Iterative Channel Estimation, Detection, and Decoding in Large CDMA Systems

Mikko Vehkaperä*, Keigo Takeuchi†, Ralf R. Müller* and Toshiyuki Tanaka‡
*Dept. Electronics & Telecommunications, Norwegian University of Science & Technology, NO-7491 Trondheim, Norway.
†Dept. Communications Engineering & Informatics, University of Electro-Communications, Tokyo 182-8585, Japan.
‡Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.
E-mail: {mikko, ralf}@iet.ntnu.no, takeuchi@ice.uec.ac.jp, tt@i.kyoto-u.ac.jp

***Abstract*—The large system analysis of randomly spread direct-sequence code-division multiple-access systems operating over frequency-selective fading channels is considered. Iterative multiuser detection and decoding (MUDD) based on generalized posterior mean estimation and single-user sum-product decoding is assumed to be used at the receiver. The channel state information (CSI) at the MUDD is mismatched and obtained by a linear channel estimator whose initial decisions are iteratively refined with the help of information feedback provided by the MUDD. Furthermore, a new training method by means of probability-biased signaling is proposed. The results indicate that in the large system limit and under certain threshold loads, a near single-user performance with perfect CSI can be achieved using a vanishing training overhead. It is also found that for the considered setups, the iterative linear minimum mean square error based channel estimator is near optimal for relatively slowly time-varying multipath fading channels.**

I. Introduction

The factor-graph [1], [2] based iterative multiuser detection and decoding (MUDD) framework [3]–[5] provides a practical method for joint multiuser decoding in code-division multiple-access (CDMA) systems. So far, analytical analyses have either assumed perfect channel state information (CSI) at the receiver [4], [5], or considered non-iterative channel estimation (CE) [6], [7] with highly idealized assumptions about the system model. Here, we extend the scope of earlier works and consider the large system analysis of iterative receivers that use the extrinsic information fedback from single user error control decoders to refine the initial channel and data estimates [8], [9][1]. Furthermore, a new training method by means of probability biased signaling [14] is proposed.

Instead of pre-defining the channel and data estimators explicitly before the analysis, the building blocks of our iterative receiver are represented by generalized posterior mean estimators (GPMEs) [15], [16]. The class of estimators described by the latter comprises, e.g., linear and non-linear minimum mean square error (MMSE) estimators, the decorrelator, and the single-user matched filter (SUMF), with parallel interference cancellation (PIC), as special cases. By virtue of this general Bayesian approach, in addition to a wide variety of estimation algorithms, we are also able to analyze schemes based on both hard and soft feedback under a single unified framework.

To assess the performance of iterative receivers, we track the feedback statistics using density evolution with a Gaussian approximation (GA-DE) [4], [5], [17], and obtain the input-output statistics of the GPMEs in the large system limit via the replica method [18], that has been recently applied with great success to a multitude of problems in telecommunications (see for example [6], [7], [15], [16], [19]–[23] and the discussion therein). Due to space limitations, however, the decoupling results [15], [16] as well as all proofs and derivations are omitted in this paper (see [8], [9], [14]).

[1]Although several studies have shown via numerical simulations that iterative channel and data estimation can reduce the pilot overhead and improve the reliability of the CSI significantly (see for example [10]–[12]), the only effort to analytically analyze the performance of such a receiver is to our knowledge the approximate study carried out in [13].

II. Preliminaries

A. Notation

Calligraphic and boldface symbols denote sets and vectors (matrices), respectively. The vectorization of a matrix $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_N\end{bmatrix} \in \mathbb{C}^{M\times N}$ is given by $\mathrm{vec}(\boldsymbol{A}) = \begin{bmatrix}\boldsymbol{a}_1^\mathsf{T} & \boldsymbol{a}_2^\mathsf{T} & \cdots & \boldsymbol{a}_N^\mathsf{T}\end{bmatrix}^\mathsf{T} \in \mathbb{C}^{MN}$. For $\boldsymbol{a} \in \mathbb{C}^M$ we let $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{a}) \in \mathbb{C}^{M\times M}$ be a diagonal matrix defined by the vector $\boldsymbol{a}$. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ return the real and imaginary part of the argument, respectively.

Throughout this paper, we write $\boldsymbol{x} \sim \mathbb{P}$ and $\tilde{\boldsymbol{x}} \sim \mathbb{Q}$ for a random vector (RV) drawn according to the true ($\mathbb{P}$) and postulated ($\mathbb{Q}$) probability distribution (or measure), respectively. The postulated distributions represent the receiver's (possibly mismatched) information about the random variables in the system. If the distribution of the RV $\boldsymbol{x}$ depends on the iteration index $\ell = 1, 2, \ldots$ we write $\mathbb{P}^{(\ell)}(\boldsymbol{x})$. We omit the index ℓ otherwise. The mean and covariance of $\mathbb{P}^{(\ell)}(\boldsymbol{x})$ are $\boldsymbol{\mu}_{\boldsymbol{x}}^{(\ell)}$ and $\boldsymbol{\Omega}_{\boldsymbol{x}}^{(\ell)}$, and the corresponding mean and covariance of the postulated RV $\tilde{\boldsymbol{x}}$ read $\tilde{\boldsymbol{\mu}}_{\boldsymbol{x}}^{(\ell)}$ and $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{x}}^{(\ell)}$. The posterior mean estimate of the RV $\boldsymbol{x}$ with respect to the postulated distribution $\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{x}}|\cdots)$ is denoted by $\langle\tilde{\boldsymbol{x}}\rangle_{(\ell)}$ and the related error covariance matrix by $\boldsymbol{\Omega}_{\Delta\boldsymbol{x}}^{(\ell)}$, unless stated otherwise. If $\boldsymbol{x}$ is a proper complex Gaussian RV [24] with mean $\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\}$ and covariance $\boldsymbol{\Omega}_{\boldsymbol{x}} = \mathbb{E}\{(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})^\mathsf{H}\}$, we write in shorthand $\boldsymbol{x} \sim \mathsf{CN}(\boldsymbol{\mu}_{\boldsymbol{x}}; \boldsymbol{\Omega}_{\boldsymbol{x}})$ or $\mathbb{P}(\boldsymbol{x}) = \mathsf{CN}(\boldsymbol{\mu}_{\boldsymbol{x}}; \boldsymbol{\Omega}_{\boldsymbol{x}})$.

B. Randomly Spread Uplink CDMA in Multipath Fading

Consider a synchronous uplink (reverse link) DS-CDMA system, operating over a block fading [25] multipath channel

978-1-4244-6746-4/10/$26.00 ©2010 IEEE

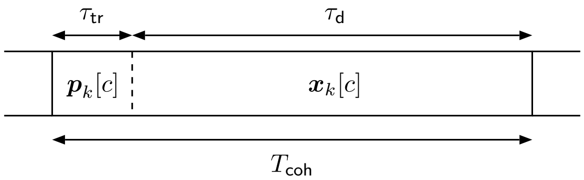

Figure 1. Frame structure of the considered system.

with a coherence time of T_{coh} symbols. We make the simplifying assumption that the inter-symbol interference (ISI) induced by the multipath fading has negligible effect on the system performance and omit it from the analysis.

