Soft-Decision-Driven Channel Estimation for Pipelined Turbo Receivers

Daejung Yoon, *Student Member, IEEE* and Jaekyun Moon, *Fellow, IEEE*

Abstract—We consider channel estimation specific to turbo equalization for multiple-input multiple-output (MIMO) wireless communication. We develop a soft-decision-driven sequential algorithm geared to the pipelined turbo equalizer architecture operating on orthogonal frequency division multiplexing (OFDM) symbols. One interesting feature of the pipelined turbo equalizer is that multiple soft-decisions become available at various processing stages. A tricky issue is that these multiple decisions from different pipeline stages have varying levels of reliability. This paper establishes an effective strategy for the channel estimator to track the target channel, while dealing with observation sets with different qualities. The resulting algorithm is basically a linear sequential estimation algorithm and, as such, is Kalman-based in nature. The main difference here, however, is that the proposed algorithm employs puncturing on observation samples to effectively deal with the inherent correlation among the multiple demapper/decoder module outputs that cannot easily be removed by the traditional innovations approach. The proposed algorithm continuously monitors the quality of the feedback decisions and incorporates it in the channel estimation process. The proposed channel estimation scheme shows clear performance advantages relative to existing channel estimation techniques.

Index Terms—**channel estimation, MIMO-OFDM, turbo equalization, sequential estimator**

I. INTRODUCTION

Combining the multiple-input multiple-output (MIMO) antenna method with orthogonal frequency division multiplexing (OFDM) and spatial multiplexing is a well-established wireless communication technique. Bit-interleaved coded modulation (BICM) [1] used in conjunction with MIMO-OFDM and spatial multiplexing (SM) is particularly effective in exploring both spatial diversity and frequency selectivity without significant design efforts on specialized codes [2], [3]. Turbo equalization [4], also known as iterative detection and decoding (IDD) in wireless applications [6], is well-suited for BICM-MIMO-OFDM for high data rate transmission with impressive performance potentials [5], [6].

A critical issue in realizing the full performance potential of a MIMO-OFDM system is significant performance degradation due to imperfect channel state information (CSI). The detrimental impact of imperfect CSI on MIMO detection is well known (see, for example, [7], [8]) and continues to be a great challenge in wireless communication system design. Previous works have identified desirable training patterns or pilot tones for estimating channel responses for MIMO systems [9]–[13]. However, with these methods the achievable

data rate is inevitably reduced, especially when the number of channel parameters to be estimated increases (e.g., caused by an increased number of antennas).

Decision-directed (DD) channel estimation algorithms can be applied to the turbo receivers to improve channel estimation accuracy [14]–[17]. However, inaccurate feedback decisions degrade the estimator performance [18]. Maximum-a-posteriori (MAP)-based DD algorithms discussed in [14], [15] can improve the estimation accuracy, but they require additional information like the channel probability density function. The DD channel estimation algorithm jointly working with IDD has been actively researched [19]–[23]. Among the existing research works, several papers have been devoted to iterative expectation-maximization (EM) channel estimation algorithms using extrinsic or a posteriori information fed back from the outer decoder [19]–[21]. Although the traditional EM-based estimation algorithms typically show outstanding performance, the heavy computation complexity and the iteration latency can be problematic for many practical applications. While an approximation scheme as discussed in [20] can reduce complexity, the performances of these approaches suffer from performance degradation as the number of antennas increases [19]. Also, the EM estimation algorithms need to be aided by pilot-based EM algorithms to guarantee a good performance [20], [21].

As an alternative approach to iterative EM channel estimation, Kalman-based channel estimators have been discussed that are effective against error propagation [22], [23]. The authors of [22], [23] have introduced a soft-input channel estimator that adaptively updates the channel estimates depending on feedback decision quality. The soft-input channel estimator of [22] evaluates the feedback decision quality by tracking the noise variance that includes the potential soft-decision error impact in its effort to improve the update process for the Kalman filter.

In the present work, we develop a Kalman-based channel estimator for MIMO-OFDM based on a specific pipelined turbo equalizer receiver architecture. Before setting up the Kalman estimator, a novel method for reducing decision error correlation is introduced. The proposed method constructs a refined innovation sequence by irregularly puncturing certain soft decisions that are deemed to be correlated with the previous decisions. The resulting algorithm is basically a linear sequential estimation algorithm and, as such, is Kalman-based in nature. We also weigh the estimated channel responses in the detection process according to the quality of the estimation.

A critical issue in turbo receiver design is long processing latency due to inherent iterative processing of information. Pipelined architecture reduces the latency and improves pro-

cessing throughput in turbo receivers and thus is the prevailing choice of the implementation architecture [24], [25]. One interesting feature of the pipelined turbo equalizer is that multiple sets of soft-decisions become available at various processing stages. A tricky issue is that these multiple decisions from different pipeline stages have varying levels of reliability. Therefore, an adequate optimization strategy is required for the estimator to track the target channel while dealing with observation sets with different qualities. An optimum channel estimator is derived based on this principle for the pipelined turbo receiver.

In demonstrating the viability of the proposed schemes, a SM-MIMO-OFDM system is constructed to comply with the IEEE 802.11n high speed WLAN standard [26]. Section II discusses the channel and system model, and briefly touches upon the high-throughput pipelined IDD architecture. Section III discusses the method to set up an improved innovation sequence via puncturing. Next, the proposed soft-DD Kalman-based channel estimation methods are presented in section III. Mean squared error (MSE) analysis is provided in Section IV that validates the performance merits of the proposed schemes. In Section V, the convergence behavior is investigated via the extrinsic information transfer (EXIT) charts [28], and packet error rate (PER) simulation results are presented for performance evaluation. Finally conclusions are drawn in Section VI.

II. CHANNEL AND SYSTEM MODEL

We assume a SM-MIMO-OFDM transmitter where a data bit sequence is encoded by a convolutional channel encoder, and the encoded bit stream is divided into N_t spatial streams by a serial-to-parallel demultiplexer. Each spatial stream is interleaved separately, and the interleaved streams are modulated using an M-ary quadrature amplitude modulation (M-QAM) symbol set $\mathcal{A}$ based on the Gray mapping. Since Q binary bits form an M-QAM symbol, a binary vector $\mathbf{b} = [b_0, b_1, \cdots, b_{QN_t-1}]^T$ is mapped to a transmitted symbol vector $\mathbf{s} = [s_1, s_2, \cdots, s_{N_t}]^T$ (with $s_i \in \mathcal{A}$), taken from the set $\mathcal{A}^{N_t}$, a Cartesian product of M-QAM constellations. The M-QAM symbol sequence in each spatial stream is transmitted by an OFDM transmitter utilizing a fixed number of frequency subcarriers. For a particular subcarrier for the nth OFDM symbol, the received signal at the discrete Fourier transform (DFT) output can be written as

$$\mathbf{z}_n = \mathbf{H}\mathbf{s}_n + \mathbf{n}_n \ , \tag{1}$$

where $\mathbf{z}_n = [z_1(n), z_2(n), \cdots, z_{N_r}(n)]^T$ is the received signal vector observed at the N_r receive antennas, and $\mathbf{H}$ is the channel response matrix associated with all wireless links connecting N_t transmit antennas with N_r receive antennas antennas, and $\mathbf{n}_n$ is a vector of uncorrelated, zero-mean additive white Gaussian noise (AWGN) samples of equal variance set to $\mathcal{N}_o$.

The IDD technique of [5] that performs turbo equalization for MIMO systems is assumed at the receiver. The extrinsic information on the coded-bit stream is exchanged in the form of log-likelihood ratio (LLR) between the soft-input soft-output (SISO) decoder and the SISO demapper as shown
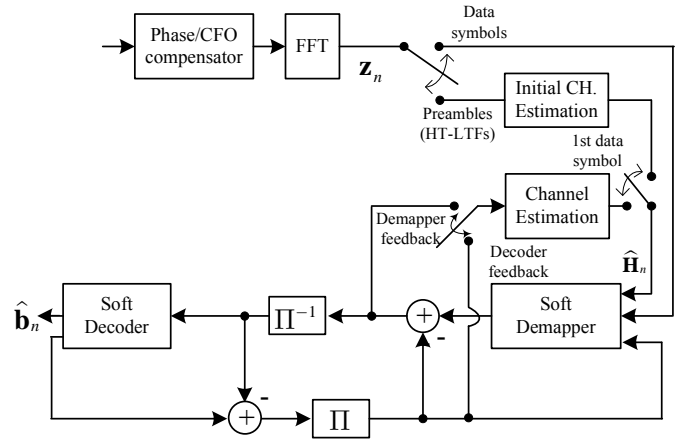

Fig. 1: Block diagram of the turbo receiver and the soft-decision-directed channel estimator
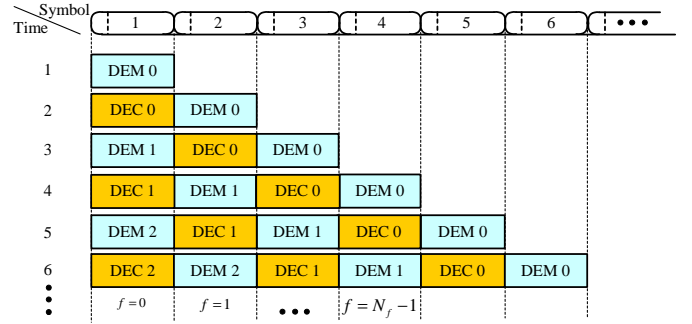

