REPORTS IN INFORMATICS

ISSN 0333-3590

Avoiding Cycles in Combined Turbo Decoding and Channel Estimation for Correlated Fading Channels

Eirik Rosnes and Øyvind Ytrehus

REPORT NO 241 **January 2003**

Department of Informatics

UNIVERSITY OF BERGEN

Bergen, Norway

This report has URL http://www.ii.uib.no/publikasjoner/texrap/ps/2003-241.ps

Reports in Informatics from Department of Informatics, University of Bergen, Norway, is available at http://www.ii.uib.no/publikasjoner/texrap/.

Requests for paper copies of this report can be sent to:

Department of Informatics, University of Bergen, Høyteknologisenteret,
P.O. Box 7800, N-5020 Bergen, Norway

Avoiding Cycles in Combined Turbo Decoding and Channel Estimation for Correlated Fading Channels

Eirik Rosnes and Øyvind Ytrehus*
Department of Informatics
University of Bergen
Norway
{eirik,oyvind}@ii.uib.no

Abstract

We consider combined turbo decoding and channel estimation using *pilot symbol assisted modulation* (PSAM). When transmitting information on correlated fading channels, a channel interleaver is required to break up deep channel bursts. Our primary concern in this work is not the estimation scheme, but the interleaver pair (consisting of the channel interleaver and the turbo code interleaver), and its effect on the overall system performance. If no precautions are taken, the interleaver pair will generate short cycles in the system's *graph description*. In this report, different types of cycles are described, and the expected length distribution of *primary combined cycles* in a uniform channel-coding scheme is derived analytically. We further discuss how to design interleavers that avoid, or reduce the number of (short) primary combined cycles. Simulation results indicate that these cycles to some extent degrade the overall system performance in the high signal-to-noise ratio (SNR) region. Both low and high rate turbo codes are considered in the simulation study.

1 Introduction

Turbo codes, first introduced in [1], achieve near-capacity performance with relatively low decoding complexity on a wide range of communication channels. These codes have been applied to the additive white Gaussian noise (AWGN) channel and the Rayleigh flat-fading channel with coherent detection and perfect channel state information (CSI) with great success [2]. In fading environments the assumption of perfect CSI at the receiver is not always a realistic assumption, especially in coherent detection schemes. For turbo coded systems, the receiver requires accurate estimates of both fading amplitude and phase, and noise variance.

Differential detection schemes have been proposed in the literature [3]. However, the system performance is severely degraded due to the non-coherent combining penalty. The non-coherent combining loss is reduced when using multiple-symbol differential detection schemes in connection with differential phase-shift keying (DPSK) signaling [4].

In [5], Valenti gave a thorough discussion of combined turbo decoding and fading channel estimation using pilot symbols. The concept of inserting pilot symbols at the transmitter for channel estimation and synchronization is called *pilot symbol assisted modulation* (PSAM). As demonstrated in [5], a significant performance improvement is achieved when channel estimation is performed between each decoder iteration using soft decisions from the decoder. Under the assumption of perfect knowledge of the transmitted sequence, the linear mean square (LMS) fading sequence estimator is the well-known Wiener filter [6]. The method was further improved in [7] using a Viterbi decoder combined with pilot symbols for initial estimation. A more advanced

*This research was supported by the Norwegian Research Council (NFR) Grants 137783/221 and 119390/431.

method was proposed in [8], combining the method in [5] with block-wise *maximum a posteriori* (MAP) sequence estimation.

In [9], Komninakis and Wesel proposed to incorporate the statistics of the channel into the decoder to improve the overall performance of the system. Obviously, the penalty is an increased decoding complexity. In [9], the phase process was modeled as a finite state, discrete time Markov chain. A *supertrellis* was constructed as the *product* of the trellis of the phase process with the code trellis. The trellis complexity of the supertrellis is upper bounded by the product of the component trellis complexities. In this context, fading amplitude estimation was assumed perfect. A major drawback of this approach is that a channel interleaver cannot be used. As demonstrated in [10], a channel interleaver with sufficiently large spreading factors (to combat channel correlation) will improve the performance significantly.

In [11], the issue of imperfect channel estimation was addressed in a different manner. The method in [5], using a Wiener filter, was analyzed based on the different components of fading estimation error, i.e., the amplitude and phase estimation errors were separated and analyzed individually. Furthermore, an upper bound on the optimum pilot symbol spacing was derived. In [5], the optimum pilot symbol spacing was determined by extensive simulations.

In [12], a new turbo decoding metric was derived to include the error variance of the fading estimate. The channel reliability factor was derived under the assumption of a Gaussian estimation error.

In this report we consider combined turbo decoding and channel estimation using PSAM. The channel estimation scheme used is based on Valenti's work in [5]. However, our primary concern is not the estimation scheme, but the interleaver pair (consisting of the turbo (code) interleaver and the channel interleaver), and its effect on the overall system performance verified by simulation.

The report is organized as follows: In Section 2, the system model is described. Detailed descriptions of the transmitter, channel, and receiver are given. The combined decoding and channel estimation scheme is reviewed in Section 3. In Section 4, the combined decoding and channel estimation scheme is described from a graphical point of view. Cycles in the system's *graph description* due the interleaver pair are described and classified, along with analytical computations of the expected length distribution of *primary combined cycles* in a uniform channel-coding scheme. We further discuss how to design interleavers that avoid, or reduce the number of (short) primary combined cycles. Simulations results along with a discussion are presented in Section 6, and some conclusions are drawn in Section 7.

2 System model

2.1 Transmitter

In Fig. 1, a schematic overview of the transmitter is given, indicating the basic components. A random column vector $\mathbf{d}$ of dimension m over the binary field $F = GF(2)$ is encoded by a rate r turbo encoder. The codeword $\mathbf{x}$ (of dimension m/r) is permuted or interleaved by a channel interleaver. Furthermore, pilot symbols from F are inserted into the interleaved codeword. Define two injective mappings as

$$\begin{aligned} \pi_{\text{channel}} : \mathcal{Z}_{m/r} &\rightarrow \mathcal{Z}_{m/r} \\ i &\mapsto i\pi_{\text{channel}} \end{aligned} \tag{1}$$

and

$$\begin{aligned} \pi_{\text{pilot}} : \mathcal{Z}_{m/r} &\rightarrow \mathcal{Z}_{m/r+N_p} \\ i &\mapsto i + \lfloor i/M_p \rfloor \end{aligned} \tag{2}$$

where $\mathcal{Z}_n = \{0, 1, \ldots, n-1\}$ for any integer $n \geq 1$. The pilot symbol spacing is denoted by the integer M_p, $M_p \geq 1$, and the number of pilot symbols by the integer N_p, where $N_p = \lfloor (m/r)/M_p \rfloor$.

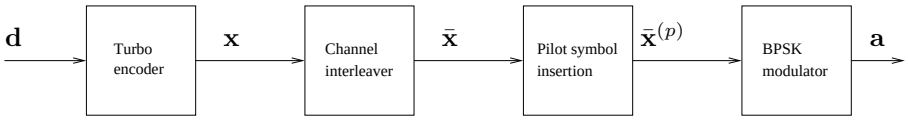

Fig. 1. *Schematic overview of transmitter*

In general, the codeword symbol in position k, $0 \leq k \leq m/r - 1$, is mapped to position $(k\pi_{\text{channel}})\pi_{\text{pilot}}$ in the *channel word* by the composition of π_{channel} and π_{pilot}. Note that in the following the composition of π_{channel} and π_{pilot} will be denoted by π_{word}. From (2), it follows that pilot symbols are inserted periodically with exactly M_p codeword symbols between each pair of consecutive pilot symbols. Further, prior to the first pilot symbol there are M_p codeword symbols.

The channel word is a column vector $\bar{\mathbf{x}}^{(p)} = (\bar{x}_0^{(p)}, \ldots, \bar{x}_{m/r+N_p-1}^{(p)})^T$ of dimension $m/r + N_p$, where the kth component ($0 \leq k \leq m/r + N_p - 1$) is given as

$$\bar{x}_k^{(p)} = \begin{cases} P & \text{if } k = lM_p + l - 1, \text{ for some integer } l \geq 1, \\ x_{(k\pi_{\text{pilot}}^{-1})\pi_{\text{channel}}^{-1}} & \text{otherwise,} \end{cases} \tag{3}$$

where π_{pilot}^{-1} and $\pi_{\text{channel}}^{-1}$ denote the inverse mappings of π_{pilot} and π_{channel}, respectively, and P is the pilot symbol from F[1]. The channel word is fed to a binary phase-shift keying (BPSK) modulator. The BPSK modulator performs the following mapping: $0 \rightarrow 1$ and $1 \rightarrow -1$ on the components of $\bar{\mathbf{x}}^{(p)}$. Note that the mapping preserves the group structure when addition is substituted with multiplication. The output from the BPSK modulator is denoted by the column vector $\mathbf{a} = (a_0, \ldots, a_{m/r+N_p-1})^T$ of dimension $m/r + N_p$, which is transformed into a continuous time signal $s(t)$ of the form

$$s(t) = \sum_{k=0}^{m/R+N_p-1} a_k g(t - kT_s), \tag{4}$$

where $g(t)$ is an arbitrary pulse shape chosen to avoid intersymbol interference (ISI) at the receiver, and T_s is the symbol interval. The energy of the transmitter pulse shape is denoted by E_s. Note that the information bit energy E_b has to include the energy of pilot symbols. Thus, E_b is calculated as

$$E_b = \frac{m/r + N_p}{m} E_s. \tag{5}$$

A channel interleaver is required on a correlated fading channel due to the bursty nature of the channel. An effective channel interleaver should be designed to break up the deep fades in the channel since turbo codes are more effective with random errors. In general, the design criteria for the channel interleaver are different from the design criteria for the turbo code interleaver. A turbo code interleaver should give a large minimum distance, and short interleaver cycles should be avoided.