The discrete time signal model after matched filtering and chip-rate sampling for the tth received vector $\boldsymbol{y}_t[c] \in \mathbb{C}^L$ within fading block $c = 1, \ldots, C$ can be written as

$$\boldsymbol{y}_t[c] = \begin{cases} \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \boldsymbol{h}_k[c] p_{k,t}[c] + \boldsymbol{w}_t[c] \in \mathbb{C}^L, & t \in \mathcal{T}, \\ \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \boldsymbol{h}_k[c] x_{k,t}[c] + \boldsymbol{w}_t[c] \in \mathbb{C}^L, & t \in \mathcal{D}, \end{cases} \tag{1}$$

where $\mathcal{T} = \{1, \ldots, \tau_{\mathsf{tr}}\}$ and $\mathcal{D} = \{\tau_{\mathsf{tr}} + 1, \ldots, T_{\mathsf{coh}}\}$ contain the time indices related to the training and data transmission phases when the $\tau_{\mathsf{tr}} = |\mathcal{T}|$ training symbols $\{p_{k,t}[c]\}_{t \in \mathcal{T}}$ and the $\tau_{\mathsf{d}} = |\mathcal{D}|$ information symbols $\{x_{k,t}[c]\}_{t \in \mathcal{D}}$ of user $k \in \mathcal{K} = \{1, \ldots, K\}$ are transmitted, respectively. The frame structure of this transmission scheme is illustrated in Fig. 1. The spreading matrix for the kth user at time index t is given by $\boldsymbol{S}_{k,t} \in \mathbb{C}^{L \times M}$. For notational simplicity, we let the number of multipaths M and the spreading factor L be identical for all users, since generalizations are straightforward. The samples $\{\boldsymbol{w}_t[c] \mid \forall t, c\}$ of thermal noise at the receiver are assumed to be IID and drawn according to the complex Gaussian distribution $\mathbb{P}(\boldsymbol{w}_t[c]) = \mathsf{CN}(\boldsymbol{0}; \sigma^2 \boldsymbol{I}_L)$.

Due to random spreading the matrices $\mathcal{S} = \{\boldsymbol{S}_{k,t} = [\boldsymbol{s}_{k,t,1} \cdots \boldsymbol{s}_{k,t,M}] \in \mathbb{C}^{L \times M} \mid \forall k, t\}$, where $\boldsymbol{s}_{k,t,m}$ is the spreading sequence corresponding to the mth resolvable path, are independent and identically distributed (IID). However, for a fixed time index t, the signature sequences $\{\boldsymbol{s}_{k,t,m}\}_{m=1}^{M}$ of the kth user are not IID RVs. In fact, the spreading sequences for each path are cyclically shifted replicas of each other. For the following analysis we make the assumption that Theorem 4 of [26] holds for our system and, thus, the sequences $\{\boldsymbol{s}_{k,t,m}\}_{m=1}^{M}$ can be modified to have IID entries with zero mean, unit variance and finite moments for all $t = 1, \ldots, T_{\mathsf{coh}}$ without loss of generality.

We consider the special case of frequency selective Rayleigh fading. We assume for simplicity that the users' signals are received with equal average power[2], that the transmitters are well separated in space and the environment creates rich scattering such that the channel vectors are elementwise IID with distribution $\mathbb{P}(\boldsymbol{h}_k[c]) = \mathsf{CN}(\boldsymbol{0}; \boldsymbol{\Omega}_{\boldsymbol{h}_k[c]})$, where $\boldsymbol{\Omega}_{\boldsymbol{h}} = \operatorname{diag}(\bar{\boldsymbol{t}})$ and $\bar{\boldsymbol{t}} = [\bar{t}_1 \cdots \bar{t}_M]^{\mathsf{T}} \in \mathbb{R}^M$ is the power delay profile (PDP) of the multipath channel. The average received signal-to-noise ratio is defined as $\overline{\mathsf{snr}} = \bar{t}/\sigma^2$, where $\bar{t} = \sum_m \bar{t}_m$.

Following the framework of [4], [5], we let the code word

$$\boldsymbol{x}_k = \operatorname{vec}\big([\boldsymbol{x}_k[1] \cdots \boldsymbol{x}_k[C]]\big) \in \mathcal{M}^T, \tag{2}$$

where $\mathcal{M} = \{\pm \frac{1}{\sqrt{2}} \pm \frac{\mathrm{j}}{\sqrt{2}}\}$ is the QPSK signal set and $\boldsymbol{x}_k[c] = [x_{k,\tau_{\mathsf{tr}}+1}[c] \cdots x_{k,T_{\mathsf{coh}}}[c]]^{\mathsf{T}} \in \mathcal{M}^{\tau_{\mathsf{d}}}$ are the information symbols transmitted during the cth fading block, be the output of a BICM encoder [27], [28]. In the following, we denote the code rate by R and the code book of the kth user as $\mathcal{C}_k$. We also assume that binary trellis codes with trellis termination are used for error control coding (ECC) and all users derive the ECC from the same ensemble of binary codes, while the random bit-interleavers are IID for all $k = 1, \ldots, K$. Gray mapping is always employed and the code word length T is assumed to be sufficiently long for the fading to be an ergodic process over the code word.

C. Training Via Biased Signaling

In the previous discussion it was assumed that in order to perform the initial channel estimation, each fading block $c = 1, \ldots, C$, contained τ_{tr} known training symbols for all users. As we shall see next, this is not the only option to initiate the channel estimation.

Let $\Theta = \{\theta_{k,t}[c] \in \mathbb{C} \mid \forall k, t, c\}$ be a set of design variables known to both the transmitter and the receiver. Define the conditional prior distribution of $x_{k,t}[c]$ as

$$\mathbb{P}(x_{k,t}[c] \mid \theta_{k,t}[c]) = \mathbb{P}'(\Re\{x_{k,t}[c]\} \mid \Re\{\theta_{k,t}[c]\}) \times \mathbb{P}'(\Im\{x_{k,t}[c]\} \mid \Im\{\theta_{k,t}[c]\}), \tag{3}$$

where

$$\mathbb{P}'(x \mid \theta) = \frac{1 + \sqrt{2}\theta}{2} \delta_x(1/\sqrt{2}) + \frac{1 - \sqrt{2}\theta}{2} \delta_x(-1/\sqrt{2}). \tag{4}$$

Similarly, we let $\mathbb{P}(\theta_{k,t}[c]) = \mathbb{P}'(\Re\{\theta_{k,t}[c]\}) \mathbb{P}'(\Im\{\theta_{k,t}[c]\})$ be the prior of $\theta_{k,t}[c]$ with

$$\mathbb{P}'(\theta) = \frac{\Delta_{\mathsf{tr}}}{2} \delta_\theta(1/\sqrt{2}) + \frac{\Delta_{\mathsf{tr}}}{2} \delta_\theta(-1/\sqrt{2}) + \frac{1 - \Delta_{\mathsf{tr}}}{2} \delta_\theta(\sigma_{\mathsf{bias}}) + \frac{1 - \Delta_{\mathsf{tr}}}{2} \delta_\theta(-\sigma_{\mathsf{bias}}), \tag{5}$$

where $\sigma_{\mathsf{bias}} \in [0, 1/\sqrt{2})$ and $\Delta_{\mathsf{tr}} \in [0, 1)$ are fixed design parameters for all k, t and c. Thus, $\mathbb{E}\{x_{k,t}[c] \mid \theta_{k,t}[c]\} = \theta_{k,t}[c]$, where $\Re\{\theta_{k,t}[c]\}, \Im\{\theta_{k,t}[c]\} \in [-1/\sqrt{2}, 1/\sqrt{2}]$. Since Θ is assumed to be known at the receiver, setting $\sigma_{\mathsf{bias}} = 0$ gives the traditional pilot assisted transmission scheme. For large T_{coh}, we may assume without loss of generality that each fading block has then $\tau_{\mathsf{tr}} = \Delta_{\mathsf{tr}} T_{\mathsf{coh}}$ modulated "hard" pilot symbols, denoted as before by $\boldsymbol{p}_k[c] \in \mathcal{M}^{\tau_{\mathsf{tr}}}$, and the number of data symbols $\tau_{\mathsf{d}} = T_{\mathsf{coh}} - \tau_{\mathsf{tr}}$ is fixed for all fading blocks. If, on the other hand, we set $\Delta_{\mathsf{tr}} = 0$, the optimum hyperprior for unconstrained receivers is retrieved (see [14, Prop. 2]).

