Reduced-Complexity Near-Capacity Joint Channel Estimation and Three-Stage Turbo Detection for Coherent Space-Time Shift Keying

Peichang Zhang, Sheng Chen, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—We propose a low-complexity joint channel estimation (CE) and three-stage iterative demapping-decoding scheme for near-capacity coherent space-time shift keying (CSTSK) based multiple-input multiple-output (MIMO) systems. In the proposed scheme, only a minimum number of space-time shift keying training blocks are employed for generating an initial least square channel estimate, which is then used for initial data detection. As usual, the detected soft information is first exchanged a number of times within the inner turbo loop between the unity-rate-code (URC) decoder and the CSTSK soft-demapper, and the information gleaned from the inner URC decoder is then iteratively exchanged with the outer decoder in the outer turbo loop. Our CE scheme is embedded into the outer turbo loop, which exploits the *a posteriori* information produced by the CSTSK soft-demapper to select a sufficient number of high-quality decisions only for CE. Since the CE is embedded into the iterative three-stage demapping-decoding process, no additional iterative loop is required for exchanging information between the decision-directed channel estimator and the three-stage turbo detector. Hence, the computational complexity of the proposed joint CE and three-stage turbo detection remains similar to that of the three-stage turbo detection-decoding scheme with the given channel estimate. Moreover, our proposed low-complexity semi-blind scheme is capable of approaching the optimal maximum likelihood turbo detection performance attained with the aid of perfect channel state information, with the same low number of turbo iterations as the latter, as confirmed by our extensive simulation results.**

***Index Terms*—Coherent space-time shift keying, three-stage concatenated turbo detector and decoder, joint channel estimation and turbo detection, low-complexity near-capacity performance.**

I. Introduction

RECENTLY, multiple-input multiple-output (MIMO) wireless communication systems [1] have attracted substantial attention due to their potential of providing spatial diversity and/or multiplexing gains. Inspired by the spatial modulation (SM) [2] and space shift keying (SSK) [3], the concept of space-time shift keying (STSK) was introduced in [4], [5], which includes SM and SSK as its special cases. In STSK systems, since no inter-channel-interference (ICI) is imposed by the equivalent STSK system model, the employment of low-complexity single-stream maximum likelihood (ML) detection becomes realistic. Additionally, STSK is capable of striking an arbitrary trade-off between the required spatial- and time-diversity gain as well as multiplexing gain. This is achieved by optimizing both the number and the size of the dispersion matrices as well as the number of transmit and receive antennas [5]. Non-coherent STSK (NCSTSK) systems enjoy significant advantages, as they do not require channel state information (CSI). Hence they have found wide-ranging applications [6]. However, NCSTSK systems suffer from the usual 3 dB signal to noise ratio (SNR) penalty, and their design freedom is also restricted [5]. By contrast, coherent STSK (CSTSK) systems offer a better performance as well as enjoying a high grade of design flexibility in comparison to their non-coherent counterparts, but impose a high CSI estimation complexity.

Naturally, the performance of a coherent MIMO system heavily relies on the accuracy of CSI. Therefore, reliable yet low-complexity channel estimators are desired. Training-based channel estimators [7] are capable of obtaining accurate CSI at the expense of reducing the system's throughput, since employing a large number of training symbols is necessary to generate an accurate channel estimate. Therefore, blind channel estimation (CE) techniques have attracted significant attention [8], [9]. However, it is well-known that blind methods not only impose a high complexity, while exhibiting a slow convergence, but also suffer from inherent estimation and decision ambiguities [10]. A possible solution to these problems is the employment of semi-blind CE algorithms [11]–[26], where only a small number of training symbols is employed for generating an initial least squares channel estimate (LSCE). Then the channel estimator and data detector iteratively exchange their information for updating the channel estimate.

The motivation of our current work accrues from a closer examination of these existing schemes in the context of joint CE and data detection. The decision-directed CE (DDCE) and the low-complexity single-stream ML data detector form an iterative loop for the uncoded CSTSK system of [16] and for the coded CSTSK system of [17]. A similar joint CE and data detection strategy was employed for the coded direct sequence (DS) code-division multiple-access (CDMA) system of [18]. Iterative joint channel estimation and multiuser

Manuscript received April 23, 2012; revised September 26, 2012 and January 11, 2013. The associate editor coordinating the review of this paper and approving it for publication was L. K. Rasmussen.

The authors are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK (e-mail: {pz3g09, sqc, lh}@ecs.soton.ac.uk).

This work was supported by the India-UK Advanced Technology Centre, and in part by the European Union under the Concerto project, as well as by the European Research Council's Advanced Fellow Grant.

Digital Object Identifier 10.1109/TCOMM.2013.031213.120283

detection schemes were proposed in [19]–[21] for CDMA systems. In [22]–[26], an iterative loop involving a channel estimator and a turbo detector-decoder was introduced for forming semi-blind joint CE and turbo detection-decoding for MIMO aided orthogonal frequency division multiplexing (OFDM) systems. However, in all these existing works, joint CE and turbo (or non-turbo) detection-decoding was carried out by introducing an extra iterative loop between the channel estimator and the original turbo (or non-turbo) detection-decoding system, which requires an additional number of iterations to achieve convergence and therefore imposes a considerable extra computational complexity. Additionally, all these contributions employ all the detected bits for updating the channel estimate, which has the following effects. Firstly, the sequence of transmitted bits is usually very long, particularly for turbo coded systems, and this imposes an unnecessarily high complexity on the CE. Secondly and more profoundly, some of the detected bits are erroneous, which will inflict a performance degradation, especially in the low SNR region.

In order to reduce the effects of the erroneous decisions imposed on DDCE, various soft-decision aided CE schemes have been proposed [20], [27]–[31]. These schemes exploit the soft estimates of the detected symbols and have been shown to be capable of improving the accuracy of CE. Additionally, due to the nature of soft CE schemes, they are commonly combined with turbo detection-decoding schemes for forming joint soft-decision aided CE and turbo detection-decoding structures. For example, an iterative CE scheme using soft-decision feedback was proposed in [28], which weights the probabilities of the decisions arriving from the equaliser. In [29], a soft-input Kalman channel estimator and a weighted turbo recursive least square channel estimator were proposed. In [20], an iterative soft-decision aided CE and symbol detection technique was proposed for coded CDMA systems. However, most of these soft-decision aided CE schemes only consider binary phase-shift keying (BPSK) as well as quadrature phase-shift keying (QPSK) and are quite complex compared to their hard-decision aided CE counterparts [20], [28]. A general soft-decision aided scheme designed for high-order quadrature amplitude modulation (QAM), was presented in [31]. However, all these joint soft-decision aided CE and turbo detection-decoding structures rely on the entire sequence of the detected decisions for CE, and they are incapable of approaching the perfect-CSI based performance bound, as shown in the simulation results of [20], [28], [29]. The reason for this performance loss is plausible. At low SNRs, the initial bit error ratio (BER) is practically 50%. Hence the soft-decision aided scheme fails to eliminate the deleterious effects of erroneous decisions. Additionally, these schemes suffer from a potentially excessive CE complexity, owing to the fact that the length of a turbo coded frame is usually high.

A low-complexity yet reliable decision selection algorithm was proposed in [32], where it was shown that the system's performance can be improved by selecting only the most reliable decisions based on their estimated probabilities. This scheme of exclusively selecting the reliable decisions has an additional benefit of significantly reducing the computational complexity of the channel estimator. However, only a BPSK scenario was considered in [32], since for BPSK the bit probability is identical to the symbol probability. Therefore, the decision selection algorithm of [32] has not been extended to high-order modulation schemes.

Against this background, our novel contribution is that we propose a low-complexity joint block-of-bits selection based channel estimation (BBSBCE) and three-stage iterative demapping-decoding scheme for near-capacity CSTSK systems, which does not impose an extra iterative loop between the channel estimator and the three-stage turbo detector. Specifically, in order to maintain a high system throughput, only the minimum number of STSK training blocks is utilized for obtaining an initial LSCE. Naturally, the number of training blocks is related to the number of transmit antennas [16]. Then the low-complexity single-antenna based ML soft-demapping of [5] is carried out and the soft decisions are exchanged between the unity-rate-code (URC) decoder and CSTSK soft-demapper within the inner turbo loop, before they are forwarded to the outer recursive systematic code's (RSC) decoder. Moreover, the "high quality" or "more reliable" blocks-of-bits are selected based on the *a posteriori* information produced by the CSTSK soft-demapper within the original inner turbo loop of the URC decoder and CSTSK soft-demapper, which are re-modulated concurrently with each outer iteration of the original outer turbo loop for the sake of facilitating decision-directed LSCE updates. Since the CE is naturally embedded into the original iterative three-stage demapping-decoding scheme, no extra iterative loop is required between the CE and the three-stage CSTSK demapper-decoder. Moreover, as our proposed BBSBCE scheme only selects a sufficient number of high-quality decisions for CE, the complexity of our channel estimator is dramatically lower than any existing DDCE scheme which employs the entire frame of the detected symbols. All these features result in a low system complexity. In other words, the proposed joint BBSBCE and three-stage turbo demapping-decoding scheme has a similar computational complexity to that of the original three-stage turbo demapping-decoding scheme with the known CSI. Furthermore, the proposed semi-blind joint CE and turbo detection-decoding scheme is capable of fully exploiting the "turbo effects" of the joint CE and three-stage demapping-decoding for approaching the optimal performance obtained by the idealised three-stage turbo demapping-decoding receiver furnished with the perfect CSI, despite using only the same low number of turbo iterations as the latter.