2.2 Channel model

The signal $s(t)$ from the modulator is transmitted through a frequency-flat fading channel. In a frequency-flat fading environment every spectral component of the transmitted signal is affected in a similar manner. This transmission model is appropriate for narrowband systems where the transmitted signal's bandwidth is much smaller than the channel's coherence bandwidth f_c. The channel's coherence bandwidth measures the frequency range where the fading process is correlated,

[1] In a practical system pilot symbols are used for several synchronization purposes. In particular, the pilot sequence should have good *correlation properties.*

and is related to the maximum delay spread $\tau_{\max}$ by

$$f_c = \frac{1}{\tau_{\max}}. \tag{6}$$

In addition, white Gaussian noise is added to the faded signal at the receiver. The received signal $r(t)$ is given by

$$r(t) = f(t)s(t) + n(t), \tag{7}$$

where $f(t)$ is a Gaussian low-pass fading process and $n(t)$ is the additive white Gaussian noise process. In general, the fading process and the noise process are complex-valued stochastic processes due to combined amplitude and phase distortion in the transmission medium. The noise process has zero mean and two-sided power spectral density $N_0/2$ per dimension. We further assume that the fading process is statistically independent of the noise process.

In mobile communication systems, Jakes' model [13] for fading correlation yields a power spectral density of

$$S(f) = \frac{1}{\pi f_d \sqrt{1 - (\frac{f}{f_d})^2}}, \tag{8}$$

where f_d is the maximum Doppler frequency due to the relative motion between the transmitter and receiver antennas. In addition, it follows from Jakes' model that the real and imaginary parts of $f(t)$ are statistically independent.

2.3 Receiver

At the receiver end of the communication system the received signal $r(t)$ is sent through a matched-filter and then sampled at the symbol rate $1/T_s$. With perfect symbol synchronization, there is no ISI, and the equivalent discrete time transmission model is $(0 \le k \le m/r + N_p - 1)$

$$r_k = f_k a_k + n_k, \tag{9}$$

where $r_k = r(kT_s)$, $f_k = f(kT_s)$, and $n_k = n(kT_s)$. The samples n_k are independent, identical distributed complex-valued Gaussian random variables with zero mean and variance $\sigma^2 = N_0/2E_s$ per dimension. In addition, the real and imaginary parts are independent for each time index k. The received word, or column vector $\mathbf{r}$, consisting of the samples r_k, is sent to the turbo decoder.

The column vector $\mathbf{f} = (f_0, \ldots, f_{m/r+N_p-1})^T$ consists of correlated complex-valued Gaussian random variables with zero mean, where the envelope, or absolute value for each time index k, $0 \le k \le m/r + N_p - 1$, is Rayleigh distributed. Assuming Jakes' model for fading correlation, the discrete time correlation matrix $\mathbf{R}_{ff}$ becomes

$$\mathbf{R}_{ff} = \begin{pmatrix} r_{ff}[0] & r_{ff}[-1] & \ldots & r_{ff}[-m/r - N_p + 1] \\ r_{ff}[1] & r_{ff}[0] & \ldots & r_{ff}[-m/r - N_p + 2] \\ \ldots & \ldots & \ldots & \ldots \\ r_{ff}[m/r + N_p - 1] & r_{ff}[m/r + N_p - 2] & \ldots & r_{ff}[0] \end{pmatrix} \tag{10}$$

with $r_{ff}[k] = E[f_l f^*_{l-k}] = J_0(2\pi f_d T_s k)$, where $E[\cdot]$ denotes the mean of its argument, $f_d T_s$ is the normalized Doppler frequency between transmitter and receiver, and $J_0(\cdot)$ is the zeroth order Bessel function of the first kind.

The received word $\mathbf{r}$ is sent to a channel estimation algorithm, which computes an initial estimate of the fading vector $\mathbf{f}$ and the channel noise variance σ^2. See Sections 3.2.1 and 3.3.1 for details. The column vector $\tilde{\mathbf{r}} = (\tilde{r}_0, \ldots, \tilde{r}_{m/r+N_p-1})^T$, where $\tilde{r}_k = 2r_k f^*_k/\sigma^2$ for all k, $0 \le k \le m/r + N_p - 1$, can be computed from these estimates. The column vector $\Re(\tilde{\mathbf{r}})$, where $\Re(\cdot)$ denotes the real part of its complex-valued argument, is the standard channel metric. Next, the channel metrics corresponding to pilot symbols are removed using the inverse mapping π^{-1}_{pilot}, and then de-interleaved using the inverse mapping $\pi^{-1}_{\text{channel}}$. Finally, the metrics are sent to the turbo decoder, which performs a single iteration. A conventional turbo decoder that produces

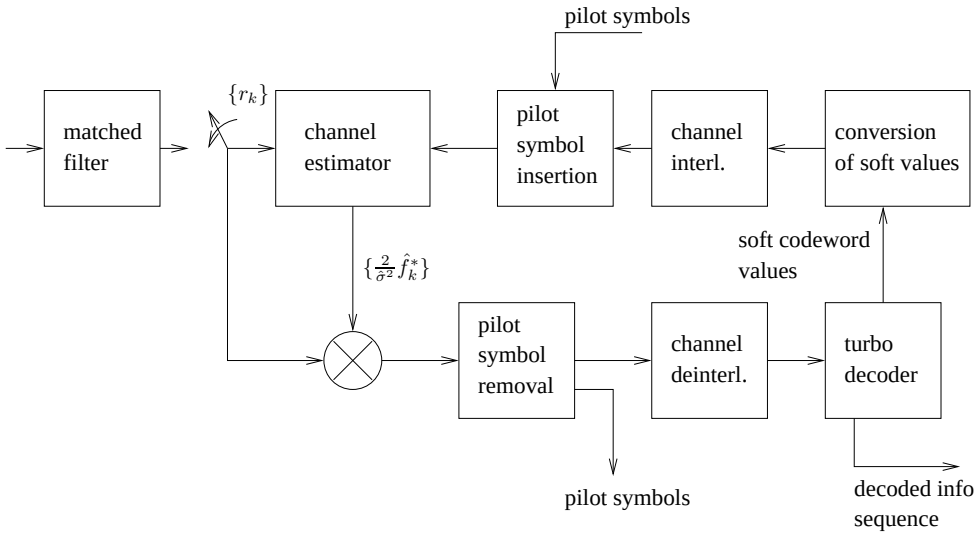

Fig. 2. *Schematic overview of receiver*

probabilistic values of information symbols, can easily be modified to produce probabilistic values of the whole codeword, from which an estimate of the codeword can be derived. See Section 3.2.2 for details. The estimate of the codeword is interleaved using π_{channel}, and pilot symbols are inserted using π_{pilot}. The resulting estimate of the channel word is again sent to a channel estimation algorithm, see Sections 3.2.2 and 3.3.2, and the procedure repeats. A schematic overview of the receiver is given in Fig. 2.

3 Combined turbo decoding and channel estimation

In this section, LMS estimation is described and applied to the problem of combined turbo decoding and channel estimation on correlated flat-fading channels.