With the above training scheme, the total training overhead of the system as a fraction of the total transmission is

$$\Delta_{\mathsf{tot}} = \Delta_{\mathsf{tr}} + (1 - \Delta_{\mathsf{tr}}) \Delta_{\mathsf{d}} \in [0, 1), \tag{6}$$

[2]This is not a necessary assumption and unequal power users can be readily considered in an analogous manner. We consider this special case to simplify notation in the following analysis.

Table I
ITERATIVE CHANNEL AND DATA ESTIMATION

1) Fix the time index $\vartheta \in \mathcal{D}$ and iteration index ℓ. Let the channel estimator postulate the priors $\{\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta}[c]) \mid \forall k \in \mathcal{K}\}$ and assume that the approximate a posteriori probabilities (APPs) $\{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D} \setminus \vartheta\}$ have been received from the iterative MUDD during the previous iteration. Given its knowledge about the system model (1) and the information
$$\mathcal{I}_{\vartheta}^{(\ell)}[c] = \Big\{\mathcal{I}_c, \mathcal{Y}_c \setminus \boldsymbol{y}_{\vartheta}[c], \{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D} \setminus \vartheta\}\Big\}, \tag{7}$$
the iterative CE calculates $\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t}[c] \mid \mathcal{I}_t^{(\ell)}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D}\}$ for all $c = 1, \dots, C$, and sends the CSI to the iterative MUDD.
2) Let the data estimator assign the postulated prior $\mathbb{Q}(\tilde{x}_{\xi,t}[c])$ to the data symbol of user $\xi \in \mathcal{K}$ at time instant $t \in \mathcal{D}$. Given
$$\mathcal{I}_{\xi,t}^{(\ell)}[c] = \Big\{\mathcal{I}_c, \boldsymbol{y}_t[c], \{\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t}[c]) \mid \forall j \in \mathcal{K} \setminus \xi\}, \{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t}[c] \mid \mathcal{I}_t^{(\ell)}[c]) \mid \forall k \in \mathcal{K}\}\Big\}, \tag{8}$$
and its knowledge about the channel (1), the data estimator calculates the posterior probabilities $\{\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t}[c] \mid \mathcal{I}_{\xi,t}^{(\ell)}[c]) \mid \forall t \in \mathcal{D}\}$, for all $c = 1, \dots, C$, and sends them to the single-user sum-product decoders.
3) For $k = 1, \dots, K$, the posterior probabilities of the data symbols $\{\mathbb{Q}^{(\ell)}(\tilde{x}_{k,t}[c] \mid \mathcal{I}_{k,t}^{(\ell)}[c]) \mid \forall c, t \in \mathcal{D}\}$ and the code book $\mathcal{C}_k$ are used by the sum-procuct decoder to calculate the approximate a posteriori probabilities $\mathbb{P}_{\mathsf{app}}^{(\ell)}(\boldsymbol{x}_k)$ and the extrinsic probabilities $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(\boldsymbol{x}_k)$ of the data symbols.
4) The operators φ_{ext} and φ_{app} are applied to the outputs of the sum-product decoders $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t}[c])$ and $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t}[c])$, respectively, to produce the corresponding feedback probabilities $\mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t}[c])$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell)}(\tilde{x}_{k,t}[c])$. The former are sent to the channel estimator while the latter are stored and used by the data estimator during the next iteration.

where we have assumed that $\Delta_{\mathsf{d}} = 1 - H\big((1-\sqrt{2}\sigma_{\mathsf{bias}})/2\big)$, is the amount of pilot information embedded in the data symbols, and $H(p)$ is the binary entropy function. This corresponds to an ideal method of biasing that incurs no additional overhead by itself. We also assume that the bit error rate (BER) performance of the BICM is not affected by the a priori bias.

III. ITERATIVE MULTIUSER RECEIVERS

Let us write the set of all received vectors during the cth fading block in (1) as $\mathcal{Y}_c = \{\boldsymbol{y}_t[c] \mid t = 1, \cdots, T_{\mathsf{coh}}\}$, and similarly let $\mathcal{H}_c = \{\boldsymbol{h}_k[c] = \big[h_{k,1}[c] \ \cdots \ h_{k,M}[c]\big]^{\mathsf{T}} \in \mathbb{C}^M \mid \forall k\}$ denote the fading coefficients of all users in the cth fading block. The postulated channel and data symbols for users $k = 1, \dots, K$, at time instant $t = \tau_{\mathsf{tr}} + 1, \dots, T_{\mathsf{coh}}$, are written as $\tilde{\boldsymbol{h}}_{k,t}[c] \in \mathbb{C}^M$ and $\tilde{x}_{k,t}[c]$, respectively[3]. We also denote $\tilde{\boldsymbol{x}}_k[c] = \big[\tilde{x}_{k,\tau_{\mathsf{tr}}+1}[c] \ \cdots \ \tilde{x}_{k,T_{\mathsf{coh}}}[c]\big]^{\mathsf{T}} \in \mathbb{C}^{\tau_{\mathsf{d}}}$ for all postulated data symbols of the kth user in the cth fading block, and assign the priors (to be defined later) $\mathbb{Q}(\tilde{\boldsymbol{x}}_k[c])$ and $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,t}[c])$ to the above RVs for all $k = 1, \dots, K$.

By assumption, the channel and the data estimator have knowledge of the received vectors $\mathcal{Y}_c$ and the set $\mathcal{I}_c = \{\mathcal{P}_c, \mathcal{S}, \Theta\}$, where $\mathcal{P}_c = \{p_{k,t}[c] \mid \forall k, t = 1, \dots, \tau_{\mathsf{tr}}\}$ are the training symbols transmitted during the fading block $c = 1, \dots, C$, and the spreading matrices in $\mathcal{S}$ are modified as stated in Section II-B. In addition, the estimators may have also received some information via feedback from the single-user decoders during the previous iteration.

Consider iteration $\ell = 1, 2, \dots$ and let $\mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x_{k,t})$ and $\mathbb{P}_{\mathsf{app}}^{(\ell-1)}(x_{k,t})$, be the extrinsic and approximate a posteriori probabilities, respectively, of the transmitted symbol $x_{k,t} \in \mathcal{M}$. For convolutional codes, both probabilities are easy to obtain by using the BCJR algorithm [29]. Let the feedback from the single-user decoders be in the form of probabilities $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{k,t})$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t})$, where again $\tilde{x}_{k,t} \in \mathcal{M}$. The relation between the decoder outputs and the feedback is defined by the operators φ_{ext} and φ_{app} that transform the probability measures (or distributions) $\mathbb{P}_{\mathsf{ext}}$ and $\mathbb{P}_{\mathsf{app}}$ to $\mathbb{Q}_{\mathsf{ext}}$ and $\mathbb{Q}_{\mathsf{app}}$, respectively, i.e.,
$$\varphi_{\mathsf{ext}} : \mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x) \mapsto \mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x} = x), \qquad x, \tilde{x} \in \mathcal{M}, \tag{9}$$
$$\varphi_{\mathsf{app}} : \mathbb{P}_{\mathsf{app}}^{(\ell-1)}(x) \mapsto \mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x} = x), \qquad x, \tilde{x} \in \mathcal{M}. \tag{10}$$
Throughout this paper we assume that the operators (9) and (10) do not depend on the iteration index ℓ and the feedback $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}$ is a well-defined probability measure (or distribution) over $\mathcal{M}$. The specific forms of φ_{ext} and φ_{app} define the type of feedback used. In particular, the identity operator gives soft feedback whereas defining (similarly for φ_{app})
$$\varphi_{\mathsf{ext}} : \mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t}) \mapsto \delta_{\tilde{x}_{k,t}}(\hat{x}_{k,t}), \qquad \tilde{x}_{k,t} = x_{k,t}, \tag{11}$$
where $\hat{x}_{k,t} = \arg\max_{x_{k,t} \in \mathcal{M}} \mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t})$ results in hard feedback. Note that above we used the nomenclature common to iterative ISI cancellation and MUDD, where the extrinsic probabilities of the coded bits do not contain channel information, whereas the approximate APPs do (see for example, [4]). Both probabilities are obtained using the knowledge of $\mathcal{C}_k$ and the feedback is extrinsic information in the sense defined for message passing algorithms in factor graphs [1], [2]. With the above in mind, a high-level algorithm for iterative channel and data estimation is given in Table I and the block diagram of the corresponding receiver structure is depicted in Fig. 2, where we omitted the iteration index $\ell = 1, 2, \dots$, for clarity. To initiate the iterative process, we let $\mathbb{Q}_{\mathsf{app}}^{(0)}(\tilde{x}_{k,t}[c])$ and $\mathbb{Q}_{\mathsf{ext}}^{(0)}(\tilde{x}_{k,t}[c])$ be equal to (3). By modifying the postulated prior probabilities and the receiver's knowledge about the signal model (1), we obtain the following special cases:

- ***Iterative channel estimators:*** linear MMSE (LMMSE) channel estimator with soft feedback, approximate maximum likelihood (ML) estimator using hard feedback;
- ***Iterative data estimators:*** maximum a posteriori (MAP), LMMSE and SUMF MUDDs with soft feedback, SUMF with hard feedback.