Fig. 2: OFDM-symbol processing procedure in the pipelined IDD

in Fig. 1. The demapper takes advantage of the reliable soft-symbol information made available by the outer SISO decoder. A soft-output Viterbi algorithm (SOVA) is used for the SISO decoder implementation [29]. Each data packet transmitted typically contains many OFDM symbols, and they are processed sequentially by the demapper and the decoder as they arrive at the receiver. The feedback decisions used for channel estimation must be interleaved coded-bit decisions. The extrinsic information from the demapper are rearranged accordingly and made available to the channel estimation block. The pipelined architecture is adopted to reduce the iteration latency [24], [25]. Fig. 2 illustrates OFDM symbols processed in the pipelined IDD, and Fig. 3 shows the structure of the pipelined IDD receiver and its interface with the channel estimator. Multiple demapper-decoder pairs process multiple OFDM symbols at different iteration stages. Let N_{itr} denote the number of the IDD iterations required to achieve satisfactory error rate performance. The N_{itr}-stage pipelined IDD receiver is equipped with N_{itr} demappers and N_{itr} decoders that are serially connected as in Fig. 3. The decoder forwards its extrinsic information output to the demapper in the next iteration stage. Simultaneously, the demapper and the decoder in the previous iteration stage start to process a new OFDM symbol. The pipelined IDD operation is functionally equivalent to the original IDD scheme [24]. The extrinsic
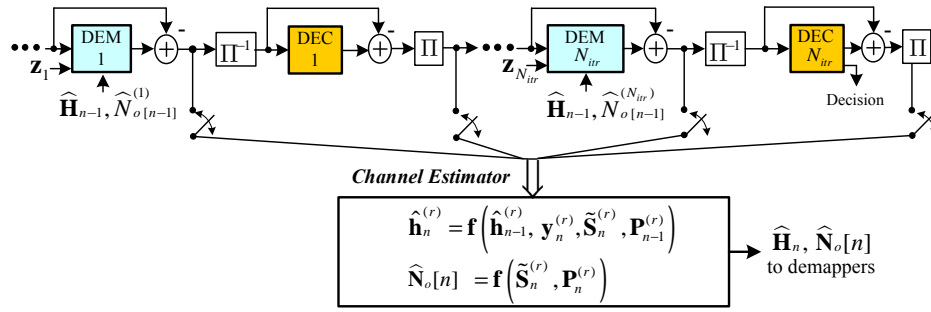
Fig. 3: Block diagram of the proposed optimum channel estimation algorithm geared to the pipelined IDD

LLRs released from the pipelined demappers and decoders are utilized for the channel estimation. Let N_{sym} denote the number of total OFDM symbols in a packet and N_f the number of the feedback symbols available for channel estimation. Note, however, that not all N_f symbols are used for the estimation. If the receiver requires N_{itr} IDD iterations, then a maximum of $2N_{itr}$ OFDM symbols are processed in the pipelined IDD receiver as illustrated in Fig. 3. Because the LLR outputs from the initial demapper and decoder have low reliability, they are not used for the channel estimation. Let index n indicate the time. In this pipelined IDD setup, when $2 \leq n \leq 2N_{itr}$, the channel estimator can get $(n-2)$ feedback decisions (i.e. $N_f = n-2$). When the number of the processed symbols increases to $2N_{itr}$ ($2N_{itr} \leq n \leq N_{sym}$), N_f is equal to $2N_{itr} - 2$. After all the OFDM symbols in the packet have arrived at the receiver front-end, it will take sometime until all symbols will clear out of the pipeline. For $n \geq N_{sym}$, N_f is equal to $N_{sym} + 2N_{itr} - n$.

III. SEQUENTIAL AND SOFT-DECISION-DIRECTED CHANNEL ESTIMATION

A. Derivation of the Kalman-Based Sequential Channel Estimation Algorithm

The sequential form of the estimator is useful in improving the quality of the channel estimate as the observed symbols arrive in a sequential fashion, as OFDM symbols do in the system of our interest. It is assumed that the channel is quasi-static over N_f OFDM symbol periods. For the pipelined IDD receiver at hand, the observation equation is set up at the rth receiver (RX) antenna as

$$\mathbf{z}_n^{(r)} = \mathbf{S}_n \mathbf{h}^{(r)} + \mathbf{n}_n^{(r)}, \tag{2}$$

where $\mathbf{z}_n^{(r)}$ is the $N_f \times 1$ received signal vector, $\mathbf{S}_n$ is a $N_f \times N_t$ matrix, $\mathbf{h}^{(r)}$ is a $N_t \times 1$ vector that is a multi-input-single-output (MISO) channel vector specific to the rth RX antenna. The goal is to do a sequential estimation of $\mathbf{h}^{(r)}$ as n progresses. The estimation process is done in parallel to obtain channel estimates for all N_r RX antennas. With an understanding that we focus on a specific RX antenna, the RX antenna index r is dropped to reduce notation cluttering.

A mean symbol decision $\tilde{s}$ is defined as the average of the constellation symbols according $\tilde{s} = \sum_{s_i \in \mathcal{A}} s_i P(s_i)$, where $P(s_i)$ is the "extrinsic probability" obtained from a direct conversion of the available extrinsic LLR.

1) Innovation Sequence Setup: The pipeline architecture can be viewed as a buffer large enough to accommodate N_f OFDM symbols, but we take into account in our channel estimator design the different levels of reliability for the soft decisions coming out of the demapper or decoder modules at different iteration stages. First, defining the soft decision error $\mathbf{E} \triangleq \mathbf{S} - \widetilde{\mathbf{S}}$, (2) can be rewritten as

$$\mathbf{z}_n = \{\widetilde{\mathbf{S}}_n + \mathbf{E}_n\}\mathbf{h} + \mathbf{n}_n. \tag{3}$$

Note that this type of soft decision representation has been used previously [22]. We emphasize, however, that unlike in [22], our derivation of a linear sequential estimator is based on the attempt to explicitly generate the innovation sequence. As will be clear in the sequel, this approach has led us to a realization that the standard steps taken to generate the innovations do not work in our set up; this in turn allowed us to devise corrective measures. Let us first see if we can find $\mathbf{x}_n$, the innovations of $\mathbf{z}_n$ (i.e., the whitened sequence that is a causal, as well as a casually invertible, linear transformation of $\mathbf{z}_n$). We write:

$$\mathbf{x}_n \triangleq \mathbf{z}_n - \widetilde{\mathbf{S}}_n \hat{\mathbf{h}}_{n-1} \tag{4}$$

$$= \widetilde{\mathbf{S}}_n(\mathbf{h} - \hat{\mathbf{h}}_{n-1}) + \mathbf{E}_n \mathbf{h} + \mathbf{n}_n. \tag{5}$$

Ideally, the vector sequence $\mathbf{x}_n$ would represent an innovation sequence in the sense that any given component of the vector $\mathbf{x}_{n-k}$ is orthogonal to any component of $\mathbf{x}_n$ as long as $k \neq 0$. In this scenario we would have

$$E\left[x_{n-k}[i]x_n^*[j]\right] = E\left[\mathbf{e}_{n-k}[i]\widehat{\mathbf{h}}_{n-k-1}\mathbf{h}^H\mathbf{e}_n^H[j]\right] + \mathcal{N}_o$$

$$= \begin{cases} \sum_{t=1}^{N_t} \hat{\rho}_{n-1}^{(t)}\sigma_s^2[t,i] + \mathcal{N}_o & \text{when } k=0 \text{ and } i=j \\ \epsilon(\approx 0) & \text{otherwise,} \end{cases} \tag{6}$$

where $\mathbf{e}_n[i]$ indicates the ith row vector $\mathbf{E}_n$, $\hat{\rho}_{n-1}^{(t)} \triangleq E[\hat{h}_{n-1}^{(t)} h^{*(t)}]$ and $\sigma_s^2 \triangleq E[|s - \tilde{s}|^2]$, the symbol decision error variance. The superscript 'H' and the symbol '$*$' denote the Hermitian transpose and the complex-conjugate, respectively. In deriving (6), we assumed: $E[x_{n-k}[i](\mathbf{h} - \widehat{\mathbf{h}}_{n-1})^H] = \mathbf{0}$, $E[s[i] e^*[j]] = 0$ for any k, i and j. In order for this to be true, though, the following must hold:

1) Links in the MISO channel are uncorrelated.
2) The channel estimate and decision error are independent.
3) The decision errors are uncorrelated.
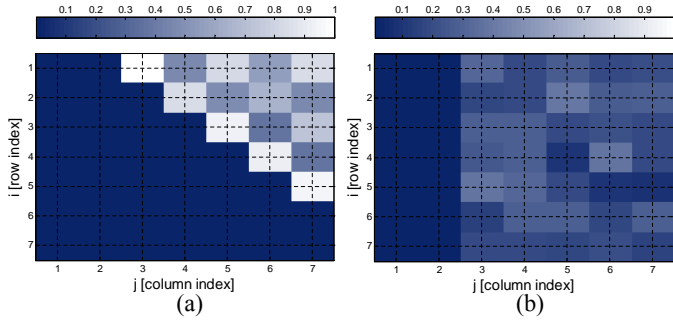   (i.e. $E[e_{n-k}[i]e_n^*[j]] = \epsilon, k \neq 0$ or $i \neq j$)

Fig. 4: Correlations in the 'innovation' sequence: (a) $E[\mathbf{x}_{n-2}\mathbf{x}_n^H]$ (b) $E[\mathbf{y}_{n-2}\mathbf{y}_n^H]$, $c = 0.8$ (normalized by $E[|y_n[0]|^2]$, averaging 50 erroneous packets)

Under these three assumptions, the vector $\mathbf{x}_n$ reasonably represents an innovation sequence.

2) Innovation Sequence with Punctured Feedback: Assumption (1) is reasonable, if the RX antennas maintain reasonable physical separation. However, assumptions (2) and (3) are not convincing. A poor channel estimate generates a poor decision, which in turn affects the ability to make reliable channel estimation. This makes both assumptions (2) and (3) invalid. As such, the Kalman filter is not optimum any more, and the correlated error circulates in the IDD and channel estimator loop. Our goal here is to provide a refined innovation sequence to reduce this error propagation. First, we observe that there is no significant correlation between the decision errors of the demapper and decoder thanks to the interleaver/deinterleaver. An issue is the demapper-demapper or decoder-decoder output correlations for a given received signal (OFDM symbol), especially when a packet is bad (i.e., certain tones cause errors despite persistent IDD efforts). In the pipelined IDD setup, it takes $n = 2$ time steps for a demapper decision to shift to the next-stage demapper, and likewise for the decoder outputs. Consequently, components in observation vectors with even time difference has correlation, as seen in Fig.4 (a) between $\mathbf{x}_n$ and $\mathbf{x}_{n-2}$. In addition, we cannot assume that the noise is random as long as identical observations are reused during the iterative channel estimation.