The rest of this contribution is organized as follows. Section II describes the three-stage serial-concatenated turbo coding aided CSTSK system. The proposed low-complexity near-capacity semi-blind joint BBSBCE and three-stage data demapping-decoding scheme is detailed in Section III, while the Cramér-Rao lower bound (CRLB) is formulated in Section IV for benchmarking the performance of the CE in our proposed scheme. The achievable performance of the system is investigated in Section V, while our conclusions are offered in Section VI.

The following notational conventions are adopted throughout our discussions. A CSTSK system employing an $\mathcal{L}$-phase shift keying (PSK) or $\mathcal{L}$-QAM scheme is denoted as $CSTSK(N_T, N_R, T_n, Q, \mathcal{L})$, where N_T and N_R indicate the

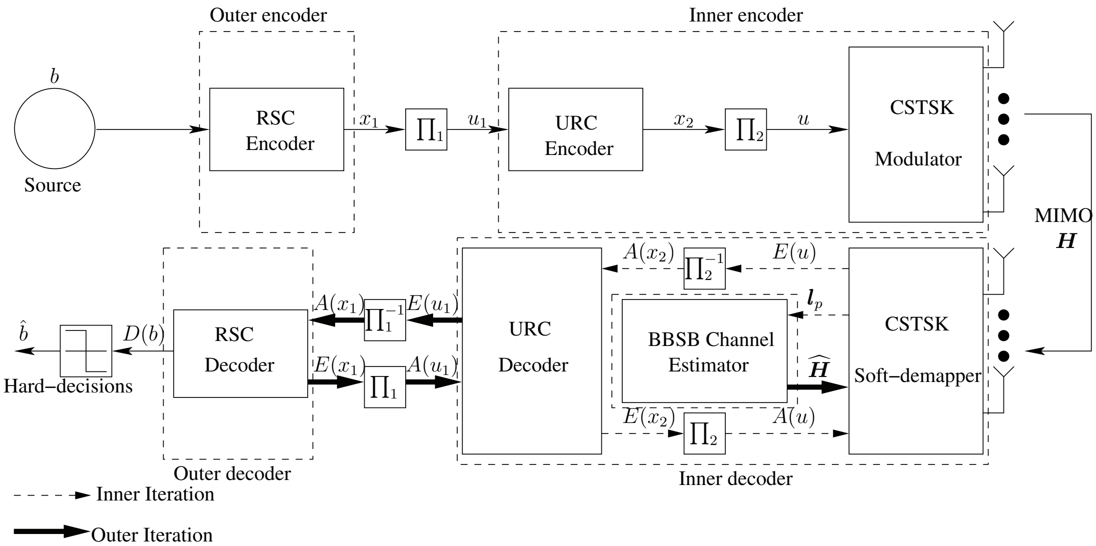

Fig. 1. Three-stage serial-concatenated CSTSK MIMO system.

numbers of transmit and receive antennas, respectively, while T_n denotes the number of time slots occupied by the CSTSK signal block and Q is the number of dispersion matrices employed. Boldface capital and lower-case letters stand for matrices and column vectors, respectively, while $\mathbb{C}$ denotes the field of complex numbers. The inverse operation is denoted by $(\ )^{-1}$, while $vec(\ )$ denotes the vector stacking operator and $\lceil\ \rceil$ is the integer ceiling operator. Furthermore, $(\ )^{\mathrm{T}}$ and $(\ )^{\mathrm{H}}$ represent the transpose and conjugate transpose operators, respectively, whereas $\|\ \|$ and $|\ \ |$ denote the norm and magnitude operators, respectively. Additionally, $\mathrm{E}\{\ \}$ denotes the expectation operator and $\mathrm{tr}\{\ \}$ the trace operator. Finally, the Kronecker product is denoted by $\otimes$, the conjugate operator by $(\ )^*$, and the $(M \times M)$-element identity matrix by $\boldsymbol{I}_M$.

II. Three-Stage Serial-Concatenated Turbo Coding Aided CSTSK System

For the sake of achieving a near-capacity performance, a three-stage serial-concatenated turbo coding scheme was invoked for a CSTSK based MIMO system in [5]. The structure of this three-stage serial-concatenated turbo encoder and decoder is shown in Fig. 1. For the time being, we will assume that perfect CSI is available at the receiver, namely, that the output of the channel estimator is replaced by the actual CSI in Fig. 1. We consider the $CSTSK(N_T, N_R, T_n, Q, \mathcal{L})$ scheme in a frequency-flat Rayleigh fading environment. Let i denote the STSK block index. At the transmitter, the information bits are firstly encoded by a half-rate RSC outer encoder. Then a low-complexity memory-1 URC encoder is incorporated for allowing the system to beneficially spread the *extrinsic* information across the iterative decoder components. As a benefit, a vanishingly low BER may be attained [33]. At the CSTSK modulator[1], the channel-coded sequence of bits is firstly converted to a number of parallel bit sequences, each containing a certain number of bits-per-block (BPB) given by $N = \log_2(Q) + \log_2(\mathcal{L})$. Then the first $\log_2(Q)$ bits are used for choosing a single dispersion matrix $\boldsymbol{A}(i)$ from the Q pre-assigned dispersion matrices $\{\boldsymbol{A}_q \in \mathbb{C}^{N_T \times T_n}, 1 \leq q \leq Q\}$, while the remaining $\log_2(\mathcal{L})$ bits are mapped to the complex-valued symbol $s(i) \in \{s_l, 1 \leq l \leq \mathcal{L}\}$ of a conventional modulation scheme, such as $\mathcal{L}$-PSK/QAM [4], [5]. In this way, a total of N forward-error-correction (FEC) coded source bits are mapped to a single STSK signalling block $\boldsymbol{S}(i) \in \mathbb{C}^{N_T \times T_n}$ with

$$\boldsymbol{S}(i) = s(i)\boldsymbol{A}(i). \tag{1}$$

Note that the mth row of $\boldsymbol{S}(i)$ is transmitted through the mth transmit antenna in the T_n time slots, and the average transmission power in each time slot is normalized to unity. Therefore, the dispersion matrices are designed to obey the power constraints of $\mathrm{tr}\big[\boldsymbol{A}_q^{\mathrm{H}}\boldsymbol{A}_q\big] = T_n$ for $1 \leq q \leq Q$ [4], [5]. The normalized throughput R per time-slot of this STSK system is given by

$$R = \frac{\log_2(Q \cdot \mathcal{L})}{T_n}\ \text{[bits/symbol]}. \tag{2}$$

The corresponding three-stage turbo decoder adopted at receiver is also shown in Fig. 1, where the received signal block $\boldsymbol{Y}(i) \in \mathbb{C}^{N_R \times T_n}$ can be expressed as [4], [5]

$$\boldsymbol{Y}(i) = \boldsymbol{H}\boldsymbol{S}(i) + \boldsymbol{V}(i), \tag{3}$$

and $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$ is the corresponding MIMO channel matrix. The elements of $\boldsymbol{H}$ obey the complex-valued Gaussian distribution of zero-mean and unit variance $\mathcal{CN}(0,1)$, while $\boldsymbol{V}(i) \in \mathbb{C}^{N_R \times T_n}$ is the additive white Gaussian noise (AWGN) matrix, whose components obey $\mathcal{CN}(0, N_0)$ with N_0 being the AWGN variance. By applying the vector stacking operation to the received signal block $\boldsymbol{Y}(i)$ in Eq. (3), the equivalent system model can be expressed as [4], [5]

$$\overline{\boldsymbol{y}}(i) = \overline{\boldsymbol{H}}\boldsymbol{\Upsilon}\boldsymbol{k}(i) + \overline{\boldsymbol{v}}(i), \tag{4}$$

[1]Here, the detailed CSTSK system model has been omitted for the sake of space economy, noting that they can be found in [4] and [5].

where we have:

$$\overline{\boldsymbol{y}}(i) = vec\big(\boldsymbol{Y}(i)\big) \in \mathbb{C}^{N_R T_n \times 1}, \tag{5}$$
$$\overline{\boldsymbol{H}} = \boldsymbol{I}_{T_n} \otimes \boldsymbol{H} \in \mathbb{C}^{N_R T_n \times N_T T_n}, \tag{6}$$
$$\boldsymbol{\Upsilon} = \big[vec(\boldsymbol{A}_1)\ vec(\boldsymbol{A}_2) \cdots vec(\boldsymbol{A}_Q)\big] \in \mathbb{C}^{N_T T_n \times Q}, \tag{7}$$
$$\overline{\boldsymbol{v}}(i) = vec\big(\boldsymbol{V}(i)\big) \in \mathbb{C}^{N_R T_n \times 1}, \tag{8}$$

while the equivalent transmitted signal vector $\boldsymbol{k}(i) \in \mathbb{C}^{Q\times 1}$ can be formulated as

$$\boldsymbol{k}(i) = \big[\underbrace{0\ \cdots\ 0}_{q-1}\ s(i)\ \underbrace{0\ \cdots\ 0}_{Q-q}\big]^{\mathrm{T}}. \tag{9}$$

Explicitly, the qth element of $\boldsymbol{k}(i)$ in Eq. (9) is the conventionally modulated signal $s(i)$ and its remaining elements are zero, where q indicates the corresponding dispersion matrix that is activated for the ith STSK block. Since the constellation size is $\mathcal{L}$ and the number of dispersion matrices is Q, the total number of legitimate transmit signal vectors $\boldsymbol{k}(i)$ is $(\mathcal{L}\cdot Q)$. Therefore, we have

$$\boldsymbol{k}(i) \in \mathbb{K} = \{\boldsymbol{k}_{q,l}, 1 \le q \le Q, 1 \le l \le \mathcal{L}\} \tag{10}$$

with

$$\boldsymbol{k}_{q,l} = \big[\underbrace{0\ \cdots\ 0}_{q-1}\ s_l\ \underbrace{0\ \cdots\ 0}_{Q-q}\big]^{\mathrm{T}}, \tag{11}$$

where s_l is the lth symbol in the $\mathcal{L}$-point constellation of conventional $\mathcal{L}$-PSK or $\mathcal{L}$-QAM.