Due to space limitations, however, the details of the steps in Table I are omitted in this paper. Next, we only give a brief example of how to derive linear MUDDs based on the above GPME framework.

A. Example: Iterative Data Estimation With PIC

Consider the ℓth iteration and let $\xi \in \mathcal{K}$ be the user of interest. Recall that $\mathcal{I}_{\xi,t}^{(\ell)}$ defined in (8) is available to the data

[3]Note that the postulated channel depends on t, while the true channel $\boldsymbol{h}_k[c]$ does not. The reason for this will become clear subsequently.

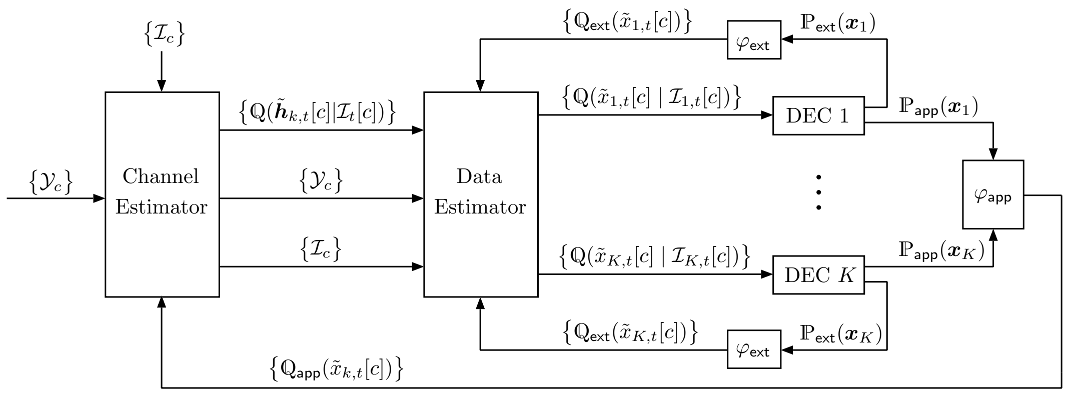

Figure 2. Simplified block diagram of a receiver employing iterative channel and data estimation.

estimator and assume that the posterior mean estimates of the channel coefficients of all users and the data symbols of the interfering users, i.e.,

$$\langle\tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} = \int \tilde{\boldsymbol{h}}_{k,t} \mathrm{d}\mathbb{Q}^{(\ell)}\big(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}_t^{(\ell)}\big), \qquad \forall k \in \mathcal{K}, \tag{12}$$

$$\langle\tilde{x}_{j,t}\rangle_{\mathsf{ext}}^{(\ell-1)} = \sum_{\tilde{x}_{j,t}\in\mathcal{M}} \tilde{x}_{j,t}\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t}), \qquad \forall j \in \mathcal{K}\setminus\xi, \tag{13}$$

respectively, have been calculated with the help of $\mathcal{I}_{\xi,t}^{(\ell)}$. Define also the RVs

$$\Delta x_{j,t} = x_{j,t} - \langle\tilde{x}_{j,t}\rangle_{\mathsf{ext}}^{(\ell-1)}, \qquad \forall j \in \mathcal{K}\setminus\xi, \tag{14}$$

$$\Delta\boldsymbol{h}_{k,t} = \boldsymbol{h}_k - \langle\tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)}, \qquad \forall k \in \mathcal{K}, \tag{15}$$

$$\Delta\boldsymbol{v}_{k,t} = \Delta\boldsymbol{h}_{k,t}x_{k,t}, \qquad \forall k \in \mathcal{K}, \tag{16}$$

which, given $\mathcal{I}_{\xi,t}^{(\ell)}$, are all zero-mean in the limit of large code word length.

Let $\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}$ and $\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{k,t}}^{(\ell)}$ be the receiver's knowledge about the true covariances $\Omega_{\Delta x_{j,t}}^{(\ell-1)}$ and $\boldsymbol{\Omega}_{\Delta\boldsymbol{v}_{k,t}}^{(\ell)}$ of (14) and (16), respectively. Denote the postulated noise variance in (1) by $\tilde{\sigma}^2$. The marginalized posterior probability of the data symbol $x_{\xi,t}$ is given by

$$\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}_{\xi,t}^{(\ell)}) = \frac{\mathbb{Q}(\tilde{x}_{\xi,t})\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}_{\xi,t}^{(\ell)})}{\mathbb{E}_{\tilde{x}_{\xi,t}}\big\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}_{\xi,t}^{(\ell)})\big\}}, \tag{17}$$

where $\mathbb{Q}(\tilde{x}_{\xi,t})$ is the postulated a priori distribution for the desired user's data symbol. The postulated channel model reads

$$\tilde{\boldsymbol{y}}_t = \frac{1}{\sqrt{L}}\boldsymbol{S}_{\xi,t}\langle\tilde{\boldsymbol{h}}_{\xi,t}\rangle_{(\ell)}\tilde{x}_{\xi,t} + \frac{1}{\sqrt{L}}\sum_{j\in\mathcal{K}\setminus\xi}\boldsymbol{S}_{j,t}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}\langle\tilde{x}_{j,t}\rangle_{(\ell)} + \tilde{\boldsymbol{w}}_{\xi}^{\mathsf{pic},(\ell)}, \tag{18}$$

where $\tilde{\boldsymbol{w}}_{\xi}^{\mathsf{pic},(\ell)} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)})$ and the modified noise covariance is given by

$$\tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)} = \tilde{\sigma}^2\boldsymbol{I}_L + \frac{1}{L}\boldsymbol{S}_{\xi,t}\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{\xi,t}}^{(\ell)}\boldsymbol{S}_{\xi,t}^{\mathsf{H}} + \frac{1}{L}\sum_{j\in\mathcal{K}\setminus\xi}\boldsymbol{S}_{j,t}\Big(\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{j,t}}^{(\ell)} + \langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}\langle\tilde{\boldsymbol{h}}_{j,t}^{\mathsf{H}}\rangle_{(\ell)}\Big)\boldsymbol{S}_{j,t}^{\mathsf{H}}. \tag{19}$$

Note that when the second term on the right hand side of (18) is moved to the left hand side of the equation, we get the PIC operation. The GPME based on (17) is given by

$$\langle\tilde{x}_{\xi,t}\rangle_{(\ell)} = \int \tilde{x}_{\xi,t}\,\mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}_{\xi,t}^{(\ell)}), \qquad \xi \in \mathcal{K}. \tag{20}$$