This correlation in $\mathbf{x}_n$ is definitely problematic for any Kalman estimator design. Imagine removing correlation in $\mathbf{x}_n$ using the Gram-Schmidt procedure:

$$x_n^{'}[f] = x_n[f] - \frac{\langle x_n[f], \ x_{n-2}[f-2] \rangle}{|x_{n-2}[f-2]|^2} x_{n-2}[f-2] \quad (7)$$

where $< a, b >$ denotes the inner product: $< a, b > = Re(a)Re(b) + Im(a)Im(b)$. Now, (6) can be rewritten (for $k = 2$ and dropping indices to simplify notation) as

$$E[x_{n-2}\,x_n^*] = E\left[x_{n-2}\,x_n^{'*}\right] + E\left[|x_{n-2}|^2\,\frac{\langle x_n, x_{n-2}\rangle}{|x_{n-2}|^2}\right]$$
$$= E[\langle x_n, \ x_{n-2}\rangle] \quad (8)$$

which suggests that using only those samples of x_n for which $\langle x_n, \ x_{n-2}\rangle \leq \epsilon$, where ϵ is an adjustable threshold, we can limit the amount of correlation in the overall observation samples utilized.

Before delving into the proposed "puncturing" process, we note that the amount of puncturing needs be decided judiciously, as removing observation samples also tends to "harden" the decisions, making the overall system approach one of hard decision feedback, a situation we need to avoid. Also, one may be tempted to use a more conceptually straightforward approach of subtracting out the correlated component as suggested by (7) or its generalized version including subtraction of less correlated components, but we had no meaningful success in reducing correlated errors with approaches along this direction.

Equation (8) suggests the following as a measure of correlation between the previous demapper and current demapper outputs (or between the previous decoder and current decoder outputs):

$$\beta_n(f) \triangleq \langle x_{n-2}[f-2], \ x_n[f] \rangle. \quad (9)$$

Now redefine N_d as the number of components among $x_n(f)$'s satisfying a threshold condition of

$$|\beta_n(f)| \ \leq \ c\mathcal{N}_o. \quad (10)$$

With this condition, let index d now denote the number of selected components among N_f feedback symbols (i.e. $d = 0, ..., N_d - 1$). The constant $c \ (\geq 0)$ is an important parameter that controls the puncturing threshold. An improved innovation sequence $\mathbf{y}_n$ can be written as

$$\mathbf{y}_n = \mathbf{G}_n\mathbf{x}_n, \quad (11)$$

where $\mathbf{G}_n$ is defined as a $N_d \times N_f$ puncturing matrix. For the dth row vector $\mathbf{g}_n^{(d)}$ of the puncturing matrix, elements are given as

$$g_n^{(d)}[f] = \begin{cases} 1, & if \ \beta_n(f) \leq c\mathcal{N}_o \ or \ d = f = 0 \ or \ d = f = 1 \\ 0, & \text{otherwise.} \end{cases}$$
$$(12)$$

Note $x_n[0]$ and $x_n[1]$ are new input elements from the first demapper and decoder outputs, which are automatically included in the refined innovation vector. As long as the observations are reused during the iterative process, the noise correlation is also problematic in the channel estimation. To resolve this issue, a scaled noise variance is adopted as a threshold criterion to judge minimum correlation, because the noise variance term in (6) is inevitable. Highly correlated signal and noise components are punctured out depending on the constant c.

It is insightful to consider a simple argument based on random puncturing. Suppose the observation samples are dropped in a random fashion. Then, the element $x_{n-2}[f-2]$ can of course be excluded from $\mathbf{y}_{n-2}$ by puncturing, and so can $x_n[f]$ from $\mathbf{y}_n$. With random puncturing, the innovation process on each element can be analyzed as (dropping index f)

$$E[y_{n-2}y_n^*] = E[g_{n-2}x_{n-2}x_n^*g_n]$$
$$= P(g_{n-2} = 1)E[x_{n-2}x_n^*]P(g_n = 1)$$
$$= \frac{N_d^{(n-2)}N_d^{(n)}E[x_{n-2}x_n^*]}{\left(N_f^{(n-2)} - N_d^{(n-2)} + 1\right)\left(N_f^{(n)} - N_d^{(n)} + 1\right)}, \quad (13)$$

where $P(g_n = 1)$ is the probability that the corresponding component x_n exists in $\mathbf{y}_n$. As N_f increases and/or N_d decreases in (13), the correlation $E[y_{n-2}y_n^*]$ decreases (likewise for the variance-normalized correlation). The same is true for the noise correlation.

Fig. 4 shows the example of correlation in the innovation sequence before and after the refinement through puncturing: $E[\mathbf{x}_{n-2}\mathbf{x}_n^H]$ vs $E[\mathbf{y}_{n-2}\mathbf{y}_n^H]$. The sequence $\mathbf{y}_n$ may have a smaller number of observation samples, but its correlation is low as seen in Fig.4 (b), which is useful to maintain the optimality of the Kalman filter. The parameter c controls trade-off : if c is large, the number of observation samples increases, which can be beneficial for ML estimation. However a large c can feed biased decision errors to the Kalman-based estimator.

Note that the actual puncturing process is not fully random as our assumption made in (13). However, the puncturing happens irregularly, and an interesting observation we make is that irregular puncturing activity become more pronounced in broken (bad) packets. Once the decisions are incorrect, correlation between the components of $\mathbf{x}_n$ appears, and puncturing becomes active. In order to salvage a bad packet from biased errors, the puncturing attempts to "innovate" the sequence $\mathbf{x}_n$. Moreover, in high SNR, random puncturing is not necessary to produce reliable decisions, because the signal term itself in (5) (without the noise and estimation error terms) is an innovation sequence. Also, the puncturing process in this context can also be viewed as an effort to prevent redundant information from circulating in the iterative signal processing. We observe that although the puncturing cannot completely remove the correlated errors, a significant portion of the biased-errors gets eliminated before the channel estimation step resumes.

B. Kalman-Based Sequential Channel Estimation Algorithm with Punctured Innovation Sequence

Once the punctured innovation sequence $\mathbf{y}_n$ is generated, a linear channel estimator can be specified as a matrix $\mathbf{A}$, that is, $\hat{\mathbf{h}} = \mathbf{A}\mathbf{y}_n$. The Kalman estimator is now derived as

$$
\begin{aligned}
\hat{\mathbf{h}}_n &= \widehat{E}[\mathbf{h}|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n] \\
&= \widehat{E}[\mathbf{h}|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{n-1}] + \widehat{E}[\mathbf{h}|\mathbf{y}_n] \\
&= \hat{\mathbf{h}}_{n-1} + \mathbf{A}_n \mathbf{y}_n,
\end{aligned} \tag{14}
$$

where $\widehat{E}[\mathbf{a}|\mathbf{b}]$ denotes the optimal linear estimator of $\mathbf{a}$ given $\mathbf{b}$. To find the linear estimator matrix $\mathbf{A}_n$, the orthogonality principle is applied:

$$
\begin{aligned}
\overline{(\mathbf{h} - \mathbf{A}_n \mathbf{y}_n)\mathbf{y}_n^H} &= \mathbf{0} \\
\mathbf{A}_n \overline{\mathbf{y}_n \mathbf{y}_n^H} &= \overline{\mathbf{h}\mathbf{y}_n^H},
\end{aligned} \tag{15}
$$

where an overbar also indicates statistical expectation. The right-hand-side of the last line in (15) is given by

$$
\overline{\mathbf{h}\mathbf{y}_n^H} = \underbrace{\overline{(\mathbf{h} - \hat{\mathbf{h}}_{n-1})(\mathbf{h} - \hat{\mathbf{h}}_{n-1})^H}}_{\triangleq \mathbf{P}_{n-1}} \widetilde{\mathbf{S}}_n^H, \tag{16}
$$

where $\mathbf{P}_{n-1}$ is defined as the channel estimation error variance matrix, and the term $\overline{\mathbf{y}_n \mathbf{y}_n^H}$ in (15) can be written as

$$
\begin{aligned}
\overline{\mathbf{y}_n \mathbf{y}_n^H} &= \widetilde{\mathbf{S}}_n \overline{(\mathbf{h} - \hat{\mathbf{h}}_{n-1})(\mathbf{h} - \hat{\mathbf{h}}_{n-1})^H} \widetilde{\mathbf{S}}_n^H \\
&\quad + \underbrace{\overline{\mathbf{E}_n \mathbf{h}\mathbf{h}^H \mathbf{E}_n^H}}_{\triangleq \mathbf{Q}_n} + \mathcal{N}_o \mathbf{I}_{N_d}.
\end{aligned} \tag{17}
$$

Now using (15), (16) and (17), the matrix $\mathbf{A}_n$ is obtained as

$$
\begin{aligned}
\mathbf{A}_n &= \overline{\mathbf{h}\mathbf{y}_n^H}(\overline{\mathbf{y}_n \mathbf{y}_n^H})^{-1} \\
&= \mathbf{P}_{n-1} \widetilde{\mathbf{S}}_n^H (\widetilde{\mathbf{S}}_n \mathbf{P}_{n-1} \widetilde{\mathbf{S}}^H + \mathbf{Q}_n + \mathcal{N}_o \mathbf{I}_{N_d})^{-1} \\
&= \left(\widetilde{\mathbf{S}}_n^H (\mathbf{Q}_n + \mathcal{N}_o \mathbf{I}_{N_d})^{-1} \widetilde{\mathbf{S}}_n + \mathbf{P}_{n-1}^{-1} \right)^{-1} \\
&\quad \cdot \widetilde{\mathbf{S}}_n^H (\mathbf{Q}_n + \mathcal{N}_o \mathbf{I}_{N_d})^{-1}.
\end{aligned} \tag{18}
$$

The next steps to complete the derivation process are to express $\mathbf{P}_{n-1}$ and $\mathbf{Q}_n$ in a recursive fashion. Noticing $(\mathbf{h} - \hat{\mathbf{h}}_n) = \mathbf{h} - (\hat{\mathbf{h}}_{n-1} + \mathbf{A}_n \mathbf{y}_n)$ from (14), the channel estimation error variance at time n can be rewritten as

$$
\begin{aligned}
\mathbf{P}_n &= \overline{\{\mathbf{h} - (\hat{\mathbf{h}}_{n-1} + \mathbf{A}_n \mathbf{y}_n)\}\{\mathbf{h} - (\hat{\mathbf{h}}_{n-1} + \mathbf{A}_n \mathbf{y}_n)\}^H} \\
&= \overline{(\mathbf{h} - \hat{\mathbf{h}}_{n-1})(\mathbf{h} - \hat{\mathbf{h}}_{n-1})^H} - \overline{(\mathbf{h} - \hat{\mathbf{h}}_{n-1})\mathbf{y}_n^H}\mathbf{A}_n^H \\
&\quad - \mathbf{A}_n \overline{\mathbf{y}_n(\mathbf{h} - \hat{\mathbf{h}}_{n-1})^H} + \mathbf{A}_n \overline{\mathbf{y}_n \mathbf{y}_n^H}\mathbf{A}_n^H \\
&= \mathbf{P}_{n-1} - \mathbf{A}_n \widetilde{\mathbf{S}}_n \mathbf{P}_{n-1}^H \\
&= (\mathbf{I}_{N_t} - \mathbf{A}_n \widetilde{\mathbf{S}}_n)\mathbf{P}_{n-1},
\end{aligned} \tag{19}
$$

where we utilized the relation $\overline{\mathbf{y}_n \mathbf{y}_n^H}\mathbf{A}_n^H = \widetilde{\mathbf{S}}_n \mathbf{P}_{n-1}^H$ which is obvious from (15) and (16). Also note $\mathbf{P}_n$ is a symmetric matrix of which pivot has non-negative real values.

Finally, $\mathbf{Q}_n$ needs to be found. The symbol decision error variance $\sigma_s^2 = E[|s - \tilde{s}|^2]$ can be found by using the extrinsic probabilities (i.e. $\sigma_s^2 = \sum_{s_i \in \mathcal{A}} |s_i - \tilde{s}|^2 P(s_i)$). Under the reasonable assumption of $\overline{(s_j - \tilde{s}_j)(s_i - \tilde{s}_i)^*} = 0$ when $i \neq j$, the $N_d \times N_d$ diagonal matrix $\mathbf{Q}_n$ is given as

$$
\overline{\mathbf{E}_n \mathbf{h}\mathbf{h}^H \mathbf{E}_n^H} = \text{diag}\left[\sum_{t=1}^{N_t} \rho^{(t)} \sigma_s^2(n, 0, t), ..., \right.
$$
$$
\left. \sum_{t=1}^{N_t} \rho^{(t)} \sigma_s^2(n, N_d - 1, t) \right]_{N_d \times N_d} \tag{20}
$$

where $\rho^{(t)} \triangleq |h^{(t)}|^2$. However, finding $\rho^{(t)}$ is a bit tricky as the channel state information is unknown to the receiver. The channel correlation matrix $\mathbf{h}\mathbf{h}^H$, on the other hand, can be found from $\overline{\mathbf{h}\mathbf{h}^H} = \overline{\{(\mathbf{h} - \hat{\mathbf{h}}_n) + \hat{\mathbf{h}}_n\}\{(\mathbf{h} - \hat{\mathbf{h}}_n) + \hat{\mathbf{h}}_n\}^H}$, which reduces to $\mathbf{P}_n + \hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^H$. Utilizing this expression, we can write

$$
\begin{aligned}
\mathbf{Q}_n &= \overline{\mathbf{E}_n \left(\mathbf{P}_n + \hat{\mathbf{h}}\hat{\mathbf{h}}_n^H \right) \mathbf{E}_n^H} \\
&= \text{diag}\left[\sum_{t=1}^{N_t} \left(p_n(t,t) + |\hat{h}_{n-1}^{(t)}|^2 \right) \sigma_s^2(n, 0, t), ..., \right. \\
&\quad \left. \sum_{t=1}^{N_t} \left(p_n(t,t) + \hat{h}_{n-1}^{(t)}|^2 \right) \sigma_s^2(n, N_d - 1, t) \right],
\end{aligned} \tag{21}
$$

where $\hat{h}_t[n-1]$ is from the previous estimate $\hat{\mathbf{h}}_{n-1}$, $p_n(t,t)$ is the tth diagonal element of $\mathbf{P}_{n-1}$, and $\sigma_s^2(n, j, t)$ is the decision error variance of the (j, t) element of $\widetilde{\mathbf{S}}_n$.

Putting it all together, for the receive antenna r, the proposed estimator is summarized as a set of equations :

$$\mathbf{Q}_n^{(r)} = \text{diag}\left[\sum_{t=1}^{N_t}\left(p_n(t,t)+|\hat{h}_{n-1}^{(t)}|^2\right)\sigma_s^2(n,0,t),\cdots,\right.$$

$$\left.\sum_{t=1}^{N_t}\left(p_n(t,t)+|\hat{h}_{n-1}^{(t)}|^2\right)\sigma_s^2(n,N_d-1,t)\right]_{N_d\times N_d} \quad (22)$$

$$\mathbf{A}_n^{(r)} = \left(\widetilde{\mathbf{S}}_n^H\left(\mathbf{Q}_n^{(r)}+\mathcal{N}_o\mathbf{I}_{N_d}\right)^{-1}\widetilde{\mathbf{S}}_n+\mathbf{P}_{n-1}^{(r)-1}\right)^{-1}$$

$$\widetilde{\mathbf{S}}_n^H\left(\mathbf{Q}_n^{(r)}+\mathcal{N}_o\mathbf{I}_{N_d}\right)^{-1} \quad (23)$$

$$\mathbf{P}_n^{(r)} = (\mathbf{I}_{N_t}-\mathbf{A}_n^{(r)}\widetilde{\mathbf{S}}_n)\mathbf{P}_{n-1}^{(r)} \quad (24)$$

$$\hat{\mathbf{h}}_n^{(r)} = \hat{\mathbf{h}}_{n-1}+\mathbf{A}_n\mathbf{y}_n, \quad (25)$$

where $\hat{\mathbf{h}}_{-1}$ corresponding to the initial time $n=0$ can be given by an initial channel estimator based on the use of known preambles. Also the initial matrix $\mathbf{P}_{-1}^{(r)}$ can be derived from the MMSE analysis [30] as $\mathbf{P}_{-1}^{(r)}=\text{diag}[|\hat{h}_{-1}^{(r,t)}|^2/(\gamma|\hat{h}_{-1}^{(r,t)}|^2+1)]$ for $t=1,..,N_t$ where $\gamma=E_s/(N_t\mathcal{N}_o)$. We note that the channel estimation algorithm summarized in (22)-(25) takes into account the quality of the soft decisions that are generated at various stages in the pipeline for a given processing time n. When t=1, the resulting algorithm becomes similar to the one presented in [22] for the inter-symbol interference channel, as the gist of the algorithm of [22] is in incorporating the quality of the soft decisions as part of effective noise in the Kalman sequential updating process. The difference, however, is that in our algorithm, we do not assume that the operation of $\mathbf{z}_n-\widetilde{\mathbf{S}}_n\hat{\mathbf{h}}_{n-1}$ makes the observation sequence automatically white, which, as argued above, would be faulty. Also, in our algorithm, varying qualities of the decisions generated from different processing modules at *a given time* are taken into account in the update process. More specifically, the effective noise covariance matrix of (22) is a function not only of n but also of N_d which itself is a growing function of n initially (up to $2N_{itr}$).

C. Noise Variance Update for the Soft Detectors

A Kalman-based estimation algorithm, as the one proposed here, has the advantage (compared with, e.g., EM-like algorithms) that the channel estimation error variance is available for free and it is continually updated as a part of the recursive process. Realizing that the channel estimation error variance is a reasonable measure of how accurate the channel estimate is, this information somehow should play a beneficial role in the detection (or demapping) process.

As the first step in utilizing the available channel estimation error variance, the observation equation of (1) is recast with the channel estimation error shown explicitly:

$$\mathbf{z}_n^{(k)} = \widehat{\mathbf{H}}\mathbf{s}_n^{(k)} + \underbrace{\left(\mathbf{H}-\widehat{\mathbf{H}}_n\right)\mathbf{s}_n^{(k)}}_{\triangleq\mathbf{a}_n^{(k)}}+\mathbf{n}_n^{(k)}, \quad (26)$$

where superscript k points to a specific demapper out of the N_{itr} demappers operating in the pipeline stages ($k=$

$1,..,N_{itr}$). Accordingly, $\mathbf{s}_n^{(k)}$ here corresponds to each odd row of $\mathbf{S}_n$ in (2). For the kth demapper in the pipeline, the noise variance is updated to include the channel estimation error:

$$\widehat{\mathbf{N}}_o^{(k)}[n] = E\left\{\left\|\mathbf{s}_n^{(k)}(\mathbf{H}-\widehat{\mathbf{H}}_n)+\mathbf{n}^{(k)}\right\|^2\right\}$$

$$= \mathbf{Cov}(\mathbf{a}_n^{(k)},\mathbf{a}_n^{(k)})+\mathcal{N}_o\mathbf{I}_{N_r}, \quad (27)$$

where $\|\ \|$ indicates vector norm operation. The $N_r\times N_r$ covariance matrix $\mathbf{Cov}(\mathbf{a}_n^{(k)},\mathbf{a}_n^{(k)})$ can be obtained (with an understanding we are focusing on the kth demapper in the pipeline at time n, drop the indices k and n to simplify notation) as

$$\mathbf{Cov}(\mathbf{a},\mathbf{a}) = E\left\{(\mathbf{H}-\widehat{\mathbf{H}})^H\mathbf{s}^H\mathbf{s}(\mathbf{H}-\widehat{\mathbf{H}})\right\}$$

$$\approx \text{diag}\left[\sum_{t=1}^{N_t}p^{(1)}(t,t)|\widetilde{s}^{(t)}|^2,..,\sum_{t=1}^{N_t}p^{(N_r)}(t,t)|\widetilde{s}^{(t)}|^2\right], \quad (28)$$

where the approximation is due to the assumption that channel estimation errors and transmitted symbols are independent and that $E[\mathbf{s}_n^H\mathbf{s}_n]\approx E[\widetilde{\mathbf{s}}_n^H\widetilde{\mathbf{s}}_n]$. Note that the updated noise variance is specified in matrix form because different RX antennas are subject to different channel estimate errors in the Kalman estimator. This is the same as saying each RX antenna is subject to a different amount of observation noise. Therefore, the demapper algorithm must properly be optimized for the given equivalent noise covariance matrix.

