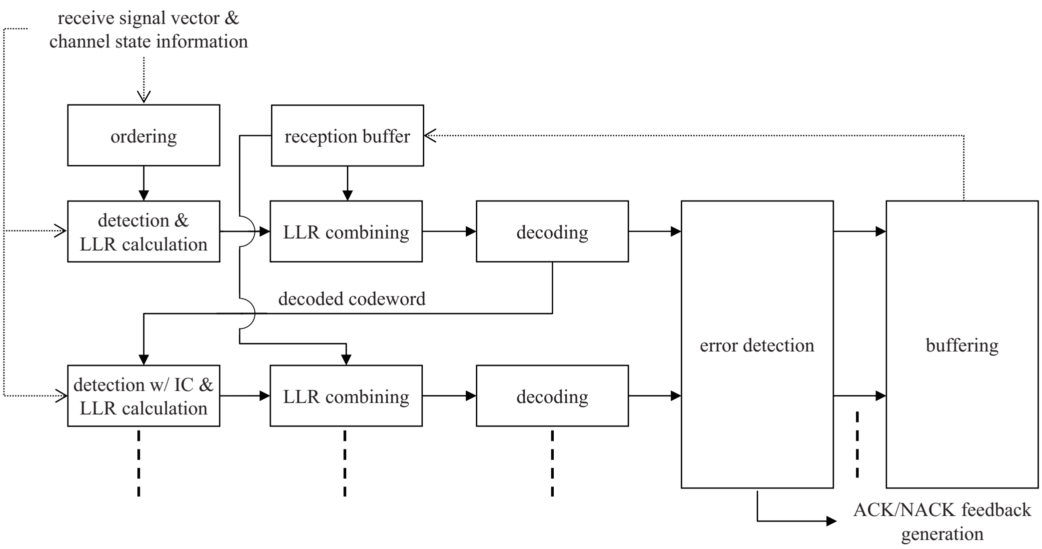

Figure 3. The reception procedure of the CSHIC-based receiver.

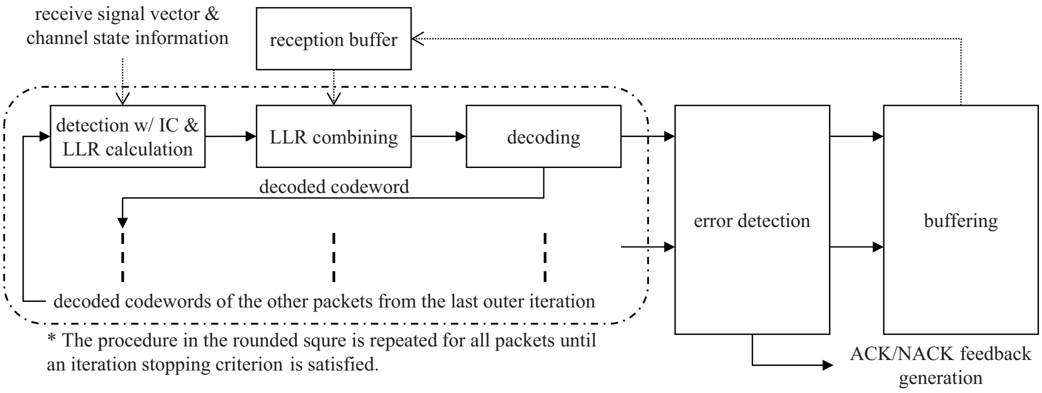

Figure 4. The reception procedure of the CIHIC-based receiver.

3.2. Proposed LLR Refining Scheme

In this subsection, the proposed LLR refining scheme for MMARQ systems is explained in detail. To describe the procedure of the proposed scheme, we define the following notation. Let $e_{t,n}$ for $1 \leq n \leq N$ be the error detection result of the packet sent from the nth transmit antenna during the tth TTI, where $e_{t,n} = 0$ and 1 denote the cases without or with errors, respectively. Then, let $\mathcal{N}_{t,\text{s}}$ be the set including the transmit antennas of the packets with $e_{t,n} = 0$ (i.e., successfully decoded packets), and let $\mathcal{N}_{t,\text{n}}$ be the set including the transmit antennas of the packets with $e_{t,n} = 1$ and $r_{t,n} < R$ (i.e., non-terminated packets). Further, let $N_{t,\text{s}}$ and $N_{t,\text{n}}$ be the size of $\mathcal{N}_{t,\text{s}}$ and $\mathcal{N}_{t,\text{n}}$, respectively, i.e., $N_{t,\text{s}} = |\mathcal{N}_{t,\text{s}}|$ and $N_{t,\text{n}} = |\mathcal{N}_{t,\text{n}}|$.

Figure 5 illustrates the block diagram of the proposed LLR refining scheme. The proposed scheme consists of initialization, packet cancellation, LLR update, and buffering stages. If there is a non-terminated packet retransmitted in the current TTI, i.e., $r_{t,n} > 1$, the LLR refining for the packet can also be performed for the previous HARQ rounds of the packet. Thus, to enable LLR refining for the previous HARQ rounds of the non-terminated packets, $D(0 \leq D \leq R-1)$ is defined as the maximum depth of the proposed LLR refining scheme, i.e., LLR refining can be performed from the $(r_{t,n} - D + 1)$th to the $r_{t,n}$th HARQ rounds of each non-terminated packet at its $r_{t,n}$th HARQ round during the tth TTI. Because the non-terminated packets should be retransmitted during the future TTIs, their HARQ rounds in the current TTI are at a maximum of $(R-1)$. Consequently, D should be set to $(R-1)$ at maximum. Further, $D = 0$ implies the cases of no LLR refining, because it is impossible to deal with the $(r_{t,n} + 1)$th HARQ round of a packet when the packet is at its $r_{t,n}$th HARQ round.

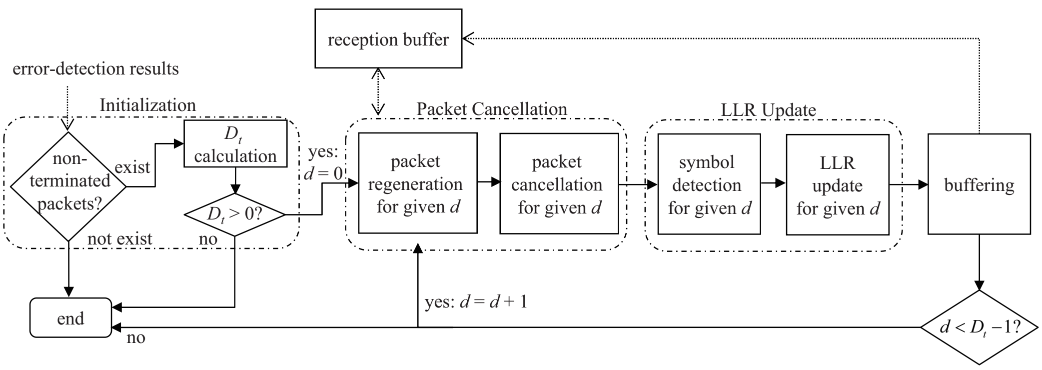

Figure 5. The block diagram of the proposed LLR refining scheme.

In the proposed LLR refining scheme, the initialization stage is performed prior to the following stages. First, an initial condition check is performed to find out whether the initial condition regarding the necessity of the LLR refining is satisfied or not. Because the LLR refining is performed only for non-terminated packets, there should be non-terminated packets that will be sent during the next TTI. Therefore, the value of $N_{t,\mathrm{n}}$ (the number of the non-terminated packets) is checked, and the following procedures are performed only when $N_{t,\mathrm{n}}$ is positive, i.e., $N_{t,\mathrm{n}} > 0$. If $N_{t,\mathrm{n}} = 0$, no additional procedures are performed for the LLR refining during the current tth TTI.

If the initial condition of $N_{t,\mathrm{n}} > 0$ is satisfied, the depth for the current tth TTI, D_t, can be determined as

$$D_t = \min(D, D_{t,\mathrm{p}}), \tag{2}$$

where $D_{t,\mathrm{p}}$ is the depth of the proposed LLR refining scheme for the tth TTI calculated by only the HARQ rounds of packets, which can be obtained as

$$D_{t,\mathrm{p}} = \min(\max_{n \in \mathcal{N}_{t,\mathrm{s}}} r_{t,n}, \max_{n \in \mathcal{N}_{t,\mathrm{n}}} r_{t,n}). \tag{3}$$

In (3), $\max_{n \in \mathcal{N}_{t,s}} r_{t,n}$ and $\max_{n \in \mathcal{N}_{t,\mathrm{n}}} r_{t,n}$ denote the maximum values of the HARQ rounds of the successfully decoded packets and non-terminated packets at the tth TTI, respectively. $\max_{n \in \mathcal{N}_{t,s}} r_{t,n}$ represents the depth at where the cancellation of the successfully decoded packets is possible (i.e., the cancellation gain is provided), and $\max_{n \in \mathcal{N}_{t,\mathrm{n}}} r_{t,n}$ represents the depth at where the LLRs of the non-terminated packets can be refined. Therefore, $D_{t,\mathrm{p}}$ is the practical depth considering the HARQ rounds of the packets in $\mathcal{N}_{t,\mathrm{s}}$ and $\mathcal{N}_{t,\mathrm{n}}$, and the final depth for the current tth TTI D_t can be determined by $D_{t,\mathrm{p}}$ and the system parameter D as in (2).

Note that if there are no successfully decoded packets in the current TTI, i.e., $\mathcal{N}_{t,\mathrm{s}} = \emptyset$ and $N_{t,\mathrm{s}} = 0$, instead of (2) and (3), both $D_{t,\mathrm{p}}$ and D_t should be set to 1 for IC, e.g., CSHIC and CIHIC. That implies that the LLR refining of the non-terminated packets for their current HARQ rounds should be performed for IC in order to eliminate the inter-packet error propagation occurring in the reception procedure for decoding during the current TTI. In addition, if $N_{t,\mathrm{s}} = 0$, and no error propagation occurs in the reception procedure for decoding, e.g., linear detection-based receiver without IC, $D_{t,\mathrm{p}}$ and D_t are set to 0 to avoid unnecessary computational operations, i.e., the refined LLRs will be identical to the initially calculated LLRs obtained in the reception procedure for decoding. Therefore, when $D_{t,\mathrm{p}}$ and D_t are 0, no additional procedures are performed for the LLR refining during the current tth TTI, as in the case of $N_{t,\mathrm{n}} = 0$

After the decision of D_t, the remaining packet cancellation, LLR update, and buffering stages are repeatedly performed for $0 \leq d \leq D_t - 1$. According to d, the LLRs for the $(r_{t,n} - d)$th HARQ round of a non-terminated packet are refined if $r_{t,n} - d > 0$. For a given d, the remaining stages are sequentially performed, and, after the end of the stages for the current d, the stages for the next depth (e.g., $d + 1$) are performed. This is repeated for all d with $0 \leq d \leq D_t - 1$.

At the beginning, for a given d, the packet cancellation stage is performed, where the successfully decoded packets, i.e., the packets with $e_{t,n} = 0$, are canceled from the receive signal vector $\mathbf{r}^*_{t-d}$. When $d = 0$ (i.e., the current TTI), $\mathbf{r}^*_{t-d} = \mathbf{r}_t$. Otherwise, if $d > 0$ (i.e., previous TTIs), $\mathbf{r}^*_{t-d}$ stored in the buffer is used.