Since the CSTSK MIMO scheme is free from ICI, as seen from the equivalent system model of Eqs. (4) to (9), a low-complexity single-stream ML detector/demapper can be employed [4], [5]. Upon applying the max-log approximation [33], the extrinsic log-likelihood ratio (LLR) value of bit u_n, $n \in \{1, 2, \cdots, N\}$ can be expressed as

$$\begin{aligned} L_e(u_n) = &\max_{\boldsymbol{k}_{q,l}\in\mathbb{K}_1^n} \Big\{ -\|\bar{\boldsymbol{y}} - \overline{\boldsymbol{H}}\boldsymbol{\Upsilon}\boldsymbol{k}_{q,l}\|^2/N_0 + \sum_{j\neq n} u_j L_a(u_j) \Big\} \\ &- \max_{\boldsymbol{k}_{q,l}\in\mathbb{K}_0^n} \Big\{ -\|\bar{\boldsymbol{y}} - \overline{\boldsymbol{H}}\boldsymbol{\Upsilon}\boldsymbol{k}_{q,l}\|^2/N_0 + \sum_{j\neq n} u_j L_a(u_j) \Big\}, \end{aligned} \tag{12}$$

where $L_a(u_n)$ represents the *a priori* information expressed in terms of the LLR of the corresponding bit u_n, $\mathbb{K}_1^n = \{\boldsymbol{k}_{q,l} \in \mathbb{K} | u_n = 1\}$ and $\mathbb{K}_0^n = \{\boldsymbol{k}_{q,l} \in \mathbb{K} | u_n = 0\}$ are the sub-sets of the legitimate equivalent signals, when the corresponding bits are $u_n = 1$ and $u_n = 0$, respectively. Note that the complexity of the optimal full ML detection conceived for STSK increases linearly with respect to $Q \cdot \mathcal{L}$, not exponentially. The *a posteriori* information output by the CSTSK demapper is then given by

$$L_p(n) = L_p(u_n) = L_e(u_n) + L_a(u_n). \tag{13}$$

Additionally, the associated *a priori* information and *extrinsic* information are interleaved and exchanged I_{inner} times within the composite inner decoder, which is formed by CSTSK soft-demapper and the URC decoder. The outer decoder of Fig. 1 is constituted by the RSC decoder, where the information gleaned from the inner decoder is iteratively exchanged I_{outer} times. As a benefit of the URC decoder's infinite impulse response, the corresponding extrinsic information transfer (EXIT) curve [34] is capable of reaching the (1.0, 1.0) point of perfect convergence to a vanishingly low BER, implying that no error floor occurs [33]. Therefore, the convergence performance of the system is improved by the URC decoder.

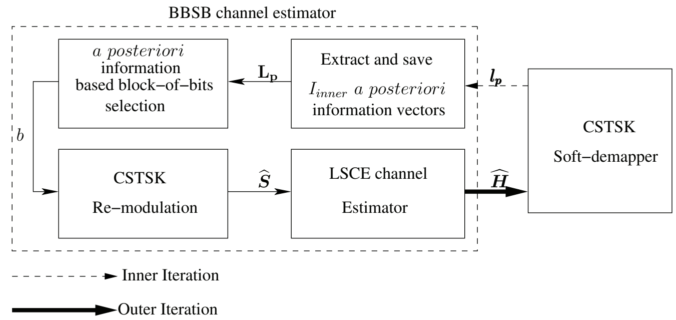

Fig. 2. The proposed block-of-bits selection based semi-blind iterative channel estimation aided system.

III. Joint BBSBCE and Three-Stage Demapping-Decoding

The *a posteriori* information output by the CSTSK soft-demapper formulated in Eq. (12) provides the confidence levels. i.e. the probabilities of binary 1s and binary 0s [33]. Therefore, based on this confidence level, we may opt for selecting reliable decisions from the CSTSK soft-demapper's output sequence for the DDCE update. Again, our proposed joint block-of-bits selection based (BBSB) CE and three-stage turbo demapper-decoder is shown in Fig. 1. Note that in contrast to all the existing joint CE and turbo detection-decoding schemes [20]–[31], as a benefit of selecting high-confidence decisions, the DDCE does not have to wait for the accurate final convergence of the turbo detector-decoder. Quite the contrary, our scheme updates the DDCE within the original outer loop of the three-stage turbo demapper-decoder, as explicitly emphasised in Figs. 1 and 2. In other words, unlike all the existing schemes [20]–[31], no additional iterative loop involving the CE and the three-stage turbo demapper-decoder is required. Consequently, the computational complexity of our joint scheme remains similar to that of the original three-stage turbo demapper-decoder, which was furnished with the Least Square (LS) CSI-estimate. Moreover, as it will be confirmed later in our simulation study, our joint BBSBCE and three-stage turbo demapper-decoder is capable of approaching the near-capacity optimal performance associated with the perfect CSI. Let us now detail our proposed scheme.

Similar to the conventional semi-blind scheme of [16], [17], our proposed scheme also relies on a low number of training blocks. Let us assume that the number of available training blocks is M. If we arrange the training data as

$$\boldsymbol{Y}_{tM} = \big[\boldsymbol{Y}(1)\ \boldsymbol{Y}(2) \cdots \boldsymbol{Y}(M)\big], \tag{14}$$
$$\boldsymbol{S}_{tM} = \big[\boldsymbol{S}(1)\ \boldsymbol{S}(2) \cdots \boldsymbol{S}(M)\big], \tag{15}$$

the initial LSCE of the MIMO channel matrix $\boldsymbol{H}$ is then given by

$$\widehat{\boldsymbol{H}}_{LSCE} = \boldsymbol{Y}_{tM}\boldsymbol{S}_{tM}^{\mathrm{H}}\big(\boldsymbol{S}_{tM}\boldsymbol{S}_{tM}^{\mathrm{H}}\big)^{-1}. \tag{16}$$

For the sake of maintaining a high system throughput, only a low number of STSK training blocks should be used. On the other hand, in order to ensure for $S_{tM}S_{tM}^{\mathrm{H}}$ to have the full rank of N_T, it is necessary to guarantee that $M \cdot T_n \geq N_T$ and this leads to a lower bound of the number of training blocks, which is

$$M \geq \left\lceil \frac{N_T}{T_n} \right\rceil. \tag{17}$$

For instance, if we have $N_T = 4$ and $T_n = 2$, then the lower bound is $M = 2$ and we may choose to use as few as two STSK training blocks for initial CE. With such a low number of initial training blocks, the accuracy of the LSCE of Eq. (16) will be poor and hence the achievable BER performance based on the initial channel estimate will also remain poor. The task of the joint BBSB channel estimator and three-stage turbo demapper-decoder is to iteratively refine the channel estimate for the sake of achieving a near-capacity BER performance. Referring to Fig. 2, let the number of iterations in the two-stage inner turbo loop be I_{inner} and the number of iterations in the outer turbo loop be I_{outer}, respectively. Let us now denote the observation data at the input of the CSTSK soft-demapper of Fig. 2 as

$$\boldsymbol{Y}_{\mathrm{d}\tau} = \left[\boldsymbol{Y}(1)\ \boldsymbol{Y}(2)\cdots\boldsymbol{Y}(\tau)\right]. \tag{18}$$

The joint BBSBCE and three-stage demapper-decoder is described as follows.

Semi-Blind Iterative Algorithm

1) Set the iteration index to $t = 0$ and the initial channel estimate to $\widehat{\boldsymbol{H}}^{(t)} = \widehat{\boldsymbol{H}}_{LSCE}$.
2) Given the initial channel estimate $\widehat{\boldsymbol{H}}^{(t)}$, perform CSTSK ML soft-demapping based on the observation data $\boldsymbol{Y}_{\mathrm{d}\tau}$ of Eq. (18). The CSTSK soft-demapper then exchanges its soft information with the URC inner decoder for I_{inner} iterations, which yields the I_{inner} vectors of *a posteriori* information formulated in Eq. (13) and arranged as

$$\boldsymbol{L}_p = \left[\boldsymbol{l}_p^1\ \boldsymbol{l}_p^2\cdots\boldsymbol{l}_p^{I_{\mathrm{inner}}}\right]^{\mathrm{T}}, \tag{19}$$

where $\boldsymbol{L}_p \in \mathbb{C}^{I_{\mathrm{inner}}\times(N\cdot\tau)}$ is referred to as the equivalent *a posteriori* information matrix and $\boldsymbol{l}_p^i = \left[L_p^i(1)\ L_p^i(2)\cdots L_p^i(N\cdot\tau)\right]^{\mathrm{T}} \in \mathbb{C}^{(N\cdot\tau)\times 1}$ denotes the *a posteriori* information vector obtained by the CSTSK soft-demapper during the ith inner iteration, while N is the BPB. Based on the fact that the nth column of $\boldsymbol{L}_p$ represents the I_{inner} soft decisions of the nth information bit, where $n \in \{1, 2, \cdots, (N\cdot\tau)\}$, the block-of-bits invoked for CE is selected in either of the following two cases:

Case 1: If the soft decisions in the same column share similar values, these soft decisions may result in a reliable bit decision, which are hence invoked for CE. Specifically, the criterion for the nth information bit to be selected is expressed in Eq. (20), where μ is the mean of the soft decisions in the nth column of the matrix $\boldsymbol{L}_p$, while T_h denotes the block-of-bits selection threshold used.