3.1 LMS estimation

A complex-valued zero mean random column vector $\mathbf{v} = (v_0, \ldots, v_{p-1})^T$ of dimension p is given. Consider the problem of finding the best linear estimate, or prediction $\hat{\mathbf{w}}_{\text{lms}}$ of a complex-valued zero mean random column vector $\mathbf{w} = (w_0, \ldots, w_{q-1})^T$ of dimension q in the mean square error sense. In other words, we want to find the estimate that minimizes the mean square error $E[\|\tilde{\mathbf{w}}_{\text{lms}}\|^2]$, where $\tilde{\mathbf{w}}_{\text{lms}} = \mathbf{w} - \hat{\mathbf{w}}_{\text{lms}}$, and $\|\cdot\|$ denotes the Euclidean norm of its column vector argument. The solution is well-known and can be found in any introductory text in digital signal processing, e.g., [6]. The solution is

$$\hat{\mathbf{w}}_{\text{lms}} = \mathbf{C}_{\text{lms}}^H \mathbf{v}, \tag{11}$$

where $(\cdot)^H$ denotes Hermitian transpose, and

$$\mathbf{C}_{\text{lms}} = E[\mathbf{v}\mathbf{v}^H]^{-1} E[\mathbf{v}\mathbf{w}^H] = \mathbf{R}_{vv}^{-1}\mathbf{R}_{vw} \tag{12}$$

is a $p \times q$ matrix over the field of complex numbers. In (12), $\mathbf{R}_{vv} = E[\mathbf{v}\mathbf{v}^H]$ and $\mathbf{R}_{vw} = E[\mathbf{v}\mathbf{w}^H]$. The minimum mean square error is

$$E[\|\tilde{\mathbf{w}}_{\text{lms}}\|^2] = \text{tr}\left(\mathbf{R}_{ww} - \mathbf{R}_{vw}^H \mathbf{R}_{vv}^{-1}\mathbf{R}_{vw}\right), \tag{13}$$

where $\mathbf{R}_{ww} = E[\mathbf{w}\mathbf{w}^H]$, and $\text{tr}(\cdot)$ denotes the trace of its square matrix argument. The trace of a square matrix of any order is the sum of the entries along the diagonal. One important property

of the LMS estimation error $\tilde{\mathbf{w}}_{\text{lms}}$ is the orthogonality property. In general,

$$E[\tilde{\mathbf{w}}_{\text{lms}}(\mathbf{B}\mathbf{v})^H] = \mathbf{0} \tag{14}$$

for any $q \times p$ matrix $\mathbf{B}$ over the field of complex numbers. As a special case $\tilde{\mathbf{w}}_{\text{lms}}$ is orthogonal to the data $\mathbf{v}$, and to the LMS estimate $\hat{\mathbf{w}}_{\text{lms}}$. Thus, $\mathbf{w} = \tilde{\mathbf{w}}_{\text{lms}} + \hat{\mathbf{w}}_{\text{lms}}$ is an orthogonal decomposition. If $\mathbf{v}$ is taken from a *wide sense stationary* random process [6], then the correlation matrix $\mathbf{R}_{vv}$ is a Toeplitz matrix with the following structure

$$\mathbf{R}_{vv} = \begin{pmatrix} r_{vv}[0] & r_{vv}[-1] & \dots & r_{vv}[-p+1] \\ r_{vv}[1] & r_{vv}[0] & \dots & r_{vv}[-p+2] \\ \dots & \dots & \dots & \dots \\ r_{vv}[p-1] & r_{vv}[p-2] & \dots & r_{vv}[0] \end{pmatrix}, \tag{15}$$

where $r_{vv}[k] = E[v_l v_{l-k}^*]$ for all integers l and k satisfying $0 \leq l \leq p-1$ and $l-p+1 \leq k \leq l$. In general, it holds that $r_{vv}[k] = r_{vv}[-k]^*$.

3.2 Fading channel estimation

Prior to the first decoder iteration the channel word is unknown, except for the values and positions of pilot symbols. After the first decoder iteration, an estimate $\hat{\mathbf{a}}^{(q)}$, $q \geq 1$, of the BPSK modulated channel word is provided to the channel estimator. Consequently, initial estimation is considered explicitly.

3.2.1 Initial fading channel estimation

Consider the problem of initial estimation of f_k for some integer k, $0 \leq k \leq m/r + N_p - 1$. Define $y_k = r_k a_k = f_k + \tilde{n}_k$, where $\tilde{n}_k = n_k a_k$ for all k. The statistical properties of the samples $\tilde{n}_k$ are identical to the statistical properties of the samples n_k. Let $(1 \leq l \leq N_p)$

$$\mathbf{v} = \left(y_{(l-N_l^L)M_p+l-N_l^L-1}, y_{(l-N_l^L+1)M_p+l-N_l^L}, \dots, y_{(l+N_l^H)M_p+l+N_l^H-1} \right)^T, \tag{16}$$

corresponding to pilot symbol positions, and let

$$\mathbf{w} = \left(f_{L_l}, f_{L_l+1}, \dots, f_{U_l} \right)^T, \tag{17}$$

where $L_l = lM_p + l - 1 - \lfloor M_p/2 \rfloor$ and $U_l = lM_p + l - 1 + \lceil M_p/2 \rceil$. N_l^L and N_l^H are estimation specific integers satisfying

$$0 \leq N_l^L \leq \left\lfloor l - \frac{1}{M_p+1} \right\rfloor \quad \text{and} \quad 0 \leq N_l^H \leq \left\lfloor -l + \frac{m/r+N_p}{M_p+1} \right\rfloor. \tag{18}$$

In follows that

$$\mathbf{R}_{vv} = \begin{pmatrix} r_{ff}[0] + \sigma_n^2 & r_{ff}[-M_p - 1] & \dots & r_{ff}[-N_l(M_p+1)] \\ r_{ff}[M_p+1] & r_{ff}[0] + \sigma_n^2 & \dots & r_{ff}[-(N_l-1)(M_p+1)] \\ \dots & \dots & \dots & \dots \\ r_{ff}[N_l(M_p+1)] & r_{ff}[(N_l-1)(M_p+1)] & \dots & r_{ff}[0] + \sigma_n^2 \end{pmatrix} \tag{19}$$

and

$$\mathbf{R}_{vw} = \begin{pmatrix} r_{ff}[-N_l^L M_p^+ + M_p^f] & r_{ff}[-N_l^L M_p^+ + (M_p^f)^-] & \dots & r_{ff}[-N_l^L M_p^+ - M_p^c] \\ r_{ff}[-(N_l^L)^- M_p^+ + M_p^f] & r_{ff}[-(N_l^L)^- M_p^+ + (M_p^f)^-] & \dots & r_{ff}[-(N_l^L)^- M_p^+ - M_p^c] \\ \dots & \dots & \dots & \dots \\ r_{ff}[N_l^H M_p^+ + M_p^f] & r_{ff}[N_l^H M_p^+ + (M_p^f)^-] & \dots & r_{ff}[N_l^H M_p^+ - M_p^c] \end{pmatrix}, \tag{20}$$

where $N_l = N_l^L + N_l^H$, $M_p^f = \lfloor M_p/2 \rfloor$, $M_p^c = \lceil M_p/2 \rceil$, and $\sigma_n^2 = 2\sigma^2$. Above, $(\cdot)^+$ and $(\cdot)^-$ denote their integer arguments added and subtracted with 1, respectively. The optimal sets of weight coefficients are computed from (12) with $\mathbf{R}_{vv}$ and $\mathbf{R}_{vw}$ taken from (19) and (20), respectively. Note that end effects are disregarded in the derivation above. Thus, estimation on the frame boundaries should be treated separately.

3.2.2 Iterative fading channel estimation

A properly modified turbo decoder can produce probabilistic information of the channel word, which can be used to produce an estimate of $\mathbf{y}$. Let $(0 \leq l \leq m/r + N_p - 1)$

$$\mathbf{v} = \left(y_{l-N_l^L}, y_{l-N_l^L+1}, \ldots, y_{l-1}, y_{l+1}, \ldots, y_{l+N_l^H}\right)^T, \tag{21}$$

corresponding to positions around the lth symbol, where it is required that $N_l^L \leq l$ and $N_l^H \leq m/r + N_p - 1 - l$ for all l. Further, let $\mathbf{w} = (f_l)^T$. Note that y_l is excluded from the estimation procedure, in accordance with the "turbo principle" [14]. After rewriting (12), the optimal set of weight coefficients is the solution to the following system of linear equations:

$$\begin{pmatrix} r_{ff}[0]+\sigma_n^2 & r_{ff}[-1] & \ldots & r_{ff}[-N_l^L+1] & r_{ff}[-N_l^L-1] & \ldots & r_{ff}[-N_l] \\ r_{ff}[1] & r_{ff}[0]+\sigma_n^2 & \ldots & r_{ff}[-N_l^L+2] & r_{ff}[-N_l^L] & \ldots & r_{ff}[-N_l+1] \\ \ldots & \ldots & \ldots & \ldots & & & \\ r_{ff}[N_l] & r_{ff}[N_l-1] & \ldots & r_{ff}[N_l^H+1] & r_{ff}[N_l^H-1] & \ldots & r_{ff}[0]+\sigma_n^2 \end{pmatrix} \mathbf{C}_{\text{lms}} = \mathbf{R}_{vw} \tag{22}$$

with

$$\mathbf{R}_{vw} = \left(r_{ff}[-N_l^L], r_{ff}[-N_l^L+1], \ldots, r_{ff}[-1], r_{ff}[1], \ldots, r_{ff}[N_l^H]\right)^T. \tag{23}$$

The estimate of f_l after $q \geq 1$ iterations is

$$\hat{f}_l^{(q)} = \mathbf{C}_{\text{lms}}^H \begin{pmatrix} \hat{y}_{l-N_l^L}^{(q)} \\ \hat{y}_{l-N_l^L+1}^{(q)} \\ \vdots \\ \hat{y}_{l-1}^{(q)} \\ \hat{y}_{l+1}^{(q)} \\ \vdots \\ \hat{y}_{l+N_l^H}^{(q)} \end{pmatrix}, \tag{24}$$

where $\hat{y}_k^{(q)}$ is the estimate of y_k, provided by the turbo decoder, after q iterations. From (22) and (23), it follows (ignoring end effects) that $\mathbf{C}_{\text{lms}}$ reduces to a column vector independent of l if N_l^L and N_l^H are chosen as constants independent of l. Consequently, the channel estimate is the output of a time-invariant finite impulse response (FIR) filter. In digital signal processing literature this is a Wiener filter [6]. The system of equations in (22) can be solved efficiently with asymptotic complexity of order $O(N_l^2)$ due to the Toeplitz structure by the Levinson-Durbin algorithm [6].