During the packet cancellation stage, packet regeneration is performed at first. Because only the packets decoded successfully at the current tth TTI are canceled, the transmit symbols of the packets with $e_{t,n} = 0$ should be regenerated from their decoded code words. Let $\tilde{\mathbf{s}}^t_{t-d}$ be the $N \times 1$ transmit signal vector regenerated from the decoded codewords for the given depth d at the tth TTI. If $e_{t,n} = 1$ (i.e., decoding errors) or $r_{t,n} \leq d$ (i.e., not sent during the $(t-d)$th TTI), the nth element of $\tilde{\mathbf{s}}^t_{t-d}$, $\tilde{s}^t_{t-d}(n)$ is 0. Otherwise, if $e_{t,n} = 0$ and $r_{t,n} > d$, $\tilde{s}^t_{t-d}(n)$ is the symbol regenerated from a decoded code word of the packet from the nth transmit antenna, where the decoded codeword is obtained from the reception procedure for decoding during the tth TTI.

Using the regenerated symbols in $\tilde{\mathbf{s}}^t_{t-d}$, the packet cancellation can be performed as

$$\mathbf{r}^*_{t-d} = \mathbf{r}^*_{t-d} - \mathbf{H}^*_{t-d}\tilde{\mathbf{s}}^t_{t-d}. \tag{4}$$

In (4), $\mathbf{H}^{*}_{t-d}$ is the $M \times N$ channel matrix for the $(t-d)$th TTI stored in the buffer, where the nth column of $\mathbf{H}^{*}_{t-d}$ is all-zero if the packet sent from the nth transmit antenna during the $(t-d)$th TTI was already canceled during the past LLR refining procedures already performed from the $(t-d)$th to the $(t-1)$th TTIs. This is similar to the case of $\mathbf{r}^{*}_{t-d}$: $\mathbf{H}^{*}_{t-d} = \mathbf{H}_t$ if $d = 0$.

$\tilde{\mathbf{s}}^{t}_{t-d}$ in (4) includes only the regenerated symbols for the packets successfully decoded during the tth TTI and sent during the $(t-d)$th TTI. Thus, assuming perfect error detection, the effective $M \times 1$ receive signal vector $\tilde{\mathbf{r}}^{*}_{t-d}$ for the proposed scheme does not include any error propagation, including the inter-transmission error propagation for MMARQ systems. This assumption for error detection can be justified by employing the powerful error detection codes used in the practical systems [12–15]. Meanwhile, this packet cancellation operation can be omitted when $D_t = 1$ and $N_{t,s} = 0$ because $\tilde{\mathbf{s}}^{t}_{t-d}$ becomes the all-zero vector in this case.

After (4), the columns in $\mathbf{H}^{*}_{t-d}$ corresponding to the regenerated non-zero symbols in $\tilde{\mathbf{s}}^{t}_{t-d}$ are all set to zero. This is the end of the packet cancellation stage.

Next, the LLR update stage is performed for the non-terminated packets sent during the tth TTI, which were also sent during the $(t-d)$th TTI. For that, first, the estimates for the non-terminated packets in $\mathcal{N}_{t,\mathrm{n}}$, which were also sent during the $(t-d)$th TTI (i.e., the packets with $r_{t-d,n} \geq 1$ at the $(t-d)$th TTI), are obtained from $\tilde{\mathbf{r}}^{*}_{t-d}$ using the detection method employed in the convention reception procedure for decoding, e.g., LZF and LMMSE. Let $\overline{\mathbf{H}}^{*}_{t-d}$ denote the $M \times N^{t}_{t-d}$ matrix identical to $\mathbf{H}^{*}_{t-d}$ except the all-zero columns, where N^{t}_{t-d} is the number of the non-all-zero columns in $\mathbf{H}^{*}_{t-d}$. $\overline{\mathbf{H}}^{*}_{t-d}$ includes the channel responses for the packets which were sent during the $(t-d)$th TTI and are not decoded successfully until the decoding during the current tth TTI. Thus, $\overline{\mathbf{H}}^{*}_{t-d}$ includes the channel responses for the packets terminated with decoding failures before the tth TTI (e.g., the packets not sent during the tth TTI). Then, the estimates can be obtained from $\mathbf{r}^{*}_{t-d}$ using $\overline{\mathbf{H}}^{*}_{t-d}$ as

$$\bar{\mathbf{s}}^{t}_{t-d} = \begin{cases} \left(\overline{\mathbf{H}}^{*\,H}_{t-d}\overline{\mathbf{H}}^{*}_{t-d}\right)^{-1}\overline{\mathbf{H}}^{*\,H}_{t-d}\mathbf{r}^{*}_{t-d} & \text{for LZF} \\ \left(\overline{\mathbf{H}}^{*\,H}_{t-d}\overline{\mathbf{H}}^{*}_{t-d} + \sigma^2\mathbf{I}_{N^{t}_{t-d}}\right)^{-1}\overline{\mathbf{H}}^{*\,H}_{t-d}\mathbf{r}^{*}_{t-d} & \text{for LMMSE} \end{cases}. \tag{5}$$

Although $\bar{\mathbf{s}}^{t}_{t-d}$ in (5) is the $N^{t}_{t-d} \times 1$ vector, the LLRs of the coded bits for all N^{t}_{t-d} symbols do not need to be updated when there are some packets terminated during the previous TTIs with errors and not sent during the tth TTI. Thus, only the LLRs of the coded bits for the symbols corresponding to the current non-terminated packets need to be updated. Let $l^{t,n}_{t-d}(q)$ denote the refined LLR of the $q(1 \leq k \leq \log_2 Q)$th coded bit of the packet for its $r_{t-d,n}$th HARQ round, where the packet is sent from the nth transmit antenna during the tth TTI and $\log_2 Q$ is the number of bits assigned for a modulated symbol. Then, for a given n corresponding to one of the current non-terminated packets, $l^{t,n}_{t-d}(q)$ can be calculated for $1 \leq q \leq \log_2 Q$ as [22,30]

$$l^{t,n}_{t-d}(q) = \ln \frac{\sum_{\forall s \in \mathcal{S}^1_q} \exp\left(\frac{-\left|\bar{s}^{t}_{t-d}(\mathrm{f}^{t,n}_{t-d}) - \mu^{t,n}_{t-d}s\right|^2}{\eta^{t,n}_{t-d}}\right)}{\sum_{\forall s \in \mathcal{S}^0_q} \exp\left(\frac{-\left|\bar{s}^{t}_{t-d}(\mathrm{f}^{t,n}_{t-d}) - \mu^{t,n}_{t-d}s\right|^2}{\eta^{t,n}_{t-d}}\right)}. \tag{6}$$

In (6), $\mathcal{S}^b_q$ with $b \in \{0,1\}$ is the subset of $\mathcal{S}$, where the qth bit of the elements in $\mathcal{S}^b_q$ is b, $\mathrm{f}^{t,n}_{t-d}$ is the column index in $\overline{\mathbf{H}}^{*}_{t-d}$ corresponding to the channel response for the packet sent from the nth transmit antenna during the tth TTI, and $\bar{s}^{t}_{t-d}(\mathrm{f}^{t,n}_{t-d})$ is the $\mathrm{f}^{t,n}_{t-d}$th element of $\bar{\mathbf{s}}^{t}_{t-d}$, which is the estimated symbol in $\bar{\mathbf{s}}_{t-d}^{t}$ for the packet sent from the nth transmit antenna during the tth TTI . Further, $\mu_{t-d}^{t,n}$ and $\eta_{t-d}^{t,n}$ in (6) can be calculated as [27,35,36]

$$\mu_{t-d}^{t,n} = \begin{cases} 1 & \text{for LZF} \\ [\mathbf{F}_{\text{LMMSE},t-d}^{t,n}]_{f_{t-d}^{t,n},:} \cdot [\overline{\mathbf{H}}_{t-d}^{*}]_{:,f_{t-d}^{t,n}} & \text{for LMMSE} \end{cases} \quad (7)$$

and

$$\eta_{t-d}^{t,n} = \begin{cases} \sigma^2 \left[\left(\overline{\mathbf{H}}_{t-d}^{*}{}^{H} \overline{\mathbf{H}}_{t-d}^{*} \right)^{-1} \right]_{f_{t-d}^{t,n}, f_{t-d}^{t,n}} & \text{for LZF} \\ \mu_{t-d}^{t,n}(1-\mu_{t-d}^{t,n}) & \text{for LMMSE} \end{cases}, \quad (8)$$

where $\mathbf{F}_{\text{LMMSE},t-d}^{t,n} = (\overline{\mathbf{H}}_{t-d}^{*}{}^{H} \overline{\mathbf{H}}_{t-d}^{*} + \sigma^2 \mathbf{I}_{N_{t-d}^{t}})^{-1} \overline{\mathbf{H}}_{t-d}^{*}{}^{H}$ in (5).

If the max-log approximation is used, (6) can be simplified as [22,36]

$$l_{t-d}^{t,n}(q) = \min_{\forall s \in \mathcal{S}_q^0} \frac{|\bar{s}_{t-d}^{t}(f_{t-d}^{t,n}) - \mu_{t-d}^{t,n} s|^2}{\eta_{t-d}^{t,n}} - \min_{\forall s \in \mathcal{S}_q^1} \frac{|\bar{s}_{t-d}^{t}(f_{t-d}^{t,n}) - \mu_{t-d}^{t,n} s|^2}{\eta_{t-d}^{t,n}}. \quad (9)$$

After the LLR update, during the buffering stage, all the refined LLRs $l_{t-d}^{t,n}(q)$ are stored in the buffer to replace the existing LLRs of the corresponding packets for their $(r_{t,n}-d)$th HARQ round. Further, when $d = R-2$ (the maximum d because the maximum possible D is $R-1$), $\mathbf{r}_{t-d}^{*}$ and $\mathbf{H}_{t-d}^{*}$ will not be used for the proposed scheme from the next $(t+1)$th TTI, because all the non-terminated packets sent in the $(t-d)$th TTI should be terminated until the next $(t+1)$th TTI by reaching R. Thus, if $d < R-2$, the obtained $\mathbf{r}_{t-d}^{*}$ and $\mathbf{H}_{t-d}^{*}$ are stored in the buffer. Otherwise, if $d = R-2$, no $\mathbf{r}_{t-d}^{*}$ and $\mathbf{H}_{t-d}^{*}$ need to be stored in the buffer. This is the end of the LLR refining procedure for the given d.

Table 1 summarizes the procedure of the proposed LLR refining scheme. The proposed scheme can benefit from the cancellation of successfully decoded packets only. Thus, the proposed post-cancellation-based LLR refining scheme prevents error propagation for the conventional hard-decision IC-based receivers and provides cancellation gain for the conventional linear-detection-based receiver. Consequently, the proposed scheme can compensate for the interference problems in MMARQ systems and improve both the error performance and throughput of MMARQ systems. In addition, the proposed scheme is a post-processing scheme that should be employed after the conventional reception procedure for decoding, while no modification of the original reception procedure for decoding is required. Therefore, the proposed scheme can be employed to any conventional LLR-level combining receivers for MMARQ systems.

In Table 2, the computational complexity of the proposed LLR refining scheme per receive signal vector is shown and compared with those of the conventional receivers for MMARQ systems. For each scheme, the computational complexities of the IC (e.g., (4)), detection (e.g., (5)), and LLR calculation (e.g., (6)–(9)) operations are calculated, and the higher-order terms for each scheme are considered. Note that the complexity for the ordering in CSHIC is not considered in Table 2.