Case 2: If the absolute values of soft decisions in the nth column appear to be in monotonically ascending order and share the same polarity (i.e. the decisions are all positive or negative), the nth information bit can be regarded as correct and hence it is selected for CE.

After checking through all the columns of $\boldsymbol{L}_p$, only "high confidence" decisions are selected and the corresponding STSK block indices may be obtained by a sliding window based method using a window size of N. More explicitly, if N consecutive information bits are all regarded as correct, the corresponding information block will be selected for CE. This yields a "high confidence" integer-valued index vector, which is denoted as $\boldsymbol{x}^t = \left[x^t(1)\ x^t(2)\cdots x^t(\tau_s^t)\right]^{\mathrm{T}}$, with the number of the selected elements τ_s^t varying within $(0,\ \tau_{\mathrm{sel}}]$, where τ_{sel} ($\ll \tau$) is the imposed maximum number of selected decision-based signal blocks. By using this index vector, the corresponding observation data can be selected from Eq. (18), and re-arranged as

$$\boldsymbol{Y}_{\mathrm{sel}}^{(t)} = \left[\boldsymbol{Y}(x^t(1))\ \boldsymbol{Y}(x^t(2))\cdots\boldsymbol{Y}(x^t(\tau_s^t))\right]. \tag{21}$$

3) By re-modulating the selected detected blocks-of-bits with the aid of $\boldsymbol{x}^t$, we have

$$\widehat{\boldsymbol{S}}_{\mathrm{sel}}^{(t)} = \left[\widehat{\boldsymbol{S}}(x^t(1))\ \widehat{\boldsymbol{S}}(x^t(2))\cdots\widehat{\boldsymbol{S}}(x^t(\tau_s^t))\right]. \tag{22}$$

The resultant decision-directed LSCE update is then given by [2]

$$\widehat{\boldsymbol{H}}^{(t+1)} = \boldsymbol{Y}_{\mathrm{sel}}^{(t)}\left(\widehat{\boldsymbol{S}}_{\mathrm{sel}}^{(t)}\right)^{\mathrm{H}}\left(\widehat{\boldsymbol{S}}_{\mathrm{sel}}^{(t)}\left(\widehat{\boldsymbol{S}}_{\mathrm{sel}}^{(t)}\right)^{\mathrm{H}}\right)^{-1}. \tag{23}$$

At the same time the corresponding soft information is further exchanged with the outer RSC decoder.
4) Set $t = t + 1$. If $t < I_{\mathrm{outer}}$, repeat steps 2) and 3); otherwise, stop.

Remark 1: The idea behind **Case 1** is that, if the decisions are relatively stable during several consecutive turbo iterations, it may be regarded as reliable decisions. This makes sense because a stable state may be achieved for turbo coding after a few iterations and the stable decisions may be high-quality ones. It can be seen that **Case 1** mainly occurs during the later turbo iterations. With regard to **Case 2**, we note that if the soft values for a specific bit are in monotonically ascending order and share the same polarity in consecutive iterations, the corresponding bit decision may also be deemed reliable. This also makes sense, because the correct decisions may experience a gradual iteration gain and this will lead to increasing absolute values of the soft-decisions as the number of iterations increases, especially during the early turbo iteration stages.

Remark 2: Selecting the high-confidence detected blocks-of-bits is carried out within the inner turbo loop consisting of the CSTSK soft-demapper and URC decoder of Fig. 1. This is because the selected $x^t(i)$-th detected block-of-bits can be linked to the corresponding observation block $\boldsymbol{Y}(x^t(i))$ after it was re-modulated to generate the CSTSK signal block $\widehat{\boldsymbol{S}}(x^t(i))$. The selection cannot be carried out within the outer turbo loop of Fig. 1, since the RSC encoder will disperse the

[2] For STSK, DDCE has to be based on hard-decisions. This is because there exists no proper soft-estimate for the STSK signal block $\boldsymbol{S}(i)$ of Eq. (1). The soft-decision aided scheme [31] for example will destroy the STSK structure.

$$\frac{|L_p^1(n)-L_p^2(n)|+|L_p^2(n)-L_p^3(n)|+\cdots+|L_p^{I_{\rm inner}-1}(n)-L_p^{I_{\rm inner}}(n)|}{|\mu|}\in(0,\ T_h), \tag{20}$$

bits of the blocks. It is worth emphasising that the decision-directed LSCE update of Eq. (23) takes place concurrently, when the information is exchanged between the two-stage inner CSTSK soft-demapper/URC decoder and the outer RSC decoder [3]. Therefore, no extra iterative loop is needed between the CE and the three-stage turbo demapper-decoder.

Remark 3: The value of the block-of-bits selection threshold T_h employed in step 2) **Case 1** should be carefully chosen. Too small a value may lead to an insufficient number of blocks selected for CE even after examining the entire sequence of $N\cdot\tau$ bit decisions. By contrast, too large a value may result in an excessive number of selected blocks τ_s^t, potentially approaching the maximum affordable value $\tau_{\rm sel}$ after only examining a small initial portion of the $N\cdot\tau$ bit decisions and the selected blocks may contain many "low confidence" decisions. Both of these two situations will result in a performance degradation. However, apart from these relatively extreme cases, our experience suggests that the performance of our semi-blind scheme is insensitive to the value of T_h. Specifically, there exists a relatively wide range of values for T_h, which allows our scheme to approach its optimal performance without increasing the number of turbo iterations $I_{\rm outer}$ and $I_{\rm inner}$. The limit $\tau_{\rm sel}$ imposed on the number of data blocks selected in step 2) is to avoid imposing an unnecessary CE complexity. Our empirical results show that $\tau_{\rm sel}=100$ is sufficient for converging to the optimal ML detection performance associated with the perfect CSI.

Let us denote the complexity of the RSC decoder, of the URC decoder, and of the CSTSK ML soft-demapper by $C_{\rm RSC}$, $C_{\rm URC}$ and $C_{\rm ML}$, respectively. The overall system complexity of the idealised three-stage turbo detector-decoder, given the perfect CSI can readily be expressed as

$$C_{\rm ideal}=I_{\rm outer}\Big(C_{\rm RSC}+I_{\rm inner}\big(C_{\rm ML}+C_{\rm URC}\big)\Big). \tag{24}$$

Typically, $I_{\rm inner}$ and $I_{\rm outer}$ are small, i.e. a few inner turbo iterations and a few outer iterations are often sufficient. Since our BBSBCE scheme selects no more than $\tau_{\rm sel}$ data blocks for CE, our decision-directed LSCE of Eq. (23) has a complexity upper bound on the order of $O\big(\tau_{\rm sel}^3\big)$. Furthermore, as it will be confirmed later in our simulation study, our proposed amalgamated scheme approaches the perfect-CSI based performance bound with the same number of turbo iterations, $I_{\rm inner}$ and $I_{\rm outer}$, while the decision-directed LSCE of Eq. (23) approaches the Cramér-Rao lower bound (CRLB) [35] with the aid of $I_{\rm outer}$ iterations. Therefore, the overall system complexity of our proposed joint BBSBCE and three-stage demapping-decoding scheme can be expressed as

$$C_{\rm proposed}\le I_{\rm outer}\cdot O\big(\tau_{\rm sel}^3\big)+C_{\rm ideal}. \tag{25}$$

By contrast, all the existing conventional schemes utilise the entire sequence of τ detected data blocks and, therefore, the complexity of the decision-directed LSCE in these conventional schemes is on the order of $O\big(\tau^3\big)$. Moreover, these conventional schemes introduce an extra iterative loop between the channel estimator and the three-stage turbo detector-decoder, which takes $I_{\rm ce}$ iterations to converge[4]. Let us make the optimistic assumption that the three-stage turbo detector-decoder within these conventional joint schemes also require $I_{\rm inner}$ inner turbo iterations and $I_{\rm outer}$ outer turbo iterations. Then the overall system complexity of these existing conventional joint CE and three-stage turbo detector-decoder structures can be written as

$$C_{\rm conventional}=I_{\rm ce}\cdot O\big(\tau^3\big)+I_{\rm ce}\cdot C_{\rm ideal}. \tag{26}$$

Note that we have $I_{\rm ce}>I_{\rm outer}$. Moreover, the length τ of a turbo coded frame is typically in the thousands, while it is sufficient to set $\tau_{\rm sel}$ to be 10% of τ in our scheme, i.e. $\tau_{\rm sel}$ is typically in the hundreds. Consequently, the complexity of the CE in our proposed scheme is at least three orders of magnitude lower than that of the existing conventional schemes. By comparing Eqs. (25) and (26), it can be seen that our proposed joint BBSBCE and three-stage turbo detector-decoder scheme dramatically reduces the required computational complexity.