The conventional turbo decoder calculates only *extrinsic* values of the information symbols, which after interleaving/de-interleaving are used by the other constituent decoder as *a priori* values. The constituent decoders can easily be modified to produce extrinsic values of the codeword symbols as well. These codeword extrinsic values are not required for proper decoding, but contain additional information for the channel estimator. After the turbo decoder has started its iterative decoding process, probabilistic values of codeword symbols are used by the channel estimator to obtain improved estimates of the fading process. Probabilistic values from the turbo decoder after each decoder iteration could be used in different ways. Clearly, hard decisions would produce an estimate of the transmitted codeword. This estimate could be used in (24) to obtain improved fading estimates for the next decoder iteration. We will refer to this estimation scheme as *hard decision feedback* estimation.

As proposed in [15], the BPSK modulated channel word could be estimated by its expectation conditioned on the received word. This conditional expectation is

$$E[a_k \mid \mathbf{r}] = 2P(a_k = +1 \mid \mathbf{r}) - 1 = \tanh(L(a_k \mid \mathbf{r})), \tag{25}$$

where

$$L(a_k \mid \mathbf{r}) = \log \frac{P(a_k = +1 \mid \mathbf{r})}{P(a_k = -1 \mid \mathbf{r})}. \tag{26}$$

Replacing a_k with $E[a_k \mid \mathbf{r}]$ in (24), produces the following probabilistic channel fading estimator:

$$\hat{f}_l^{(q)} = \mathbf{C}_{\text{lms}}^H \begin{pmatrix} r_{l-N_l^L} \tanh(L^{(q)}(a_{l-N_l^L} \mid \mathbf{r})) \\ r_{l-N_l^L+1} \tanh(L^{(q)}(a_{l-N_l^L+1} \mid \mathbf{r})) \\ \vdots \\ r_{l-1} \tanh(L^{(q)}(a_{l-1} \mid \mathbf{r})) \\ r_{l+1} \tanh(L^{(q)}(a_{l+1} \mid \mathbf{r})) \\ \vdots \\ r_{l+N_l^H} \tanh(L^{(q)}(a_{l+N_l^H} \mid \mathbf{r})) \end{pmatrix}. \tag{27}$$

The log-likelihood ratio values needed in (27) are provided to channel estimator by the turbo decoder. Henceforth, we denote the estimation scheme in (27) as *soft decision feedback* estimation.

3.3 Channel noise estimation

The turbo decoder requires an estimate of the channel noise variance σ_n^2. If the channel word and the fading vector are known at the receiver, compute $z_k = y_k - f_k$ for all k, $0 \le k \le m/r + N_p - 1$. The samples z_k are independent, identical distributed complex-valued Gaussian random variables with zero mean and variance σ_n^2. The minimum mean square error estimator of σ_n^2 is the sample variance defined as follows:

$$\hat{\sigma}_n^2 = \frac{1}{m/r + N_p - 1} \sum_{k=0}^{m/r+N_p-1} (z_k - \bar{z})^2, \tag{28}$$

where $\bar{z}$ is the average of the samples z_k.

3.3.1 Initial channel noise estimation

Initially, the BPSK modulated channel word $\mathbf{a}$ is known only at time indices corresponding to pilot symbols. In addition, an estimate of the fading vector $\mathbf{f}$ is provided by initial fading channel estimation. Thus, the estimator in (28) is used with time indices corresponding to pilot symbols.

3.3.2 Iterative channel noise estimation

After the first decoder iteration, an estimate (either through hard or soft decision feedback estimation) of the BPSK modulated channel word $\mathbf{a}$ is provided by the turbo decoder. In addition, an estimate of the fading vector $\mathbf{f}$ is provided by the fading channel estimator. Thus, the estimator in (28) is used to estimate the channel noise variance σ_n^2.

4 Avoiding cycles

In this section, a graphical description of the combined turbo coding and channel estimation scheme is given. We focus primarily on the interleaver pair and its contribution to short cycles in the system's graph description.

For convenience, a conventional rate 1/2 turbo code (employing a turbo code interleaver $\pi_{\text{turbo}} : \mathcal{Z}_m \to \mathcal{Z}_m$ and two constituent codes A and B) is used as an example. The code is punctured and multiplexed by the multiplexer map $\mu : \mathcal{Z}_{m_{\max}} \times \mathcal{Z}_2 \to \mathcal{Z}_{m/r}$ into an intermediate word, where $m_{\max} = \max(m_A, m_B)$, and m_A and m_B denote the lengths of constituent word A and B, respectively. The multiplexer map μ maps a pair of integers, where the first integer denotes

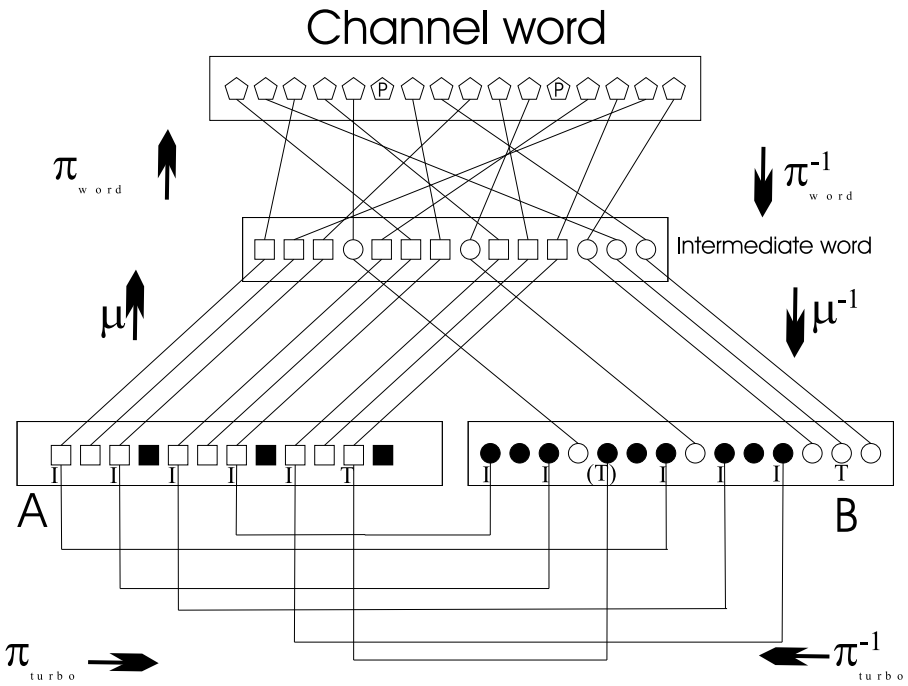

Fig. 3. *Cycles in combined turbo coding and channel estimation: Coding and interleaving overview. The information symbols (marked "I") are encoded once in codeword A, and , after interleaving with* π_{turbo}*, encoded again in codeword B. The termination symbols (marked "T") from constituent code A are included in the interleaving. The black code symbols are punctured to obtain a suitable code rate, while the remaining symbols are multiplexed by use of the multiplexer map* μ *into the intermediate word. Finally, the channel word is formed from the intermediate word by use of the channel interleaver* π_{channel} *and the pilot mapping* π_{pilot}*.*

position and the second integer denotes constituent code (constituent code A is represented by 0 and constituent code B by 1), into a position in the intermediate word. The channel interleaver π_{channel} and the pilot mapping π_{pilot} map the intermediate word together with appropriately spaced pilot symbols into the channel word, which is BPSK modulated and transmitted on a correlated fading channel. The relationship between these entities is shown in Fig. 3. We note that the intermediate word serves no practical function, other than clarifying the structure of the system. Thus, in practice, the multiplexer map μ, the channel interleaver π_{channel}, and the pilot mapping π_{pilot} could be combined.

When the channel is subjected to fading, we need to estimate the channel parameters for each received symbol. The estimation procedure can be carried out in an iterative fashion together with turbo decoding as described previously in Section 3.