The computational complexity burden of the proposed scheme comes from the detection ((4) and (5)) and LLR update ((6)–(9)) operations. Because of the multiplication of the $M \times N$ matrix $\mathbf{H}_{t-d}^{*}$ and the $N \times 1$ vector $\tilde{\mathbf{s}}_{t-d}^{t}$ with $N_{t,s}$ non-zero elements at maximum, (4) requires a complexity of $\mathcal{O}(MN_{t,s})$ at maximum. Further, because the size of the effective channel matrix $\overline{\mathbf{H}}_{t-d}^{*}$ is $M \times N_{t-d}^{t}$, (5) requires an identical complexity to linear detection in $M \times N_{t-d}^{t}$ MIMO systems requiring the multiplication of $N_{t-d}^{t} \times M$ and $M \times N_{t-d}^{t}$ matrices and the inverse of the $N_{t-d}^{t} \times N_{t-d}^{t}$ matrix, which yields a complexity of $\mathcal{O}((N_{t-d}^{t})^3 + (N_{t-d}^{t})^2 M)$. Therefore, the proposed scheme requires a complexity of $\mathcal{O}(\sum_{d=0}^{D_t-1} ((N_{t-d}^{t})^3 + (N_{t-d}^{t})^2 M))$ for detection per receive signal vector for a given TTI. Note that $N_{t-d}^{t} \leq N$ because of the successfully decoded packets sent from the $(t-d)$th to the tth TTIs.

Table 1. Summary of Proposed LLR Refining Procedure.

[0. Initialization] Perform the following: (0-A) If $N_{t,\text{n}} > 0$, proceed to the next step. Otherwise, if $N_{t,\text{n}} = 0$, stop the LLR refining procedure for the current tth TTI. (0-B) If $N_{t,\text{s}} = 0$ and there is no possibility of error propagation during the current reception procedure for decoding, stop the LLR refining procedure for the current tth TTI. Otherwise, if $N_{t,\text{s}} = 0$ and there is a possibility of the error propagation, set the depth D_t for the current TTI to 1. Otherwise, if $N_{t,\text{s}} > 0$, calculate D_t as in (2) and (3).
From $d = 0$ to $d = D_t - 1$, perform the following:
[1. Packet Cancellation] Perform the following: (1-A) Generate an $N \times 1$ vector $\tilde{\mathbf{s}}_{t-d}^t$. The nth element of $\tilde{\mathbf{s}}_{t-d}^t$, $\tilde{s}_{t-d}^t(n)$, is zero if $e_{t,n} = 1$ or $r_{t,n} \le d$. Otherwise, if $e_{t,n} = 0$ and $r_{t,n} > d$, $\tilde{s}_{t-d}^t(n)$ is the regenerated symbol from the decoded codeword of the packet sent from the nth transmit antenna, where the decoded codeword is obtained from the reception procedure during the tth TTI. (1-B) Cancel the regenerated transmit signal vector $\tilde{\mathbf{s}}_{t-d}^t$ from $\mathbf{r}_{t-d}^*$ as in (4). (1-C) Set the columns in $\mathbf{H}_{t-d}^*$ corresponding to the regenerated symbols in $\tilde{\mathbf{s}}_{t-d}^t$ to all-zero.
[2. LLR Update] Perform the following: (2-A) Generate $\overline{\mathbf{H}}_{t-d}^*$ from $\mathbf{H}_{t-d}^*$, which contains the channel responses of the packets sent during the $(t-d)$th TTI and remained (not canceled until the current TTI) in $\tilde{\mathbf{r}}_{t-d}^*$. (2-B) Obtain an $N_{t-d}^t \times 1$ vector $\overline{\mathbf{s}}_{t-d}^t$ as in (5). (2-C) Calculate the refined LLR $l_{t-d}^{t,n}(q)$ $(1 \le q \le \log_2 Q)$ for the non-terminated packets with $n \in \mathcal{N}_{t,\text{n}}$ and sent during the $(t-d)$th TTI, as in (6)–(9).
[3. Buffering] The refined LLRs for non-terminated packets (i.e., $n \in \mathcal{N}_{t,\text{n}}$) sent during the $(t-d)$th TTI are stored to replace the previous LLRs in the buffer. Further, $\mathbf{r}_{t-d}^*$ and $\mathbf{H}_{t-d}^*$ are stored in the buffer if $d < R - 2$.

Table 2. Computational Complexity of the Proposed LLR Refining Procedure

Scheme	Detection and IC (If Necessary)	LLR Calculation
Proposed LLR Refining [1,2]	$\mathcal{O}(\sum_{d=0}^{D_t-1}((N_{t-d}^t)^3 + (N_{t-d}^t)^2 M))$	$\mathcal{O}(\sum_{d=0}^{D_t-1} N_{t-d,\text{n}}^t Q \log_2 Q)$
Linear Detection	$\mathcal{O}(N^3 + N^2 M)$	$\mathcal{O}(NQ\log_2 Q)$
Coded Successive Hard-Decision IC	$\mathcal{O}(N^4 + N^3 M)$	$\mathcal{O}(NQ\log_2 Q)$
Coded Iterative Hard-Decision IC [3]	$\mathcal{O}(N^3 + N^2 M + I_{\text{out}} M N^2)$	$\mathcal{O}(I_{\text{out}} NQ\log_2 Q)$

[1] This is the worst-case complexity for $N_{t,\text{s}} > 0$ and $N_{t,\text{n}} > 0$. [2] If $N_{t,\text{n}} = 0$ or $D_t = 0$, no computational operations are performed. [3] I_{out} is the number of outer iterations in CIHIC.

Meanwhile, the LLR calculation of each bit requires the use of all Q symbols in the constellation $\mathcal{S}$, which should be performed for $\log_2 Q$ bits consisting of each symbol of the non-terminated packets remained in $\overline{\mathbf{s}}_{t-d}^t$. Let $N_{t-d,\text{n}}^t$ be the number of the non-terminated packets (the number that will be transmitted during the next $(t+1)$th TTI) remaining in $\overline{\mathbf{s}}_{t-d}^t$, i.e., the number of the packets experiencing LLR refining from $\overline{\mathbf{s}}_{t-d}^t$. Then, for LLR calculation, both (6) and (9) require the complexity of $\mathcal{O}(N_{t-d,\text{n}}^t Q \log_2 Q)$ at maximum, while (7) requires the additional complexity of $\mathcal{O}(MN_{t-d}^t)$ at maximum in cases of LMMSE. Consequently, considering the higher-order terms, the proposed scheme requires the complexity of $\mathcal{O}(\sum_{d=0}^{D_t-1} N_{t-d,\text{n}}^t Q \log_2 Q)$ for LLR calculation per receive signal vector for a given TTI.

The computational complexity of the proposed scheme is derived for when the initial condition to operate the LLR refining is satisfied, e.g., $N_{t,\text{n}} > 0$. If $N_{t,\text{s}} > 0$, $N_{t-d}^t < N$. Thus, the proposed scheme for the given d yields a smaller computational complexity than the detection and LLR calculation of the conventional receivers. In addition, no computational operations are performed for the proposed scheme if $N_{t,\text{s}} = 0$ (i.e., no possibility of error propagation) or $N_{t,\text{n}} = 0$ (no non-terminated packets benefiting from LLR refining). Furthermore, unlike the conventional receivers, the proposed scheme does not require a decoding operation, which also yields large computational complexity. Consequently, the proposed scheme can provide the LLR refining with a little extra complexity, which will be verified further in Section 4 via numerical simulations.

In terms of the memory requirement, the proposed LLR refining scheme with $D \geq 2$ requires additional memory units to store the $M \times N$ matrix $\mathbf{H}^*_{t-d}$ and the $M \times 1$ vector $\mathbf{r}^*_{t-d}$. Thus, assuming one memory unit as the memory size to store one complex value, the proposed scheme can require $(D-1)(MN+M)$ additional memory units. Meanwhile, the proposed scheme with $D=1$, i.e., the LLR refining for only the current HARQ rounds of the non-terminated packets using $\mathbf{r}_t$ and $\mathbf{H}_t$, requires no additional memory units for the buffer compared with the conventional LLR-level combining-based MMARQ receivers. In addition, the $M \times N^t_{t-d}$ matrix $\overline{\mathbf{H}}^*_{t-d}$ can be stored instead of $\mathbf{H}^*_{t-d}$ because $\overline{\mathbf{H}}^*_{t-d}$ is identical to $\mathbf{H}^*_{t-d}$ except the all-zero columns. In this case, the additional memory units required for the proposed scheme can be further reduced.

4. Simulation Results

For numerical simulations, the following environment is considered. The maximum HARQ rounds per packet, R is set to 3, and Chase combining retransmission strategy is considered. Because $R=3$, D can be set to 1 or 2 for the proposed scheme as explained in Section 3.2, while $D=0$ indicates the cases without the proposed scheme. For each packet, the data bits of a length $K=352$ are encoded by a CRC-32 for error detection and a 384×576 LDPC code in [13] for error correction. Thus, a code word length of $J=576$ is generated as the mother codeword for each packet. After encoding, the modulation is performed using gray-coded 4-QAM (i.e., $Q=4$) or 16-QAM (i.e., $Q=16$). As the channel model, the independent and quasi-static Rayleigh fading channels are considered. In the independent fading channel, the channel response is independently varied for transmit signal vector, and, in the quasi-static Rayleigh fading channel, the channel response is static in a given TTI but independently varied for the next TTI. In CSHIC, the detection order of packets is decided by the channel gain in the current TTI, e.g., $\mathrm{diag}(\mathbf{H}_t^H\mathbf{H}_t)$. The min-sum algorithm is used for decoding, where the number of decoding iterations for each packet is 40 at maximum during each TTI. After decoding, the syndrome of CRC-32 is used to detect errors in the decoded codeword of each packet and ACK/NACK feedback of each packet is generated according to its error detection result. Further, the number of outer iterations for CIHIC, I_{out} is set to 4. Finally, the average signal-to-noise ratio (SNR) is defined as $1/\sigma^2$.

For the performance evaluation, the average packet error rate (PER), average block error rate (BLER), and average throughput are considered. The PER is defined as the decoding error probability of packets considering up to the maximum Rth HARQ round, i.e., the decoding error probability of terminated packets. Further, the BLER of the $r_{t,n}$th HARQ round is defined as the decoding error probability of a packet when the HARQ round of the packet is $r_{t,n}$. Thus, using the average BLERs, the average PER can be written as

$$PER = \prod_{i=1}^{R} BLER_i, \tag{10}$$

where PER is the average PER and $BLER_i$ is the average BLER of the ith HARQ round. In addition, the average throughput is defined as the average number of successfully decoded packets per TTI, which can be written as [27]

$$TP = \frac{N(1-PER)}{1+\sum\limits_{k=1}^{R-1}\prod\limits_{i=1}^{k} BLER_i}, \tag{11}$$

where TP denotes the average throughput.

In Figures 6–9, the average PERs in 16×32 systems are evaluated. Figures 6 and 7 consider 4-QAM and Figures 8 and 9 consider 16-QAM. Further, LMMSE-based receivers are considered in Figures 6 and 8, while LZF-based receivers are considered in Figures 7 and 9.