The demappers in the pipeline utilize (26). An M-QAM symbol vector transmitted from N_t-TX streams is demapped to one binary vector $\mathbf{b}=[b_0,b_1,\cdots,b_{QN_t-1}]^T$. Using the updated noise variance, the likelihood function of the MIMO demapper is given as

$$P(\mathbf{z}|\mathbf{s}) = \prod_{r=1}^{N_r}\frac{1}{\sqrt{2\pi\widehat{\mathcal{N}}_o^{(r)}}}\exp\left(-\frac{|z^{(r)}-\widehat{\mathbf{h}}^{(r)}\mathbf{s}|^2}{\widehat{\mathcal{N}}_o^{(r)}}\right)$$

$$= \frac{1}{(\sqrt{2\pi})^{N_r}\prod_{r=1}^{N_r}\widehat{\mathcal{N}}_o^{(r)}}\exp\left(-\sum_{r=1}^{N_r}\frac{|z^{(r)}-\widehat{\mathbf{h}}^{(r)}\mathbf{s}|^2}{\widehat{\mathcal{N}}_o^{(r)}}\right), \quad (29)$$

where $\widehat{\mathcal{N}}_o^{(r)}$ is the noise variance corresponding to $z^{(r)}$, that is, the (r,r) diagonal element of matrix $\widehat{\mathbf{N}}_o$. The kth soft MAP demapper in the pipeline directly gives out the posteriori LLR output L_P:

$$L_P(b_i) = \ln\frac{P(b_i=1|\mathbf{z})}{P(b_i=0|\mathbf{z})}$$

$$= \ln\frac{\sum_{\mathbf{s}\in\mathcal{A}^{Nt}|b_i=1}P(\mathbf{z}|\mathbf{s})\prod_{j\neq i}P(b_j)}{\sum_{\mathbf{s}\in\mathcal{A}^{Nt}|b_i=0}P(\mathbf{z}|\mathbf{s})\prod_{j\neq i}P(b_j)}+\ln\frac{P(b_i=1)}{P(b_i=0)}, \quad (30)$$

where $i=0,...,QN_t-1$ for the individual bits in the transmitted symbol vector.

The MMSE demapper solution can also be derived from the modified observation equation (26). The MMSE demapper can be shown to yield

$$\widehat{\mathbf{s}} = E[\mathbf{s}]+\mathbf{\Sigma_s}\widehat{\mathbf{H}}^H\left(\widehat{\mathbf{H}}\mathbf{\Sigma_s}\widehat{\mathbf{H}}^H+\mathbf{Cov}(\mathbf{a},\mathbf{a})+\mathcal{N}_o\mathbf{I}\right)^{-1}$$

$$\cdot\left(\mathbf{z}-\widehat{\mathbf{H}}E[\mathbf{s}]\right), \quad (31)$$

where $E[\mathbf{s}]$ is a mean-symbol vector based on the *apriori* probabilities, and $\boldsymbol{\Sigma}_\mathbf{s}$ is given as $\text{diag}[\sigma_{s_0}^2, \ldots, \sigma_{s_{N_t-1}}^2]$.

IV. MEAN SQUARED ERROR (MSE) ANALYSIS

In the MSE analysis, we try to understand 1) the impact of biased soft decision errors, and 2) when the soft decision error is unbiased, the performance impact of mismatching the soft decision error variance in the estimation channel process. Through the MSE analysis, we also investigate the impacts of the number and quality of decisions used in the estimation process.

The soft decisions fed back from the detectors and decoders are assumed to have potential errors and are written as

$$\tilde{s}[d,n] = s[d,n] + e[d,n], \tag{32}$$

where $d = 0, .., N_d - 1$. As discussed in Section III-B, the feedback soft-decisions may contain biased decision errors. So the decision error e_d is modeled as

$$e[d,n] = m[n] + q[d,n], \tag{33}$$

where m is a non-zero-mean random variable, and q is a white Gaussian noise with zero mean and variance σ_q^2. Also, denote $\sigma_s^2 = E[|e|^2]$. For both biased and unbiased cases, assume that the total decision error power σ_s^2 is identical. Also assume correlations of the bias mean with the symbol as well as with the channel are zero (i.e. $E[s\,m] = 0$ and $E[h\,m] = 0$).

The proposed estimator is designed based on the linear MMSE (LMMSE) criterion. For the MISO communication channel of (2), the LMMSE estimator is expressed as

$$\hat{\mathbf{h}}_n^{(r)} = \mathbf{A}_n^{(r)}\mathbf{z}_n^{(r)} \tag{34}$$

$$\mathbf{A}_n^{(r)} = \mathbf{R}_h^{(r)}\widetilde{\mathbf{S}}_n^H \left(\widetilde{\mathbf{S}}_n\mathbf{R}_h^{(r)}\widetilde{\mathbf{S}}_n^H + \mathbf{V}_n^{(r)}\right)^{-1}$$

$$= \underbrace{\left(\widetilde{\mathbf{S}}_n^H\widetilde{\mathbf{S}}_n + v_n^{(r)}\mathbf{R}_h^{(r)-1}\right)^{-1}}_{\triangleq\boldsymbol{\Psi}^{-1}}\widetilde{\mathbf{S}}_n^H, \tag{35}$$

where $\mathbf{V}_n^{(r)} = v_n^{(r)}\mathbf{I}_{N_d}$ with $v_n^{(r)} = (\sigma_{s,n}^2\sum_{t=1}^{N_t}\rho^{(r,t)} + \mathcal{N}_o)$, and $\mathbf{R}_h^{(r)-1} = E\left\{\mathbf{h}^{(r)}\mathbf{h}^{(r)H}\right\}$. Also, denote $\boldsymbol{\Psi} \triangleq \widetilde{\mathbf{S}}_n^H\widetilde{\mathbf{S}}_n + v_n\mathbf{R}_h^{(r)-1}$. The estimator of (35) is optimum under unbiased decision errors ($m = 0$), and the minimum estimation error variance of the MIMO LMMSE estimator is obtained as

$$\varepsilon_{unbiased}^2[n] = \sum_{r=1}^{N_r} E\left\{\|\mathbf{h}^{(r)} - \hat{\mathbf{h}}_n^{(r)}\|^2\right\} \tag{36}$$

$$= \sum_{r=1}^{N_r} \mathbf{tr}\left\{\left(\widetilde{\mathbf{S}}_n^H\mathbf{V}_n^{(r)-1}\widetilde{\mathbf{S}}_n + \mathbf{R}_h^{(r)-1}\right)^{-1}\right\},$$

where $\varepsilon_{unbiased}^2$ is the estimation error variance of the optimum MMSE estimator (35). As N_d increases, it is reasonable to write $\widetilde{\mathbf{S}}^H\widetilde{\mathbf{S}} = N_d E\{\widetilde{\mathbf{S}}^H\widetilde{\mathbf{S}}\} = N_d(E_s + \sigma_s^2)\mathbf{I}_{N_t}$. Accordingly, we have

$$\varepsilon_{opt}^2 = \varepsilon_{unbiased}^2 = \sum_{r=1}^{N_r}\sum_{t=1}^{N_t}\frac{1}{N_d(E_s + \sigma_s^2)/v^{(r)} + 1/\rho^{(r,t)}}. \tag{37}$$
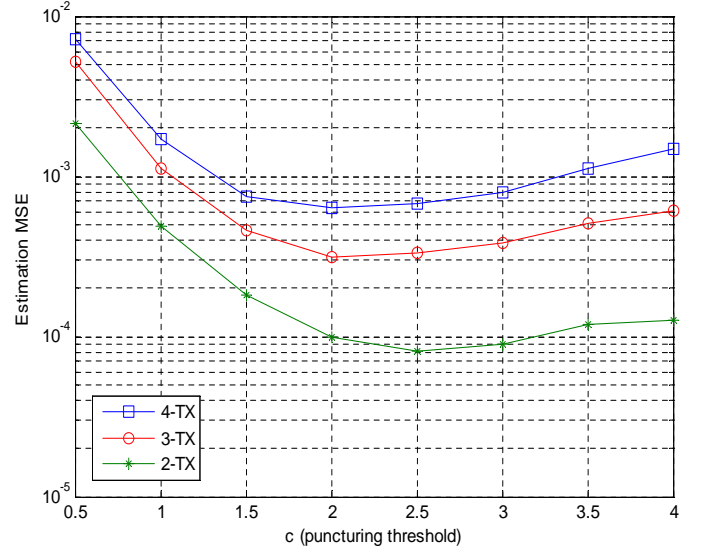

Fig. 5: Threshold parameter c optimization

Meanwhile, when $m \neq 0$ with the same decision error power σ_s^2, the MSE with the biased decision error is calculated as

$$\varepsilon_{biased}^2 = \sum_{r=1}^{N_r} E\left\{\|\mathbf{h}^{(r)} - \hat{\mathbf{h}}_{biased}^{(r)}\|^2\right\}, \tag{38}$$

where $\hat{\mathbf{h}}_{biased} = \mathbf{A}_{biased}\mathbf{z}$ and $\mathbf{A}_{biased} = (\widetilde{\mathbf{S}}_{biased}^H\widetilde{\mathbf{S}}_{biased} + v\mathbf{R}_h^{-1})^{-1}\widetilde{\mathbf{S}}_{biased}^H$ that utilizes soft-decisions with correlated error. Note that the correlation matrix of decision errors is $E[\mathbf{E}^H\mathbf{E}] = N_d\sigma_s^2\mathbf{I}_{Nt} + N_d\boldsymbol{\Phi}_{N_t}$, where $\boldsymbol{\Phi}_{N_t}$ is a matrix with all diagonal elements set to zeros and all non-diagonal elements to $|m|^2$. Assuming a very large N_d and applying a matrix inversion lemma $(\mathbf{X} + \mathbf{Y})^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1}(\mathbf{X}^{-1} + \mathbf{Y}^{-1})\mathbf{X}^{-1}$, we can write

$$\left(\widetilde{\mathbf{S}}_{biased}^H\widetilde{\mathbf{S}}_{biased} + v_n\mathbf{R}_h^{-1}\right)^{-1}$$
$$= \underbrace{(N_d\left(E_s + \sigma_s^2\right)\mathbf{I}_{Nt} + v_n\mathbf{R}_h^{-1}}_{=\boldsymbol{\Psi}} + N_d\boldsymbol{\Phi}_{Nt})^{-1} \tag{39}$$
$$= \boldsymbol{\Psi}^{-1} - \underbrace{\boldsymbol{\Psi}^{-1}\left(\frac{1}{N_d}\boldsymbol{\Phi}_{N_t}^{-1} + \boldsymbol{\Psi}^{-1}\right)^{-1}\boldsymbol{\Psi}^{-1}}_{=\boldsymbol{\Lambda}}.$$