IV. Cramér-Rao Lower Bound

Since the CRLB provides the best attainable performance for an unbiased estimator [35]–[40], we can also evaluate the achievable performance of our joint BBSBCE and three-stage turbo demapping/decoding scheme by comparing its mean square error (MSE) to the CRLB. Let the number of available signal blocks invoked for training based CE be $\widetilde{\tau}=\tau_{\rm sel}$. Then the CSTSK model of Eq. (3) over the $\widetilde{\tau}$ blocks can be written as

$$\boldsymbol{Y}_{t\widetilde{\tau}}=\boldsymbol{H}\,\boldsymbol{S}_{t\widetilde{\tau}}+\boldsymbol{V}_{t\widetilde{\tau}}, \tag{27}$$

or equivalently

$$\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}=\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}^{\rm T}\widetilde{\boldsymbol{h}}+\widetilde{\boldsymbol{v}}_{\widetilde{\tau}}, \tag{28}$$

where $\boldsymbol{Y}_{t\widetilde{\tau}}=\big[\boldsymbol{Y}(1)\ \boldsymbol{Y}(2)\cdots\boldsymbol{Y}(\widetilde{\tau})\big]$, $\boldsymbol{S}_{t\widetilde{\tau}}$ and $\boldsymbol{V}_{t\widetilde{\tau}}$ are similarly defined, while

$$\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}=vec\big(\boldsymbol{Y}_{t\widetilde{\tau}}^{\rm T}\big)\in\mathbb{C}^{N_RT_n\widetilde{\tau}\times 1}, \tag{29}$$

$$\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}=\boldsymbol{I}_{N_R}\otimes\boldsymbol{S}_{t\widetilde{\tau}}\in\mathbb{C}^{N_RN_T\times N_RT_n\widetilde{\tau}}, \tag{30}$$

$$\widetilde{\boldsymbol{h}}=vec\big(\boldsymbol{H}^{\rm T}\big)\in\mathbb{C}^{N_RN_T\times 1}, \tag{31}$$

$$\widetilde{\boldsymbol{v}}_{\widetilde{\tau}}=vec\big(\boldsymbol{V}_{t\widetilde{\tau}}^{\rm T}\big)\in\mathbb{C}^{N_RT_n\widetilde{\tau}\times 1}. \tag{32}$$

Given the training signal matrix $\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}$, the conditional probability $p\big(\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}|\widetilde{\boldsymbol{h}}\big)$ is expressed as

$$p\big(\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}|\widetilde{\boldsymbol{h}}\big)=\frac{1}{(\pi N_0)^{N_RT_n\widetilde{\tau}}}\exp\left(-\frac{\|\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}-\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}^{\rm T}\widetilde{\boldsymbol{h}}\|^2}{N_0}\right). \tag{33}$$

[3]In this context the question arises that whether it is necessary at all to estimate the channel for each outer iteration. The direct answer is yes because the CE performance improves upon increasing the number of (outer) iterations. This may be seen from the simulation results shown in Fig. 5, Fig. 7 and Fig. 11.

[4]We would like to emphasise that since a conventional joint CE and turbo receiver utilises all the symbol decisions of a transmitted frame in channel estimation regardless of whether they are reliable, it is necessary to wait until the turbo detector/decoder converges before the channel update can take place, in order to benefit from the error correction capability of the turbo detection/decoding.

The Fisher information matrix (FIM) [35] is defined as

$$\boldsymbol{F}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big) = -\mathrm{E}\left\{\frac{\partial^2 \log\big(p(\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}|\widetilde{\boldsymbol{h}})\big)}{\partial\widetilde{\boldsymbol{h}}\ \partial\widetilde{\boldsymbol{h}}^{\mathrm{H}}}\right\} = -\mathrm{E}\left\{\frac{\partial}{\partial\widetilde{\boldsymbol{h}}^{\mathrm{H}}}\left(\frac{\partial \log\big(p(\widetilde{\boldsymbol{y}}_{\widetilde{\tau}}|\widetilde{\boldsymbol{h}})\big)}{\partial\widetilde{\boldsymbol{h}}^{\mathrm{H}}}\right)^{\mathrm{H}}\right\}, \quad (34)$$

where the expectation is with respect to the noise $\widetilde{\boldsymbol{v}}_{\widetilde{\tau}}$. Substituting (33) into (34) results in

$$\boldsymbol{F}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big) = \frac{1}{N_0}\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}^{\mathrm{H}}. \quad (35)$$

Let

$$\widetilde{\boldsymbol{S}}_{\widetilde{\tau}\mathrm{opt}} = \arg\max_{\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}} \mathrm{tr}\big\{\boldsymbol{F}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big)\big\}. \quad (36)$$

The ith diagonal element of $\boldsymbol{F}^{-1}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}\mathrm{opt}}\big)$ provides a lower bound of the variance for the unbiased estimate of the ith element of $\widetilde{\boldsymbol{h}}$ based on the training signal block length of $\widetilde{\tau}$, where

$$\mathrm{CRLB}(\widetilde{\tau}) = \mathrm{tr}\big\{\boldsymbol{F}^{-1}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}\mathrm{opt}}\big)\big\} \quad (37)$$

is the minimum MSE achievable by an unbiased estimator based on the training block length $\widetilde{\tau}$. Note that we have $\widetilde{\boldsymbol{S}}_{\widetilde{\tau}\mathrm{opt}} = \boldsymbol{I}_{N_R} \otimes \boldsymbol{S}_{t\widetilde{\tau}\mathrm{opt}}$ with

$$N_0 \cdot \boldsymbol{S}_{t\widetilde{\tau}\mathrm{opt}} = \arg\max_{\boldsymbol{S}_{t\widetilde{\tau}}} \mathrm{tr}\big\{\boldsymbol{S}_{t\widetilde{\tau}}\boldsymbol{S}_{t\widetilde{\tau}}^{\mathrm{H}}\big\}, \quad (38)$$

where

$$\mathrm{tr}\big\{\boldsymbol{S}_{t\widetilde{\tau}}\boldsymbol{S}_{t\widetilde{\tau}}^{\mathrm{H}}\big\} = \mathrm{tr}\Big\{\sum_{i=1}^{\widetilde{\tau}}\boldsymbol{S}(i)\boldsymbol{S}(i)^{\mathrm{H}}\Big\}. \quad (39)$$

Since $\boldsymbol{S}(i) = s(i)\boldsymbol{A}(i)$ with $s(i) \in \{s_l, 1 \le l \le \mathcal{L}\}$ and $\boldsymbol{A}(i) \in \big\{\boldsymbol{A}_q, 1 \le q \le Q\big\}$, the determination of the "optimal" training signal matrix $\widetilde{\boldsymbol{S}}_{\widetilde{\tau}\mathrm{opt}}$ is nontrivial. We therefore use the CRLB for the given training signal matrix $\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}$, defined as

$$\mathrm{CRLB}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big) = \mathrm{tr}\big\{\boldsymbol{F}^{-1}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big)\big\}, \quad (40)$$

to approximate $\mathrm{CRLB}(\widetilde{\tau})$. For a large block length of $\widetilde{\tau}$, this approximation is accurate.

Let $\widehat{\widetilde{\boldsymbol{h}}}$ be an unbiased estimate of $\widetilde{\boldsymbol{h}}$ produced by our semi-blind scheme with the selected decisions limited to no more than τ_{sel}. The MSE of the channel estimate $\widehat{\widetilde{\boldsymbol{h}}}$ is expressed as

$$J_{\mathrm{MSE}}\Big(\widehat{\widetilde{\boldsymbol{h}}}\Big) = \mathrm{E}\Big\{\big\|\widehat{\widetilde{\boldsymbol{h}}} - \widetilde{\boldsymbol{h}}\big\|^2\Big\} \approx \frac{1}{K_{\mathrm{run}}}\sum_{k=1}^{K_{\mathrm{run}}}\big\|\widehat{\widetilde{\boldsymbol{h}}}^{(k)} - \widetilde{\boldsymbol{h}}\big\|^2, \quad (41)$$

where K_{run} is the number of estimation experiments and $\widehat{\widetilde{\boldsymbol{h}}}^{(k)}$ is the channel estimate obtained during the kth experiment. Clearly, we have

$$J_{\mathrm{MSE}}\Big(\widehat{\widetilde{\boldsymbol{h}}}\Big) \ge \mathrm{CRLB}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big). \quad (42)$$

In the following simulation study, we will demonstrate that the MSE of the channel estimate produced by our proposed semi-blind joint BBSBCE and three-stage demapping-decoding algorithm associated with as few as $M = 2$ initial training blocks and with the limit of $\tau_{\mathrm{sel}} = 100$ imposed on selected decision blocks attains the CRLB $\mathrm{CRLB}\big(\widetilde{\boldsymbol{S}}_{\widetilde{\tau}}\big)$ with $\widetilde{\tau} = 100$.

V. Simulation Results

Let us now investigate the achievable performance of our low-complexity semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme. A quasi-static Rayleigh fading environment was considered. An interleaver length of 10^6 bits was used by the three-stage serial-concatenated turbo encoder/decoder of Fig. 1. The binary generator polynomials of the RSC encoder were $G_{RSC} = [1,0,1]_2$ and $G_{RSC}^r = [1,1,1]_2$, while these of the URC encoder were $G_{URC} = [1,0]_2$ and $G_{URC}^r = [1,1]_2$, where G_{RSC}^r and G_{URC}^r are the feedback polynomials of the RSC and URC encoders, respectively. The number of inner iterations was set to $I_{\mathrm{inner}} = 3$. The transmitted signal power of all the simulated systems was normalized to unity and, therefore, the SNR was defined as $\frac{1}{N_0}$, with N_0 being the AWGN power. The achievable performance was assessed in terms of three metrics: the MSE of the CE defined in Eq. (41), the achievable BER and the EXIT charts [33]. All the results were averaged over $K_{\mathrm{run}} = 100$ channel realisations. The set of dispersion matrices was designed based on the criterion of maximizing the discrete-input continuous-output memoryless channel capacity, as detailed in [5]. The up-bounded fraction of most reliable detected STSK symbols selected for updating the BBSB channel estimate was set to 10%, which corresponds to an up-bounded number of STSK blocks of $\tau_{\mathrm{sel}} = 100$.

