Kalman Filtering Based Combining for MIMO Systems With Hybrid ARQ

Sangjoon Park, *Member, IEEE*

***Abstract*—This paper proposes a Kalman filtering (KF)-based combining scheme for multiple-input multiple-output (MIMO) systems with hybrid automatic repeat request (ARQ). In the state-space model, the change of a transmit signal vector in consecutive transmission time intervals (TTIs) is interpreted as the transition of a state vector. Accordingly, the proposed KF operation performs symbol-level combining (SLC) based on linear minimum mean-square-error (LMMSE) detection using the information aggregated up to the current TTI. The state-space model can represent various configurations for MIMO systems with hybrid ARQ (HARQ) via a simple modification of a transition indicating vector. Therefore, compared to the existing SLC schemes designed for a specific system in a dedicated manner, the proposed scheme can be applied to various MIMO-HARQ systems, similar to bit-level combining (BLC). In addition, by employing a low-complexity correction step for the KF operation, the computational complexity of the proposed scheme is lower than that of existing SLC schemes and is comparable to that of BLC. Simulation results and analyses confirm that the proposed scheme outperforms BLC and achieves near-identical error performance to the LMMSE-based direct SLC scheme with a brute-force aggregation of all related information up to the current TTI for the transmitted packet(s).**

***Index Terms*—HARQ, MIMO systems, Kalman filtering, packet combining, LMMSE, SLC, BLC.**

I. Introduction

IN WIRELESS communication systems, the receiver can request retransmission of a packet with detection errors. By retransmitting the forward error correction (FEC) encoded packet, the hybrid automatic repeat request (HARQ), which is a combined scheme of FEC and ARQ, can overcome such detection errors and can achieve an extremely low error probability [1]. In addition, because the packet in each HARQ round can have a higher code rate than the original mother codeword [2], HARQ can enhance the system throughput in comparison with a system without HARQ. Therefore, as an essential technique for packet transmission, HARQ is employed in most modern communication standards [3]–[5].

To completely utilize HARQ, the receiver must combine the information of the packet obtained throughout the HARQ rounds. The basic combining scheme for HARQ is bit-level combining (BLC) [6]. In BLC, the log-likelihood ratio (LLR) of a packet in each HARQ round is separately calculated, followed by combining the LLRs obtained from the HARQ rounds of the packet for FEC decoding. This BLC scheme has high flexibility and can be applied to any system employing an HARQ regardless of the number of HARQ processes [7] and HARQ retransmission strategy [8]. In addition to BLC, symbol-level combining (SLC) is a combining scheme for HARQ, which combines the received signals related to a packet and calculates the LLR for FEC decoding from the combined signal model [9]–[23]. In particular, in multiple-input multiple-output (MIMO) systems, SLC performance can be considerably improved from BLC. However, the direct SLC scheme with a brute-force aggregation of all related information of a packet is impractical in MIMO systems with HARQ owing to the large computational complexity and memory loads. Therefore, several alternative SLC schemes have been investigated for MIMO systems with HARQ.

Despite the simplification of the direct SLC scheme, existing SLC schemes still require higher computational complexity and memory units than BLC. In addition, the existing SLC schemes are dedicated for a specific MIMO system with HARQ according to the number of HARQ processes (e.g., MIMO single ARQ (MSARQ) and MIMO multiple ARQ (MMARQ) systems) and the HARQ retransmission strategy (e.g., Chase combining (CC) and incremental redundancy (IR)). Thus, employing these strategies to practical systems supporting various configurations in terms of the number of HARQ processes and HARQ retransmission strategies is infeasible. In contrast, despite its flexibility for various systems and low overhead, BLC shows significantly worse error performance and system throughput than SLC.

Therefore, to overcome the limitations of conventional SLC and BLC schemes, this paper proposes a Kalman filtering (KF)-based combining scheme for MIMO systems with HARQ. The Kalman filter [24] has been widely studied for signal detection in MIMO systems, although the applications have usually been limited to equalization under frequency selective fading channels [25]. In the proposed KF-based combining (KC) scheme, considering HARQ retransmissions, changes in a transmitted signal vector from two consecutive transmission time intervals (TTIs) are modeled as the transition of a state vector, which has not been considered in previous studies for MIMO systems with HARQ. Then, the KF operation is performed to obtain an estimate of each transmitted signal vector according to the corresponding state-space model. Because KF can process all

Manuscript received May 17, 2021; accepted August 22, 2021. Date of publication September 14, 2021; date of current version September 24, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sangarapilla Lambotharan. This work was supported in part by Ministry of Science and ICT under Grant NRF-2019R1C1C1003202. *(Corresponding author: Sangjoon Park.)*

The author is with the Department of Electronic Engineering, Kyonggi University, Suwon 16227, Korea (e-mail: sj.park@kgu.ac.kr).

Digital Object Identifier 10.1109/TSP.2021.3110655

information collected up to a given point and obtain a linear minimum mean-square-error (LMMSE) estimate of a given state vector [24], the proposed KC scheme enables LMMSE detection of the transmit symbols based on the SLC by a single KF operation. In this operation, the related information (e.g., channel matrices and received signal vectors related to the transmitted packet(s)) for SLC is aggregated by the state transition for the KF operation.

In the proposed scheme, the sizes of the state and observation vectors are equal to the numbers of transmit and receive antennas of the system, respectively. In addition, a low-complexity correction step is developed for the KF operation of the proposed scheme. Therefore, the computational complexity of the proposed KC scheme is similar to that of the conventional LMMSE-based BLC scheme, which is significantly lower than that of the conventional LMMSE-based SLC schemes. Furthermore, the state transition model for each transmit signal vector is determined by system configurations, such as the number of HARQ processes, HARQ retransmission strategy, bit-to-symbol mapping for modulation, and spatial multiplexing of modulated symbols. These configurations should already be known to the receiver to decode the packet(s). Therefore, by modifying the transition indicating vector for a given system configuration, the proposed scheme can be applied to various MIMO systems with HARQ, regardless of the number of HARQ processes and the retransmission strategy, similar to BLC.

The remainder of this paper is organized as follows. Section II describes the system model including MSARQ and MMARQ systems with CC and IR. Section III explains the proposed scheme in detail, including the state-space model for the KF operation, and compares the computational complexity and memory requirements of the proposed scheme with the conventional BLC and SLC schemes. Section IV presents the simulation results for various system configurations, and Section V concludes the paper.

A. Related Work

For MSARQ systems, the modeling of HARQ procedures is straightforward and several SLC schemes have been developed. In [9], pre-combining and post-combining schemes with linear detection were proposed. The pre-combining scheme was extended to maximal-ratio combining to be used with maximum-likelihood (ML) detection and linear detection for CC in [10]. In [11], an SLC scheme with sphere decoding was developed for CC. In [12], SLC schemes with QR decomposition of the aggregated channel matrix were investigated for CC. In [13], a joint utilization of SLC and turbo equalization under frequency selective fading channels was proposed for CC. Meanwhile, in [14], an SLC scheme with ML and linear detections was investigated for IR, which was extended to the system with acknowledgement feedback bundling [15]. In [16], an SLC scheme under an interference channel was proposed for CC. In [17], an SLC scheme with ML detection was proposed for partial retransmission considering orthogonal space-time block codes. In [18], an SLC scheme utilizing a subset of the channel matrix was studied for CC.

In MMARQ systems, because of multiple ARQ processes, SLC schemes become significantly complex compared to those in MSARQ systems. In [19], a joint detection approach for CC was proposed based on the interference cancellation (IC) of successfully decoded packets. In [20], SLC schemes for CC were investigated according to the availability of IC and packet blanking. In [21], an SLC scheme based on QR decomposition with IC was proposed for CC. Meanwhile, in [22], an SLC scheme with IC of all terminated packets was investigated for CC. In [23], the SLC scheme without IC was developed for CC by considering the interference from the terminated packets as noise.

B. Notations

Vectors and matrices are denoted by lower-case and upper-case boldface letters, respectively. The superscripts T, $*$, H, and -1 are the transpose, conjugate, conjugate-and-transpose, and inverse operators, respectively. $\odot$ is an element-wise multiplication operator for vectors and matrices. $\mathbf{I}_j$ is the $j \times j$ identity matrix, and $\mathbf{0}_{i\times j}$ and $\mathbf{1}_{i\times j}$ are the $i \times j$ all-zero and all-one matrices, respectively. $\mathbf{A} = \mathrm{diag}(\mathbf{a})$ is the diagonal matrix $\mathbf{A}$ whose diagonal elements are $\mathbf{a}$. Finally, $\mathbb{E}[\cdot]$ is the mathematical expectation.

II. System Model

We consider a spatially multiplexed MIMO system with packet transmission and retransmission by HARQ, where N and M denote the numbers of transmit and receive antennas, respectively. R is a system limit for the number of transmissions (HARQ round) of a packet, and each packet has an independent ARQ process. Each transmit symbol of a packet is modulated using a 2^Q-ary constellation $\mathcal{S}$ with $\sum_{s\in\mathcal{S}} s = 0$ and $\sum_{s\in\mathcal{S}} |s|^2 = 2^Q$. Each packet contains J transmit symbols for its $r(1 \le r \le R)$th HARQ round; these J transmit symbols are classified as data and parity symbols, including the data and parity bits of the packet for the current HARQ round, respectively. Let C be the code rate of a packet for each HARQ round. Then, assuming an integer JC for simplicity, each packet contains JC data and $J(1-C)$ parity symbols for each HARQ round. For CC, by sending the same codeword for every retransmission, both data and parity symbols become identical regardless of the HARQ round. Meanwhile, for IR, JC data symbols are identical regardless of the HARQ round, whereas the remaining $J(1-C)$ parity symbols are regenerated according to the HARQ round because new redundancy information is sent for retransmission.

Let P denote the number of ARQ processes in each TTI. We consider the following two cases: $P = 1$ for MSARQ systems with single-packet transmission and $P = N$ for MMARQ systems with multiple-packet transmission. When $P = 1$, the transmit symbols of a packet are spatially multiplexed across the transmit antennas, and the number of the transmit signal vectors for each TTI L is J/N, assuming J is a multiple of N for simplicity. Meanwhile, when $P = N$, each transmit antenna is utilized to send transmit symbols of each packet; therefore, $L = J$. Then, for both MSARQ and MMARQ systems, the $l(1 \le l \le L)$th receive signal vector at the tth TTI can be expressed as follows:

$$\mathbf{y}_{t,l} = \mathbf{H}_{t,l}\mathbf{x}_{t,l} + \mathbf{n}_{t,l}. \tag{1}$$

In (1), $\mathbf{x}_{t,l}$ is the $N \times 1$ transmit signal vector, $\mathbf{y}_{t,l}$ is the $M \times 1$ receive signal vector, $\mathbf{H}_{t,l}$ is the $M \times N$ channel matrix, and $\mathbf{n}_{t,l}$ is the $M \times 1$ additive white Gaussian noise (AWGN) vector with zero-mean and a covariance matrix $\mathbb{E}[\mathbf{n}_{t,l}\mathbf{n}_{t,l}^H] = \sigma^2\mathbf{I}_M$.

At the receiver, LLRs for the packet(s) transmitted in the tth TTI are generated considering the HARQ rounds of the packet(s), and the generated LLRs are utilized for decoding each packet. A packet that fails at decoding when its HARQ round is less than R is retransmitted at the next $(t+1)$th TTI. Otherwise, if a packet is successfully decoded or its HARQ round is R, the packet is terminated, and a new packet is transmitted from the $(t+1)$th TTI.

In the following subsections, a multiplexing strategy for mapping the transmit symbols of the packet(s) to the transmit signal vectors is explained for MSARQ and MMARQ systems.

A. MSARQ Systems With Single-Packet Transmission

Let $\mathbf{s}_r = [s_r(1), s_r(2), \ldots, s_r(J)]$ denote the $1 \times J$ transmit symbol vector of the packet for its $r(1 \le r \le R)$th HARQ round, where $s_r(j)$ is the jth transmit symbol of $\mathbf{s}_r$. Therefore, $\{s_r(1), \ldots, s_r(JC)\}$ and $\{s_r(JC+1), \ldots, s_r(J)\}$ denote the data and parity symbols of the packet for the rth HARQ round, respectively. Then, the elements in $\mathbf{s}_r$ can be multiplexed to transmit signal vectors of the current tth TTI as follows:

$$\begin{aligned}\mathbf{x}_{t,l} &= [x_{t,l}(1), \ldots, x_{t,l}(N)]^T \\ &= [s_r(l), \ldots, s_r(l+(N-1)L)]^T.\end{aligned} \tag{2}$$

In (2), $x_{t,l}(n)(= s_r(l+(n-1)L))$ is the nth element of $\mathbf{x}_{t,l}$, which is sent from the nth transmit antenna. Therefore, $x_{t,l}(n)$ with $1 \le l+(n-1)L \le JC$ is the repeatedly transmitted data symbol, where $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 1$, regardless of the HARQ retransmission strategy, if $r > 1$. Meanwhile, $x_{t,l}(n)$ for $JC+1 \le l+(n-1)L \le J$ is the parity symbol. Therefore, if $r > 1$, $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 0$ for IR, whereas $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 1$ for CC. Furthermore, if $r = 1$, indicating that a new packet is transmitted from the tth TTI, $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 0$, regardless of l, n, and HARQ retransmission strategy.

B. MMARQ Systems With Multiple-Packet Transmission

Let $\mathbf{s}_{t,n} = [s_{t,n}(1), s_{t,n}(2), \ldots, s_{t,n}(J)]$ denote the $1 \times J$ transmit symbol vector of the packet transmitted from the nth transmit antenna at the tth TTI. Because $L = J$, the elements in $\{\mathbf{s}_{t,1}, \ldots, \mathbf{s}_{t,N}\}$ can be multiplexed to $\{\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,L}\}$ as follows:

$$\mathbf{x}_{t,l} = [x_{t,l}(1), \ldots, x_{t,l}(N)]^T = [s_{t,1}(l), \ldots, s_{t,N}(l)]^T. \tag{3}$$

Therefore, $x_{t,l}(n) = s_{t,n}(l)$, regardless of t, n, and l. Let $\mathbf{r}_t = [r_t(1), \ldots, r_t(N)]^T$ denote the $N \times 1$ vector, indicating the HARQ round of the active packets at the tth TTI, where $r_t(n)$ is the HARQ round of the packet from the nth transmit antenna with $1 \le r_t(n) \le R$. Therefore, if $r_t(n) > 1$ and $1 \le l \le JC$, $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 1$, regardless of the HARQ retransmission strategy. In contrast, if $r_t(n) > 1$ and $JC+1 \le l \le J$, $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 1$ for CC and 0 for IR. Furthermore, similar to the case of MSARQ systems, if $r_t(n) = 1$, indicating that a new packet is sent from the nth transmit antenna, $\mathbb{E}[x_{t-1,l}(n)x_{t,l}^*(n)] = 0$ for the given n, regardless of l and HARQ retransmission strategy.

TABLE I
EXAMPLES OF TRANSITION INDICATING VECTOR $\mathbf{d}_{t,l}$

MSARQ		$r = 1$	$r > 1$
	CC	$\mathbf{0}_{N\times 1}$	$\mathbf{1}_{N\times 1}$
	IR		$[\mathbf{1}_{1\times NC}, \mathbf{0}_{1\times N(1-C)}]^T$
MMARQ		$1 \le l \le JC$	$JC+1 \le l \le J$
	CC	$\mathbf{p}_t$	$\mathbf{p}_t$
	IR		$\mathbf{0}_{N\times 1}$

III. PROPOSED KALMAN FILTERING BASED COMBINING

A. State-Space Model

First, we develop a state-space model for the proposed KC scheme based on the system model described in Section II. In this model, each transmit signal vector $\mathbf{x}_{t,l}$ is considered to be the state vector, and the state is transited from $\mathbf{x}_{t-1,l}$ to $\mathbf{x}_{t,l}$ by packet transmission and retransmission at the consecutive TTIs. Therefore, the process equation for $\mathbf{x}_{t,l}$ $(1 \le l \le L)$ can be written as follows:

$$\mathbf{x}_{t,l} = \mathbf{F}_{t,l}\mathbf{x}_{t-1,l} + \mathbf{w}_{t,l}, \tag{4}$$

where $\mathbf{F}_{t,l}$ and $\mathbf{w}_{t,l}$ are the $N \times N$ state transition matrix and $N \times 1$ process noise vector, respectively. Let $\mathbf{d}_{t,l} = [d_{t,l}(1), \ldots, d_{t,l}(N)]^T = \mathbb{E}[\mathbf{x}_{t-1,l} \odot \mathbf{x}_{t,l}^*]$ be the $N \times 1$ transition indicating vector for $\mathbf{x}_{t,l}$, where $d_{t,l}(n) = 1$ if $x_{t-1,l}(n)$ is identical to $x_{t,l}(n)$ (e.g., a retransmitted data symbol) and 0 otherwise. Then, $\mathbf{F}_{t,l}$ and $\mathbf{w}_{t,l}$ can be written as follows:

$$\mathbf{F}_{t,l} = \operatorname{diag}(\mathbf{d}_{t,l}) \tag{5}$$

and

$$\mathbf{w}_{t,l} = \mathbf{x}_{t,l} \odot (\mathbf{1}_{N\times 1} - \mathbf{d}_{t,l}). \tag{6}$$

In Table I, examples of $\mathbf{d}_{t,l}$ according to the system configuration are shown, where an integer NC is assumed for MSARQ systems. $\mathbf{p}_t = [p_t(1), \ldots, p_t(N)]^T$ denotes the retransmission status in MMARQ systems, where $p_t(n)$ is 0 for $r_t(n) = 1$ and 1 for $2 \le r_t(n) \le R$. As shown in Table I, $\mathbf{d}_{t,l}$ can be easily determined according to the system configuration that should be shared between the transmitter and receiver for successful packet transmission.[1] Therefore, the process equation in (4) can

[1]For normal functioning of reception procedures, in addition to the number of HARQ processes and the retransmission strategy, the receiver should know (i) the position in the mother codeword of the coded bits consisting each symbol and (ii) the packet to which each symbol belongs. For packet transmission, this information is usually shared between the transmitter and receiver via control channel. Therefore, the receiver and transmitter can calculate $\mathbf{d}_{t,l} = \mathbb{E}[\mathbf{x}_{t-1,l} \odot \mathbf{x}_{t,l}^*]$ by deciding whether each $x_{t-1,l}(n)$ is retransmitted as $x_{t,l}(n)$, as described in Sections II-A and II-B.

represent the transition behavior of $\mathbf{x}_{t,l}$ in both MSARQ and MMARQ systems with CC or IR.

Further, the observation equation is equivalent to (1), which is

$$\mathbf{y}_{t,l} = \mathbf{H}_{t,l}\mathbf{x}_{t,l} + \mathbf{n}_{t,l}, \tag{7}$$

where $\mathbf{y}_{t,l}$ and $\mathbf{n}_{t,l}$ become the $M \times 1$ observation vector and observation noise vector, respectively, and $\mathbf{H}_{t,l}$ becomes the $M \times N$ observation matrix.

Note that the state-space model in (4)–(7) is valid for a general MIMO system, while each term in (4)–(7) can be changed according to the specific system configuration.

B. Proposed Kalman Filtering Based Combining

The proposed KC scheme includes three sequentially performed stages: KF stage, LLR calculation stage, and optional LLR combining stage for IR. The output LLRs of the proposed scheme are used as the input for decoding. The following section presents the detailed procedures of each stage.

1) Kalman Filtering Stage: Let $\hat{\mathbf{x}}_{t,l}$ denote the $N \times 1$ filtered estimate of $\mathbf{x}_{t,l}$ using all observations up to the tth TTI, and let $\mathbf{P}_{t,l} = \mathbb{E}[(\mathbf{x}_{t,l} - \hat{\mathbf{x}}_{t,l})(\mathbf{x}_{t,l} - \hat{\mathbf{x}}_{t,l})^H]$ denote its $N \times N$ error covariance matrix. $\mathbf{x}_{0,l}$ and $\mathbf{P}_{0,l}$ can be initialized to all-zero for $1 \le l \le L$. In addition, let $\hat{\mathbf{x}}_{t,l}^{(-)}$ denote the $N \times 1$ predicted estimate of $\mathbf{x}_{t,l}$ using all observations up to the $(t-1)$th TTI and $\mathbf{P}_{t,l}^{(-)} = \mathbb{E}[(\mathbf{x}_{t,l} - \hat{\mathbf{x}}_{t,l}^{(-)})(\mathbf{x}_{t,l} - \hat{\mathbf{x}}_{t,l}^{(-)})^H]$. Then, without a priori information regarding $\mathbf{x}_{t,l}$, $\hat{\mathbf{x}}_{t,l}^{(-)}$ and $\mathbf{P}_{t,l}^{(-)}$ can be obtained from $\hat{\mathbf{x}}_{t-1,l}$ and $\mathbf{P}_{t-1,l}$ as [24], [25]

$$\hat{\mathbf{x}}_{t,l}^{(-)} = \mathbf{F}_{t,l}\hat{\mathbf{x}}_{t-1,l} \tag{8}$$

and

$$\mathbf{P}_{t,l}^{(-)} = \mathbf{F}_{t,l}\mathbf{P}_{t-1,l}\mathbf{F}_{t,l}^T + \mathbf{W}_{t,l}, \tag{9}$$

where $\mathbf{W}_{t,l} = \mathbb{E}[\mathbf{w}_{t,l}\mathbf{w}_{t,l}^H] = \mathbf{I}_N - \mathbf{F}_{t,l}$ is the $N \times N$ process noise covariance matrix.

After the prediction step in (8) and (9), a correction step is performed to obtain $\hat{\mathbf{x}}_{t,l}$ and $\mathbf{P}_{t,l}$ from $\hat{\mathbf{x}}_{t,l}^{(-)}$ and $\mathbf{P}_{t,l}^{(-)}$. Based on (7), the correction step can be written as follows [24]:

$$\hat{\mathbf{x}}_{t,l} = \hat{\mathbf{x}}_{t,l}^{(-)} + \mathbf{K}_{t,l}(\mathbf{y}_{t,l} - \mathbf{H}_{t,l}\hat{\mathbf{x}}_{t,l}^{(-)}) \tag{10}$$

and

$$\mathbf{P}_{t,l} = \mathbf{P}_{t,l}^{(-)} - \mathbf{K}_{t,l}\mathbf{H}_{t,l}\mathbf{P}_{t,l}^{(-)}, \tag{11}$$

where $\mathbf{K}_{t,l}$ is the $N \times M$ Kalman gain matrix given by

$$\mathbf{K}_{t,l} = \mathbf{P}_{t,l}^{(-)}\mathbf{H}_{t,l}^H(\mathbf{H}_{t,l}\mathbf{P}_{t,l}^{(-)}\mathbf{H}_{t,l}^H + \sigma^2\mathbf{I}_M)^{-1}. \tag{12}$$

Therefore, $\hat{\mathbf{x}}_{t,l}$ is obtained by the KF operation in (8)–(12), which is an LMMSE estimate of $\mathbf{x}_{t,l}$ using the information aggregated up to the tth TTI by the state transition.

Although the correction step in (10)–(12) provides an LMMSE estimate of $\mathbf{x}_{t,l}$ based on SLC, it requires several matrix multiplication and inverse operations. To avoid such matrix operations, a low-complexity correction step is developed, which generates the same $\hat{\mathbf{x}}_{t,l}$ and $\mathbf{P}_{t,l}$ to the correction step in (10)–(12).

For the low-complexity correction step, first, the single $M \times 1$ observation vector $\mathbf{y}_{t,l}$ in (7) is modeled as the M consecutive scalar observations, i.e.,

$$y_{t,l}(m) = \mathbf{h}_{t,l,m}^T\mathbf{x}_{t,l} + n_{t,l}(m), 1 \le m \le M, \tag{13}$$

where $y_{t,l}(m)$ and $n_{t,l}(m)$ are the mth elements of $\mathbf{y}_{t,l}$ and $\mathbf{n}_{t,l}$, respectively, and $\mathbf{h}_{t,l,m}^T$ is the $1 \times N$ vector corresponding to the mth row of $\mathbf{H}_{t,l}$.

Then, the state transition of $\mathbf{x}_{t-1,l} \to \mathbf{x}_{t,l}$ with an $M \times 1$ observation vector based on (7) can be interpreted as M sub-state transitions of $\mathbf{x}_{t-1,l} \to \mathbf{x}_{t,l,1} \to \cdots \to \mathbf{x}_{t,l,M}$ with M scalar observations based on (13), where $\mathbf{x}_{t,l,m}$ is the mth sub-state vector considering the mth scalar observation $y_{t,l}(m)$. Because $\mathbf{x}_{t,l,m} = \mathbf{x}_{t,l}$ for $1 \le m \le M$, the state-transition matrix for each sub-state transition from $\mathbf{x}_{t,l,1}$ to $\mathbf{x}_{t,l,M}$ is $\mathbf{I}_N$. Consequently, the prediction steps for the sub-state transitions can be omitted, except $\mathbf{x}_{t-1,l} \to \mathbf{x}_{t,l,1}$, where the prediction of $\mathbf{x}_{t,l,1} (= \mathbf{x}_{t,l})$ from $\mathbf{x}_{t-1,l}$ can be performed via the prediction step in (8) and (9). Therefore, with $\hat{\mathbf{x}}_{t,l,0} = \hat{\mathbf{x}}_{t,l}^{(-)}$ and $\mathbf{P}_{t,l,0} = \mathbf{P}_{t,l}^{(-)}$, the low-complexity correction step is sequentially performed from $m = 1$ to M as follows:

$$\hat{\mathbf{x}}_{t,l,m} = \hat{\mathbf{x}}_{t,l,m-1} + \mathbf{k}_{t,l,m}(y_{t,l}(m) - \mathbf{h}_{t,l,m}^T\hat{\mathbf{x}}_{t,l,m-1}) \tag{14}$$

and

$$\mathbf{P}_{t,l,m} = \mathbf{P}_{t,l,m-1} - \mathbf{k}_{t,l,m}\mathbf{h}_{t,l,m}^T\mathbf{P}_{t,l,m-1}, \tag{15}$$

where $\mathbf{k}_{t,l,m}$ is the $N \times 1$ Kalman gain vector expressed as

$$\mathbf{k}_{t,l,m} = \frac{\mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^*}{\mathbf{h}_{t,l,m}^T\mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^* + \sigma^2}. \tag{16}$$

The calculated $\hat{\mathbf{x}}_{t,l,M}$ and $\mathbf{P}_{t,l,M}$ are then set to $\hat{\mathbf{x}}_{t,l}$ and $\mathbf{P}_{t,l}$, respectively. Therefore, the matrix-matrix multiplications and matrix inversion in (10)–(12) are replaced by matrix-vector multiplications and vector-scalar division, respectively. Furthermore, because the observation equation in (7) is linear and the observation noise $\mathbf{n}_{t,l}$ is AWGN, $\hat{\mathbf{x}}_{t,l}$ and $\mathbf{P}_{t,l}$ from the low-complexity correction step in (14)–(16) are identical to those from the correction step in (10)–(12) [24].

In Algorithm 1, the proposed KF stage with the low-complexity correction step is summarized. $\mathbf{v}_{t,l,m} = \mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^*$ in Algorithm 1 is defined to reduce the number of calculations of $\mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^*$ in both $\mathbf{P}_{t,l,m}$ in (15) and $\mathbf{k}_{t,l,m}$ in (16).

Next, we compare $\hat{\mathbf{x}}_{t,l}$ to the estimate of other combining schemes with LMMSE detection. When $\mathbf{F}_{t,l} = \mathbf{0}_{N\times N}$ regardless of t and l (e.g., full IR), $\hat{\mathbf{x}}_{t,l}$ is equivalent to the estimate of $\mathbf{x}_{t,l}$ by BLC with LMMSE detection because $\mathbf{K}_{t,l}$ becomes $(\mathbf{H}_{t,l}^H\mathbf{H}_{t,l} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{H}_{t,l}^H$. In addition, for general $\mathbf{F}_{t,l}$, we have the following lemma:

Lemma 1 (Equivalence of the estimated $\mathbf{x}_{t,l}$ in the proposed and direct SLC schemes): $\hat{\mathbf{x}}_{t,l}$ is identical to the corresponding estimate of $\mathbf{x}_{t,l}$ obtained by the direct SLC scheme with LMMSE detection at the tth TTI.

Proof: See Appendix A. ■

Therefore, the proposed and direct SLC schemes at the tth TTI obtain the identical estimate for $\mathbf{x}_{t,l}$ regardless of the number

Algorithm 1: KF Stage with Low-Complexity Correction Step.

Input: l, σ^2, $\mathbf{y}_{t,l}$, $\mathbf{H}_{t,l}$, $\mathbf{F}_{t,l}$, $\mathbf{W}_{t,l}$, $\hat{\mathbf{x}}_{t-1,l}$, $\mathbf{P}_{t-1,l}$

1: **Prediction**: $\hat{\mathbf{x}}_{t,l}^{(-)} = \mathbf{F}_{t,l}\hat{\mathbf{x}}_{t-1,l}$, $\mathbf{P}_{t,l}^{(-)} = \mathbf{F}_{t,l}\mathbf{P}_{t-1,l}\mathbf{F}_{t,l}^T + \mathbf{W}_{t,l}$;

2: **Correction Step Initialization**: $\hat{\mathbf{x}}_{t,l,0} = \hat{\mathbf{x}}_{t,l}^{(-)}$, $\mathbf{P}_{t,l,0} = \mathbf{P}_{t,l}^{(-)}$;

3: **for** $m = 1$ **to** M **do**

4: $\quad \mathbf{v}_{t,l,m} = \mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^*$;

5: $\quad \mathbf{k}_{t,l,m} = \mathbf{v}_{t,l,m}/(\mathbf{h}_{t,l,m}^T\mathbf{v}_{t,l,m} + \sigma^2)$;

6: $\quad \hat{\mathbf{x}}_{t,l,m} = \hat{\mathbf{x}}_{t,l,m-1} + \mathbf{k}_{t,l,m}(y_{t,l}(m) - \mathbf{h}_{t,l,m}^T\hat{\mathbf{x}}_{t,l,m-1})$;

7: $\quad \mathbf{P}_{t,l,m} = \mathbf{P}_{t,l,m-1} - \mathbf{k}_{t,l,m}\mathbf{v}_{t,l,m}^H$;

8: **end for**

9: **Output**: $\hat{\mathbf{x}}_{t,l} = \hat{\mathbf{x}}_{t,l,M}$, $\mathbf{P}_{t,l} = \mathbf{P}_{t,l,M}$;

of HARQ processes and retransmission strategy. However, the direct SLC scheme requires a significantly larger computational complexity and has to be designed in a dedicated manner for a specific system configuration.

2) LLR Calculation Stage: After the KF operation, LLRs of the bits consisting of each element of $\mathbf{x}_{t,l}$ must be calculated from $\hat{\mathbf{x}}_{t,l}$ and $\mathbf{P}_{t,l}$ for FEC decoding. The residual interference with noise in $\hat{x}_{t,l}(n)$, the nth element of $\hat{\mathbf{x}}_{t,l}$, can be approximated as a Gaussian distribution with variance $p_{t,l}(n)$, which is the nth diagonal element of $\mathbf{P}_{t,l}$ [25]. Let $b_{t,l,n}^q$ denote the $q(1 \le q \le Q)$th bit comprising $x_{t,l}(n)$. Then, $l_{t,l,n}^q$, the LLR of $b_{t,l,n}^q$ can be calculated as follows:

$$l_{t,l,n}^q = \log \frac{\sum_{\forall s \in \mathcal{S}, b_{t,l,n}^q = 1} \exp\left(\frac{-|\hat{x}_{t,l}(n) - s|^2}{p_{t,l}(n)}\right)}{\sum_{\forall s \in \mathcal{S}, b_{t,l,n}^q = 0} \exp\left(\frac{-|\hat{x}_{t,l}(n) - s|^2}{p_{t,l}(n)}\right)}. \tag{17}$$

3) LLR Combining Stage: For LLRs of the bits comprising repeatedly transmitted symbols included in $\mathbf{x}_{t,l}$, e.g., the transmit symbols of the packet(s) with CC and data symbols of the packet(s) with IR, LLR combining is not required. Meanwhile, the parity symbols of the packet sent until the last $(t-1)$th TTI are not included in $\mathbf{x}_{t,l}$ if IR is employed. Therefore, for IR, the LLRs of the coded bits comprising the parity symbols sent until the $(t-1)$th TTI should be stored to be combined with the currently obtained LLRs for decoding.[2]

[2]In general, all LLRs of the bits comprising the symbols that are not sent at the next TTI should be stored in the proposed scheme for combining in future TTIs in which the packet is retransmitted. For example, if the symbol $x_{t,l}(n)$ for a packet is sent at the $(t-3)$th, $(t-2)$th, and tth TTIs, then $l_{t-2,l,n}^q$ should be pre-stored and combined with $l_{t,l,n}^q$ at the tth TTI to utilize the information for $x_{t,l}(n)$ obtained at the $(t-3)$th and $(t-2)$th TTIs.

C. Computational Complexity and Memory Units

In this subsection, the complexity and memory size of the proposed scheme for detection (i.e., the KF stage) are derived and compared with those of other combining schemes.

By the structure of $\mathbf{F}_{t,l}$, the prediction step in (8) and (9) becomes the selection of subsets in $\hat{\mathbf{x}}_{t-1,l}$ and $\mathbf{P}_{t-1,l}$. Therefore, the correction step in (14)–(16) mainly yield the complexity. Considering the highest-order terms only, the complexity is governed by calculations of $\mathbf{P}_{t,l,m-1}\mathbf{h}_{t,l,m}^*$ ($= \mathbf{v}_{t,l,m}$ in Algorithm 1) and $\mathbf{k}_{t,l,m}\mathbf{h}_{t,l,m}^T\mathbf{P}_{t,l,m-1}$ ($= \mathbf{k}_{t,l,m}\mathbf{v}_{t,l,m}^H$), and each of them entails the complexity of $\mathcal{O}(N^2)$. Because (14)–(16) should be performed M times, the complexity of the KF stage is $\mathcal{O}(2\ N^2\ M)$ for each $\mathbf{x}_{t,l}$. Meanwhile, the complexity with the correction step in (10)–(12) is $\mathcal{O}(2\ N^2\ M + 2NM^2 + M^3)$ for each $\mathbf{x}_{t,l}$. Therefore, the low-complexity correction step can significantly reduce the complexity of the KF stage.

Next, the memory requirement for the KF stage is derived, in which the memory size to store one complex number is denoted as one memory unit. At the KF stage, only $\hat{\mathbf{x}}_{t-1,l}$ and $\mathbf{P}_{t-1,l}$ need to be stored for the current tth TTI. Therefore, $N(N+1)$ memory units are required for each transmit signal vector, which can be further reduced because $\mathbf{P}_{t,l}$ is Hermitian.

Table II compares the computational complexity, memory units, and error performance of the combining schemes with LMMSE detection. The existing SLC schemes require lower complexity than the brute-force direct SLC scheme; however, they require additional overhead from BLC to aggregate information. In contrast, the complexity of the proposed scheme is comparable to that of BLC and is significantly smaller than that of the conventional SLC schemes.[3] In addition, the proposed scheme requires lesser memory units than most conventional SLC schemes. Even with the low computational complexity and memory units, the proposed scheme can be applied to both MSARQ and MMARQ systems regardless of the retransmission strategy, unlike conventional SLC schemes. Furthermore, the proposed scheme can outperform BLC and achieve near-identical error performance to the direct SLC scheme, as shown in Section IV and Appendices A and B.

IV. Simulation Results

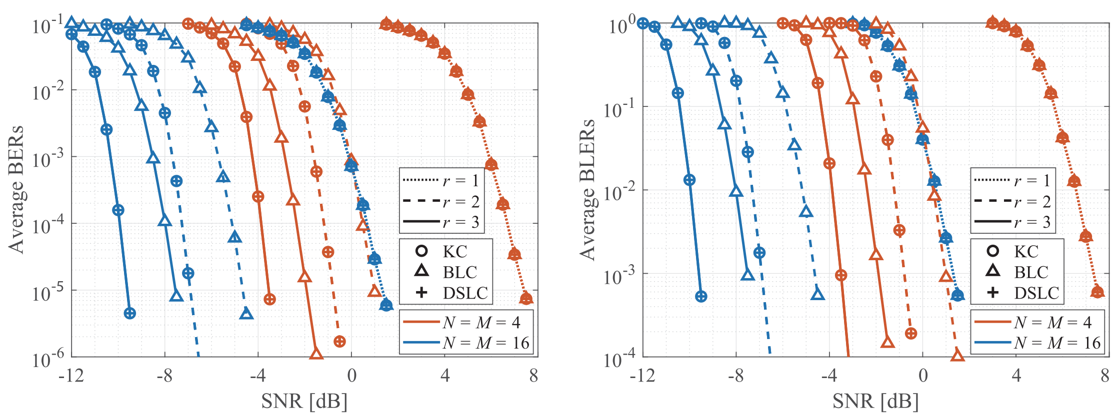
In this section, the average bit-error ratio (BER) and block-error ratio (BLER) of the combining schemes are evaluated. A block error is defined as the case in which at least one bit-error occurs among the decoded data bits of a packet. In addition to the proposed scheme, the BLC and direct SLC schemes are considered as reference schemes. In MMARQ systems, the size of the aggregated channel matrix for the direct SLC scheme increases unless all active packets are simultaneously terminated at the same TTI; this leads to an excessively large complexity for simulations. To prevent such situations, in MMARQ systems, all active packets are forcefully terminated for every 100 TTIs.

[3]Although the complexity based on the highest-order terms of the proposed scheme is lower than that of BLC, the proposed scheme has several low-order terms, whereas BLC has few low-order terms. In that sense, despite the advantage in terms of complexity for a large N, we consider the complexity of both schemes to be similar in this paper.

TABLE II
COMPUTATIONAL COMPLEXITY, MEMORY REQUIREMENT, AND PERFORMANCE OF THE COMBINING SCHEMES

System applicability	Combining scheme	Computational complexity	Memory requirement	Performance
Applicable to all	Proposed KC	$\mathcal{O}(2N^2M)$	$N(N+1)$[g]	high
	Conventional BLC	$\mathcal{O}(N^3+2N^2M)$	-[h]	low
MSARQ-CC	Direct SLC	$\mathcal{O}(N^3+2rN^2M)$	$rM(N+1)$	high
	Pre-combining [9]	$\mathcal{O}(N^3+(r+1)N^2M)$	$N(N+1)$	high
	Maximal-ratio combining [10]	$\mathcal{O}(4N^3+N^2M)$	$N(N+1)$	high
	SLC w/ direct QR [12]	$\mathcal{O}(3N^3+rN^2M)$	$rM(N+1)$	high
	SLC w/ incremental QR [12]	$\mathcal{O}(3N^3+N^2(M+\alpha_r))$[a]	$N(N+3)/2$	high
MSARQ-IR	Direct SLC	$\mathcal{O}(\beta_r^3N^3+2r\beta_r^2N^2M)$[b,j]	$rM(N+1)$	high
	Extended detection [14]	$\mathcal{O}(\gamma_rN^3+\delta_rN^2M)$[c,j]	$rM(N+1)+rN^2C(1-C)$	high
MMARQ-CC	Direct SLC	$\mathcal{O}(N_t^3+2N_t^2M_t)$[d]	$tM(N+1)$[i]	high
	SLC w/ packet elimination [19], [22]	$\mathcal{O}(N^3+2r_t^*N^2M)$[e,f]	$r_t^*M(N+1)$	high
	Aggregation-assisted combining [23]	$\mathcal{O}(N^3+2r_t^*N^2M)$[e]	$r_t^*M(N+1)$	moderate
MMARQ-IR	Direct SLC	$\mathcal{O}(N_t^3+2N_t^2M_t)$[d,j]	$tM(N+1)$[i]	high

[a] $\alpha_r=0$ for $r=1$ and N for $r\geq 2$.
[b] $\beta_r=1+(r-1)(1-C)$.
[c] $\gamma_r=1$ and $\delta_r=2$ for $r=1$, and $\gamma_r=C^3+(1-C)^3$ and $\delta_r=1+(1-C)^2+r\beta_rC$ for $r\geq 2$.
[d] N_t and M_t are the numbers of columns and rows in the utilized aggregated channel matrix, respectively. See Appendix A for details.
[e] $r_t^*=\max(\mathbf{r}_t)$.
[f] The complexity of the terminated packet elimination is additionally required.
[g] The LLRs of the previously transmitted parity bits for the current active packet(s) must be stored for IR.
[h] The LLRs obtained at the previous HARQ rounds of the current active packet(s) need to be stored.
[i] This can be reduced to $(t-t^*+1)M(N+1)$, where $t^*-1(<t)$ is the last TTI in which all active packets are terminated. See Appendix A for details.
[j] The complexity for LLR calculation is higher than the other cases by the increased number of transmit symbols $(>N)$ for LLR calculation.

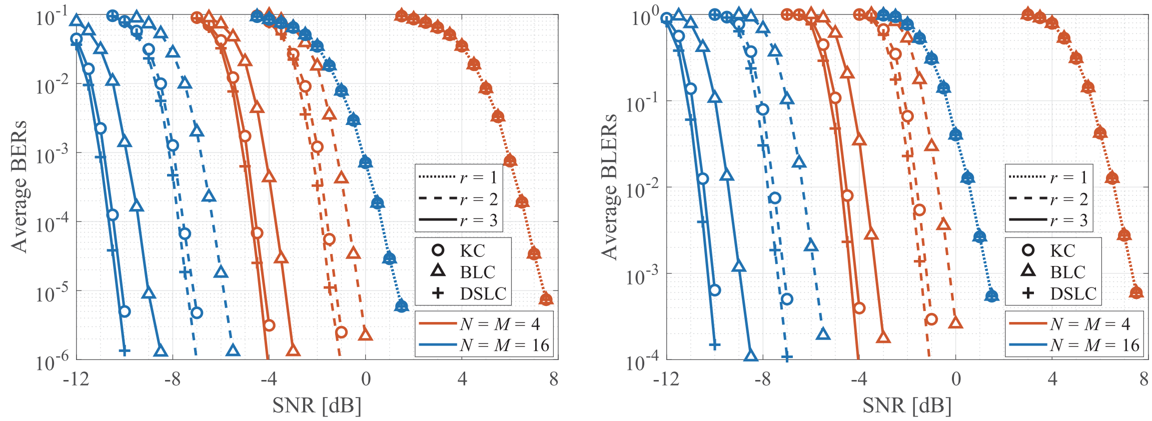


(a) (b)

Fig. 1. Error performance of combining schemes in MSARQ-CC systems: (a) average BER and (b) average BLER.

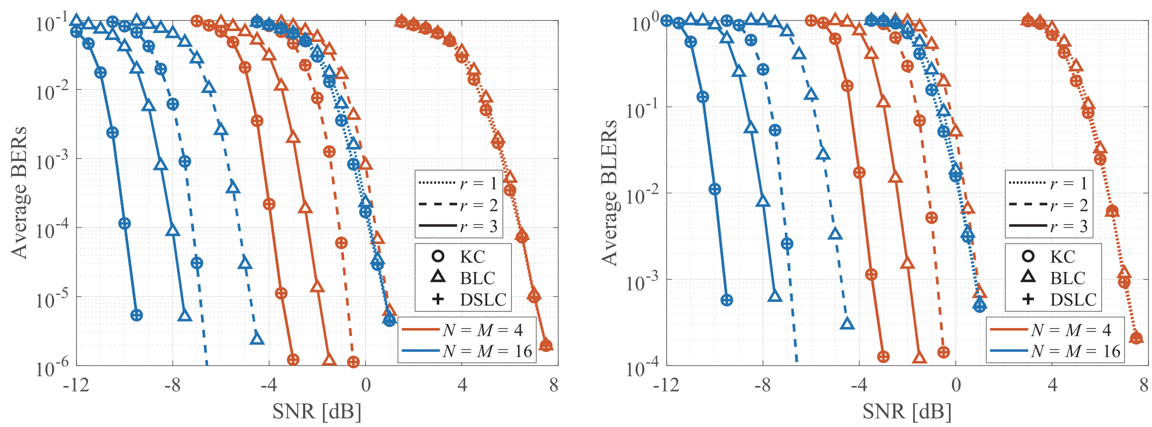
Two antenna configurations of $N=M=4$ and $N=M=16$ are considered, and $R=3$, i.e., maximum two retransmissions for each packet. A low-density parity-check (LDPC) code in [3] with a rate of 0.5 and a codeword length of 1440 is considered as the mother code. As $C=3/4$, 960 bits among the mother codeword of a packet are transmitted for each HARQ round. The quadrature phase shift keying (QPSK) modulation with $Q=2$ is considered; therefore, $J=480$. The independent Rayleigh fading channel in which the channel response independently varies for transmit signal vectors is considered. Furthermore, the maximum number of decoding iterations for a packet at each TTI is set to 20. After decoding, the syndromes of cyclic redundancy check (CRC) and LDPC codes are used to detect errors in the decoded packets for a retransmission request, where CRC-32 is considered.

Figs. 1 and 2 depict the error performance of the combining schemes in MSARQ systems with CC and IR, respectively. As analyzed in Appendix B, compared with the direct SLC scheme, the proposed scheme shows identical error performance for CC regardless of the antenna configuration. Meanwhile, the performance for IR slightly degrades with retransmission $(r\geq 2)$ for both $N=M=4$ and $N=M=16$, because the LLRs of the parity bits transmitted at the previous HARQ rounds only are not updated in the future HARQ rounds, unlike the direct SLC scheme. Nevertheless, the gain in the signal-to-noise ratio (SNR) of the direct SLC scheme over the proposed scheme is negligible compared with that of the proposed scheme over BLC. For both CC and IR with $r\geq 2$, the proposed scheme outperforms BLC with a similar complexity. Further, as all combining schemes for $r=1$ in MSARQ systems become equivalent to the LMMSE



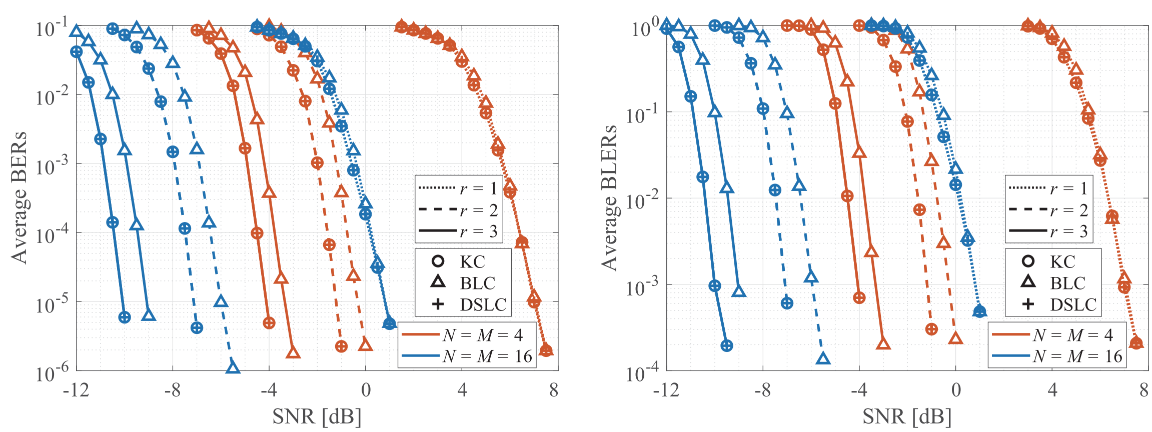
(a) (b)

Fig. 2. Error performance of combining schemes in MSARQ-IR systems: (a) average BER and (b) average BLER.



(a) (b)

Fig. 3. Error performance of combining schemes in MMARQ-CC systems: (a) average BER and (b) average BLER.



(a) (b)

Fig. 4. Error performance of combining schemes in MMARQ-IR systems: (a) average BER and (b) average BLER.

detection in MIMO systems without retransmission, the error performance of the combining schemes for $r = 1$ is identical in MSARQ systems.

Figs. 3 and 4 show the error performance of the combining schemes in MMARQ systems with CC and IR, respectively. Again, as analyzed in Appendix B, the proposed scheme achieves identical error performance to the direct SLC scheme for both CC and IR regardless of the antenna configuration and r. Meanwhile, the direct SLC scheme is impractical specifically for MMARQ systems, because of the possibility of having an extremely huge aggregated system model. In addition, the proposed and direct SLC schemes outperform BLC, and they achieve a slightly better error performance than BLC even for the initially transmitted packets ($r = 1$). This is because, unlike SLC in MSARQ systems, SLC in MMARQ systems can provide a better estimation result for initially transmitted packets as well as retransmitted packets [22].

V. Conclusion

In this study, we proposed and investigated an SLC scheme based on KF operation for MIMO systems with HARQ. The proposed scheme exhibits system applicability and computational complexity similar to the BLC scheme with degraded error performance. The error performance of the proposed scheme is nearly identical to that of the dedicated direct SLC scheme for a specific system configuration with high computational complexity. Therefore, the proposed KC scheme is considered an effective and powerful combining scheme for MIMO systems with HARQ.

Notably, the proposed scheme can be employed for other system models not considered in this paper. For example, through proper calculation of the transition indicating vector $\mathbf{d}_{t,l}$, the proposed scheme can be applied to MIMO-HARQ systems with different spatial multiplexing (e.g., $\mathbf{x}_{t,l} = [s_r((l-1)N + 1), \ldots, s_r(lN)]^T$ for MSARQ systems or a packet transmitting from multiple transmit antennas for MMARQ systems with $1 < P < N$), different HARQ retransmission strategies (e.g., full IR without repeatedly transmitted symbols [8]), acknowledgement feedback bundling (e.g., MSARQ systems with multiple packets forced to have the same ARQ process [15]), and packet blanking (e.g., sending no new packets until all the remaining packets are terminated [20]). In practical systems, the repeatedly transmitted symbols can be included in different transmit signal vectors or transmit antennas for each HARQ round [4], [5], and the proposed scheme can support such cases by simple modifications. In addition, the proposed scheme can be extended to employ iterative detection procedures for better error performance. These topics will be investigated in future works.

Appendix A
Proof of Lemma 1

Let $\breve{\mathbf{H}}_{t^*,l}$ denote the $M_{t^*} \times N_{t^*}$ aggregated channel matrix including all channel matrices up to the t^*th TTI $\{\mathbf{H}_{1,l}, \ldots, \mathbf{H}_{t^*,l}\}$, where M_{t^*} and N_{t^*} are the numbers of aggregated receive and transmit antennas in $\breve{\mathbf{H}}_{t^*,l}$, respectively. We assume that all packets transmitted at the (t^*-1)th TTI are simultaneously terminated after the reception procedures. Then, all elements of $\mathbf{x}_{t^*,l}$ are new transmit symbols, regardless of the number of HARQ processes and retransmission strategy, which should be regarded as those sent from additional transmit antennas in the aggregated system model [14], [19]–[23]. Therefore, $\breve{\mathbf{H}}_{t^*,l}$ can be written as follows:

$$\breve{\mathbf{H}}_{t^*,l} = \begin{bmatrix} \breve{\mathbf{H}}_{t^*-1,l} & \mathbf{0}_{M_{t^*-1}\times N} \\ \mathbf{0}_{M\times N_{t^*-1}} & \mathbf{H}_{t^*,l} \end{bmatrix}. \tag{18}$$

As the two column partitions of $\breve{\mathbf{H}}_{t^*,l}$ in (18) are orthogonal, the estimate of $\mathbf{x}_{t,l}$ using $\breve{\mathbf{H}}_{t^*,l}$ is identical to that using $\mathbf{H}_{t^*,l}$, regardless of the detection criterion. Therefore, if all active packets are simultaneously terminated at (t^*-1)th TTI, $\breve{\mathbf{H}}_{t^*,l}$ can be initialized to $\mathbf{H}_{t^*,l}$ of a much smaller size, and the aggregation can be performed from the initialized $\breve{\mathbf{H}}_{t^*,l}$ at future TTIs until the condition is satisfied again. Note that the packet termination in MSARQ systems always satisfies the above condition.

Based on the above strategy, instead of $\breve{\mathbf{H}}_{t,l}$, we consider the effective aggregated channel matrix $\tilde{\mathbf{H}}_{t,l}$, which includes the channel matrices $\{\mathbf{H}_{t^*,l}, \ldots, \mathbf{H}_{t,l}\}$, i.e., the (t^*-1)th TTI with $t^* \le t$ was the last TTI satisfying the condition of simultaneous packet termination. For simple notations, we assume that i) $t^* = 1$ and ii) a symbols are repeatedly transmitted from the first to the ath transmit antennas, and $b(= N - a)$ symbols are always newly transmitted from the $(a+1)$ to Nth transmit antennas. Then, the numbers of aggregated transmit and receive antennas are $N_t = a + bt$ and $M_t = Mt$, respectively, and $\tilde{\mathbf{H}}_{t,l}$ becomes the $M_t \times N_t$ matrix. Let $\tilde{\mathbf{x}}_{t,l} = [x_{1,l}(1), \ldots, x_{1,l}(a), x_{1,l}(a+1), \ldots, x_{t,l}(N)]^T$ denote the $N_t \times 1$ aggregated transmit signal vector for $\tilde{\mathbf{H}}_{t,l}$. Then, the aggregated receive signal vector for $\tilde{\mathbf{H}}_{t,l}$ can be represented as follows:

$$\tilde{\mathbf{y}}_{t,l} = \tilde{\mathbf{H}}_{t,l}\tilde{\mathbf{x}}_{t,l} + \tilde{\mathbf{n}}_{t,l}, \tag{19}$$

where $\tilde{\mathbf{y}}_{t,l} = [\mathbf{y}_{1,l}^T, \ldots, \mathbf{y}_{t,l}^T]^T$ and $\tilde{\mathbf{n}}_{t,l} = [\mathbf{n}_{1,l}^T, \ldots, \mathbf{n}_{t,l}^T]^T$ denote the $M_t \times 1$ receive signal and AWGN vectors for $\tilde{\mathbf{H}}_{t,l}$, respectively.

Based on (19), the direct SLC scheme with LMMSE detection at the tth TTI can be performed as follows:

$$\hat{\tilde{\mathbf{x}}}_{t,l} = (\tilde{\mathbf{H}}_{t,l}^H\tilde{\mathbf{H}}_{t,l} + \sigma^2\mathbf{I}_{N_t})^{-1}\tilde{\mathbf{H}}_{t,l}^H\tilde{\mathbf{y}}_{t,l} = \mathbf{G}_{t,l}\tilde{\mathbf{y}}_{t,l}, \tag{20}$$

where $\hat{\tilde{\mathbf{x}}}_{t,l}$ is the $N_t \times 1$ LMMSE estimate of $\tilde{\mathbf{x}}_{t,l}$ and $\mathbf{G}_{t,l}$ is the $N_t \times M_t$ LMMSE filter matrix.

Next, we consider the aggregated KF operation based on a state-space model using the process equation $\tilde{\mathbf{x}}_{t,l} = \tilde{\mathbf{F}}_{t,l}\tilde{\mathbf{x}}_{0,l} + \tilde{\mathbf{w}}_{t,l}$ and observation equation (19), where $\tilde{\mathbf{F}}_{t,l} = \mathbf{0}_{N_t\times N_t}$, $\tilde{\mathbf{x}}_{0,l} = \mathbf{0}_{N_t\times 1}$, and $\tilde{\mathbf{w}}_{t,l} = \tilde{\mathbf{x}}_{t,l}$. As the output of the prediction step $\{\hat{\tilde{\mathbf{x}}}_{t,l}^{(-)}, \tilde{\mathbf{P}}_{t,l}^{(-)}\}$ is $\{\mathbf{0}_{N_t\times 1}, \mathbf{I}_{N_t}\}$ and the Kalman gain matrix is $\mathbf{G}_{t,l}$, the filtered output of the aggregated KF operation at the tth TTI becomes identical to that of the direct SLC scheme, $\hat{\tilde{\mathbf{x}}}_{t,l}$, in (20).

The aggregated KF operation for $\tilde{\mathbf{x}}_{0,l} \to \tilde{\mathbf{x}}_{t,l}$ can be divided into several small KF operations for $\mathbf{x}_{0,l} \to \mathbf{x}_{1,l} \to \cdots \to \mathbf{x}_{t,l}$, which correspond to consecutive KF operations of the proposed scheme from the first to tth TTI. As (19) is a linear equation and every element of $\tilde{\mathbf{n}}_{t,l}$ is AWGN, the elements in $\hat{\mathbf{x}}_{t,l}$ (the estimate after several small KF operations) are identical to the corresponding elements in $\hat{\tilde{\mathbf{x}}}_{t,l}$ (the estimate after the aggregated KF operation) [24], i.e., $\hat{x}_{t,l}(n) = \hat{\tilde{x}}_{t,l}(n)$ with $1 \le n \le a$ for the repeatedly transmitted symbols and $\hat{x}_{t,l}(n) = \hat{\tilde{x}}_{t,l}(n + N_t - N)$ with $a + 1 \le n \le N$ for the newly transmitted symbols. Therefore, the proposed and direct SLC schemes at the tth TTI obtain the identical estimate for $\mathbf{x}_{t,l}$.

APPENDIX B
ERROR PERFORMANCE OF THE PROPOSED AND DIRECT SLC SCHEMES

Based on the analysis in Appendix A, the error performance of the proposed KC and direct SLC schemes according to the system configuration can be compared as follows:

1) MSARQ-CC: Because there are no newly transmitted symbols for retransmission, the proposed and direct SLC schemes always obtain LLRs for decoding from the estimate of $\mathbf{x}_{t,l}$ only, where the estimate of $\mathbf{x}_{t,l}$ is identical in both schemes. Therefore, the error performance of the proposed scheme is identical to that of the direct SLC scheme.

2) MSARQ-IR: The estimate for data symbols is identical in both combining schemes, as in MSARQ-CC systems. Meanwhile, the parity symbols transmitted at the previous TTIs are not re-estimated in the proposed scheme, while the direct SLC scheme re-estimates all the previously transmitted symbols. Therefore, the error performance of the proposed scheme for retransmission can be slightly degraded from that of the direct SLC scheme. However, this performance degradation is marginal, because the performance improvement of parity symbols by retransmissions in the direct SLC scheme is limited in contrast to the significant improvement of the data symbols [14].

3) MMARQ-CC: Similar to MSARQ-CC systems, the proposed scheme obtains LLRs for decoding from the current estimate $\hat{\mathbf{x}}_{t,l}$ in which the elements are identical to the corresponding elements in $\hat{\hat{\mathbf{x}}}_{t,l}$ of the direct SLC scheme. Therefore, the error performance of the proposed scheme is identical to that of the direct SLC scheme.

4) MMARQ-IR: The estimated data symbols in the proposed and direct SLC schemes are identical, as in MMARQ-CC systems. Meanwhile, as $\mathbf{F}_{t,l} = \mathbf{0}_{N \times N}$ for the transmit signal vectors of parity symbols, their estimate in the proposed scheme is identical to BLC. In addition, as the aggregated channel matrix for the transmit signal vectors of parity symbols is block-diagonal with the t'th diagonal block $\mathbf{H}_{t,'l}$ for $1 \leq t' \leq t$, the direct SLC scheme also provides an estimate of the parity symbols identical to BLC. Therefore, the proposed and direct SLC schemes exhibit identical error performance.

REFERENCES

[1] D. Chase, "Code combining-A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. COM-33, no. 5, pp. 385–393, May 1985.

[2] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 948–959, Jun. 2000.

[3] IEEE P802.16e, *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Syst.*, Feb. 2006.

[4] 3GPP, *TR 36.212: Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding*, v8.6.0, Mar. 2009.

[5] 3GPP, *TR 38.212: NR; Multiplexing and Channel Coding*, v15.2.0, Jun. 2018.

[6] J. Lee, H. Lou, D. Toumpakaris, E. W. Jang, and J. M. Cioffi, "Transceiver design for MIMO wireless systems incorporating hybrid ARQ," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 32–40, Jan. 2009.

[7] H. Zheng, A. Lozano, and M. Haleem, "Multiple ARQ processes for MIMO systems," *EURASIP J. Appl. Signal Process.*, vol. 5, pp. 772–782, 2004.

[8] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Oct. 2001, pp. 1829–1833.

[9] E. N. Onggosanusi, A. G. Dabak, Y. Hui, and G. Jeong, "Hybrid ARQ transmission and combining for MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, USA, May 2003, pp. 3205–3209.

[10] E. W. Jang, J. Lee, H. Lou, and J. M. Cioffi, "On the combining schemes for MIMO systems with hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 836–842, Feb. 2009.

[11] H. Samra and Z. Ding, "New MIMO ARQ protocols and joint detection via sphere decoding," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 473–482, Feb. 2006.

[12] E. W. Jang, J. Lee, L. Song, and J. M. Cioffi, "An efficient symbol-level combining scheme for MIMO systems with hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2443–2451, May 2009.

[13] T. Ait-Idir and S. Saoudi, "Turbo packet combining strategies for the MIMO-ISI ARQ channel," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3782–3793, Dec. 2009.

[14] S. Park, Y. Whang, and S. Choi, "Extended detection for MIMO systems with partial incremental redundancy based hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3714–3722, Oct. 2012.

[15] S. Park, "Multiple retransmission strategy-based HARQ for MIMO systems with acknowledgment bundling," *Electron. Lett.*, vol. 50, no. 8, pp. 637–639, Apr. 2014.

[16] H. Kwon, J. Lee, and I. Kang, "Symbol-level combining for hybrid ARQ on interference-aware successive decoding," in *Proc. IEEE Glob. Commun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 3650–3654.

[17] M. Zia and Z. Ding, "Bandwidth efficient variable rate HARQ under orthogonal space-time block codes," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3360–3370, Jul. 2014.

[18] Y. Whang, S. Park, and H. Liu, "Low-complexity symbol-level combining for hybrid automatic repeat request in multiple-input multiple-output systems with linear detection," *IET Commun.*, vol. 10, no. 10, pp. 1131–1137, Jul. 2016.

[19] C. Bai, W. A. Krzymien, and I. J. Fair, "Hybrid-ARQ for layered space time MIMO systems with channel state information only at the receiver," *IET Commun.*, vol. 4, no. 14, pp. 1765–1773, Sep. 2010.

[20] N. Prasad and X. Wang, "Efficient combining techniques for multi-input multi-output multi-user systems employing hybrid automatic repeat request," *IET Commun.*, vol. 5, no. 13, pp. 1785–1796, Sep. 2011.

[21] T. Yamamoto, K. Adachi, S. Sun, and F. Adachi, "Recursive QR packet combining for uplink single-carrier multi-user MIMO HARQ using near ML detection," in *Proc. Int. Wireless Commun. Mobile Comput. Conf.*, Limassol, Cyprus, Aug. 2012, pp. 407–412.

[22] S. Park and S. Choi, "Performance of symbol-level combining and bit-level combining in MIMO multiple ARQ systems," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1517–1528, Apr. 2016.

[23] S. Park, "Aggregation-assisted combining for MIMO multiple ARQ systems," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 348–351, Feb. 2018.

[24] D. Simon, *Optimal State Estimation: Kalman, $textH_{\infty}$, and Nonlinear Approaches*, New York, NY, USA: Wiley, 2006.

[25] S. Park and S. Choi, "Iterative equalizer based on Kalman filtering and smoothing for MIMO-ISI channels," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5111–5120, Oct. 2015.

Sangjoon Park (Member, IEEE) was born in Seoul, Korea, in 1981. He received the B.S., M.S., and the Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2004, 2006, and 2011, respectively. From August 2011 to February 2014, he was with LG Electronics as a Senior Research Engineer participating in research about telematics and vehicular communication systems. From April 2014 to August 2016, he was a Research Professor with the School of Electrical and Electronic Engineering, Yonsei University. From September 2016 to February 2019, he was an Assistant Professor with the Department of Information and Communication Engineering, Wonkwang University. From March 2019, he has been an Assistant Professor with the Department of Electronic Engineering, Kyonggi University, Suwon, Korea. His main research interests include communication signal processing techniques, such as detection and equalization algorithms, hybrid ARQ and MIMO applications for wireless communications, error correcting codes, algorithms for physical layer security, and other signal processing techniques.