From the average PER results, it is shown that the proposed scheme can significantly improve the average PER performance in all tested configurations, especially for CSHIC and CIHIC with 16-QAM experiencing significant error propagation. The SNR gain of $D = 1$ (the proposed scheme for $\mathbf{r}_t^*$ only) over $D = 0$ (without the proposed scheme) is larger than that of $D = 2$ (the proposed scheme for $\mathbf{r}_t^*$ and $\mathbf{r}_{t-1}^*$) over $D = 1$ for a given receiver, because the LLR refining using $\mathbf{r}_{t-1}^*$ can also be performed during the $(t-1)$th TTI when $D = 1$. Further, it is observed that the overall average PER characteristic for a given receiver with $D > 0$ is similar regardless of the channel environment (independent or quasi-static) and employed detection method (LMMSE or LZF). This implies that the proposed scheme is effective in terms of the average PER performance regardless of the channel environment and employed detection method.

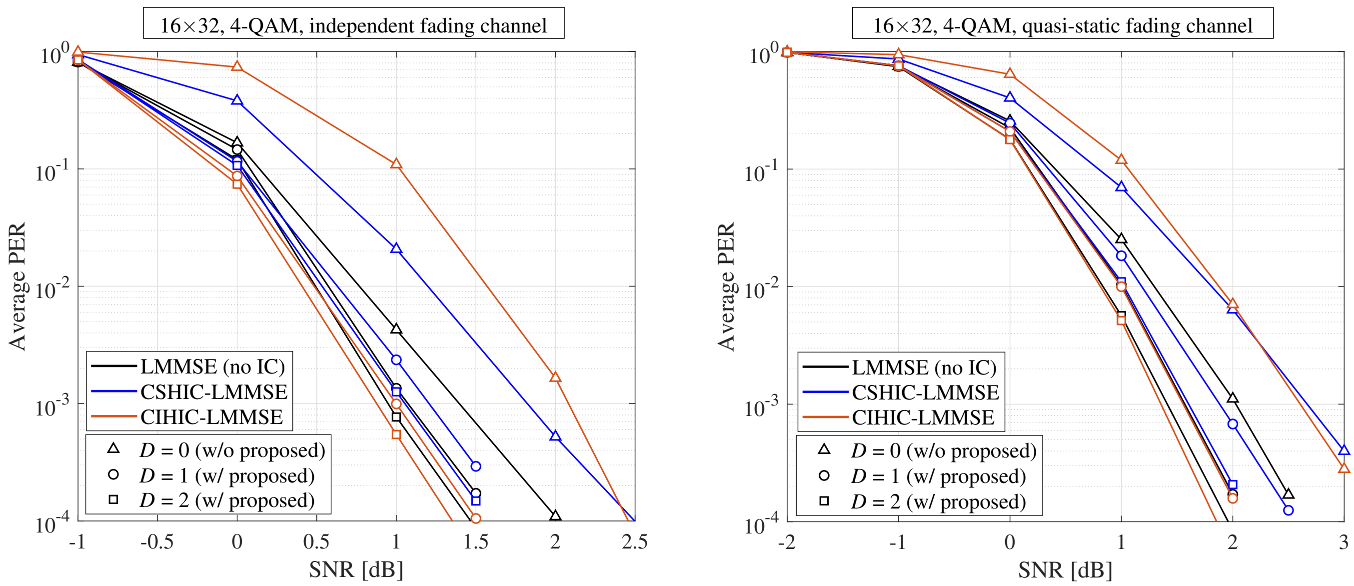

Figure 6. The average PERs for 16×32 MMARQ systems with LMMSE and 4-QAM.

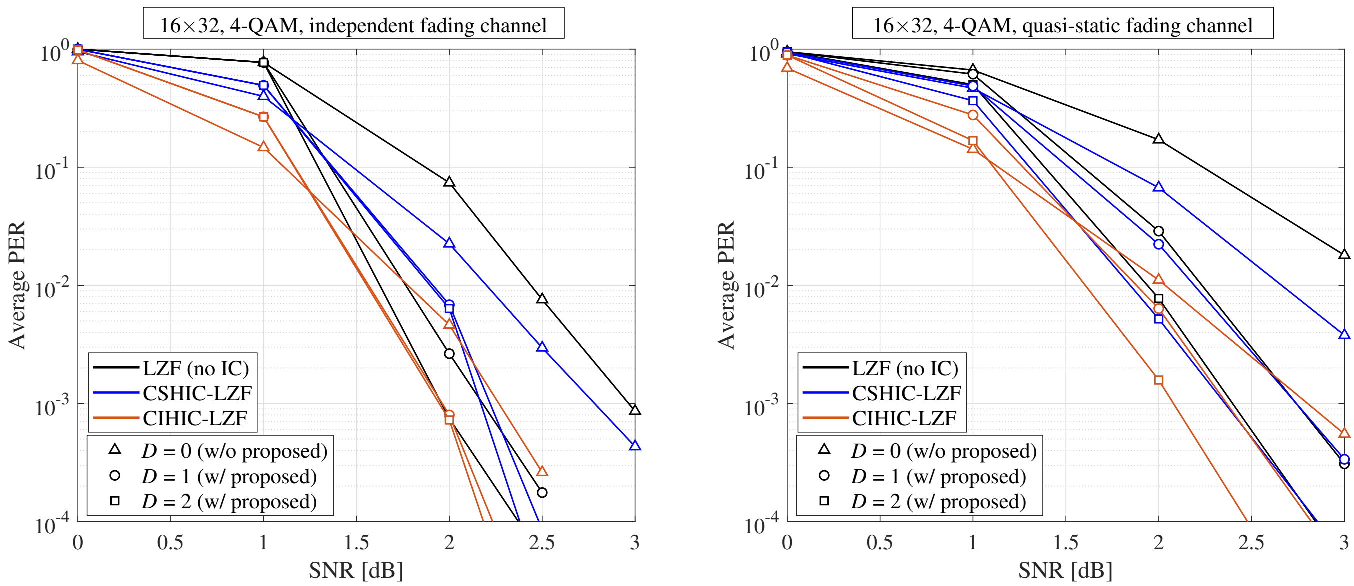

Figure 7. The average PERs for 16×32 MMARQ systems with LZF and 4-QAM.

Meanwhile, from Figure 6–9, it is observed that linear detection outperforms CSHIC and CIHIC in terms of the average PER for a given D in cases of the high modulation order ($Q = 16$), while CSHIC and CIHIC can have better average PERs than linear detection as

D increases in cases of the low modulation order ($Q = 4$). In particular, when $Q = 16$, an error floor is observed for CSHIC and CIHIC with $D = 0$ (without the proposed scheme) regardless of the detection method and channel environment. This implies that significant error propagation occurs in CSHIC and CIHIC for MMARQ systems without the proposed LLR refining scheme as Q increases. Specifically, because of iterative IC, which cancels all the other packets for a given packet, CIHIC with $D = 0$ experiences the worst error propagation and average PERs when $Q = 16$, which are greatly compensated for by employing the proposed scheme. In addition, when $Q = 16$, the performance improvement of CSHIC by employing the proposed scheme is relatively small compared to that of CIHIC for a given D, which leads CSHIC to have the worst average PER performance when the proposed scheme is employed. Meanwhile, when $Q = 4$, it is observed that CIHIC achieves the lowest average PER if the proposed scheme is employed, and CSHIC and CIHIC with $D = 0$ do not have a degraded performance as when $Q = 16$. This is because the effects of error propagation are weakened by having a low modulation order.

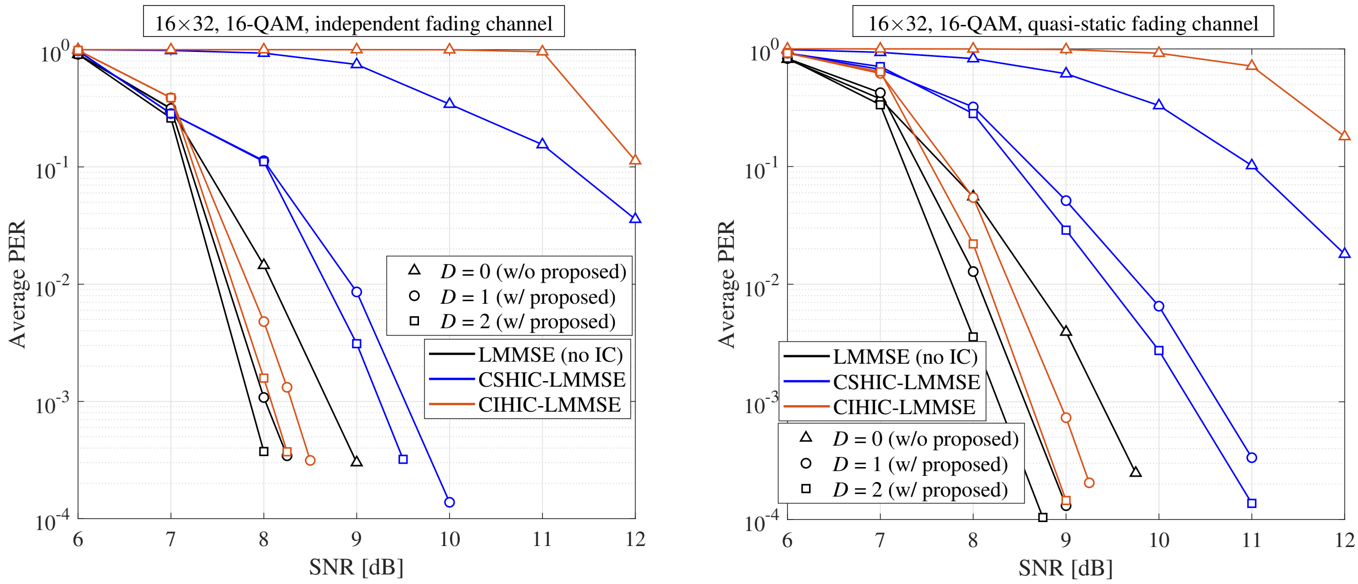

Figure 8. The average PERs for 16 × 32 MMARQ systems with LMMSE and 16-QAM.

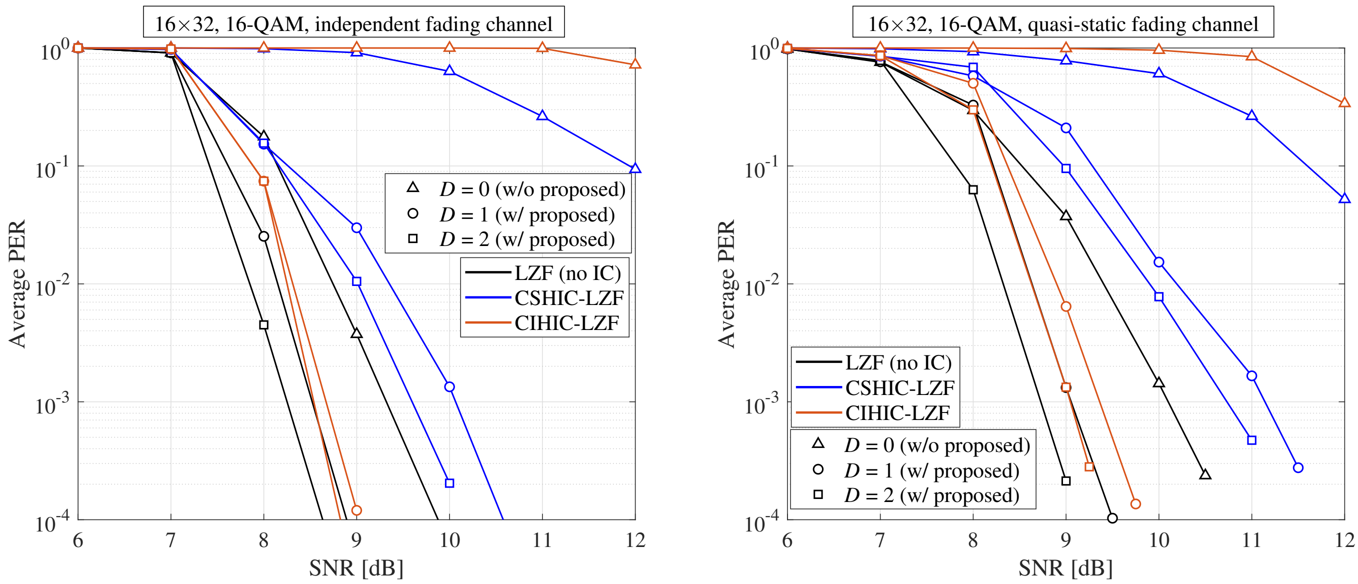

Figure 9. The average PERs for 16 × 32 MMARQ systems with LZF and 16-QAM.