Actually, in (24) it is assumed that the estimate $\hat{\mathbf{a}}^{(q)}$, $q \geq 1$, is independent of the previous estimate $\hat{\mathbf{f}}^{(p)}$, $p < q$, but this is clearly not the case in practice. Although complete independence is impossible, a proper choice of channel interleaver can remove some of the local dependency and thus potentially improve the overall system performance.

4.1 "Cycles" of length zero

First we briefly note that the correctness of the estimate $\hat{a}_l^{(q)}$, $q \geq 1$, on symbol l depends strongly on the previous channel gain estimate $\hat{f}_l^{(q-1)}$. This particular dependency cannot be removed by the interleavers. Motivated by "factor graph intuition" [14], we have removed the term y_l in (21). Indeed, in the simulations that we have carried out this appears to improve overall performance

typically by a few tenths of a dB, depending on the specific code and channel parameters.

4.2 Primary cycles

Without loss of generality, let k be a position in constituent word A of length m_A, and let j be a position in the channel word. We say (taking puncturing into account) that $(k, k+b, j+a, j)$, $a \geq 1$ and $b \geq 1$, constitutes a primary cycle of length $l = d_{\text{code}}(k, k+b) + d_{\text{word}}(j+a, j)$ if $(k,0)\mu\pi_{\text{word}} = j$ and $(k+b,0)\mu\pi_{\text{word}} = j+a$, where $0 \leq k, k+b \leq m_A - 1$ and $0 \leq j, j+a \leq m/r + N_p - 1$. In the definition above, any cyclic permutation of the entities in the 4-tuple defines an equivalent primary cycle. Note that the arguments in any practically defined *offset* functions $d_{\text{code}}(\cdot,\cdot)$ and $d_{\text{word}}(\cdot,\cdot)$ should commute[2]. We use the following definitions throughout the rest of this report:

$$d_{\text{code}}(i,j) = \left| \left\lfloor \frac{i}{n_0} \right\rfloor - \left\lfloor \frac{j}{n_0} \right\rfloor \right|, \quad 0 \leq i, j \leq m_{\max} - 1, \tag{29}$$

$$d_{\text{word}}(i,j) = |i-j|, \quad 0 \leq i, j \leq m/r + N_p - 1, \tag{30}$$

where n_0 is the number of codeword symbols in each Viterbi trellis section of the constituent codes. It is assumed that both constituent codes have the same nominal code rate. Short primary cycles will cause feedback from one iteration to the next, and should of course be avoided. If the channel interleaver is a suitable block interleaver, such primary cycles will be eliminated. An S-random interleaver [16] will also to some extent achieve this.

4.3 Secondary cycles

In a similar way, we can define a secondary cycle as a cycle describing a "number eight figure", touching the channel word twice and one or both of the constituent words twice. These cycles tend to be longer than the primary cycles and thus less serious, but should also be avoided. Note that block interleavers suffer from a large number of such secondary cycles. S-random interleavers, on the other hand, have fewer of these cycles and can even be designed to avoid them.

4.4 Primary combined cycles

Short cycles can also arise from bad connections between the turbo interleaver and the channel interleaver. Let k and l be positions in the constituent words A and B, respectively, and let j be a position in the channel word. Then we say (taking puncturing into account) that $(j = (k,0)\mu\pi_{\text{word}}, k,\ k+b,\ l = (k+b)\pi_{\text{turbo}},\ l+c,\ (l+c,1)\mu\pi_{\text{word}} = j+a)$, $a \geq 1$, b and c non-negative, and $b \geq 1$ or $c \geq 1$, constitutes a primary combined cycle of length $l = d_{\text{code}}(k, k+b) + d_{\text{code}}(l, l+c) + d_{\text{word}}(j+a, j)$, where $0 \leq k, k+b \leq m_A - 1$, $0 \leq l, l+c \leq m_B - 1$, and $0 \leq j, j+a \leq m/r + N_p - 1$. The offset functions are defined as in (29) and (30). As before, any cyclic permutation of the entities in the 6-tuple defines an equivalent primary combined cycle.

5 Channel interleaving

In this section we present analytical results on the expected cycle length distribution of primary combined cycles in an interleaver pair consisting of a uniform turbo interleaver and a uniform channel interleaver. Furthermore, we look at a special channel block interleaver construction, and interleaver pairs containing no short primary combined cycles.

[2]The practical relevance of the length l of the cycle depends on parameters of both the channel (the $d_{\text{word}}(j+a, j)$ part) and the constituent codes (the $d_{\text{code}}(k, k+b)$ part). Similar remarks can be made about the other types of cycles.

Scheme	Puncturing map	Turbo code rate	p_C
I	(1110),(0001)	1/2	3/8
II	(1001),(0110)	1/2	1/2
III	(11),(01)	1/3	4/9

Table 1. *Puncturing schemes common in turbo codes constructed from two identical, rate* 1/2 *constituent codes.*

5.1 Uniform channel-coding scheme

Consider the case of a uniform turbo code interleaver combined with a uniform channel interleaver. In general, we will assume a very long block length.

Two given (non-pilot) positions in the channel word are on a combined cycle (of some length) if they correspond to symbols from different constituent codes. The probability p_C of this event depends on the actual puncturing scheme. Table 1 gives p_C for some examples of puncturing schemes common in turbo codes constructed from two identical, rate 1/2 constituent codes.

Now let us estimate the expected number of primary combined cycles of length $l = a + b + c$, $a \geq 1$, b and c non-negative, and $b \geq 1$ or $c \geq 1$, through positions i and $i + a$ in the channel word. For short cycles, that is, $l < M_p$, the number of such cycles is given as the coefficient of the x^l term of the generating function

$$1/(m/r) \cdot p_C \cdot \frac{M_p - 1}{M_p} \cdot g_{\text{Channel}}(x) \cdot g_{\text{A}}(x) \cdot g_{\text{B}}(x), \tag{31}$$

where

$$g_{\text{Channel}}(x) = \sum_{a=1}^{\infty} x^a, \tag{32}$$

and

$$g_{\text{A}}(x) = 1 + 2\sum_{b=1}^{\infty} x^b = g_{\text{B}}(x) = 1 + 2\sum_{c=1}^{\infty} x^c. \tag{33}$$

The factor $(M_p - 1)/M_p$ follows since position i is selected as a non-pilot symbol. Thus, among the next M_p channel symbols there will be exactly one pilot symbol (which is obviously not on a combined cycle) and $M_p - 1$ non-pilot symbols. Therefore, for $a \leq M_p$, the probability of a non-pilot symbol is $(M_p - 1)/M_p$. For $a > M_p$ this probability approaches $M_p/(M_p + 1)$. For a more accurate calculation, the corresponding probability could be moved into the coefficient of the x^a term in (32), however, for practical values of M_p there is not much difference anyway.

Summing (31) over all positions in the channel word, we find that for the case of a uniform turbo interleaver and a uniform channel interleaver, the expected number of primary combined cycles of length l is given as the coefficient of the x^l term of the generating function

$$\frac{1}{r} \cdot p_C \cdot \frac{M_p - 1}{M_p} \cdot \left(\sum_{j=1}^{\infty} x^j\right) \left(1 + 2\sum_{j=1}^{\infty} x^j\right)^2. \tag{34}$$

For the example puncturing schemes I and III, with $M_p = 10$ this yields the expected (truncated) length distributions in Table 2.

5.2 Channel block interleaving

A widely used channel interleaver on fading channels is the block interleaver, which is conveniently described in terms of an $s \times t$ matrix. These interleavers are generated by a process in which the integers in $\mathcal{Z}_{\text{interleaver length}}$, in increasing order are written into the rows of matrix of memory elements (the interleaver matrix) and read out along the columns. The integers are written into

Scheme	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$
I	0.7	3.4	8.8	16.9	27.7	41.2	57.4	76.3
III	1.2	6	15.6	30	49.2	73.2	102	135.6

Table 2. *Expected (truncated) length distributions of primary combined cycles for different puncturing schemes when both interleavers are uniform. The pilot symbol spacing M_p is* 10.

the matrix row-wise from left to right and from top to bottom. In the LR/TB scheme, the columns are read from left to right and from top to bottom. Obviously, four non-equivalent schemes are possible. The following example will clarify the procedure.

Example 1 *Consider a* 3×4 *block interleaver of length 11. The interleaver matrix becomes*

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 8 & 9 & 10 & -1 \end{pmatrix}, \tag{35}$$

where the -1 *entry in the lower right corner of the matrix indicates an unused matrix cell. Using the LR/TB scheme, the following interleaver results:*

$$\pi = (0, 4, 8, 1, 5, 9, 2, 6, 10, 3, 7). \tag{36}$$

In this case the first data element is interleaved to position 0, the second data element to position 4, and so on.

As pointed out in Section 4, the number of (short) primary combined cycles from an interleaver pair should be small.