Using (39) and the facts $\mathbf{tr}\left(\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_{Nt}\mathbf{R}_h\right) = 0$ and $\mathbf{tr}\left(\boldsymbol{\Phi}_{Nt}\mathbf{R}_h\right) = 0$, the MSE of the LMMSE estimator suffering from correlated decision errors is finally expressed as

$$\varepsilon_{biased}^2 = \sum_{r=1}^{N_r}\mathbf{tr}\left\{\mathbf{R}_h^{(r)} - \hat{\mathbf{h}}_{unbiased}^{(r)}\mathbf{h}^{(r)H}\right\}$$
$$+ \mathbf{tr}\left\{\left(N_d\left(E_s + \sigma_s^2\right)\boldsymbol{\Lambda}^{(r)}\mathbf{R}_h^{(r)}\right)\right\} \tag{40}$$
$$= \varepsilon_{unbiased}^2 + \sum_{r=1}^{N_r}\mathbf{tr}\left\{\left(N_d\left(E_s + \sigma_s^2\right)\boldsymbol{\Lambda}^{(r)}\mathbf{R}_h^{(r)}\right)\right\}.$$

Note that $\boldsymbol{\Lambda}\mathbf{R}_h$ is a semi-positive definite matrix, and therefore $\varepsilon_{biased}^2 > \varepsilon_{unbiased}^2$ when $|m| \neq 0$. This confirms the loss due to correlated decision errors, even if the error power is the same. In effectively whitening the correlated decision error, the
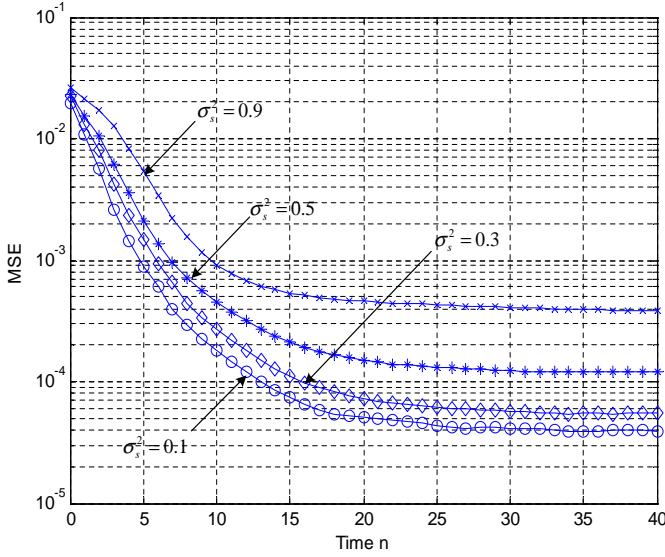
Fig. 6: Open-loop channel estimation MSE for different values of σ_s^2 ($N_d = 12$)

Fig. 7: Open-loop channel estimation MSE depending on different values of N_d ($\sigma_s^2 = 0.1$)

constant c is a crucial parameter that determines the number of selected symbols N_d and thus controls the trade-off between the observation sample size and the amount of error correlation in the channel estimator. The existence of an optimum value for c is also shown through the MSE simulation results of Fig. 5. Based on Fig. 5, we set $c = 2$ for the 3×3 and 4×4 SM-MIMO-OFDM systems, and $c = 2.5$ for the 2×2 SM-MIMO-OFDM system.

Even with unbiased decision errors, the LMMSE estimator suffers performance degradation when the noise variance is mismatched. Let us quantify the MSE penalty associated with not accounting for the uncertainty inherent in the soft decisions in the form of increased noise variance. The LMMSE estimator failing to consider the soft decision error can be described as

$$\hat{\mathbf{w}}_n^{(r)} = \mathbf{W}_n^{(r)} \mathbf{z}_n^{(r)} \tag{41}$$

$$\mathbf{W}_n^{(r)} = \mathbf{R}_h^{(r)} \widetilde{\mathbf{S}}_n^H \left(\widetilde{\mathbf{S}}_n \mathbf{R}_h^{(r)} \widetilde{\mathbf{S}}_n^H + \mathcal{N}_o \mathbf{I}_{N_d} \right)^{-1}$$

$$= \left(\widetilde{\mathbf{S}}_n^H \widetilde{\mathbf{S}}_n + \mathcal{N}_o \mathbf{R}_h^{(r)-1} \right)^{-1} \widetilde{\mathbf{S}}_n^H. \tag{42}$$

Utilizing (37) and denoting $\boldsymbol{\Delta}_n^{(r)} \triangleq \hat{\mathbf{h}}_n^{(r)} - \hat{\mathbf{w}}_n^{(r)}$, the estimation error variance ε_w^2 of the estimator (42) can be shown to be

$$\varepsilon_w^2[n] = \sum_{r=1}^{N_r} E \left\{ \| \mathbf{h}_n^{(r)} - \hat{\mathbf{w}}_n^{(r)} \|^2 \right\}$$

$$= \sum_{r=1}^{N_r} \mathbf{tr} E \left\{ \| \mathbf{h}_n^{(r)} - \left(\hat{\mathbf{h}}_n^{(r)} - \boldsymbol{\Delta}_n^{(r)} \right) \|^2 \right\}$$

$$= \varepsilon_{opt}^2[n] + \sum_{r=1}^{N_r} \mathbf{tr} \left\{ E \left\{ \mathbf{h}_n^{(r)} \hat{\mathbf{h}}_n^{(r)H} - \hat{\mathbf{w}}_n^{(r)} \hat{\mathbf{w}}_n^{(r)H} \right. \right.$$

$$\left. \left. + \hat{\mathbf{w}}_n^{(r)} \mathbf{h}_n^{(r)H} - \mathbf{h}_n^{(r)} \hat{\mathbf{w}}_n^{(r)H} \right\} \right\}. \tag{43}$$

To simplify notation, the indices r and n are temporally dropped. As the number of iteration increases, the matrix inversions in (35) and (42) can be simplified as
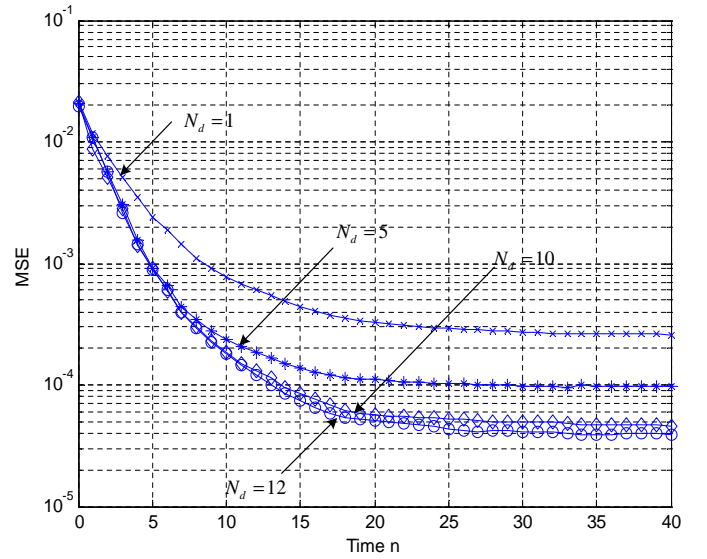
$$\left(\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}} + v \mathbf{R}_h^{-1} \right)^{-1} = \text{diag}[\ \rho_1 / \left(\rho_1 N_d (E_s + \sigma_s^2) + v \right),...,$$

$$\rho_{N_t} / \left(\rho_{N_t} N_d (E_s + \sigma_s^2) + v \right) \ ] \quad \text{and} \quad \left(\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}} + \mathcal{N}_o \mathbf{R}_h^{-1} \right)^{-1} =$$

$$\text{diag}[\rho_1 / \left(\rho_1 N_d (E_s + \sigma_s^2) + \mathcal{N}_o \right),...,\rho_{N_t} / \left(\rho_{N_t} N_d (E_s + \sigma_s^2) + \mathcal{N}_o \right)], \text{ where the subscript for } \rho \text{ for the time being indicates the TX antenna. Also, noting } E\{\hat{\mathbf{h}}(\mathbf{h} - \hat{\mathbf{h}})^H\} = 0 \text{ by the orthogonality principle, it can be shown that}$$

$$E\{\mathbf{h}\hat{\mathbf{h}}^H\} = \mathbf{R}_h \mathbf{S}^H \widetilde{\mathbf{S}} \left(\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}} + v \mathbf{R}_h^{-1} \right)^{-H}$$

$$= \text{diag} \left[\frac{\rho_1^2 N_d E_s}{\rho_1 N_d (E_s + \sigma_s^2) + v}, ..., \frac{\rho_{N_t}^2 N_d E_s}{\rho_{N_t} N_d (E_s + \sigma_s^2) + v} \right]_{N_t \times N_t}. \tag{44}$$

We also write

$$E\{\hat{\mathbf{w}}\hat{\mathbf{w}}^H\}$$

$$= \left(\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}} + \mathcal{N}_o \mathbf{R}_h^{-1} \right)^{-1} \widetilde{\mathbf{S}}^H \left(\mathbf{S} \mathbf{R}_h \mathbf{S}^H + \mathcal{N}_o \mathbf{I}_{N_d} \right)$$

$$\cdot \widetilde{\mathbf{S}} \left(\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}} + \mathcal{N}_o \mathbf{R}_h^{-1} \right)^{-H}$$

$$= \text{diag} \left[\frac{\rho_1^2 N_d \left(\rho_1 N_d E_s^2 + \rho_\Sigma \sigma_s^2 E_s + \mathcal{N}_o (E_s + \sigma_s^2) \right)}{(\rho_1 N_d (E_s + \sigma_s^2) + \mathcal{N}_o)^2}, ...,$$

$$\frac{\rho_{N_t}^2 N_d \left(\rho_{N_t} N_d E_s^2 + \rho_\Sigma \sigma_s^2 E_s + \mathcal{N}_o (E_s + \sigma_s^2) \right)}{(\rho_{N_t} N_d (E_s + \sigma_s^2) + \mathcal{N}_o)^2} \right]_{N_t \times N_t}, \tag{45}$$

where $\rho_\Sigma = \sum_{t=1}^{N_t} \rho_t$. Finally, substituting (44) and (45) in (43) and also noting $E\{\hat{\mathbf{w}}\mathbf{h}^H\} = E\{\mathbf{h}\hat{\mathbf{w}}^H\}$, the MSE convergence behavior of the estimator (43) can be shown to be

$$\lim_{N_d \to \infty} \varepsilon_w^2[n]$$

$$= \varepsilon_{opt}^2[n] + \sum_{r=1}^{N_r} \frac{\rho_\Sigma^{(r)} E_s}{E_s + \sigma_s^2[n]} \left(1 - \frac{E_s}{E_s + \sigma_s^2[n]} \right), \tag{46}$$

from which it is easy to see that the mismatched MSE is an increasing function of the soft decision error variance σ_s^2.

To develop insights into the performance sensitivity off the sequentially updated channel estimator against the variations of the parameters σ_s^2 and N_d, we resort to an open-loop
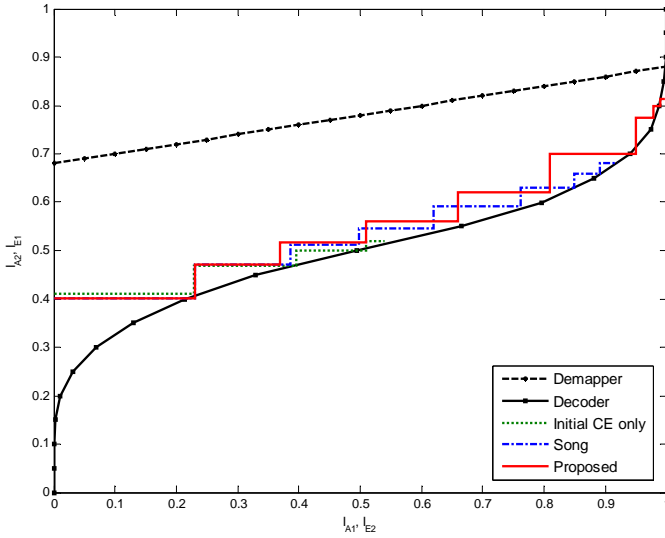
Fig. 8: EXIT charts for the 2×2 SM-MIMO-OFDM turbo receiver at SNR=14 dB

Fig. 9: PERs with different channel estimators: 2x2 SM-MIMO-OFDM (7 iterations)

investigation. For this, the decision-feedback channel estimator is modified in such a way that unbiased soft decisions with various σ_s^2 and N_d combinations are artificially generated for the channel estimator. A 7-iteration IDD receiver for the 2×2 16-QAM MIMO-OFDM system is used for this test, but instead of using actual feedback from the demappers and decoders, artificially generated soft-decisions are provided to the channel estimator of (22)-(25).

Fig. 6 and Fig. 7 show the MSE performance depending on the decision quality σ_s^2 and the number of feedback decisions N_d with an assumption of uncorrelated feedback decisions. The signal power is fixed at $E_s = 1$ and the channel SNR at 14 dB. With $N_d = 12$, the packet error rate (PER) due to imperfect CSI became negligible when $\sigma_s^2 \approx 0.1$. In Fig. 7, it is seen that reducing the number of feedback decisions, N_d, while fixing the decision quality causes the MSE to increase.

V. PERFORMANCE EVALUATION

The proposed algorithm is investigated through an extrinsic information transfer (EXIT) chart analysis and packet error rate (PER) simulation. Performances are evaluated for 2×2, 3×3 and 4×4 16-QAM SM-MIMO-OFDM systems. The transmitter sends a packet with 1000 bytes of information. The SISO MAP-demapper is used for the 2×2 SM-MIMO-OFDM system, whereas the SISO MMSE-demapper is used for the 3×3 and 4×4 SM-MIMO-OFDM system [6] due to complexity. A rate-$1/2$ convolutional code is used with generator polynomials $g_o = 133_8$ and $g_1 = 171_8$, complying with the IEEE 802.11n specifications [26]. The SOVA is used for decoding. The MIMO multi-path channel is modeled with an exponentially-decaying power profile with $T_{rms} = 50ns$ uncorrelated across the TX-RX links established.

A. EXIT and PER Performance Comparisons

The EXIT chart is a well-established tool that allows the understanding of the average convergence behavior of the
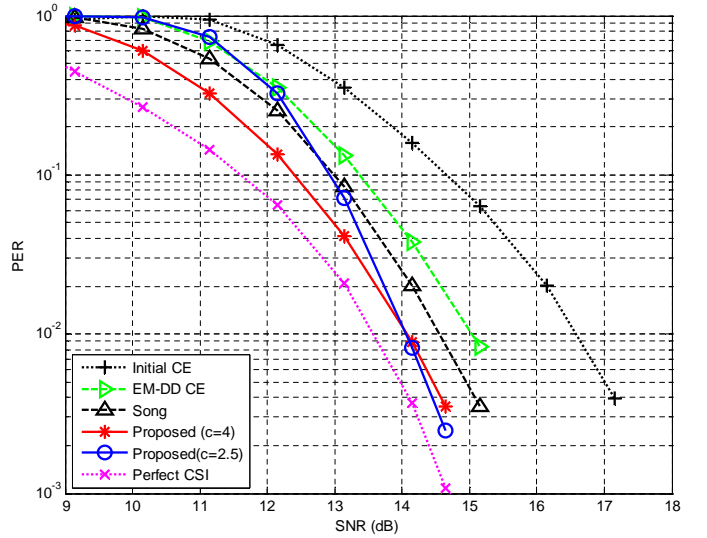
mutual information (MI) in iterative soft-information processing systems [28]. Fig. 8 shows the results of an EXIT chart analysis on various competing schemes. A 2×2 SM-MIMO-OFDM system is used for this, and an SNR of 14 dB is chosen. I_{A_1} and I_{E_1} measure the MI at the input and output of the demapper, respectively, whereas I_{A_2} and I_{E_2} are the respective MI at the input and output of the decoder. At the next iteration stage, I_{E_1} becomes I_{A_2} and I_{E_2} turns to I_{A_1}.

In the figure, the top-most curve indicates the average transfer function of MI through the demapper and the bottom-most curve is the same function for the decoder. Both the demapper and decoder EXIT chart curves correspond to Gaussian-distributed input LLRs, and the demapper EXIT curve is also based on the assumption of perfect channel estimation. The stair-case MI plots represent actual MI measured during IDD simulation runs and shows how the MI improves through the iterative process for three different channel estimation schemes. The gap between each stair-case MI trajectory and the demapper EXIT curve represents the performance loss due to imperfect-CSI. The solid stair-case line represents the proposed channel estimation algorithm. The dashed-line (labeled "Song") corresponds to the Kalman channel estimator of [23] applied to the conventional-IDD setting (non-pipelined IDD with a demapper utilizing the noise-variance update of (27) with channel estimation using only the decoder output decision). The dotted line is for the demapper utilizing only the preamble-based initial channel estimation (following the IEEE 802.11n format, where a fixed number of initial preamble symbols in the high-throughput long training field is utilized). For the proposed scheme, the MI trajectory measurement is taken from the last demapper in the pipeline, as the last demapper block best reflects the quality of the final decisions. It is clear that the proposed punctured-feedback Kalman estimation with pipelined-IDD shows superior MI convergence characteristics. The scheme of [23] fails to improve MI beyond nine iterations. With the demapper utilizing only initial channel estimation, the trajectory fails to advance earlier in the iteration.
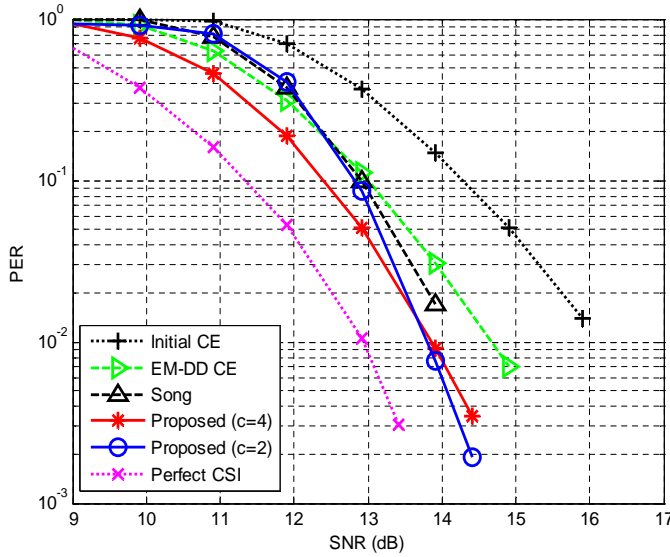
Fig. 10: PERs with different channel estimators: 3x3 SM-MIMO-OFDM (7 iterations)

Fig. 11: PERs with different channel estimators: 4x4 SM-MIMO-OFDM (9 iterations)

Fig. 9 shows PER performances of the receivers with different channel estimators in the 2×2 SM-MIMO-OFDM system. Seven iterations are applied beyond which the iteration gain is plateaued. The performance gap between perfect CSI and preamble-based initial CE only is nearly 3 dB at low PERs. It can be seen that at low PER the proposed estimator almost compensates for the loss due to imperfect-CSI when the threshold parameter is set at $c = 2.5$. Although the performance with small c has inferior performance at low SNRs, the proposed Kalman CE curve with $c = 2.5$ crosses the $c = 4$ curve as SNR gets higher. The large c is effective in averaging noise in low SNR, but allows relatively large correlated errors. As expected from the EXIT chart analysis results, the Kalman estimator of [23] that utilizes only the decoder output in a non-pipelined setting does not perform as well. As one of the algorithms considered for comparison, the decision-directed EM estimator (referred to as EM-DD here) introduced as a variant of the EM estimator in [19] is applied with $\hat{\mathbf{h}}_{o,n}^{(r)} = \left(\widetilde{\mathbf{S}}_n^H \widetilde{\mathbf{S}}_n \right)^{-1} \widetilde{\mathbf{S}}_n^H \mathbf{z}_n^{(r)}$. In addition, the EM estimate is blended with the preamble-based channel estimate by a combining method (i.e., $\hat{h}_n^{(t,r)} = a_n \hat{h}_{preamble}^{(t,r)} + b_n \hat{h}_o^{(t,r)}[n]$) [20]. A method to find the combining coefficients a_n and b_n is discussed in [20]. The EM noise variance update method is presented in [19] as $\widehat{\mathcal{N}}_o[n] = 1/N_r N_d \sum_{r=1}^{N_r} \sum_{d=0}^{N_d-1} \left(z_n^{(r)} - \widetilde{\mathbf{S}}_n \hat{\mathbf{h}}_n^{(r)} \right)^* \left(z_n^{(r)} - \widetilde{\mathbf{S}}_n \hat{\mathbf{h}}_n^{(r)} \right)$. As can be seen, this scheme also does not perform as well as the proposed algorithm.

Fig. 10 and Fig. 11 show PER curves for 3×3 SM-OFDM-OFDM and 4×4 SM-OFDM-OFDM systems, respectively. These figures tell a consistent story. Namely, the initial-CE-only scheme suffers about a 3dB SNR loss relative to the perfect CSI case. The proposed schemes close this gap significantly, outperforming both the Kalman-based algorithm of [23] and the EM-based algorithm of [19]. As for the proposed channel estimation scheme, a more aggressive punc-
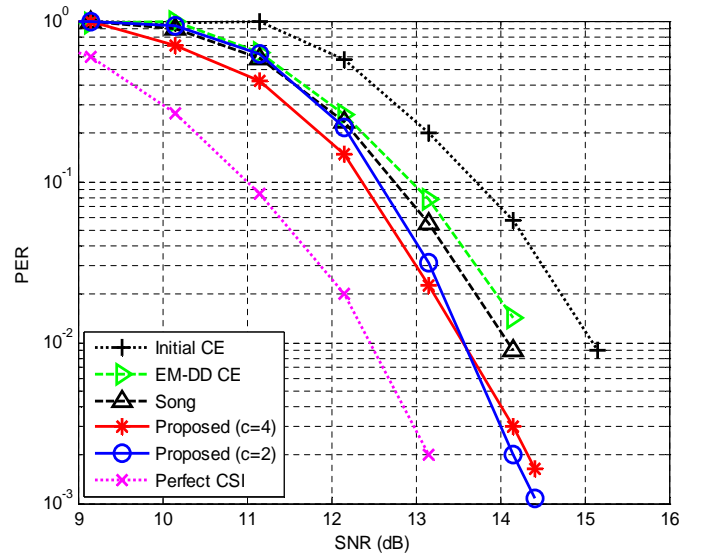
turing (corresponding to a lower c value) tends to give a lower PER as SNR increases. Before finishing this section, we briefly mention complexity. For all considered channel estimation schemes - the proposed, the Song method and the EM-DD scheme - implementation complexity largely arises from the matrix inversion operation. All schemes require matrix inversions of the same dimension. Consequently, the proposed method and the Song method require complexity that roughly grows as $2N_r \times O(N_t^3)$ whereas the EM-DD requires complexity proportional to just $O(N_t^3)$. This is due to the consequence that both our method and the Song method require matrix inversion for each receive antenna, whereas the EM-DD method needs matrix inversion just once and can be used for all receive antennas. The factor 2 accounts for the fact that two matrix inversions are required for each update of the Kalman gain in the proposed and Song methods.

VI. CONCLUSIONS

A sequential soft-decision-directed channel estimation algorithm for MIMO-OFDM systems has been proposed for the specific pipelined turbo-receiver architecture. The algorithm deals with observation sample sets with varying levels of reliability. In coping with decision errors that propagate in the pipeline, we have introduced a novel method of innovating a correlated observation sequence via puncturing. Based on the refined innovation sequence, a Kalman-based estimator has been constructed. The proposed algorithm establishes improved Kalman-based channel estimation where the traditional innovations approach cannot create a true innovation sequence due to soft-decision error propagation. The EXIT chart, MSE analysis and PER simulation results have been used to validate the performance advantage of the proposed channel estimator.

REFERENCES

[1] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927-946, May, 1998.

[2] A. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," *Proc. of IEEE Vehicular Technology Conference*, pp. 473-478, Boston, Sept., 2000.

[3] D. Park and B. Lee, "Design criteria and performance of space-frequency bit-interleaved coded modulations in frequencyselective Rayleigh fading channels," *Journal of Commun. and Networks*, vol. 5, no. 2, pp. 141-149, June, 2003.

[4] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," in *European Trans. Telecomm.*, vol. 6, pp. 507–511, September 1995.

[5] R. Koetter, A. Singer, and M. Tuchler, "Turbo equalization : an iterative equalization and decoding technique for coded data transmission," *IEEE Signal Processing Mag.*, vol. 21, pp. 67-80, Jan. 2004.

[6] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean square error equalization using a prori information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673-683, Mar, 2002.

[7] Y. Huang and J. Ritcey, "EXIT chart analysis of BICM-ID with imperfect channel state information," *IEEE Commun. Letters*, vol. 7, no. 9, pp. 434-436, Sept., 2003.

[8] Y. Huang and J. Ritcey, "16-QAM BICM-ID in fading channels with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 1000-1007, Sept., 2003.

[9] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select Areas Commun.*, vol. 17, no. 3, pp. 461-471, Mar., 1999.

[10] Y. Li, L. Cimini, and N. Sollegberger, "Robust channels estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902-915, Jul., 1998.

[11] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67-75, Jan., 2002.

[12] X. Ma, L. Yang, and G. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453-466, Mar., 2005.

[13] B. Hassibi and B. Hochwald, "Optimal training in space time systems," *Proc. 34th Asilomar Conf. on Signals, Systems and Computers*, pp. 743-747, Oct., 2000.

[14] X. Deng, A. Haimovich, and J. Garcia-Frias, "Decision directed iterative channel estimation for MIMO systems," *Proc. IEEE Int. Conf. Commun.*, vol. 4, pp. 2326-2329, Anchorage, AK, May, 2003.

[15] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1412-1417, Jul., 2005.

[16] M. Loncar, R. Muller, J. Wehinger, and T. Abe, "Iterative joint detection, decoding, and channel estimation for dual antenna arrays in frequency selective fading," *Proc. Int. Symposium on Wireless Personal Multimedia Commun.*, Honolulu, HI, Oct., 2002.

[17] M. Sandell, C. Luschi, P. Strauch, and R. Yan, "Iterative channel estimation using soft decision feedback," *Proc. Globecom*, vol. 6, pp. 3728-3733, Sydney, Australia, Nov., 1998.

[18] M. Tuchler, R. Otnes, and A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalizer," *Proc. IEEE ICC2002 Int. Conf.* vol. 3, pp. 1858-1862, New York, NY, Apr. 2002.

[19] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparision of EM-based algorithms for MIMO channel estimation," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 216-226, Jan., 2007.

[20] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Select Areas Commun.*, vol. 19, no. 8, pp. 1450-1460, Aug., 2001.

[21] M. Khalighi and J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Wireless Commun.*, vol. 5, no. 11, pp. 3165-3173, Nov., 2006.

[22] S. Song, A. Singer, and K. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Sig. Processing*, vol. 52, no. 10, pp. 2885-2894, Oct., 2004.

[23] S. Song, A. Singer, and K. Sung, "Turbo equalization with unknown channel,"*in Proc. IEEE Int. Conf. Acoustics, Speech, and Sig. Processing ICASSP*, vol. 3, pp. 2805-2808, Orlando, FL, May 2002.

[24] S. Abbasfar, "Turbo-like codes; design for high speed decoding," Springer, Netherlands, 2007.

[25] S. Lee, N. Shanbhag, and A. Singer, "Area-efficient high-throughput VLSI architecture for MAP-based turbo equalizer," *Proc. IEEE signal procssing system: design and implementation* pp. 87-92, Aug. 2003, Seoul, Korea.

[26] *IEEE P802.11n/D1.0 : Draft Amendment to STANDARD FOR 2 Information Technology-Telecommunications and 3 information exchange between systems-Local and 4 Metropolitan networks-Specific requirements-Part 5 11: Wireless LAN Medium Access Control (MAC) 6 and Physical Layer (PHY) specifications: 7 Enhancements for Higher Throughput.*

[27] H. Stark and J. Woods, "Probability and random processes with applications to signal processing," USR, NJ, Prentice-Hall, 2002.

[28] S. Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *Int. J. Electron. Commun.*, vol. 54, No. 6, pp. 389-398, Nov. 2000.

[29] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," *Globecom 1989*, vol. 3, pp. 1680-1686, Dallas,TX, Nov., 1989

[30] S. Kay, "Fundamentals of statistical signal processing-estimation theory," Englewood Cliffs, NJ, Prentice-Hall, 1993.

Daejung Yoon received the B.S. and M.S. degrees in electrical engineering from the Kyungpook National University, Daegu, Korea, in 2003 and 2005, respectively. Currently, he is working toward the Ph.D. degree in electrical engineering at the University of Minnesota at Minneapolis. Since 2009, he has worked as a senior engineer in Samsung Information Systems America, San Jose, CA. His research interests are in the general areas of communication systems, advanced signal processing for digital communications and hard-disk signal processing.

Jaekyun Moon received a BSEE degree with high honor from the State University of New York at Stony Brook in 1984 and M.S. and Ph.D. degrees in the Electrical and Computer Engineering Department at Carnegie Mellon University in 1987 and 1990, respectively. In 1990, he joined the faculty of the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities, as an Assistant Professor. He was promoted to a Tenured Associate Professor in 1995 and then to a Full Professor in 1999. Prof. Moon's research interests are in the area of channel characterization, signal processing and coding for data storage and digital communication. His recent interests are in coding and equalization for interference-dominant channels. Prof. Moon was selected to receive the IEEE-Engineering Foundation Research Initiation Award in 1991 and received the NSF Reseacrh Initialtion Award in the same year. He received the 1994-1996 McKnight Land-Grant Professorship from the University of Minnesota. He also received the IBM Faculty Development Awards as well as the IBM Partnership Awards. He was awarded the National Storage Industry Consortium (NSIC) Technical Achievement Award for the invention of the maximum transition run (MTR) code, a widely-used error-control/modulation code in commercial storage systems. Prof. Moon served as Program Chair for the 1997 IEEE Magnetic Recording Conference. In 1998, he was a Visiting Professor at Seoul National University. He is also a past elected Chair of the Signal Processing for Storage Technical Committee of the IEEE Communications Society. In 2001, he co-founded Bermai, Inc., a fabless semiconductor start-up, and served as founding President and CTO. From 2004 to 2007, Prof. Moon worked as a consulting Chief Scientist for DSP Group, Inc. He served as a guest Editor for the 2001 IEEE J-SAC issue on Signal Processing for High Density Storage. He also served as an Editor for IEEE Transactions on Magnetics in the area of signal processing and coding for 2001-2006. While on leave from the University of Minnesota, he worked as Chief Technology Officer at Link-A-Media Devices Corp. in 2008. Prof. Moon joined KAIST as a Full Professor in 2009. He is an IEEE Fellow.