If we let $\mathbb{Q}(\tilde{x}_{\xi,t}) \sim \mathsf{CN}(0;\, 1)$ and denote

$$\tilde{\boldsymbol{y}}_{\xi,t} = \tilde{\boldsymbol{y}}_t - \frac{1}{\sqrt{L}}\sum_{j\in\mathcal{K}\setminus\xi}\boldsymbol{S}_{j,t}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}\langle\tilde{x}_{j,t}\rangle_{(\ell)}, \tag{21}$$

the GPME further simplifies the to the familiar form

$$\langle\tilde{x}_{\xi,t}\rangle_{(\ell)} = \boldsymbol{m}_{\xi,t}^{\mathsf{H}}\tilde{\boldsymbol{y}}_{\xi,t} = \frac{\frac{1}{\sqrt{L}}\langle\tilde{\boldsymbol{h}}_{\xi,t}\rangle_{(\ell)}^{\mathsf{H}}\boldsymbol{S}_{\xi,t}^{\mathsf{H}}\big(\tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)}\big)^{-1}}{1 + \frac{1}{L}\langle\tilde{\boldsymbol{h}}_{\xi,t}\rangle_{(\ell)}^{\mathsf{H}}\boldsymbol{S}_{\xi,t}^{\mathsf{H}}\big(\tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)}\big)^{-1}\boldsymbol{S}_{\xi,t}\langle\tilde{\boldsymbol{h}}_{\xi,t}\rangle_{(\ell)}}\tilde{\boldsymbol{y}}_{\xi,t} \in \mathbb{C}. \tag{22}$$

A proper choice of the parameters

1) $\varphi_{\mathsf{ext}} : \mathbb{P} \mapsto \mathbb{Q}$, defines the type of PIC (soft / hard);
2) $\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}$ and $\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{k,t}}^{(\ell)}$ quantify the estimator's knowledge about the error statistics;
3) $\tilde{\sigma}^2$ defines the type of linear filtering used;

yields all the iterative data estimators utilizing linear filtering and PIC analyzed in this paper.

IV. Performance Of Large DS-CDMA Systems Using Iterative Channel And Data Estimation

We now turn to the analysis of the multiuser DS-CDMA system described in Section II-B that uses the iterative estimators discussed in Section III. We concentrate on the (linear) MMSE based estimators and omit the ML and SUMF estimators due to space constraints.

A. Density Evolution With Gaussian Approximation

Let us write the BICM encoded channel inputs (1) as $x_{k,t} = (a_{k,t,1} + \mathrm{j}a_{k,t,2})/\sqrt{2}$, where $a_{k,t,q} \in \{\pm 1\}$. The extrinsic probabilities of the data symbols at the ℓth iteration factor as $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t}) = \mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,1})\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,2})$, in the limit of large code word length $T \to \infty$. Let

$$\hat{a}_{k,t,q}^{\mathsf{ext},(\ell)} = \operatorname*{arg\,max}_{a_{k,t,q}\in\{\pm 1\}} \mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q}), \qquad q = 1, 2, \tag{23}$$

be the extrinsic information-based hard estimate of $a_{k,t,q}$, and define the error probability

$$\varepsilon_k^{\mathsf{ext},(\ell)} = \frac{1}{2T} \sum_{j=1}^{2} \sum_{t=1}^{T} \Pr\left(\hat{a}_{k,t,q}^{\mathsf{ext},(\ell)} \neq a_{k,t,q}\right). \quad (24)$$

A completely analogous development for the approximate a posteriori-based feedback can be defined.

To simplify the density evolution, we use the Gaussian approximation [4], [5], [17] for the log-likelihood ratios obtained by the sum-product decoders, i.e., given $a_{k,t,q}$,

$$\lambda_{a_{k,t,q}}^{\mathsf{ext},(\ell)} = \log\left(\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q} = +1)\right) - \log\left(\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q} = -1)\right) \sim \mathsf{N}(2a_{k,t,q}\mu_k^{\mathsf{ext},(\ell)}, 4\mu_k^{\mathsf{ext},(\ell)}), \quad (25)$$

where $\mu_k^{\mathsf{ext},(\ell)} = [Q^{-1}(\varepsilon_k^{\mathsf{ext},(\ell)})]^2$, and Q^{-1} is the functional inverse of the Q-function [30]. The approximate APPs $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t})$ are handled in a completely analogous manner.

The probabilities $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t})$ obtained through (25) are transformed via φ_{ext} to $\mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t})$ as discussed in Section III. The posterior mean estimate based on $\mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t})$ is denoted by $\langle\tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell)}$ and the MSE of the extrinsic information-based symbols conditioned on the feedback reads

$$\Omega_{\Delta x_{k,t}}^{\mathsf{ext},(\ell)} = \mathbb{E}\left\{|x_{k,t} - \langle\tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell)}|^2 \mid \langle\tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell)}\right\}. \quad (26)$$

Note that (26) depends on the type of feedback. Completely analogous notation is used for the feedback based on approximate APPs $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t})$. In the following, we also omit the user and time indices k and t when they are deemed unnecessary for the presentation.

B. Linear Channel Estimation With Information Feedback

Proposition 1. *Consider the iterative LMMSE channel estimator that uses soft feedback and has perfect knowledge about the feedback error statistics. Given the feedback symbols $\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}^{(\ell-1)}$, the conditional MSE for the mth path, user k and time index $\vartheta \in \mathcal{D}$ reads*

$$\mathsf{mse}_{\vartheta,m}^{\mathsf{fb},(\ell)} = \bar{t}_m\left(1 + \bar{t}_m\left[\frac{\tau_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \sum_{t\in\mathcal{D}\setminus\vartheta} \frac{|\langle\tilde{x}_{k,t}\rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m\Omega_{\Delta x}^{\mathsf{app},(\ell-1)}}\right]\right)^{-1}, \quad (27)$$

where $\Omega_{\Delta x_t}^{\mathsf{app},(\ell-1)} = 1 - |\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2$, and

$$C_{\mathsf{tr}}^{(\ell)} = \sigma^2 + \alpha\sum_{m=1}^{M}\mathbb{E}\{\mathsf{mse}_{\vartheta,m}^{\mathsf{fb},(\ell)}\}, \quad (28)$$

$$C_{\mathsf{d}}^{(\ell)} = \sigma^2 + \alpha\sum_{m=1}^{M}\mathbb{E}\left\{\frac{C_{\mathsf{d}}^{(\ell)}}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m\Omega_{\Delta x}^{\mathsf{app},(\ell-1)}} \times\left[\bar{t}_m\Omega_{\Delta x}^{\mathsf{app},(\ell-1)} + \frac{C_{\mathsf{d}}^{(\ell)}|\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m\Omega_{\Delta x}^{\mathsf{app},(\ell-1)}}\mathsf{mse}_{\vartheta,m}^{\mathsf{fb},(\ell)}\right]\right\}. \quad (29)$$

The average MSE is thus given by $\mathsf{mse}_{\vartheta,m}^{(\ell)} = \mathbb{E}\{\mathsf{mse}_{\vartheta,m}^{\mathsf{fb},(\ell)}\}$, where the parameters $C_{\mathsf{tr}}^{(\ell)}$ and $C_{\mathsf{d}}^{(\ell)}$ satisfy the coupled equations (27) – (29).

Remark 1. By (27) to (29), the use of soft feedback never increases the per-path MSE of this channel estimator.

Corollary 1. *Let us consider the special case where all multipaths have the same average power, i.e., $\bar{\boldsymbol{t}} = (\bar{t}/M)[1 \ \cdots \ 1]^{\mathsf{T}} \in \mathbb{R}^M$. If we define the ratios $\Delta_{\mathsf{tr}} = \tau_{\mathsf{tr}}/T_{\mathsf{coh}}$ and $\Upsilon = T_{\mathsf{coh}}/M$, taking the limit $\tau_{\mathsf{tr}}, \tau_{\mathsf{d}}, M \to \infty$ while keeping $\Delta_{\mathsf{tr}}, \Upsilon$ finite and fixed gives the average asymptotic normalized per-path MSE $\xi^{(\ell)} = \frac{\mathsf{mse}^{(\ell)}}{\bar{t}/M}$ as*

$$\xi^{(\ell)} = \left[1 + \Upsilon\bar{t}\left(\frac{\Delta_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \frac{(1-\Delta_{\mathsf{tr}})}{C_{\mathsf{d}}^{(\ell)}}\mathbb{E}\{|\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2\}\right)\right]^{-1}. \quad (30)$$

The parameters (28) *and* (29) *also reduce to the simplified fixed point equations*

$$C_{\mathsf{tr}}^{(\ell)} = \sigma^2 + \alpha\bar{t}\xi^{(\ell)}, \quad (31)$$

$$C_{\mathsf{d}}^{(\ell)} = \sigma^2 + \alpha\left[1 - \left(1 - \xi^{(\ell)}\right)\mathbb{E}\{|\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2\}\right], \quad (32)$$

respectively.

Remark 2. The scenario considered in Corollary 1 is highly ideal, but allows for simplified numerical evalution of the fixed point equations given in Proposition 1. The physical interpretation of the scenario is a broadband high data rate transmission over a uniform rich scattering environment with very long code words. Using the notation of [25] and denoting the delay and Doppler spread of the channel by T_m and B_d, respectively, Υ turns out to be the inverse of the channel spread factor $\Upsilon^{-1} = T_m B_d$. Note that the requirement for accurate channel estimation is known to be $T_m B_d \ll 1$, which in our notation means $\Upsilon \gg 1$.

C. Data Estimation With Feedback and Mismatched CSI

Proposition 2. *Consider iteration $\ell = 1, 2, \ldots$ and let the channel estimation be performed by the LMMSE estimator of Proposition 1. The SINR of the iterative data estimator for any user $k = 1, \ldots, K$, and time index $t \in \mathcal{D}$ is given by*

$$\mathsf{sinr}^{(\ell)}(D^{(\ell)}) = \sum_{m=1}^{M}\frac{|\langle\tilde{h}_m\rangle_{(\ell)}|^2}{D^{(\ell)} + \mathsf{mse}_m^{(\ell)}}, \quad (33)$$

where $|\langle\tilde{h}_m\rangle_{(\ell)}|^2 \sim \mathsf{CN}(0; \bar{t}_m - \mathsf{mse}_m^{(\ell)})$. The parameter $D^{(\ell)}$ reads for the SUMF as

$$D_{\mathsf{sumf}}^{(\ell)} = \sigma^2 + \alpha\sum_{m=1}^{M}\mathsf{mse}_m^{(\ell)} + \Omega_{\Delta x}^{\mathsf{ext},(\ell-1)}(\bar{t}_m - \mathsf{mse}_m^{(\ell)}), \quad (34)$$

where the feedback error variance $\Omega_{\Delta x}^{\mathsf{ext},(\ell-1)}$ is given in (26)*. For the MAP-MUDD or the LMMSE-PIC MUDD, on the other hand, $D^{(\ell)}$ is the solution to the fixed point equation*

$$D^{(\ell)} = \sigma^2 + \alpha\sum_{m=1}^{M}\left(\frac{D^{(\ell)}}{D^{(\ell)} + \mathsf{mse}_m^{(\ell)}}\right)^2 \times\left[\mathsf{mse}_m^{(\ell)}\left(\frac{D^{(\ell)} + \mathsf{mse}_m^{(\ell)}}{D^{(\ell)}}\right) + \mathbb{E}\left\{|\langle\tilde{h}_m\rangle_{(\ell)}|^2 V^{(\ell)}(D^{(\ell)})\right\}\right], \quad (35)$$

where for the MAP-MUDD, $V^{(\ell)}\big(D^{(\ell)}\big)$ *is given by*

$$V^{(\ell)}\big(D^{(\ell)}\big) = 1 - \sum_{a_1 \in \{\pm 1\}} \frac{1 + a_1 \langle \tilde{a}_1 \rangle_{\mathsf{ext}}^{(\ell-1)}}{2} \int \mathrm{D}\nu$$

$$\times \tanh\left(\mathsf{sinr}^{(\ell)}\big(D^{(\ell)}\big) + \nu\sqrt{\mathsf{sinr}^{(\ell)}\big(D^{(\ell)}\big)} + a_1 \frac{\lambda_{a_1}^{(\ell-1)}}{2} \right), \quad (36)$$

and for the LMMSE-PIC MUDD

$$V^{(\ell)}\big(D^{(\ell)}\big) = \Omega_{\Delta x}^{\mathsf{ext},(\ell-1)} \left(1 + \Omega_{\Delta x}^{\mathsf{ext},(\ell-1)} \mathsf{sinr}^{(\ell)}\big(D^{(\ell)}\big)\right)^{-1}. \quad (37)$$

In the following the solutions to (35) *to* (37) *for the MAP and the LMMSE MUDD are denoted by* $D_{\mathsf{map}}^{(\ell)}$ *and* $D_{\mathsf{lmmse}}^{(\ell)}$*, respectively.*

Interestingly, there is a common part in (35) for the MAP and LMMSE based data estimators that do not depend on the extrisic information-based feedback at all. Note that these terms vanish if and only if $\mathsf{mse} \to 0$. Furthermore, there is a connection with the estimator specific terms $V^{(\ell)}$ to the related terms in the case of perfect CSI.

Remark 3. Consider the terms (36) and (37) that are connected to the MAP and the LMMSE-PIC MUDDs. For fixed $D^{(\ell)}$, $\mathbb{E}\{V^{(\ell)}\} = \mathbb{E}\{|x - \langle\tilde{x}\rangle_{(\ell)}|^2\}$, where $\langle\tilde{x}\rangle_{(\ell)}$ is the estimate of the desired user's data symbol at the ℓth iteration. These MSEs of the data symbols are equal to the corresponding terms in [4], [5], [15] where perfect CSI was assumed, with the noise variance increased by the MSE of the channel estimates and the channel power reduced accordingly.

D. Multiuser Efficiency and Related Performance Measures

Let $\mathsf{snr}_k[c] = \|\boldsymbol{h}_k[c]\|^2/\sigma^2$, be the instantaneous received SNR of the kth user during the cth fading block in (1). Furthermore, let $\mathsf{sinr}_k^{(\ell)}[c]$ be the corresponding SINR of the same user at the output of the MUDD during iteration $\ell = 1, 2, \ldots$, and given in Section IV-C. Let us consider the simplest case of equal power users with uniform power delay profiles and LMMSE channel estimation. We omit the user and block indices and following the notation of [4], [5], define a "multiuser efficiency" like parameter

$$\eta_{\mathsf{ce}}^{(\ell)} = \frac{\mathsf{sinr}^{(\ell)}}{\mathsf{snr}} = \frac{\sigma^2}{\frac{(\bar{t}/M)(D^{(\ell)} + \mathsf{mse}^{(\ell)})}{\bar{t}/M - \mathsf{mse}^{(\ell)}}} \quad (38)$$

$$\xrightarrow{\text{Corollary 1}} \frac{\sigma^2}{D^{(\ell)}}(1 - \xi^{(\ell)}), \qquad 0 \le \eta_{\mathsf{ce}}^{(\ell)} \le 1, \quad (39)$$

where (39) corresponds to the simplified case of large number of multipaths, considered in Corollary 1. Furthermore, we let

$$\Psi_{\mathsf{ce}} : [0,1] \to [0,1] : \eta_{\mathsf{ce}}^{(\ell-1)} \mapsto \eta_{\mathsf{ce}}^{(\ell)}, \quad (40)$$

be a mapping function that describes the DE of (38) to (39) with Gaussian approximation for a specific iterative MUDD. Naturally $\eta_{\mathsf{ce}}^{(\ell)} \to \eta^{(\ell)}$ and $\Psi_{\mathsf{ce}} \to \Psi$ as given in [4], [5] when $\mathsf{mse}^{(\ell)} \to 0$ or $\xi^{(\ell)} \to 0$. One should note, however, that instead of describing just the MAI suppression capacity of the iterative MUDD, $\eta_{\mathsf{ce}}^{(\ell)}$ provides information about the efficiency of the entire iterative channel and data estimation scheme.

Finally, let us consider the situation when the feedback symbols tend to correct decisions, that is, $\Omega_{\Delta x}^{(\ell-1)} \to 0$ and

$$D_{\mathsf{lmmse}}^{(\ell)} \to D_{\mathsf{map}}^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \frac{D_{\mathsf{map}}^{(\ell)} \mathsf{mse}_m^{(\ell)}}{D_{\mathsf{map}}^{(\ell)} + \mathsf{mse}_m^{(\ell)}}, \quad (41)$$

$$D_{\mathsf{sumf}}^{(\ell)} \to \sigma^2 + \alpha \sum_{m=1}^{M} \mathsf{mse}_m^{(\ell)}. \quad (42)$$

Plugging the solution of (41) or (42), depending on the iterative MUDD, into (38) gives an upper bound for the maximum achievable $\eta_{\mathsf{ce}}^{(\ell)}$ for a given MSE of the channel estimates. For the iterative channel estimator, on the other hand, $\mathsf{mse}_m^{(\ell)}$ can be lower bounded by considering the corresponding non-iterative channel estimator with $T_{\mathsf{coh}} - 1$ known training symbols. Note that in contrast to the case of perfect CSI, where $D_{\mathsf{sumf}}^{(\ell)} = D_{\mathsf{lmmse}}^{(\ell)} = D_{\mathsf{map}}^{(\ell)} \to \sigma^2$ as $\Omega_{\Delta x}^{(\ell-1)} \to 0$, with CSI mismatch the performance of the SUMF and the LMMSE-PIC / MAP MUDDs can be different with genie aided PIC. For the special case $\bar{t}_{k,m} = \bar{t}/M$, it is easy to show that the maximum difference in the average post-detection SINR between the LMMSE-PIC / MAP-MUDD and SUMF with genie aided feedback is 3 dB. However, when $M \to \infty$ the maximum loss for the SUMF vanishes. Thus, for wideband channels there are, in general, little differences between the performances of the different iterative MUDDs, if they converge to their maximum values of η_{ce}. However, their convergence properties may differ strongly.

V. Numerical Examples and Discussion

Selected numerical examples of the analytical results obtained in Section IV are provided next. We remind the reader that the numerical examples given here are based on the large-system analysis and is approximate for finite systems. For all considered cases the binary ECC for the BICM is a half-rate convolutional code. Two maximum free distance codes defined by the (octal) polynomials $(5,7)_8$ and $(561,753)_8$, with respective constraint lengths 3 and 9, are used [31]. We use Gray-mapped QPSK. Thus, the BICM has code rate $R = 1$ and the average SNR per information bit is $\overline{\mathsf{snr}}_k = \bar{t}_k/\sigma^2$.

In Fig. 3, we have plotted Ψ_{ce} given in (40) for the case of LMMSE based channel estimation and LMMSE-PIC MUDD. For comparison, the upper bounds discussed in Section IV-D (dashed lines), and the corresponding curve Ψ for the case of perfect CSI are also included. Ergodic Rayleigh fading with three equal power paths is assumed and the $(561,753)_8$ code is considered. For both cases with channel estimation, the iterative MUDD converges to its maximum value of η_{ce}. Thus, the limiting factor in the performance is the mismatched CSI. The combination of iterative LMMSE-CE and LMMSE-PIC MUDD in fact achieves single-user performance for the given channel conditions, which is quite remarkable given that only a single pilot symbol was used. With non-iterative LMMSE channel estimation and ten training symbols, on the other hand, a severe loss in output SINR is observed.

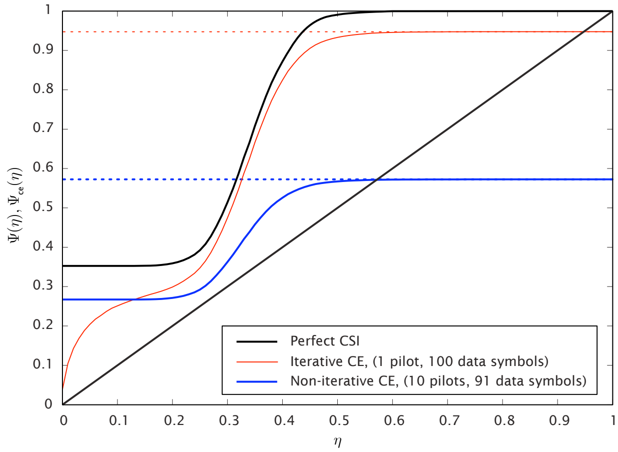

Figure 3. Mapping function Ψ for the case of perfect CSI and Ψ_{ce} when channel estimation is employed. Three equal power paths, channel coherence time of $T_{\text{coh}} = 101$ symbols, $\overline{\text{snr}} = 6$ dB, user load $\alpha = 1.2$, LMMSE-PIC MUDD and $(561, 753)_8$ convolutional code for all cases. Channel estimation by non-iterative or iterative LMMSE estimator.

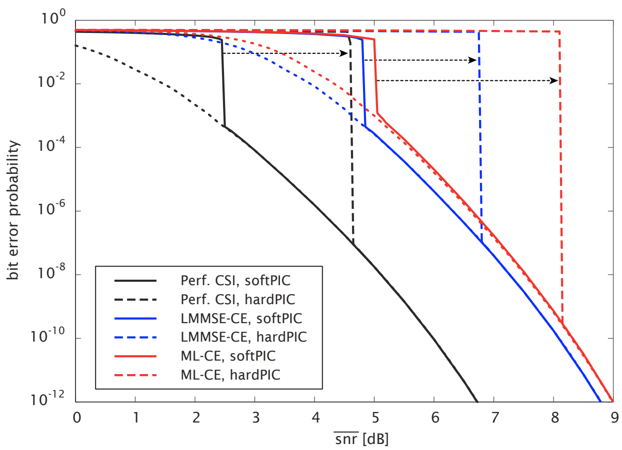

Figure 4. Bit error probability vs. SNR of the SUMF based MUDs with soft and hard feedback and LMMSE or ML channel estimation. Training overhead of 10%, user load $\alpha = 0.7$ and inverse of channel spread factor $\Upsilon = 20$. Rate-$1/2$ convolutional code $(561, 753)_8$ and Gray mapped QPSK.

Fig. 4 depicts the BEP vs. SNR for the SUMF with two linear channel estimators under the simplifying assumptions of Corollary 1. To guarantee convergence for all considered cases within the given SNR range, the user load was set to $\alpha = 0.7$ and a pilot overhead of 10% was used. Inverse channel spread factor $\Upsilon = T_{\text{coh}}/M = 20$ was assumed. The genie aided performance of the MUDs is plotted with the dotted lines. Both the soft and the hard PIC have a phase transition in BEP from one half to the minimum attainable for the given receiver and system parameters. Furthermore, there is very little difference between the non-iterative ML channel estimator (ML-CE) and the non-iterative LMMSE channel estimator (LMMSE-CE) in the latter region. There is, however, a significant difference in the threshold SNR where the phase transition occurs at, when comparing soft vs. hard feedback (see the arrows in the figure). Due to the poor performance of the hard feedback based PIC and ML channel estimation, we drop it from further discussion in this section and concentrate on presenting results for the iterative MUDDs that use soft PIC and LMMSE channel estimation.

We next consider the case when the system is operating at moderately high SNR and we are allowed to adapt the user load in order to achieve maximum total throughput. The loss in spectral efficiency due to transmission of known training symbols is taken into account in the results. In addition to the LMMSE-based data decoder, the non-linear MAP-MUDD is considered as well. Fig. 5 depicts the spectral efficiency vs. training overhead for $\Upsilon = T_{\text{coh}}/M = 80$ under the simplifying assumptions of Corollary 1. The system load is adjusted to meet the bit error rate requirement BER $\leq 10^{-5}$. Only selected combinations of system parameters are plotted for sake of clarity. Note that the iterative LMMSE-CE is suboptimal when there is uncertainty in the transmitted symbols. The upper bound corresponds to the case of genie-aided feedback, much like we did previously with the iterative MUDDs. We remark the following:

- Significant improvement over the non-iterative data estimators studied in [15], [19], [26] can be achieved by using iterative MUDD, even with the non-iterative LMMSE channel estimator. As expected, the receivers using non-linear MAP detectors show notable gains in spectral efficiency over the LMMSE-based receivers. The difference, however, is smaller in the iterative cases.
- For $\Upsilon = 80$, the highest loads with iterative MUDD and CE are obtained by using the $(5, 7)_8$ code. In this case, the performance of the iterative LMMSE channel estimator also follows closely the upper bound, showing that everything else being equal, little can be gained by using a more complex channel estimator. This does not hold for scenarios where Υ is sufficiently small.
- Iterations over the channel estimator provide only minor improvements in spectral efficiency if the $(561, 753)_8$ code is used, and its performance is quite far from the upper bounds in this case. This hints that matching the channel code to the provisional channel conditions might be very important for the iterative MUDD and CE.

Finally, total training overhead Δ_{tot} (see Section II-C) vs. the inverse channel spread factor $\Upsilon = T_{\text{coh}}/M$ for the iterative receiver with LMMSE-PIC MUDD and LMMSE-CE is shown in Fig. 6. The average SNR is $\overline{\text{snr}} = 6$ dB, user load is fixed at $\alpha = 1.8$ and the target bit error rate BER is set to 10^{-5}. Conventional pilot-aided channel estimation and the probability biased signaling introduced in Section II-C are compared and probability-biased signaling is shown to have the potential to significantly reduce training overhead..

ACKNOWLEDGMENT

The work of M. Vehkaperä was supported by the Norwegian Research Council under grant 171133/V30. K. Takeuchi was supported in part by the Grant-in-Aid for Young Scientists (Start-up) (No. 21860035) from MEXT, Japan. Tanaka and Takeuchi acknowledge support through Grant-in-Aid for Scientific Research on Priority Areas (No. 18079010) from MEXT, Japan.

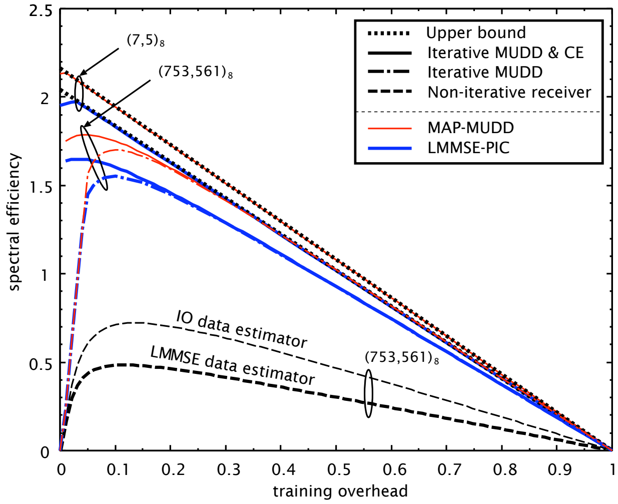

Figure 5. Spectral efficiency $\alpha R(1-\Delta_{\text{tr}})$ vs. the training overhead Δ_{tr}. Average SNR of 6 dB, inverse channel spread factor $\Upsilon = 80$, target BER $\leq 10^{-5}$, and convolutional code $(5,7)_8$ or $(561,753)_8$.

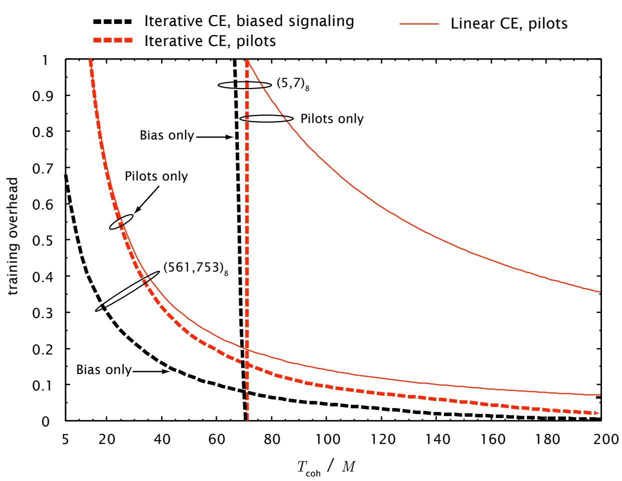

Figure 6. Minimum training overhead vs. $\Upsilon = T_{\text{coh}}/M$ for target BER $\leq 10^{-5}$. Average SNR of 6 dB, LMMSE channel estimator, LMMSE-PIC MUDD and convolutional code $(5,7)_8$ or $(561,753)_8$. Load $\alpha = 1.8$.

References

[1] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[2] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, UK: Cambridge University Press, 2008.

[3] A. Worthen and W. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.

[4] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.

[5] G. Caire, R. R. Müller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1950–1973, Sep. 2004.

[6] H. Li and H. V. Poor, "Impact of channel estimation errors on multiuser detection via the replica method," *EURASIP J. Wirel. Commun. Netw.*, vol. 2005, no. 2, pp. 175–186, Apr. 2005.

[7] P. Ting, C.-K. Wen, J.-C. Chen, and J.-T. Chen, "BER analysis of the optimum multiuser detection with channel mismatch in MC-CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 24, no. 6, pp. 1221–1235, Jun. 2006.

[8] M. Vehkaperä, K. Takeuchi, R. R. Müller, and T. Tanaka, "Iterative channel and data estimation: Framework and analysis via replica method," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun. 28 – Jul. 3 2009, pp. 2689–2693.

[9] ——, "On asymptotic performance of iterative channel and data estimation in large DS-CDMA systems," in *Proc. PHYSCOMNET / WiOPT*, Seoul, Korea, Jun. 27 2009, pp. 1–5.

[10] P. D. Alexander and A. J. Grant, "Iterative channel and information sequence estimation in CDMA," in *Proc. IEEE Int. Symp. Spread Spectrum Techniques and Applications*, vol. 2, New Jersey, USA, Sep. 6–8 2000, pp. 593–597.

[11] G. Caire, A. Tulino, and E. Biglieri, "Iterative multiuser joint decoding and parameter estimation: A factor-graph approach," in *Proc. IEEE Inform. Theory Workshop*, Cairns, Australia, Sep. 2–7 2001, pp. 36–38.

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

[13] 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 Processing*, vol. 55, no. 5, pp. 1981–1993, May 2007.

[14] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "Practical signaling with vanishing pilot-energy for large noncoherent block-fading MIMO channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, 2009, pp. 759–763.

[15] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[16] D. Guo and T. Tanaka, "Generic multiuser detection and statistical physics," in *Advances in Multiuser Detection*, M. Honig, Ed. Wiley-IEEE Press, 2009.

[17] S. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.

[18] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. New York: Oxford University Press, 2001.

[19] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[20] R. R. Müller, "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2821–2828, Nov. 2003.

[21] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[22] K. Takeuchi, T. Tanaka, and T. Yano, "Asymptotic analysis of general multiuser detectors in MIMO DS-CDMA channels," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 486–496, Apr. 2008.

[23] R. R. Müller, D. Guo, and A. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 530–540, Apr. 2008.

[24] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[25] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[26] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.

[27] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[28] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1-2, pp. 1–153, 2008.

[29] 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. 20, no. 2, pp. 284–287, Mar. 1974.

[30] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[31] K. Larsen, "Short convolutional codes with maximal free distance for rates 1/2, 1/3, and 1/4," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 371–372, May 1973.