Consider the following block interleaver construction, which reduces the number of primary combined cycles significantly: Construct an interleaver matrix such that the product of the number of rows s and the number of columns t is equal to the code length. Furthermore, the number of rows should be divisible by the Hamming weight N of the puncturing map, i.e., $s = kN$ for some positive integer k. Clearly, both s and t should be sufficiently large to break up the correlation in the channel. In general, the interleaver matrix becomes

$$\begin{pmatrix} 0 & 1 & \cdots & t-1 \\ t & t+1 & \cdots & 2t-1 \\ \cdots & \cdots & \cdots & \cdots \\ (kN-1)t & (kN-1)t+1 & \cdots & kNt-1 \end{pmatrix}. \tag{37}$$

From (37), it follows that the first t symbols in the interleaved sequence are mapped from positions j satisfying $j \equiv 0 \mod N$. The next t symbols are mapped from positions j satisfying $j \equiv 1 \mod N$, and so on. The symbols in the positions in the range $[(N-1)t, Nt-1]$ are mapped from positions j satisfying $j \equiv N-1 \mod N$. Moreover, the next t symbols are mapped from positions j satisfying $j \equiv 0 \mod N$, and the pattern repeats with a period of N. Clearly, symbols of the same *type* are grouped in blocks of length t. In general, a channel block interleaver with the properties outlined above is said to be a *well-designed* channel block interleaver. If s and t do not factor the code length perfectly, then the nice structure in (37) will disappear after the first occurrence of an empty matrix cell, and the number of primary combined cycles will increase to some extent.

Note that this construction will reduce the number of (short and longer) primary combined cycles for any given turbo code interleaver π_{turbo}.

Example 2 *Consider a conventional rate* $1/2$ *turbo code with information length* 1028 *using puncturing scheme I from Table 1. If the termination scheme described schematicly in the caption of*

Cycle length	1	2	3	4	5	6	7	8	9	10	11
RCo & RCh	0.7	3.4	8.8	16.9	27.6	41.2	57.2	76.3	97.4	121.8	148.5
RCo & BCh(48x43)	0.0	0.2	0.3	0.9	1.7	2.8	4.1	6.2	8.6	11.8	16.3
RCo & BCh(45x46)	0.7	3.8	10.9	19.5	30.5	45.4	61.7	80.6	105.4	126.5	154.8

Table 3. *Average (truncated) length distributions of primary combined cycles for different interleaver pairs when puncturing scheme I in Table 1 is used. The pilot symbol spacing M_p is* 10. *Note that the average is taken over a set of randomly generated interleaver pairs in the first row, while the average is taken over a set of randomly generated turbo interleavers in the second and third rows.*

Fig. 3 is applied, the code length will be 2064. *In this termination scheme both constituent trellises are terminated, and the termination bits of encoder A are interleaved and re-encoded in encoder B. Furthermore, termination bits and the corresponding parity bits of encoder B are transmitted unpunctured. Construct a* 48×43 *channel block interleaver. Note that the Hamming weight of the puncturing map is a divisor of 48, and that* $48 \cdot 43 = 2064$. *In Table 3, experimental results for the average number of short primary combined cycles are tabulated for a random turbo code interleaver used in conjunction with a random channel interleaver, a* 48×43 *channel block interleaver, and a* 45×46 *channel block interleaver. The pilot symbol spacing M_p is* 10. *From Table 3, we observe that the* 48×43 *channel block interleaver is superior when primary combined cycles are used as the performance measure. Note that the first row in Table 3 matches almost perfectly with the first row in Table 2, which is computed using analytical tools.*

5.3 Cycle free interleaver pair

A *cycle free* interleaver pair is informally defined as an interleaver pair that do not generate *short* primary combined cycles. We propose the following design method: 1) Construct a turbo code interleaver π_{turbo} using your *favorite* interleaver algorithm [17], [18], [19], [20]. 2) Construct a channel interleaver π_{channel} conditioned on π_{turbo} using backtracking [18]. The channel interleaver should be designed in a random fashion subjected to the constraints of 1) no primary combined cycles of length less than or equal to a given value, and 2) sufficient spreading to cope with the channel correlation. Note that a joint optimization algorithm could potentially give better (from a primary combined cycles perspective) interleaver pairs.

Example 3 *We consider the conventional rate* 1/2 *turbo code from Example 2. In Table 4 cycle length distributions of primary combined cycles are tabulated. The first row in the table shows the (truncated) cycle length distribution of a cycle free interleaver pair, which is designed to avoid primary combined cycles of length shorter than* 7. *The interleaver pair is denoted by "Cycle free int." in Table 4. The second row shows the (truncated) cycle length distribution of an interleaver pair consisting of the turbo interleaver from the first row in conjunction with a* 48×43 *channel block interleaver. Finally, in the last row, the corresponding (truncated) cycle length distribution of an interleaver pair consisting of the turbo interleaver from the first row in conjunction with a* 45×46 *channel block interleaver is tabulated. The corresponding cycle length distributions of primary cycles and secondary cycles are tabulated in Tables 5 and 6 for the same interleaver pairs, respectively. The turbo code interleaver is an S-random interleaver.*

Example 4 *Consider a conventional rate* 1/3 *turbo code with information length* 192 *using puncturing scheme III from Table 1. If the termination scheme described in the caption of Fig. 3 is applied, the code length will be* 586. *In Table 7 cycle length distributions of primary combined cycles are tabulated. The first row in the table shows the (truncated) cycle length distribution of a cycle free interleaver pair, which is designed to avoid primary combined cycles of length shorter than* 5. *The second row shows the (truncated) cycle length distribution of an interleaver pair consisting of*

Cycle length	1	2	3	4	5	6	7	8	9	10	11
Cycle free int.	0	0	0	0	0	0	74	93	100	123	161
SCo & BCh(48x43)	0	1	0	1	2	4	2	2	12	5	17
SCo & BCh(45x46)	2	5	4	18	34	51	73	80	95	106	151

Table 4. *Truncated length distributions of primary combined cycles for different interleaver pairs when puncturing scheme I in Table 1 is used. The pilot symbol spacing* M_p *is* 16.

Cycle length	18	19	20	21	22	23	24	25	26	27	28
Cycle free int.	47	55	42	42	59	62	50	56	61	57	67
SCo & BCh(48x43)	0	0	0	0	0	0	0	1887	126	0	0
SCo & BCh(45x46)	0	0	0	0	0	629	460	28	0	0	0

Table 5. *Truncated length distributions of primary cycles in both constituent codes for different interleaver pairs when puncturing scheme I in Table 1 is used. The pilot symbol spacing* M_p *is* 16.

the turbo interleaver from the first row in conjunction with a 21×28 *channel block interleaver. Note that the Hamming weight of the puncturing map is a divisor of 21, and that* $21 \cdot 28 = 588$*, which is almost the code length, from which it follows that we should expect a reduced number of primary combined cycles. Finally, in the last row, the corresponding (truncated) cycle length distribution of an interleaver pair consisting of the turbo interleaver from the first row in conjunction with a* 24×25 *channel block interleaver is tabulated. The corresponding cycle length distributions of primary cycles and secondary cycles are tabulated in Tables 8 and 9 for the same interleaver pairs, respectively. The turbo code interleaver is an S-random interleaver.*

6 Simulation results

In this section simulation results are presented to compare the performance when using different interleaver pairs. The comparison is based upon simulated frame error rate (FER) performance of both low and high rate turbo codes on a correlated Rayleigh flat-fading channel with coherent detection. The channel is correlated according to Jakes' model [13].

6.1 Low rate case

A conventional turbo code using two identical, rate 1/2, recursive systematic convolutional encoders with feedback polynomial 7 and feedforward polynomial 5 (octal notation) is used. The termination scheme described schematicly in the caption of Fig. 3 is used. In more detail, both constituent trellises are terminated, and the termination bits of encoder A are interleaved and re-encoded in encoder B. Furthermore, termination bits and the corresponding parity bits of encoder B are transmitted unpunctured. Two different schemes are simulated. In scheme I, a rate 1/2 code is obtained using puncturing scheme I in Table 1. We use an information length of 1028, which

Cycle length	2	3	4	5	6
Cycle free int.	0	0	1	6	53
SCo & BCh(48x43)	441	76	2124	412	4019
SCo & BCh(45x46)	50	410	279	468	2298

Table 6. *Truncated length distributions of secondary cycles in both constituent codes for different interleaver pairs when puncturing scheme I in Table 1 is used. The pilot symbol spacing* M_p *is* 16.

Cycle length	2	3	4	5	6	7	8	9	10
Cycle free int.	0	0	0	12	85	115	167	201	266
SCo & BCh(21x28)	0	3	3	4	12	15	19	36	44
SCo & BCh(24x25)	3	13	24	36	36	89	107	122	144

Table 7. *Truncated length distributions of primary combined cycles for different interleaver pairs when puncturing scheme III in Table 1 is used. The pilot symbol spacing M_p is* 16.

Cycle length	8	9	10	11	12	13	14	15	16	17	18
Cycle free int.	0	0	0	0	16	46	64	51	49	54	68
SCo & BCh(21x28)	520	31	0	0	0	0	0	0	468	62	0
SCo & BCh(24x25)	0	340	22	0	0	0	0	0	0	79	230

Table 8. *Truncated length distributions of primary cycles in both constituent codes for different interleaver pairs when puncturing scheme III in Table 1 is used. The pilot symbol spacing M_p is* 16.

results in a code length of 2064. In scheme II, a rate 1/3 turbo code is obtained using puncturing scheme III in Table 1. We use an information length of 192, which results in a code length of 586.

Pilot symbols are inserted into the intermediate word in both schemes when non-perfect CSI is assumed. The pilot symbol spacing M_p is 16. Iterative soft decision feedback estimation with $N_l^L = N_l^H = 30$ for all l, $0 \leq l \leq m/r + N_p - 1$, is implemented. On the frame boundary, decisions from the previous frame are used to perform *interframe* estimation [5]. Iterative MAP decoding using a maximum of 20 iterations of the log-MAP algorithm is used. To ensure statistical significance, the simulations continued until 100 independent frame errors were generated for each value of E_b/N_0, unless otherwise is pointed out. The interleaver pairs, whose cycle length distributions are discussed in Examples 3 and 4, are used in the simulations. Performance graphs are given in Figs. 4 and 5 for schemes I and II, respectively.

6.2 High rate case

Consider a rate 8/9, $\nu = 4$ convolutional code $\mathcal{C}$ defined by the parity check matrix

$$\mathbf{H}(D) = \begin{pmatrix} 17 & 21 & 23 & 25 & 27 & 33 & 35 & 37 & 31 \end{pmatrix}, \tag{38}$$

where the entries are given in octal notation in the sense that $23 = 10011 = 1 + D + D^4$. A rate 4/5 turbo code, using as constituent codes the code defined in (38), is constructed. The systematic bits from encoder B are punctured to obtain the overall code rate of 4/5. The information length is 1280, and both constituent encoders are terminated. The UMTS termination scheme [21] is used, i.e., both parity and systematic bits of both encoders are transmitted, which results in a code length of 1618.

Pilot symbols are inserted into the intermediate word when non-perfect CSI is assumed. The pilot symbol spacing M_p is 16. Iterative (interframe) estimation with $N_l^L = N_l^H = 30$ for all l,

Cycle length	2	3	4	4	6
Cycle free int.	0	5	20	64	126
SCo & BCh(21x28)	163	22	805	115	1349
SCo & BCh(24x25)	78	13	549	97	983

Table 9. *Truncated length distributions of secondary cycles in both constituent codes for different interleaver pairs when puncturing scheme III in Table 1 is used. The pilot symbol spacing M_p is* 16.

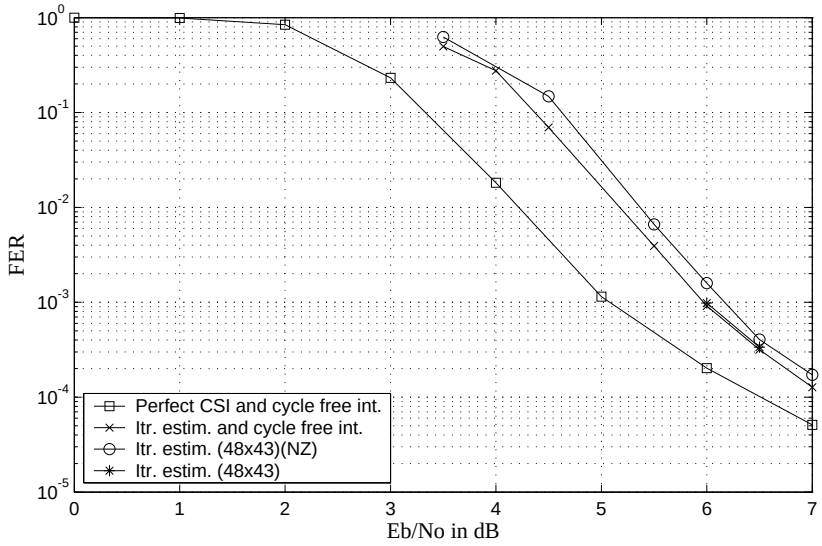

Fig. 4. *FER performance versus E_b/N_0 for different interleaver pairs on a correlated Rayleigh flat-fading channel with normalized Doppler frequency $f_dT_s = 0.02$. Scheme I with an information length of 1028 is simulated. See Example 3 for a discussion of the interleaver pairs. Note that (NZ) is an abbreviation for nonzero center tap in the estimation filters, i.e., "cycles" of length zero are not avoided.*

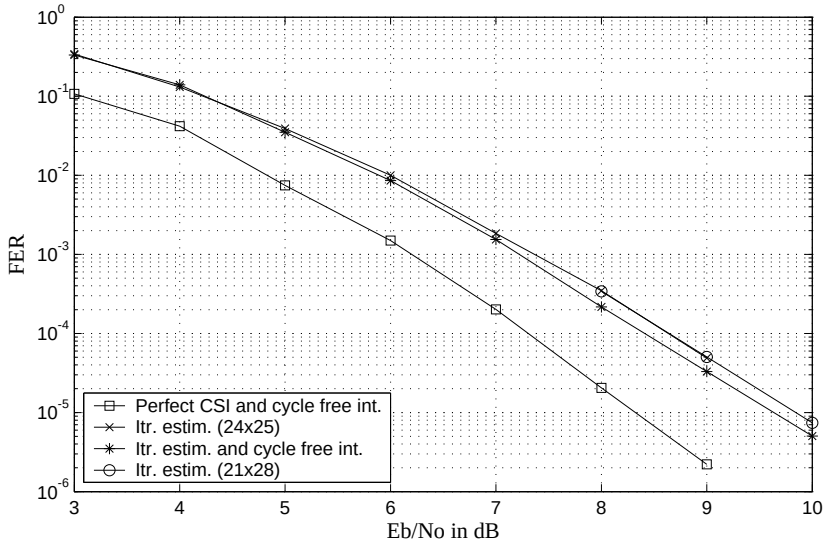

Fig. 5. *FER performance versus E_b/N_0 for different interleaver pairs on a correlated Rayleigh flat-fading channel with normalized Doppler frequency $f_dT_s = 0.01$. Scheme II with an information length of 192 is simulated. See Example 4 for a discussion of the interleaver pairs. Under the assumption of perfect CSI for $E_b/N_0 = 9$ dB, 58 independent frame errors were simulated.*

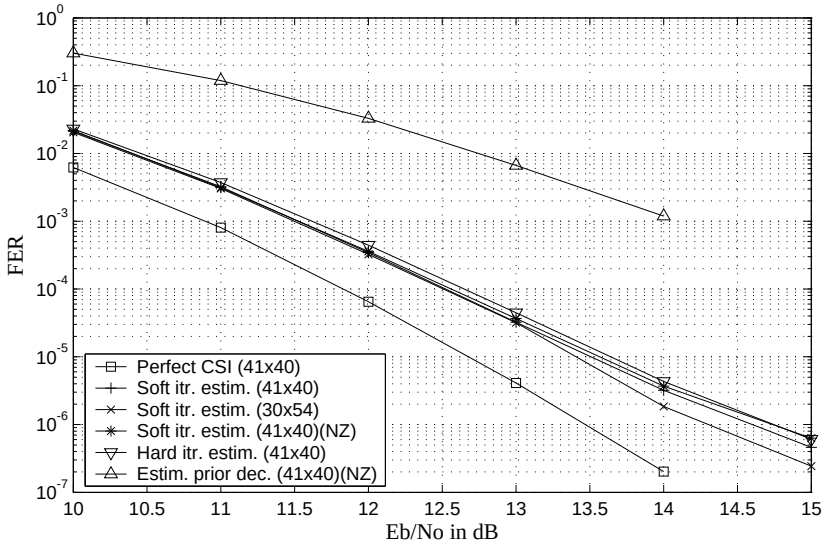

Fig. 6. *FER performance versus E_b/N_0 for different interleaver pairs on a correlated Rayleigh flat-fading channel with normalized Doppler frequency $f_dT_s = 0.01$. The information length is 1280. Note that (NZ) is an abbreviation for nonzero center tap in the estimation filters, i.e., "cycles" of length zero are not avoided. Under the assumption of perfect CSI, 100 independent frame errors were simulated. Furthermore, at $E_b/N_0 = 14$ dB, 12, 100, 86, 100, 100, and 400 frame errors were observed, respectively for each of the curves in the figure (in the order of the legend). At $E_b/N_0 = 15$ dB, 124, 61, 67, and 67 frame errors were observed.*

$0 \leq l \leq m/r + N_p - 1$, is implemented, and iterative MAP decoding using the dual code [22] with a maximum of 12 iterations is used. A short introduction to MAP decoding using the dual code is given in the Appendix. Performance graphs are given Fig. 6. Note that the simulations continued until 400 independent frame errors were generated for each value of E_b/N_0, unless otherwise is pointed out.

6.3 Discussion

Simulation results for scheme I are presented in Fig. 4. We observe a performance improvement of a few tenths of a dB, depending on the actual value of E_b/N_0, when "cycles" of length zero are avoided. Note that these "cycles" of length zero are not discussed at all in [5].

Simulation results for scheme II are presented in Fig. 5. A slight improvement in performance is demonstrated when using a cycle free interleaver pair compared to an interleaver pair with a well-designed channel block interleaver. A 21×28 channel block interleaver is a well-designed channel block interleaver for the specific puncturing scheme employed, see Example 4 for details. We believe that the large number of secondary cycles present in any block interleaver could be the reason.

For the high rate case, simulations results are presented in Fig. 6, where we have compared soft and hard decision feedback estimation. Furthermore, a 30×54 channel block interleaver, which is a well-designed channel block interleaver for the specific puncturing scheme employed (10 is a divisor of 30 and $30 \cdot 54 = 1620 \approx 1618$), is compared to a 41×40 channel block interleaver. At high SNR a performance improvement is demonstrated when a well-designed channel block interleaver is used.

7 Conclusions

In this report combined turbo decoding and channel estimation using PSAM has been reviewed. Cycles in the system's graph description have been classified. In particular, primary combined cycles have been discussed, along with analytical computations of the expected length distribution of these cycles in a uniform channel-coding scheme. The analytical derivations have been compared with experimental results. Motivated by "factor graph intuition", these primary combined cycles were identified as a potential cause for performance degradation. A specific channel block interleaver construction that reduces the number of (short and longer) primary combined cycles significantly was proposed. Furthermore, a simple algorithm to design an interleaver pair that generates no short primary combined cycles was outlined.

The report ends with a presentation of various simulation results of both low and high rate turbo codes on a correlated Rayleigh flat-fading channel. In particular, for the high rate case, performance improvements in the high SNR region were demonstrated when primary combined cycles were reduced in number.

In general, a more extensive simulation study should be conducted to determine to what extent primary combined cycles influence the performance.

Appendix

In this appendix a brief introduction to MAP decoding using the dual code is given.

Let $\mathcal{C}$ denote an (n, k) systematic linear block code over $GF(2)$ with elements $\{0, 1\}$. The codeword $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in \mathcal{C}$ is binary modulated and transmitted on a binary input memoryless channel. The modulator performs the following mapping: $0 \rightarrow 1$ and $1 \rightarrow -1$. The output of the channel is denoted by $\mathbf{y}$.

Hartmann and Rudolph [23] and Battail *et al.* [24] derived expressions for the a posteriori probabilities $P(c_l = 0, 1 \mid \mathbf{y})$ using the codewords of the dual code. Both papers presented the formula for equally likely information symbols. This work was extended by Hagenauer *et al.* [25] to provide formulas for including a priori probabilities. Recently, Riedel [22], [26] applied these ideas to provide low complexity iterative MAP decoding algorithms for high rate concatenated codes.

The log-likelihood ratio

$$L(\hat{c}_l) = \ln \frac{P(c_l = 0 \mid \mathbf{y})}{P(c_l = 1 \mid \mathbf{y})} \tag{39}$$

of the lth codeword symbol c_l can be computed using the original code $\mathcal{C}$ as follows [25]:

$$L(\hat{c}_l) = L(c_l; y_l) + \ln \frac{\sum_{\mathbf{c}_i \in \mathcal{C},\, c_{il}=0} \prod_{j=0, j\neq l}^{n-1} e^{-L(c_j; y_j) c_{ij}}}{\sum_{\mathbf{c}_i \in \mathcal{C},\, c_{il}=1} \prod_{j=0, j\neq l}^{n-1} e^{-L(c_j; y_j) c_{ij}}}, \tag{40}$$

where

$$L(c_l; y_l) = \begin{cases} L(y_l \mid c_l) + L(c_l) & \text{if } c_l \text{ is an information symbol,} \\ L(y_l \mid c_l) & \text{if } c_l \text{ is a parity check symbol.} \end{cases}$$

Above, $L(y_l \mid c_l)$ and $L(c_l)$ are the channel metric and the a priori knowledge, both in log-likelihood ratio form, of the lth codeword symbol, respectively. Alternatively, $L(\hat{c}_l)$ can also be computed using all codewords $\mathbf{c}_i^{\perp}$, $1 \leq i \leq 2^{n-k}$, of the dual code $\mathcal{C}^{\perp}$ of $\mathcal{C}$ as follows [25]:

$$L(\hat{c}_l) = L(c_l; y_l) + \ln \frac{\sum_{\mathbf{c}_i^{\perp} \in \mathcal{C}^{\perp}} \prod_{j=0, j\neq l}^{n-1} \tanh(L(c_j; y_j)/2)^{c_{ij}^{\perp}}}{\sum_{\mathbf{c}_i^{\perp} \in \mathcal{C}^{\perp}} (-1)^{c_{il}^{\perp}} \prod_{j=0, j\neq l}^{n-1} \tanh(L(c_j; y_j)/2)^{c_{ij}^{\perp}}}. \tag{41}$$

Assuming a trellis description of the dual code, a BCJR [27] type of algorithm can be used to compute the log-likelihood ratio in (41).

References

[1] C. Berrou, A. Glavieux, and P. Thitimasjshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[2] B. Vucetic and J. Yuan, *Turbo Codes: Principles and Applications*. Kluwer, 2000.

[3] E. K. Hall and S. G. Wilson, "Turbo codes for noncoherent channels," in *Proc. IEEE GLOBECOM*, (Phoenix, AZ), pp. 66–70, Nov. 1997.

[4] P. Hoeher and J. Lodge, "Turbo DPSK: Iterative differential PSK demodulation and channel decoding," *IEEE Trans. Commun.*, vol. 47, pp. 837–843, June 1999.

[5] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1697–1705, Sept. 2001.

[6] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. McGraw-Hill, 2000.

[7] B. Mielczarek and A. Svensson, "Improved iterative channel estimation and turbo decoding over flat-fading channels," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, (Vancouver, Canada), pp. 975–980, Sept. 2002.

[8] Q. Li, C. N. Georghiades, and X. Wang, "An iterative receiver for turbo-coded pilot-assisted modulation in fading channels," *IEEE Commun. Lett.*, vol. 35, Apr. 2001.

[9] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1706–1717, Sept. 2001.

[10] E. K. Hall and S. G. Wilson, "Design and analysis of turbo codes on Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 160–174, Feb. 1998.

[11] B. Mielczarek and A. Svensson, "Modelling the influence of fading channel estimation errors on turbo code performance." *IEEE Trans. Commun.*, submitted for publication.

[12] P. Frenger, "Turbo decoding for wireless systems with imperfect channel estimates," *IEEE Trans. Commun.*, vol. 48, pp. 1437–1440, Sept. 2000.

[13] W. C. Jakes, *Mobile Microwave Communication*. John Wiley & Sons, Inc., 1974.

[14] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[15] M. Sandell, C. Luschi, P. Strauch, and R. Yan, "Iterative channel estimation using soft decision feedback," in *Proc. IEEE GLOBECOM*, (Sydney, Australia), pp. 3728–3733, Nov. 1998.

[16] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations." TDA Progress Report 42-121, JPL, Pasadena, CA, Aug. 1995.

[17] S. N. Crozier, "New high-spread high-distance interleavers for turbo-codes." Preprint, 2001.

[18] M. Breiling, S. Peeters, and J. Huber, "Interleaver design using backtracking and spreading methods," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, (Sorrento, Italy), p. 451, June 2000.

[19] J. Hokfelt, O. Edfors, and T. Maseng, "Turbo codes: Correlated extrinsic information and its impact on iterative decoding performance," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 3, (Houston, Texas), pp. 1871–1875, May 1999.

[20] K. S. Andrews, C. Heegard, and D. Kozen, "Interleaver design methods for turbo codes," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, (Boston, MA), p. 420, Aug. 1998.

[21] 3rd Generation Partnership Project, *Multiplexing and channel coding (FDD)*, June 1999. 3G TS 25.212.

[22] S. Riedel, "MAP decoding of convolutional codes using reciprocal dual codes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1176–1187, May 1998.

[23] C. R. Hartmann and L. D. Rudolph, "An optimum symbol-by-symbol decoding rule for linear codes," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 514–517, Sept. 1976.

[24] G. Battail, M. C. Decouvelaere, and P. Godlewski, "Replication decoding," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 332–345, May 1979.

[25] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 429–445, Mar. 1996.

[26] S. Riedel, "Symbol-by-symbol MAP decoding algorithm for high-rate convolutional codes that use reciprocal dual codes," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 175–185, Feb. 1998.

[27] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.