A. Example One: $CSTSK(4,2,2,4,QPSK)$ having a throughput of 4 bits/symbol

We first considered the $CSTSK(4,2,2,4,QPSK)$ system with $\mathcal{L} = 4$. Our investigation commenced with the EXIT chart analysis of the proposed semi-blind iterative scheme in conjunction with the block-of-bits selection threshold of $T_h = 0.2$, in comparison to the perfect-CSI scenario. It can be seen from the EXIT charts shown in Fig. 3 that open tunnels exist between the EXIT curves of the amalgamated inner CSTSK soft-demapper-URC decoder and the outer RSC decoder for both the proposed semi-blind scheme and the optimal ML detection based on the perfect CSI at SNR= -0.1 dB. The actual Monte-Carlo simulation based stair-case shaped decoding trajectories, which closely match the EXIT curves, are also provided at SNR= -0.1 dB. Both the trajectories show that the perfect convergence point at $(1.0, 1.0)$ can be reached with the aid of $I_{\mathrm{outer}} = 10$ iterations, implying that the proposed semi-blind scheme is capable of achieving the optimal ML detection performance at the same number of turbo iterations. Additionally, due to the poor-quality LSCE used during the first few iterations, the starting point of the EXIT curves of the inner CSTSK soft-demapper-URC decoder of the semi-blind scheme is lower than that of the optimal ML detection based on the perfect CSI. However, the *a priori* information is improved as the number of iterations increased, and the two curves gradually become overlapped. In other words, an accurate CE is obtained by the semi-blind BBSB iterative scheme, even though the initial LSCE based on $M = 2$ training blocks is very poor.

The BER performance achieved by our proposed scheme relying on $M = 2$ initial training blocks is shown in Fig. 4, where the near-capacity optimal performance obtained with

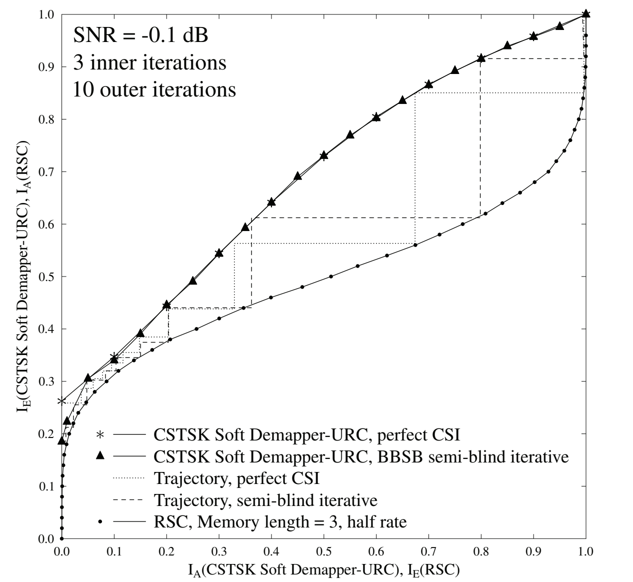

Fig. 3. EXIT chart analysis of our proposed semi-blind BBSB iterative channel estimation and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to the perfect-CSI scenario, for the $CSTSK(4, 2, 2, 4, QPSK)$ system.

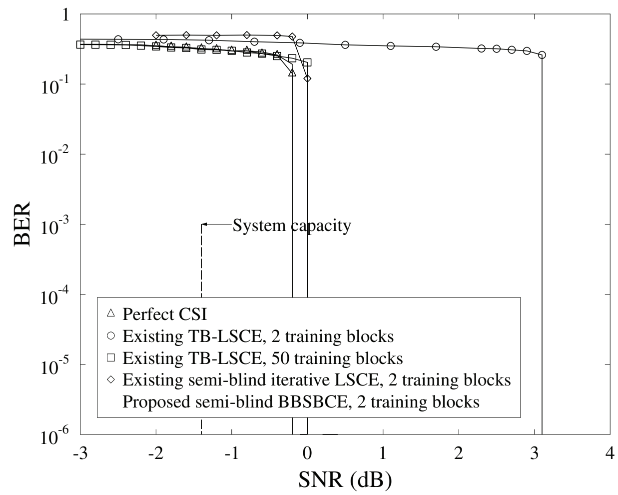

Fig. 4. BER performance of our proposed semi-blind BBSBCE and three-stage turbo demapping-decoding scheme using $I_{\text{outer}} = 10$ outer turbo iterations, $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to a) the training-based cases using $M = 2$ and 50 training blocks, respectively, b) the existing conventional joint CE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and utilising all the detected signal blocks for the decision-directed LSCE, as well as c) the perfect-CSI case, for the $CSTSK(4, 2, 2, 4, QPSK)$ system.

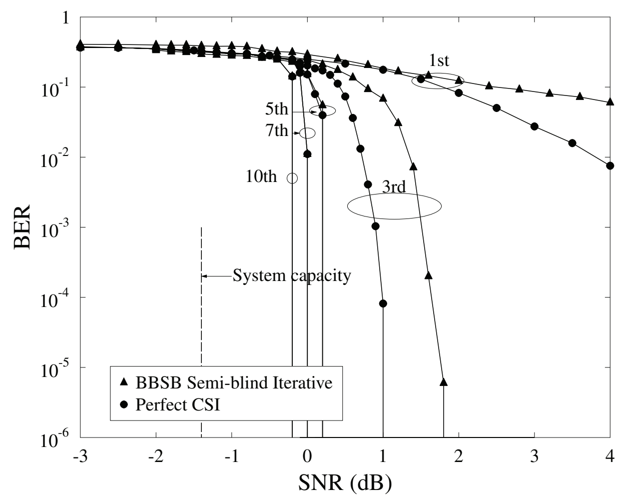

Fig. 5. BER convergence performance of our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to the perfect-CSI case, for the $CSTSK(4, 2, 2, 4, QPSK)$ system.

the aid of perfect CSI is also included as a benchmark. For the $CSTSK(4, 2, 2, 4, QPSK)$ system, there are $N_T \cdot N_R = 8$ complex-valued Rayleigh-faded channel taps. Two training blocks correspond to 8 training bits, leading to an extremely low training overhead of 1 bit per channel. Observe that our semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme associated with $M = 2$ initial training blocks approaches the perfect-CSI based performance bound, as predicted by the EXIT chart analysis of Fig. 3, while maintaining a high system throughput. The ML detection performance of the three-stage turbo demapping-decoding scheme operating with the aid of the training-based LSCE (TB-LSCE) [7] relying on 2 and 50 training blocks, respectively, is also depicted in Fig. 4. It can be seen that when using only 2 training blocks, as expected, the performance of the TB-LSCE scheme is extremely poor. Even with 50 training blocks, the TB-LSCE scheme still suffers from a 0.2 dB SNR loss in comparison to the perfect-CSI based bound. Additionally, we also designed a convectional semi-blind joint CE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks as well as utilising all the detected signal blocks for the decision-directed LSCE to represent the existing conventional approaches. The corresponding BER performance is portrayed in Fig. 4. Observe that unlike our proposed scheme, this conventional scheme cannot attain the perfect-CSI based performance despite using the entire sequence of detected data blocks for CE, albeit it imposes a substantially higher complexity than our scheme.

The convergence performance of the proposed semi-blind scheme is illustrated in Fig. 5, in comparison to that of the perfect-CSI scenario. Note that unlike all the other existing schemes, our scheme does not impose an extra iterative CE loop, since it is naturally embedded into the original three-stage serial-concatenated turbo loop. It can be seen from Fig. 5 that for both the semi-blind and perfect CSI based cases, $I_{\text{outer}} = 10$ outer iterations are sufficient for achieving near-capacity performance. In addition, it can be seen that the BER performance gap between the proposed semi-blind scheme and the optimal ML detection using the perfect CSI reduces, as the number of outer iterations increases. More specifically, during the third iteration, the BER gap is still about 0.8 dB, but at the fifth iteration, the semi-blind iterative scheme has converged to the optimal ML detection performance. In other words, the

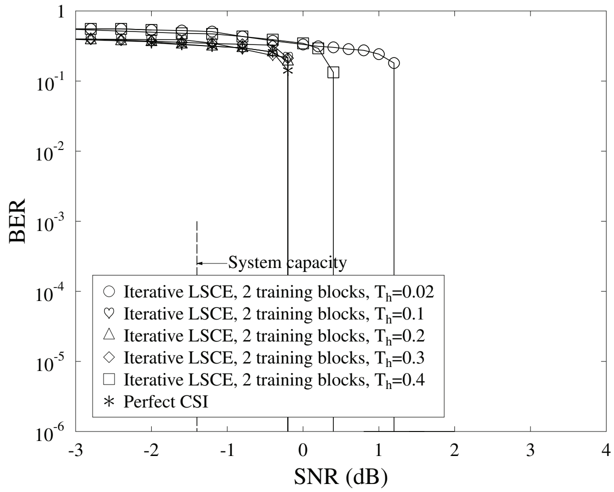

Fig. 6. Effects of the bits selection threshold T_h on the BER performance of our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and $I_{\rm outer} = 10$ outer turbo iterations, for the $CSTSK(4, 2, 2, 4, QPSK)$ system.