From the average PER results, it is observed that the overall performance characteristics are similar regardless of the channel environment (independent or quasi-static) and detection method (LZF or LMMSE). Therefore, the remaining results are obtained under the assumption of LMMSE and quasi-static fading channel. Thus, Figures 10–12 show the average BLERs of the packets at the first, second, and third HARQ rounds in 16×32 systems with LMMSE-based receivers under the quasi-static fading channels, respectively,

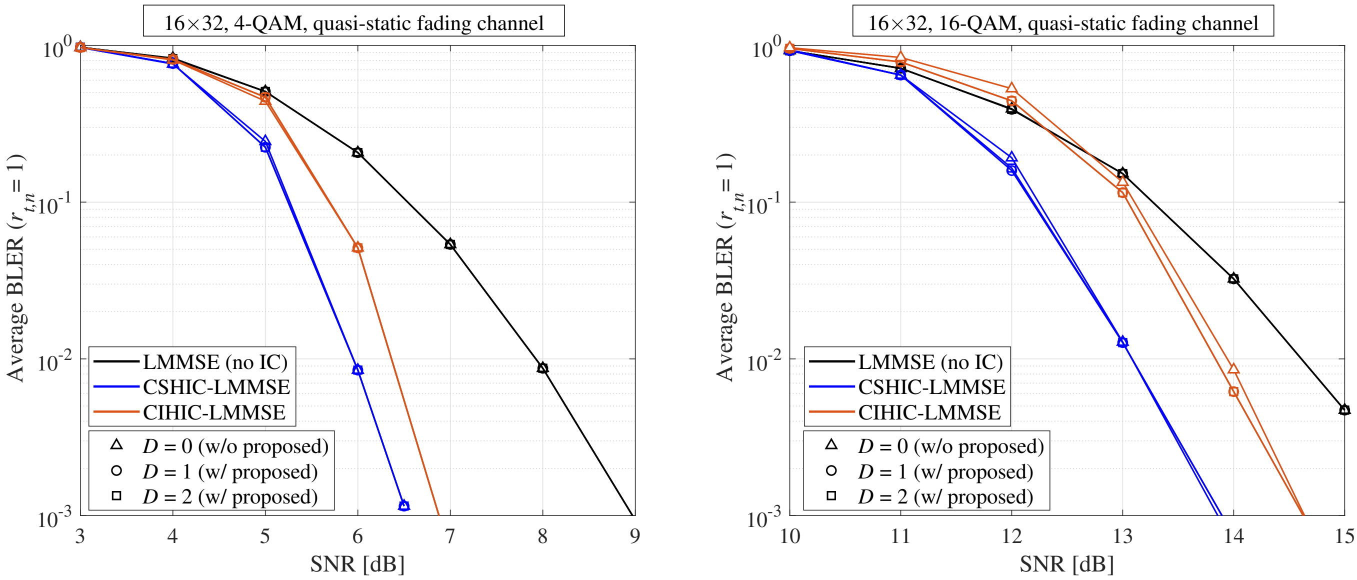

Figure 10. The average BLERs of 1st HARQ round for 16×32 MMARQ systems with LMMSE under the quasi-static fading channel.

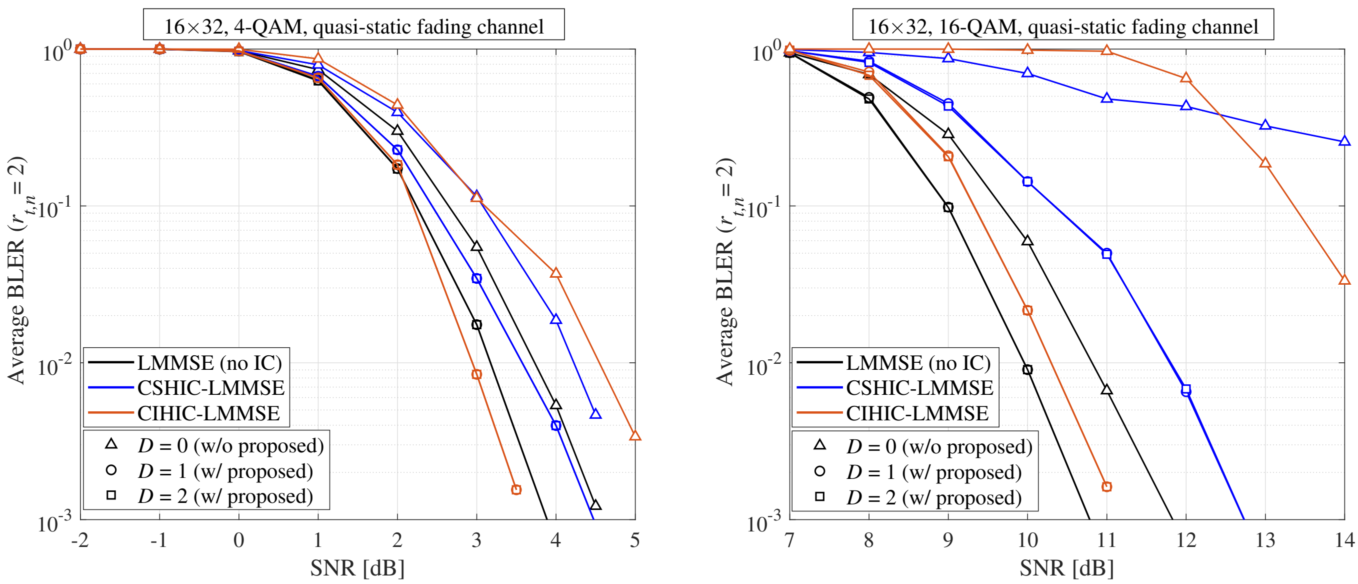

Figure 11. The average BLERs of 2nd HARQ round for 16×32 MMARQ systems with LMMSE under the quasi-static fading channel.

Figure 10 shows that the average BLER performance of the first HARQ round is near-identical for the given receiver regardless of D and Q. This is because the refined LLRs of a terminated packet will be used for the actual decoding from its next HARQ round, i.e., retransmission with $r_{t,n} > 1$. Meanwhile, CSHIC and CIHIC show a better average BLER for the packets at the first HARQ round than linear detection, which is consistent with the results of MIMO systems without HARQ [31–35]. Specifically, CSHIC achieves the lowest average BLER for a given SNR, because the ordering using the current channel response becomes accurate for initially transmitted packets (i.e., packets at the 1st HARQ round) as in MIMO systems without HARQ.

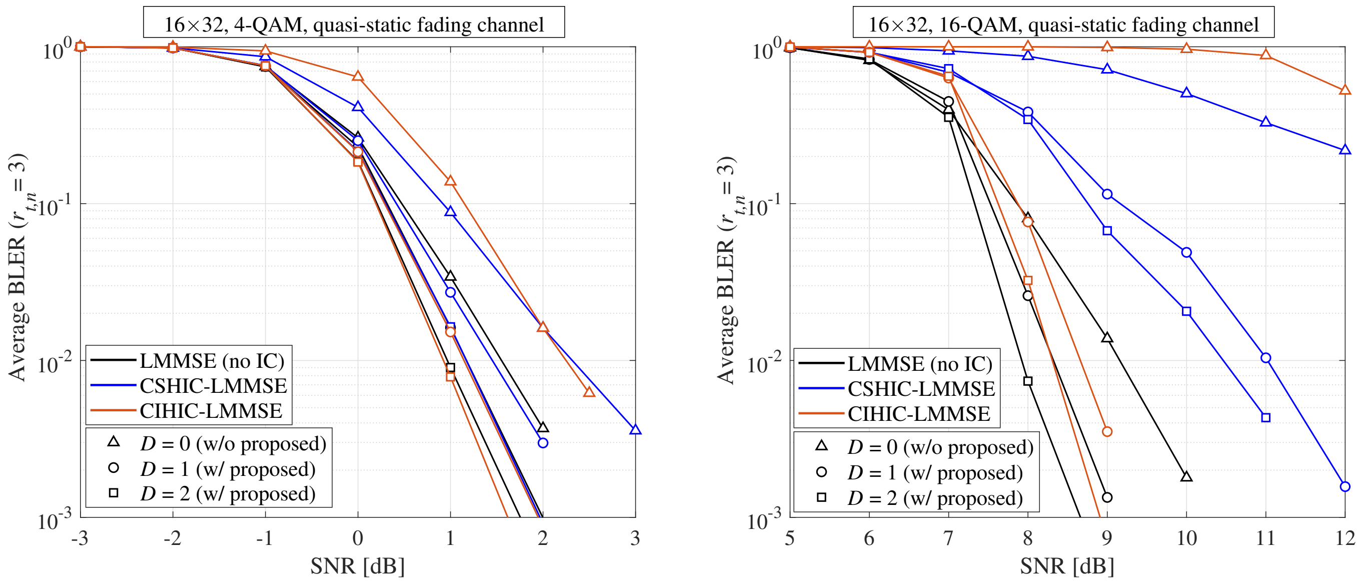

Figure 12. The average BLERs of 3rd HARQ round for 16×32 MMARQ systems with LMMSE under the quasi-static fading channel.

Figures 11 and 12 show that, for a given Q, the proposed scheme can provide a greatly improved average BLER performance for the retransmitted packets, i.e., $r_{t,n} > 1$, compared to the cases without the proposed scheme. Further, as shown in Figure 11, the cases of $D = 1$ and $D = 2$ have an identical average BLER performance for the packets at the second HARQ round. This is because the additional gain of $D = 2$ over $D = 1$ comes from the LLR refining of non-terminated packets, which will have the third HARQ round during the next TTI. Meanwhile, if $Q = 4$, CSHIC and CIHIC achieve a worse error performance than linear detection for $D = 0$ and $r_{t,n} > 1$, which shows the effects of inter-transmission error propagation. Further, if $Q = 16$, CSHIC and CIHIC achieve a worse error performance than linear detection for any given D, even when the proposed scheme is employed. This is because, for a large Q, ordering based on the current channel response becomes inaccurate for the retransmitted packets in CSHIC and the number of decoding iterations per outer iteration is insufficient to successfully decode the retransmitted packets by using the combined LLRs in CIHIC. Meanwhile, because the average PER is the multiplication of the average BLERs as shown in (10), the average BLER of a higher HARQ round has more effects on the average PER than the average BLER of a lower HARQ round [27]. Consequently, the average BLER characteristics in Figure 12 are similar to the average PER results in Figures 6 and 8 under the quasi-static fading channel.

Next, from Figures 13–16, the average system throughputs according to the antenna configuration and modulation order are provided. Figures 13 and 14 consider 8×8 and 16×16 antenna configurations with a loading factor of 1 (N/M), and Figures 15 and 16 consider 8×16 and 16×32 antenna configurations with a loading factor of 0.5. Further, $Q = 4$ for Figures 13 and 15, and $Q = 16$ for Figures 14 and 16.

It has been shown that the proposed scheme can also improve the system throughput for a given receiver, especially for CSHIC and CIHIC with a high loading factor and modulation order. With the proposed scheme ($D \geq 1$), the increase in rate is much greater in the low-SNR region than that in the high-SNR region. Further, especially when $Q = 16$, linear detection with the proposed scheme achieves the highest throughput in the low-SNR region, where the decoding success of the packets with a higher HARQ round dominates the system throughput because of no decoding success of initially transmitted packets.

Further, CSHIC and CIHIC without the proposed scheme have a significantly degraded throughput compared to linear detection without the proposed scheme in the low SNR region, especially when the loading factor is 1. This is because the number of incorrect IC operations increases due to the lack of the diversity gain. On the other hand, in the high-SNR region, CSHIC with $D \geq 1$ outperforms the other receivers with $D \geq 1$, because the decoding success of initially transmitted packet (i.e., packets at the first HARQ round) has more effects on the system throughput [27], and CSHIC shows the lowest average BLER for the initially transmitted packets, as shown in Figure 10. Similarly, CIHIC with $D \geq 1$ has a higher throughput than linear detection with the same D in the high-SNR region.

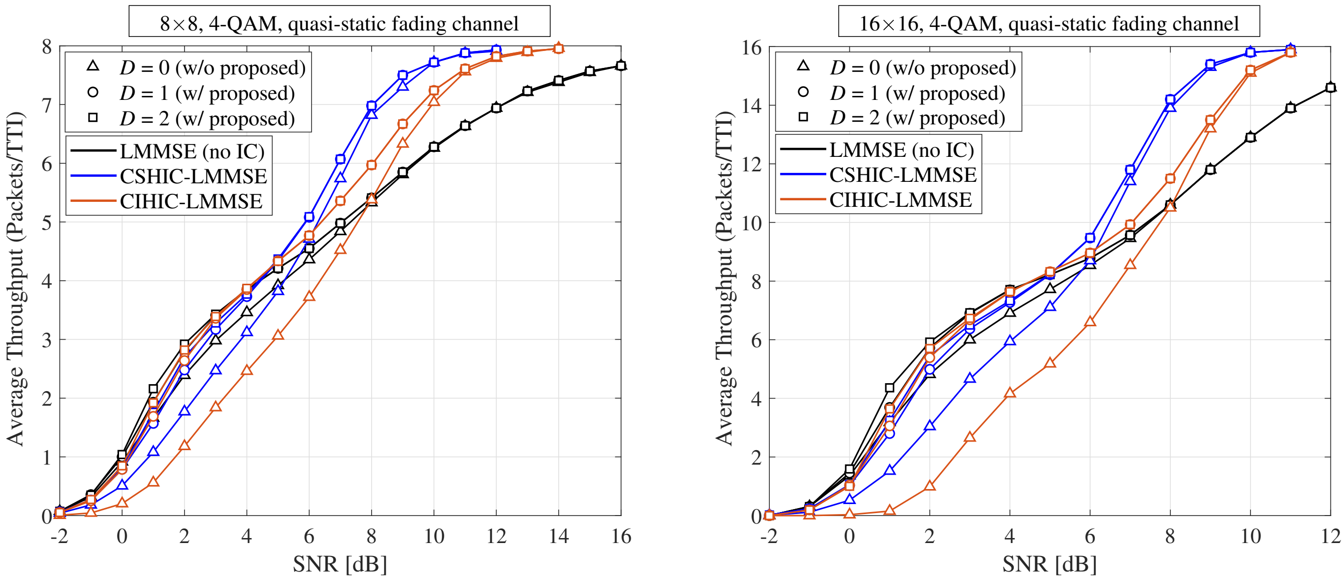

Figure 13. The average throughputs for 8×8 and 16×16 MMARQ systems with LMMSE and 4-QAM under the quasi-static fading channel.

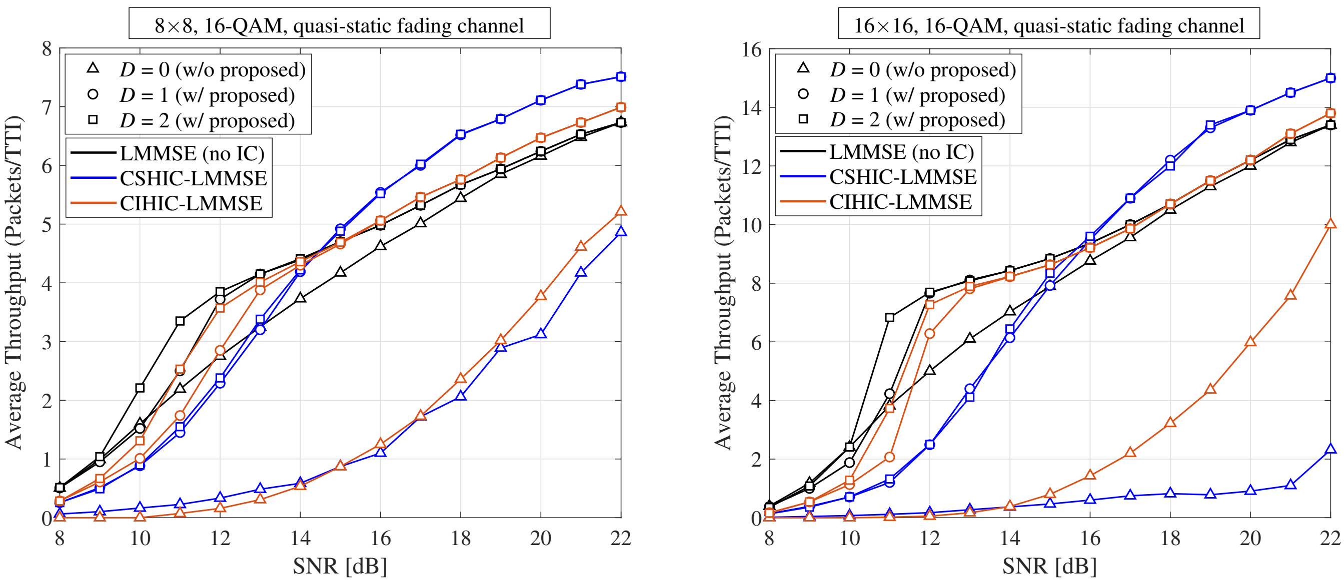

Figure 14. The average throughputs for 8×8 and 16×16 MMARQ systems with LMMSE and 16-QAM under the quasi-static fading channel.

Finally, in Figures 17 and 18, the computational complexities of the proposed LLR refining scheme and conventional receivers are evaluated according to Table 2 for 16×32 MMARQ systems with LMMSE under the quasi-static fading channel, where

Figures 17 and 18 are the detection and LLR calculations per receive signal vector, respectively. The cases of $D = 0$ are for the conventional receivers, and the cases of $D > 0$ are for the proposed LLR refining procedure only, i.e., $\sum_{d=0}^{D_t-1}\left(\left(N^t_{t-d}\right)^3 + \left(N^t_{t-d}\right)^2 M\right)$ for Figure 17 and $\sum_{d=0}^{D_t-1} N^t_{t-d,\mathrm{n}} Q\log_2 Q$ for Figure 18. Because N^t_{t-d} and $N_{t-d,\mathrm{n}}$ can be varied with t and d according to the decoding results of packets, the average values of N^t_{t-d} and $N_{t-d,\mathrm{n}}$ for the proposed scheme are obtained with numerical simulations. During the calculation of the average values, if no LLR refining is performed for a given TTI, e.g., $N_{t,\mathrm{n}} = 0$, N^t_{t-d} and $N_{t-d,\mathrm{n}}$ are set to 0 for a given TTI.

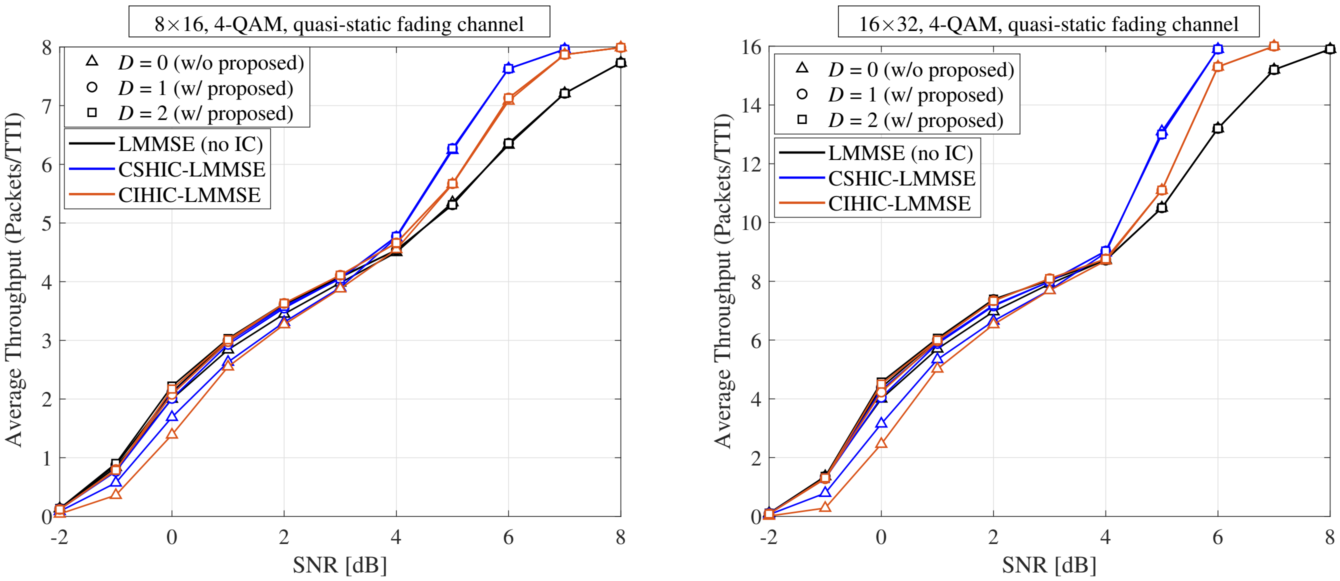

Figure 15. The average throughputs for 8 × 16 and 16 × 32 MMARQ systems with LMMSE and 4-QAM under the quasi-static fading channel.

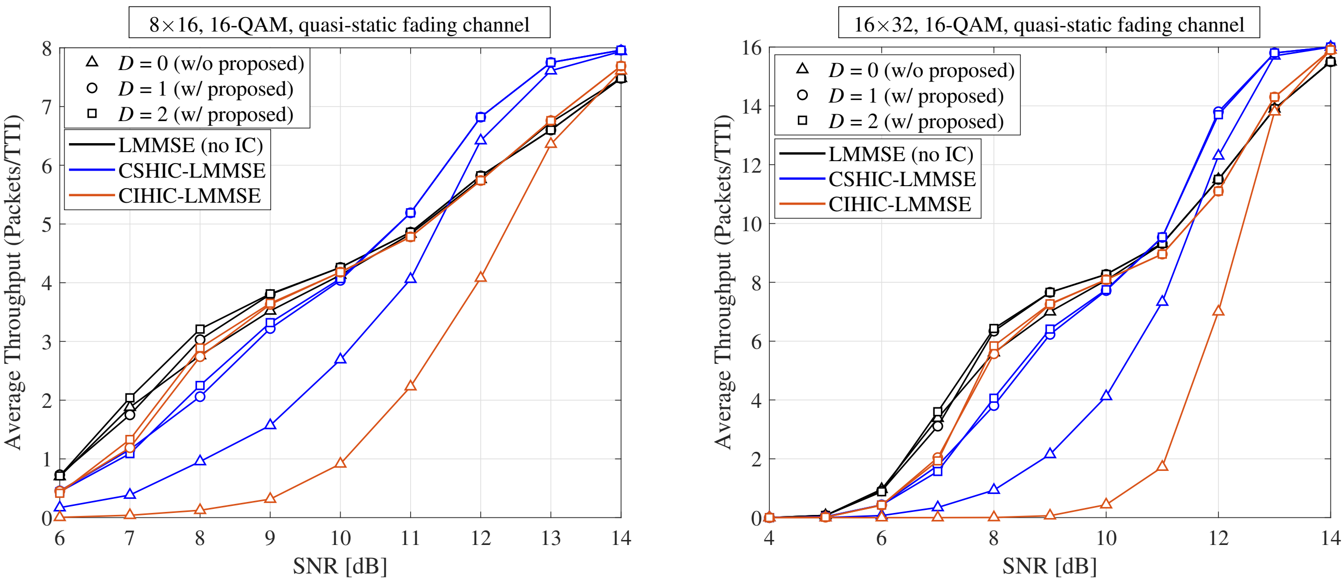

Figure 16. The average throughputs for 8 × 16 and 16 × 32 MMARQ systems with LMMSE and 16-QAM under the quasi-static fading channel.

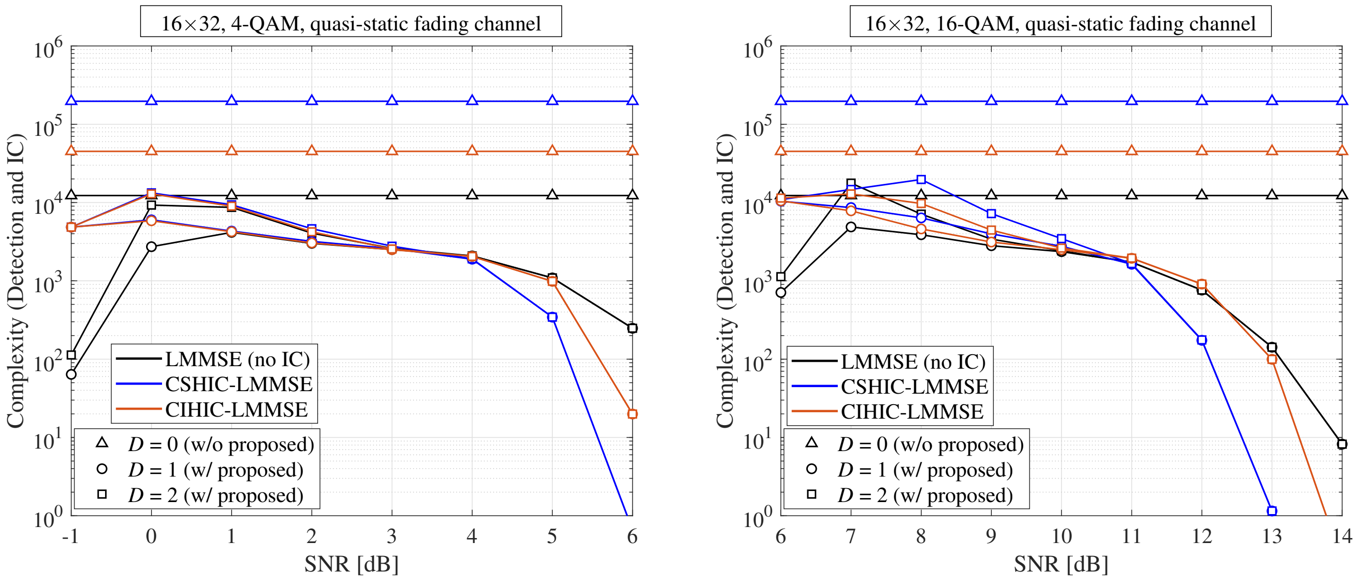

Figure 17. The complexity of detection per receive signal vector for 16 × 32 MMARQ systems with LMMSE under the quasi-static fading channel.

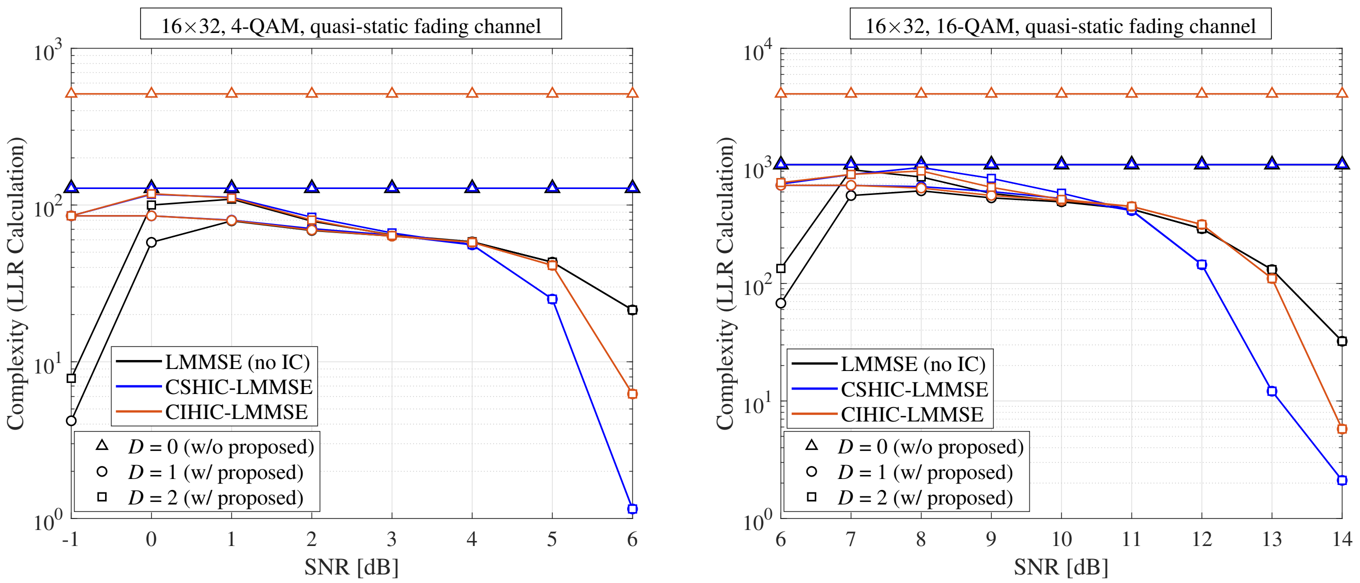

Figure 18. The complexity of LLR calculation per receive signal vector for 16 × 32 MMARQ systems with LMMSE under the quasi-static fading channel.

It is observed from Figures 17 and 18 that the additional computational complexity required by the proposed scheme is, in the worst-case scenario, similar to the complexity of linear receiver (LMMSE) for both detection and the LLR calculation regardless of Q, where linear receiver has the smallest complexity among the conventional receivers as shown in Table 2. The worst-case scenario (i.e., the maximum complexity) occurs in the low-SNR region, where the successful decoding of packets begins to occur, i.e., the SNR region, where a great deal of error propagation occurs but LLR refining is also possible. As the SNR increases, error propagation and the necessity of LLR refining decrease; thereby, the complexity of the proposed scheme also decreases, which makes the complexity of the proposed scheme considerably smaller than that of the conventional receivers in the high-SNR region. In addition, there is no significant difference in complexity for $D = 1$ and 2, especially as the SNR increases. Moreover, the proposed scheme does not require a decoding operation, which also incurs a large computational complexity for the conventional receivers. Therefore, the proposed scheme can provide LLR refining with a small amount of extra complexity in comparison to the conventional receivers.

5. Conclusions

This paper proposed an LLR refining scheme that can be employed with conventional receivers for MMARQ systems as the post-processing scheme. By utilizing the successfully decoded packets for cancellation, the proposed scheme can perform the LLR refining for non-terminated packets that will be retransmitted during future TTIs. This enables the proposed scheme to achieve better error performance and throughput with a small amount of extra complexity, as verified by numerical simulations. In addition, the proposed scheme can be employed to any conventional LLR-level combining-based receivers for MMARQ systems without modifying the original reception procedure, and the proposed scheme can be utilized regardless of the HARQ retransmission strategy. Therefore, the proposed scheme can be considered to be an effective and practical post-processing scheme for an MMARQ receiver.

Because the proposed LLR refining scheme needs to be performed after the conventional reception procedure is over, the overall latency for the receiver increases. Thus, for the HARQ protocols designed for a low-latency transmission [2,3,7,29], the overall process of the receiver with LLR refining has to be redesigned. Furthermore, because the LLR refining is performed for non-terminated packets that will be retransmitted, the proposed scheme cannot directly improve the performance of initially transmitted packets. Therefore, a transmission control algorithm can be developed to increase the simultaneous transmission of the retransmitted packets and initially transmitted packets, which can improve the performance of the initially transmitted packets in MMARQ systems [27]. In addition, although the soft IC-based receiver is not considered in this paper, the proposed LLR refining scheme can also be employed to the soft IC-based LLR-level combining receiver for MMARQ systems. In this case, the effects of previous soft IC operations need to be considered for LLR refining to achieve an improved error performance, while the computational complexity should be considered as well. These are the topics to be investigated in future works.

Funding: This work was supported by a Kyonggi University Research Grant 2022.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to a related project.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

ACK/NACK	Acknowledgment/non-acknowledgment
ARQ	Automatic repeat request
AWGN	Additive white Gaussian noise
BLER	Block error rate
CIHIC	Coded iterative hard-decision interference cancellation
CRC	Cyclic redundancy check
CSHIC	Coded successive hard-decision interference cancellation
CSI	Channel state information
FEC	Forward error correction
HARQ	Hybrid automatic repeat request
IC	Interference cancellation
IHIC	Iterative hard-decision interference cancellation
LDPC	Low-density parity-check
LLR	Log-likelihood ratio
LMMSE	Linear minimum mean-squared-error
LZF	Linear zero-forcing

MIMO	Multiple-input multiple-output
MMARQ	Multiple-input multiple-output multiple automatic repeat request
MSARQ	Multiple-input multiple-output single automatic repeat request
PER	Packet error rate
QAM	Quadrature amplitude modulation
SHIC	Successive hard-decision interference cancellation
SNR	Signal-to-noise ratio
TTI	Transmission time interval

Mathematical Symbols

The following mathematical symbols are used in this paper:

N	Number of transmit antennas/number of packets sent together in each TTI
M	Number of receive antennas
K	Length of data bits for a packet
J	Length of mother codeword for a packet
L	Length of symbol sequence for a given HARQ round of a packet
Q	Modulation order
$\log_2 Q$	Number of bits assigned for a modulated symbol
R	Maximum HARQ round of a packet
$\mathcal{S}$	Q-ary constellation set
$\mathbf{s}_t$	$N \times 1$ transmit signal vector for the tth TTI
$\mathbf{r}_t$	$M \times 1$ receive signal vector for the tth TTI
$\mathbf{w}_t$	$M \times 1$ AWGN vector for the tth TTI
$\mathbf{H}_t$	$M \times N$ channel matrix for the tth TTI
σ^2	Variance of each element of $\mathbf{w}_t$
$r_{t,n}$	HARQ round of the packet from the nth transmit antenna at the tth TTI
$e_{t,n}$	Error detection result of the packet from the nth transmit antenna at the tth TTI
$\mathcal{N}_{t,\mathrm{s}}$	Set including the transmit antennas of the successfully decoded packets
$\mathcal{N}_{t,\mathrm{n}}$	Set including the transmit antennas of the non-terminated packets
$N_{t,\mathrm{s}}$	Size of $\mathcal{N}_{t,\mathrm{s}}$
$N_{t,\mathrm{n}}$	Size of $\mathcal{N}_{t,\mathrm{n}}$
D	Maximum depth of the proposed LLR refining scheme
D_t	Depth for the current tth TTI
$D_{t,\mathrm{p}}$	Depth for the tth TTI calculated by only the HARQ rounds of packets
$\mathbf{r}^*_{t-d}$	Stored receive signal vector for the $(t-d)$th TTI used in the tth TTI
$\tilde{\mathbf{s}}^t_{t-d}$	$N \times 1$ transmit signal vector for $\mathbf{r}^*_{t-d}$ regenerated at the tth TTI
$\mathbf{H}^*_{t-d}$	Stored $M \times N$ channel matrix for the $(t-d)$th TTI used in the tth TTI
N^t_{t-d}	Number of the non all-zero columns in $\mathbf{H}^*_{t-d}$
$\overline{\mathbf{H}}^*_{t-d}$	$M \times N^t_{t-d}$ matrix identical to $\mathbf{H}^*_{t-d}$ except the all-zero columns
$\bar{\mathbf{s}}^t_{t-d}$	$N^t_{t-d} \times 1$ transmit signal vector estimated in LLR refining
$\mathrm{f}^{t,n}_{t-d}$	Column in $\overline{\mathbf{H}}^*_{t-d}$ for the packet from the nth transmit antenna at the tth TTI
$l^{t,n}_{t-d}(q)$	Refined LLR of the qth bit of $\bar{s}^t_{t-d}(\mathrm{f}^{t,n}_{t-d})$
$\mathcal{S}^b_q$	Subset of $\mathcal{S}$ with the qth bit of the elements in $\mathcal{S}^b_q$ is b
$\mu^{t,n}_{t-d}$	Mean of $\bar{s}^t_{t-d}(\mathrm{f}^{t,n}_{t-d})$
$\eta^{t,n}_{t-d}$	Residual interference plus noise variance in $\bar{s}^t_{t-d}(\mathrm{f}^{t,n}_{t-d})$
$N^t_{t-d,\mathrm{n}}$	Number of the non-terminated packets experiencing LLR refining from $\bar{\mathbf{s}}^t_{t-d}$
I_{out}	Number of outer iterations in CIHIC
PER	Average PER
$BLER_i$	Average BLER of the ith HARQ round
TP	Average throughput

References

1. Chase, D. Code Combining—A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets. *IEEE Trans. Commun.* **1985**, *33*, 385–393. [CrossRef]
2. Jung, Y.-H.; Choi, J. Hybrid ARQ Scheme with Autonomous Retransmission for Multicasting in Wireless Sensor Networks. *Sensors* **2017**, *17*, 463. [CrossRef] [PubMed]

3. Makki, B.; Svensson, T.; Caire, G.; Zorzi, M. Fast HARQ Over Finite Blocklength Codes: A Technique for Low-Latency Reliable Communication. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 194–209. [CrossRef]
4. Xiang, Z.; Yang, W.; Pan, G.; Cai, Y.; Song, Y.; Zou, Y. Secure Transmission in HARQ-Assisted Non-Orthogonal Multiple Access Networks. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 2171–2182. [CrossRef]
5. Shi, Z.; Zhang, C.; Fu, Y.; Wang, H.; Yang, G.; Ma, S. Achievable Diversity Order of HARQ-Aided Downlink NOMA Systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 471–487. [CrossRef]
6. Pan, G.; Ye, J.; Tian, Y.; Alouini, M. On HARQ Schemes in Satellite-Terrestrial Transmissions. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 7998–8010. [CrossRef]
7. Nadeem, F.; Shirvanimoghaddam, M.; Li, Y.; Vucetic, B. Nonorthogonal HARQ for URLLC: Design and Analysis. *IEEE Internet Things J.* **2021**, *8*, 17596–17610. [CrossRef]
8. Ahmed, A.; Al-Dweik, A.; Iraqi, Y.; Mukhtar, H.; Naeem, M.; Hossain, E. Hybrid Automatic Repeat Request (HARQ) in Wireless Communications Systems and Standards: A Contemporary Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 2711–2752. [CrossRef]
9. Bian, H.; Liu, R. Reliable and Energy-Efficient LEO Satellite Communication with IR-HARQ via Power Allocation. *Sensors* **2022**, *22*, 3035. [CrossRef]
10. Wu, Y.; Mu, K.; Duan, K.; Yin, S.; Yang, H. On the Secure Performance of Intelligent Reflecting Surface-Assisted HARQ Systems. *Entropy* **2023**, *25*, 519. [CrossRef]
11. Lin, S.-Y.; Yang, M.-H.; Jia, S. An Adaptive Hybrid Automatic Repeat Request (A-HARQ) Scheme Based on Reinforcement Learning. *Electronics* **2023**, *12*, 4127. [CrossRef]
12. *3GPP TS 25.848*; Physical Layer Aspects of UTRA High Speed Downlink Packet Access. 3GPP: Sophia Antipolis, France, 2001.
13. *IEEE P802.16e*; IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. IEEE: Piscataway, NJ, USA, 2006.
14. *3GPP TS 36.212*; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding. 3GPP: Sophia Antipolis, France, 2009.
15. *3GPP TS 38.212*; NR; Multiplexing and Channel Coding. 3GPP: Sophia Antipolis, France, 2018.
16. Makki, B.; Eriksson, T. On the Performance of MIMO-ARQ Systems with Channel State Information at the Receiver. *IEEE Trans. Commun.* **2014**, *62*, 1588–1603. [CrossRef]
17. Park, S. Multiple Retransmission Strategy-Based HARQ for MIMO Systems with Acknowledgment Bundling. *Electron. Lett.* **2014**, *50*, 637–639. [CrossRef]
18. Wu, W.; Mittelmann, H.; Ding, Z. Modulation Design for MIMO-CoMP HARQ. *IEEE Commun. Lett.* **2017**, *21*, 290–293. [CrossRef]
19. Kim, S.; Yu, H. Energy-Efficient HARQ-IR for Massive MIMO Systems. *IEEE Trans. Commun.* **2018**, *66*, 3892–3901. [CrossRef]
20. Radbord, A.; Harsini, J.S. Slow and Fast Adaptive Modulation and Coding for Uplink Massive MIMO Systems with Packet Retransmission. *IET Commun.* **2022**, *16*. 915–928. [CrossRef]
21. Zhang, H.; Liao, Z.; Shi, Z.; Yang, G.; Dou, Q.; Ma, S. Performance Analysis of MIMO-HARQ Assisted V2V Communications With Keyhole Effect. *IEEE Trans. Commun.* **2022**, *70*, 3034–3046. [CrossRef]
22. Park, S. Kalman Combining Based Iterative Detection and Decoding for MIMO Systems With Hybrid ARQ. *IEEE Trans. Veh. Technol.* **2023**, *72*, 2040–2050. [CrossRef]
23. Zheng, H.; Lozano, A.; Haleem, M. Multiple ARQ Processes for MIMO Systems. *EURASIP J. Adv. Signal Process.* **2004**, *5*, 772–782. [CrossRef]
24. Bai, C.; Krzymień, W.A.; Fair, I.J. Hybrid-ARQ for Layered Space Time MIMO Systems With Channel State Information Only at the Receiver. *IET Commun.* **2010**, *4*, 1765–1773. [CrossRef]
25. Prasad, N.; Wang, X. Efficient Combining Techniques for Multi-Input Multi-Output Multi-User Systems Employing Hybrid Automatic Repeat Request. *IET Commun.* **2011**, *5*, 1785–1796. [CrossRef]
26. Lee, J.W. Spectrally Efficient Multiple-ARQ Scheme for MIMO Channel. *Electron. Lett.* **2013**, *49*, 261–263. [CrossRef]
27. Park, S.; Choi, S. Performance of Symbol-Level Combining and Bit-Level Combining in MIMO Multiple ARQ Systems. *IEEE Trans. Commun.* **2016**, *64*, 1517–1528. [CrossRef]
28. Park, S. Aggregation-Assisted Combining for MIMO Multiple ARQ Systems. *IEEE Commun. Lett.* **2018**, *22*, 348–351. [CrossRef]
29. Kashif, M.; Iqbal, M.; Ullah, Z.; Dagiuklas, T.; Sarwar, S.; Ul-Qayyum, Z.; Safyan, M. Shared Hybrid ARQ With Incremental Redundancy (SHARQ IR) in Overloaded MIMO Systems to Support Energy-Efficient Transmissions. *IEEE Access* **2020**, *8*, 111653–111659. [CrossRef]
30. Park, S. Kalman Filtering Based Combining for MIMO Systems with Hybrid ARQ. *IEEE Trans. Signal Process.* **2021**, *69*, 5250–5258. [CrossRef]
31. Choi, J. Capacity/Throughput Optimization for H-BLAST With SC Receiver Over MIMO Channels. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 1016–1024. [CrossRef]
32. Ketonen, J.; Juntti, M.; Cavallaro, J.R. Performance—Complexity Comparison of Receivers for a LTE MIMO-OFDM System. *IEEE Trans. Signal Process.* **2010**, *58*, 3360–3372. [CrossRef]
33. Matache, A.; Jones, C.; Wesel, R. Reduced Complexity MIMO Detectors for LDPC Coded Systems. In Proceedings of the IEEE MILCOM, Monterey, CA, USA, 31 October–3 November 2004. [CrossRef]
34. Liu, D.N.; Fitz, M.P. Low Complexity Affine MMSE Detector for Iterative Detection-Decoding MIMO OFDM Systems. *IEEE Trans. Commun.* **2008**, *56*, 150–158. [CrossRef]

35. Paulraj, A.; Nabar, R.; Gore, D. *Introduction to Space-Time Wireless Communications*; Cambridge University Press: Cambridge, UK, 2008.
36. Park, S. Low-Complexity LMMSE-Based Iterative Soft Interference Cancellation for MIMO Systems. *IEEE Trans. Signal Process.* **2022**, *70*, 1890–1899. [CrossRef]
37. Chavali, N.K.; Kuchi, K.; Reddy, V.U. Downlink MU-MIMO With THP Combined With Pre- and Post-Processing and Selection of the Processing Vectors for Maximization of Per-Stream SNR. *IEEE Trans. Veh. Technol.* **2017**, 2223–2236. [CrossRef]
38. Silva, M.M.; Dinis, R. A Simplified Massive MIMO Implemented With Pre or Post-Processing. *Phys. Commun.* **2017**, *25*, 355–362. [CrossRef]
39. Mahdavi, M.; Edfors, O.; Öwall, V.; Liu, L. Angular-Domain Massive MIMO Detection: Algorithm, Implementation, and Design Tradeoffs. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2020**, *67*, 1948–1961. [CrossRef]
40. Moon, T.K. *Error Correction Coding: Mathematical Methods and Algorithms*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2021.