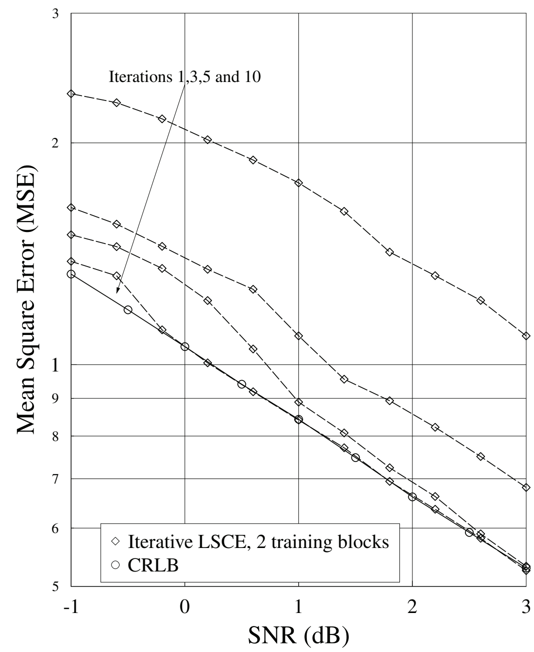

Fig. 7. MSE convergence performance of the channel estimator in our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to the CRLB, for the $CSTSK(4, 2, 2, 4, QPSK)$ system.

decision-directed channel estimate in the semi-blind BBSBCE scheme has converged to the true channel within five iterations for this $CSTSK(4, 2, 2, 4, QPSK)$ system.

The effects of the block-of-bits selection threshold T_h on the achievable BER performance were also investigated by varying the value of T_h in the set $\{0.02, 0.1, 0.2, 0.3, 0.4\}$ under the same system configuration. The corresponding results are shown in Fig. 6, where it can be seen that for $T_h = 0.1$, 0.2 and 0.3, the BER performance of the proposed semi-blind iterative scheme all converge to that of the perfect-CSI case. However, for a threshold value of $T_h = 0.02$, a performance degradation occurred, since the number of selected decision blocks for CE is probably insufficient for such a low threshold. On the other hand, given a relatively high value of $T_h = 0.4$, some unreliable decision blocks may have been selected for CE and this may lead to a performance degradation from the perfect-CSI case. The results of Fig. 6 clearly demonstrate that as long as the threshold value is not chosen to be too high or too low, the performance of the proposed semi-blind iterative scheme is not too sensitive to the actual value of T_h. Indeed, there exists a wide range of values for T_h, which allow our scheme to approach the optimal performance of the perfect-CSI case even without increasing the number of turbo iterations. For this system, $T_h \in [0.1,\ 0.3]$ are all appropriate.

The MSE performance of the CE in the proposed semi-blind iterative scheme is depicted in Fig. 7 in comparison to the CRLB. It can be seen that the MSE of the channel estimate approaches the CRLB, once the number of iterations reaches $I_{\rm outer} = 10$ for SNR≥ -0.2 dB. This corresponds to the BER cliff at SNR$= -0.2$ dB and $I_{\rm outer} = 10$ shown in Fig. 5, implying that the decision-directed CE in our scheme is most efficient for SNR≥ -0.2 dB, since it approaches the CRLB. However, it can also be seen that for SNR< -0.2 dB, the MSE of the CE was slightly degraded. This is expected because for such low SNR values, the open EXIT tunnel shown in Fig. 3 becomes closed and hence the BER remains practically to

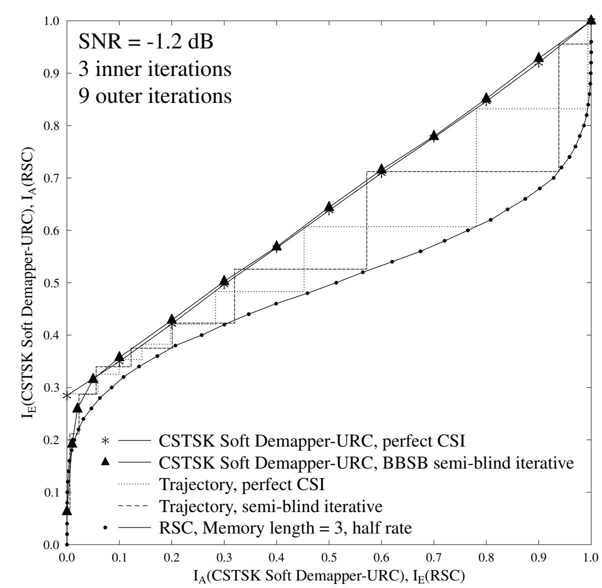

Fig. 8. EXIT chart analysis of our proposed semi-blind BBSB iterative channel estimation and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to the perfect-CSI scenario, for the $CSTSK(4, 4, 2, 4, 16QAM)$ system.

be 50% even for the perfect-CSI case. Under such adverse conditions, the decision-directed CE cannot be expected to approach the CRLB.

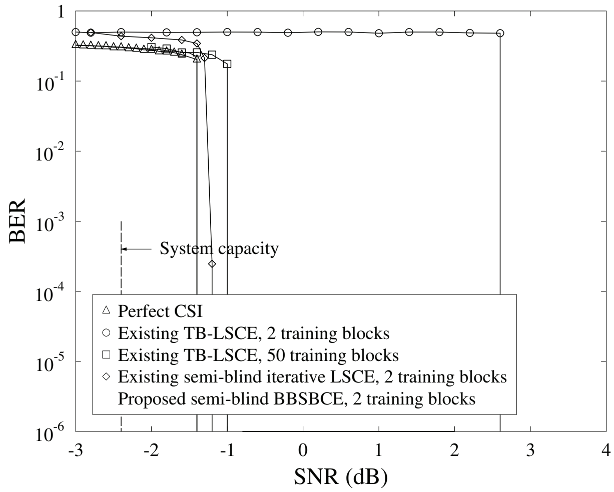

Fig. 9. BER performance of our proposed semi-blind BBSBCE and three-stage turbo demapping-decoding scheme using $I_{\text{outer}} = 9$ outer turbo iterations, $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to a) the training-based cases using $M = 2$ and 50 training blocks, respectively, b) the existing conventional joint CE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and utilising all the detected signal blocks for the decision-directed LSCE, as well as c) the perfect-CSI case, for the $CSTSK(4, 4, 2, 4, 16QAM)$ system.

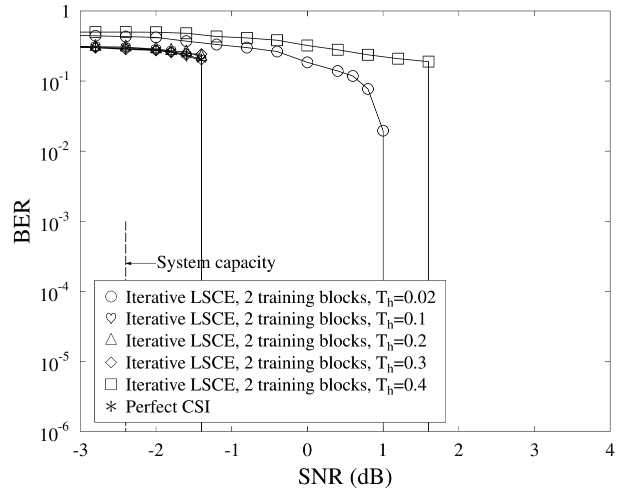

Fig. 10. Effects of the bits selection threshold T_h on the BER performance of our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and $I_{\text{outer}} = 9$ outer turbo iterations, for the $CSTSK(4, 4, 2, 4, 16QAM)$ system.

B. Example Two: $CSTSK(4, 4, 2, 4, 16QAM)$ having a throughput of 6 bits/symbol

We also considered the $CSTSK(4, 4, 2, 4, 16QAM)$ system with $\mathcal{L} = 16$. Fig. 8 depicts our EXIT chart analysis, where it can be seen that an open tunnel existed at SNR$= -1.2$ dB. The stair-case shaped decoding trajectories recorded for the proposed semi-blind joint BBSBCE scheme using $M = 2$ initial training blocks and the optimal ML detection relying on the perfect CSI are also plotted in Fig. 8 to show that for this $CSTSK(4, 4, 2, 4, 16QAM)$ system, the point of perfect convergence at (1.0, 1.0) is reached by our proposed semi-blind iterative algorithm for the same $I_{\text{outer}} = 9$ outer turbo iterations as for the perfect CSI case. Fig. 9 shows the BER performance of our joint BBSBCE and three-stage turbo detection-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$. This performance is compared to those of the three-stage turbo detector-decoder algorithm relying on the TB-LSCE based on 2 and 50 training blocks, respectively. Fig. 9 also include the performance of the existing conventional semi-blind joint CE and three-stage turbo detection-decoding scheme using $M = 2$ initial training blocks and invoking the entire detected data sequence for the decision-directed LSCE. It can be seen from Fig. 9 that the TB-LSCE scheme suffers from a performance loss of 0.4 dB in comparison to the perfect-CSI based bound even with 50 training blocks. By contrast, our semi-blind joint BBSBCE scheme using only 2 initial training blocks is capable of approaching the perfect-CSI based performance bound at the same number of turbo iterations. The results shown in Fig. 9 also confirm that the existing conventional semi-blind joint scheme suffers from a performance degradation of 0.2 dB with respect to the perfect-CSI based performance bound, despite utilising the entire detected data sequence for the DDCE, while imposing a considerably higher complexity than our proposed scheme.

Fig. 10 investigates the effects of the block-of-bits selection threshold T_h on the BER performance of our semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme. It can be seen from Fig. 10 that for this $CSTSK(4, 4, 2, 4, 16QAM)$ system, using $T_h \in [0.1,\ 0.3]$ enables our scheme using as few as $M = 2$ initial training blocks to approach the near-capacity performance of the perfect CSI using the same $I_{\text{outer}} = 9$ iterations. Finally, the MSE convergence of the CE used in our semi-blind scheme is portrayed in Fig. 11, in comparison to the CRLB for this $CSTSK(4, 4, 2, 4, 16QAM)$ system. The results of Fig. 11 confirm that the MSE of our decision-directed CE converges to the CRLB using $I_{\text{outer}} = 9$ iterations over the range of SNR≥ -1.2 dB.

VI. Conclusions

The major challenge in implementing a coherent MIMO system has been the acquisition of accurate MIMO channel estimates without sacrificing the system's throughput, while avoiding any significant increase in complexity. We have proposed a low-complexity semi-blind block-of-bits selection based joint channel estimation and three-stage data demapping-decoding scheme for near-capacity CSTSK systems. A high system throughput is maintained, since our scheme utilises an extremely low training overhead for generating an initial LSCE. Most significantly, unlike the existing methods, our proposed scheme does not require an extra iterative loop between the CE and the turbo detector-decoder, since our BBSB iterative CE is naturally embedded into the original iterative three-stage demapping-decoding turbo loop. This novel arrangement enables us to maintain a low system complexity. Furthermore, since only high-confidence decisions are selected for our DDCE, our CE approaches the CRLB. Extensive simulation results have confirmed that our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme is capable of approaching the optimal near-

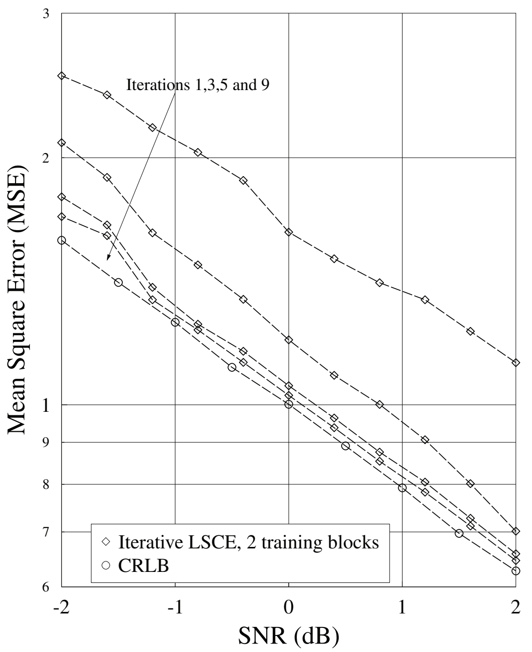

Fig. 11. MSE convergence performance of the channel estimator in our proposed semi-blind joint BBSBCE and three-stage turbo demapping-decoding scheme using $M = 2$ initial training blocks and a block-of-bits selection threshold of $T_h = 0.2$, in comparison to the CRLB, for the $CSTSK(4, 4, 2, 4, 16QAM)$ system.

capacity performance associated with the perfect CSI at the same low number of turbo iterations.

References

[1] N. Chiurtu, B. Rimoldi, and E. Telatar, "On the capacity of multi-antenna Gaussian channels," in *Proc. 2001 IEEE Int. Symp. Inf. Theory*, p. 53.

[2] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, July 2008.

[3] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3692–3703, July 2009.

[4] S. Sugiura, S. Chen, and L. Hanzo, "A universal space-time architecture for multiple-antenna aided systems," *IEEE Commun. Surveys & Tutorials*, vol. 14, no.2, pp. 401–420, 2012.

[5] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential space-time shift keying: a dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 3219–3230, Nov. 2010.

[6] S. Sugiura, S. Chen, H. Haas, P. M. Grant, and L. Hanzo, "Coherent versus non-coherent decode-and-forward relaying aided cooperative space-time shift keying," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1707–1719, June 2011.

[7] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[8] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.

[9] E. de Carcalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, Apr. 2004.

[10] L. Tang, R. W. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

[11] S. Buzzi, M. Lops, and S. Sardellitti, "Performance of iterative data detection and channel estimation for single-antenna and multiple-antennas wireless communications," *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1085–1104, July 2004.

[12] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1266–1270, Aug. 2003.

[13] A. Medles and D. T. M. Slock, "Semiblind channel estimation for MIMO spatial multiplexing systems," in *Proc. 2001 VTC – Fall*, pp. 1240–1244.

[14] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Process. Lett.*, vol. 15, pp. 202–205, 2008.

[15] H. R. Palally, S. Chen, W. Yao, and L. Hanzo, "Partical swarm optimisation aided semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," in *Proc. 2009 SSP*, pp. 309–312.

[16] S. Chen, S. Sugiura, and L. Hanzo, "Semi-blind joint channel estimation and data detection for space-time shift keying systems," *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 993–996, Dec. 2010.

[17] P. Zhang, I. Dey, S. Sugiura, and S. Chen, "Semi-blind adaptive space-time shift keying systems based on iterative channel estimation and data detection," in *Proc. 2011 VTC – Spring*, pp. 1–5.

[18] H. Li, S. M. Betz, and H. V. Poor, "Performance analysis of iterative channel estimation and multiuser detection in multipath DS-CDMA channels," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1981–1993, May 2007.

[19] B. Hu, I. Land, L. K. Rasmussen, R. Piton, and B. H. Fleury, "A divergence minimization approach to joint multiuser decoding for coded CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 432–445, Apr. 2008.

[20] B. Hu, A. Kocian, R. Piton, A. Hviid, B. H. Fleury, and L. K. Rasmussen, "Iterative joint channel estimation and successive interference cancellation using a SISO-SAGE algorithm for coded CDMA," in *Proc. 2004 Asilomar Conf. Signals, Syst. Comput.*, pp. 622–626.

[21] B. Hu, I. Land, R. Piton, and B. H. Fleury, "A Bayesian framework for iterative channel estimation and multiuser decoding in coded DS-CDMA," in *Proc. 2007 GLOBECOM*, pp. 1582–1586.

[22] M. Jiang, J. Akhtman, and L. Hanzo, "Iterative joint channel estimation and multi-user detection for multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2904–2914, Aug. 2007.

[23] P. S. Rossi and R. R. Müller, "Joint twofold-iterative channel estimation and multiuser detection for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4719–4729, Nov. 2008.

[24] J. Ylioinas and M. Juntti, "Iterative joint detection, decoding, and channel estimation in turbo-coded MIMO-OFDM," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1784–1796, May 2009.

[25] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Joint channel estimation and multi-user detection for SDMA/OFDM based on dual repeated weighted boosting search," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3265–3275, Sept. 2011.

[26] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Turbo multi-user detection for OFDM/SDMA systems relying on differential evolution aided iterative channel estimation," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1621–1633, June 2012.

[27] Y. Wu, X. Zhu, and A. K. Nandi, "Soft-input turbo channel estimation for single-carrier multiple-input-multiple-output systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3867–3873, Sept. 2009.

[28] M. Sandell, C. Luschi, P. Strauch, and R. Yan, "Iterative channel estimation using soft decision feedback," in *Proc. 1998 Globecom*, pp. 3728–3733.

[29] S. Song, A. C. Singer, and K. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2885–2894, Oct. 2004.

[30] W. Haselmayr, D. Schellander, and A. Springer, "Iterative channel estimation and turbo equalization for time-varying channels in a coded OFDM-LTE system for 16-QAM and 64-QAM," in *Proc. 2010 IEEE Int. Symp. Personal Indoor Mobile Radio Commun.*, pp. 614–618.

[31] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.

[32] T. Abe and T. Matsumoto, "Space-time turbo equalization in frequency-selective MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 3, pp. 469–475, May 2003.

[33] L. Hanzo, O. R. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation*. John Wiley & Sons, 2009.

[34] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[36] J. Zhang, L. Hanzo, and X. Mu, "Joint decision-directed channel and noise-variance estimation for MIMO OFDM/SDMA systems based on expectation-conditional maximization," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2139–2151, June 2011.

[37] A. Scherb, V. Kühn, and K. D. Kammeyer, "Cramer-Rao lower bound for semiblind channel estimation with respect to coded and uncoded finite-alphabet signals," in *Proc. 2004 Asilomar Conf. Signals, Syst., Comput.*, pp. 2193–2197.

[38] T.-H. Pham, Y.-C. Liang, and A. Nallanathan, "A joint channel estimation and data detection receiver for multiuser MIMO IFDMA systems," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1857–1865, June 2009

[39] T.-H. Pham, Y.-C. Liang, and A. Nallanathan, "Joint channel estimation and data detection for multi-input multi-output single carrier cyclic-prefix (MIMO-SCCP) systems," in *Proc. 2008 ICC*, pp. 594–598.

[40] S. Wang and A. Abdi, "Low-complexity optimal estimation of MIMO ISI channels with binary training sequences," *IEEE Signal Process. Lett.*, vol. 13, no. 11, pp.657–660, Nov. 2006.

Peichang Zhang received his BEng degree with first class honour in electronic engineering from the University of Central Lancashire, UK, in 2009 and his MSc degree with distinction in wireless communications from the University of Southampton, UK, in 2010. He is currently working towards the PhD degree with the Communications, Signal Processing and Control Research Group, Electronics and Computer Science, the University of Southampton, UK. His research interests include cooperative communications, coherent and non-coherent detection, iterative detection, as well as channel estimation.

Sheng Chen (M'1990-SM'1997-F'2008) obtained his BEng degree from the East China Petroleum Institute, Dongying, China, in January 1982, and his PhD degree from the City University, London, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with Electronics and Computer Science, the University of Southampton, UK, where he currently holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia.

Dr Chen is a Chartered Engineer (CEng) and a Fellow of IET (FIET). His recent research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 460 research papers. Dr Chen is an ISI highly cited researcher in the engineering category (March 2004).

Lajos Hanzo FREng, FIEEE, FIET, Fellow of EURASIP, DSc, received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest.

During his 35-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the Electronics and Computer Science, the University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 80 Ph.D. students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1300 research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk