Turbo Equalization/Estimation of Doubly Selective Channels using Basis Expansion and Tree Search

Sung-Jun Hwang and Philip Schniter*

***Abstract*—For turbo reception of coded transmissions over unknown doubly selective channels, such as time-varying ISI channels or frequency-varying ICI channels, we propose several soft noncoherent equalizers based on basis expansion (BE) channel modeling and tree-search, as a departure from traditional designs based on autoregressive (AR) channel modeling and/or trellis processing. By "noncoherent," we mean an equalizer that operates in the absence of channel state information. We begin by deriving the optimal BE-based soft noncoherent equalizer, whose complexity is shown to be impractical. We then propose a near-optimal approximation, based on soft tree-search and leveraging a fast recursive metric update, whose per-symbol complexity is only quadratic in the number of BE coefficients. Finally, we propose a different approach to soft noncoherent equalization that results from an application of the space-alternating generalized expectation-maximization (SAGE) algorithm. Using a tree-search-based practical implementation, the per-symbol complexity of this latter scheme is, for the multicarrier case, only linear in the number of BE coefficients. Numerical experiments demonstrate coded bit error rates near genie-aided bounds, as well as robustness to Doppler-spread mismatch.**

***Index Terms*—Turbo decoding, noncoherent decoding, equalization, channel estimation, semi-blind methods, basis expansion models, time-varying frequency-selective channels, doubly selective channels, doubly dispersive channels, expectation maximization, SAGE.**

I. INTRODUCTION

In this paper, we consider the problem of decoding a data sequence transmitted over an *unknown doubly selective* (DS) channel, such as a time-varying inter-symbol interference (ISI) channel or a frequency-varying inter-carrier interference (ICI) channel. Such channels occur in, e.g., shallow-water underwater acoustic and wideband mobile radio applications. In particular, we are interested in the case of coded transmissions with possibly long codewords (from, e.g., LDPC codes). A practical and near-optimal strategy for equalization in this scenario follows from the turbo principle [3], [4], which suggests to iterate between "soft noncoherent" equalization and soft decoding (see Fig. 1). By "soft noncoherent," we mean that the equalizer's role is to produce posterior bit probabilities from the received samples, pilots, and prior bit probabilities supplied by the soft decoder, in the absence of channel state information. Noncoherent equalizers are also referred to as "semi-blind" in the literature (e.g., [5]).

Optimal soft noncoherent equalization requires evaluating a noncoherent metric for every possible bit sequence and then summing over subsets of these metrics (as shown in [6] for Gauss-Markov channels). Since the number of possible bit sequences grows exponentially in the sequence length, practical implementation demands a suboptimal approach. Broadly speaking, suboptimal approaches fall into one of two categories: iterative channel-estimation and equalization (ICEE), or joint channel-estimation and equalization (JCEE). ICEE methods iterate between learning the channel coefficients and learning the coded bits, whereas JCEE methods attempt to simultaneously learn the channel and bits. (See [7] for a detailed discussion.) In both cases, the channel coefficients may represent time-varying ISI or frequency-varying ICI. Furthermore, one might choose to parameterize the time- or frequency-variation using a basis expansion (BE) model [8], [9] and/or an auto-regressive (AR) model [10].

Several ICEE approaches (e.g., [2], [5], [11]–[13]), have been proposed as incarnations of the expectation-maximization (EM) algorithm [14]. To our knowledge, this idea was first proposed for frequency-selective channels in [11], and later extended to doubly selective (DS) channels that use an AR model for coefficient time-variation (e.g., [13]) and a BE model for coefficient time-variation (e.g., [2], [5]). Of these works, most employ the trellis-based BCJR [15] algorithm for soft coherent[1] equalization. The principal drawback to BCJR is its complexity, $\mathcal{O}(|\mathbb{S}|^{N_h})$, which is impractical for large channel lengths N_h, even when the alphabet size $|\mathbb{S}|$ is modest. An alternative, which we explore in this paper, is tree-search based soft coherent equalization, as used in [2].

ICEE approaches have also been proposed that iterate a soft coherent equalizer (e.g., [16]–[21]) with a soft[2] DS channel estimator (e.g., [21]–[24]) without explicitly considering the optimality of their interaction. Motivated by the high cost of BCJR, reduced-complexity soft coherent equalization schemes have been proposed based on linear methods (e.g., [16], [20], [21]), soft interference cancellation (e.g., [17], [18]), reduced-state trellis techniques (e.g., [19]), and—as previously mentioned—soft tree-search (e.g., [25], [26]). Meanwhile, a number of soft DS channel estimation techniques have been proposed that support deterministic channel models, via LMS and RLS adaptation (e.g., [23]); AR time-domain variation,

The authors are with the Department of Electrical and Computer Engineering at The Ohio State University, Columbus, OH.

Please direct all correspondence to Prof. Philip Schniter, Dept. ECE, 2015 Neil Ave., Columbus OH 43210, e-mail: schniter@ece.osu.edu, phone 614.247.6488, fax 614.292.7596. Sung-Jun Hwang can be reached at shwang@qualcomm.com.

This work was supported in part by the National Science Foundation under Grant 0237037 and the Office of Naval Research grant N00014-07-1-0209.

Portions of this work were presented in [1] and [2].

[1]By "coherent" equalization, we mean that the equalizer has access to channel state estimates.

[2]By "soft" channel estimation, we mean that the estimator is able to use soft symbol estimates.

via Kalman techniques (e.g., [22]); and BE time- or frequency-domain variation (e.g., [21]). While it is also possible to use an AR model for frequency-domain variation, this approach is not as effective—as we shall see later.

As an alternative to the ICEE approaches described above, one might consider suboptimal JCEE schemes, as long as their complexity is far below that of optimal JCEE. Many of the previously proposed suboptimal JCEE schemes are trellis-based, and can be recognized as extensions of the (coherent) BCJR algorithm where the trellis is expanded to allow conditional AR-coefficient estimation at each state (e.g., [27]–[30]). In the related "fixed-lag" approach (e.g., [6], [29]–[31]), the problem is relaxed to one of computing each bit posterior using only a local subset of the observations, again using an expanded trellis. Like BCJR, the complexities of these trellis-based methods grow exponentially in the channel length N_h, and are thus impractical when N_h is large. Moreover, as mentioned earlier, AR models are not as effective as BE models for frequency-varying ICI. A very different approach to JCEE of DS channels was recently proposed in [32], [33], leveraging the fact that—when the finite-alphabet symbol property is ignored—nonlinear Kalman filtering techniques become admissible. This approach was initially proposed for AR-modeled channels [32] and later extended to BE-modeled channels [33]. A third JCEE approach, which we will elaborate on in the main body of this paper, is based on soft tree-search with per-sequence BE-coefficient estimation [1]. This third approach should not be confused with *hard* noncoherent equalization via sphere decoding (a form of tree-search) and BE-coefficient estimation [34], since turbo reception requires that the equalizer accept and produce *soft* bit estimates.

TABLE I
PER-SYMBOL COMPLEXITY OF SEVERAL SOFT NONCOHERENT EQUALIZERS, WHERE N_h IS THE CHANNEL'S DISCRETE DELAY SPREAD AND N_{D} ITS DISCRETE DOPPLER SPREAD, N IS THE BLOCK SIZE, $|\mathbb{S}|$ IS THE CONSTELLATION SIZE, AND N_f IS A LINEAR FILTER LENGTH.

algorithm	single carrier	multicarrier
INC or "$(\mathsf{sBE+cT})^K$"	$\mathcal{O}(N_{\mathsf{D}}^2 N_h)$	$\mathcal{O}(N_{\mathsf{D}} N_h \log_2 N)$
SNC or "ncT-BE"	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^2)$	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^2)$
BW-BE [5]	$\mathcal{O}(N_h^2 \lvert\mathbb{S}\rvert^{N_h} + N_h^2 N_{\mathsf{D}}^2)$	-
ICE-TE [21]	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)$
FL-EKF-BEM [33]	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^3)$	-
FL-EKF [32]	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^3)$	-
SKTE [22]	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^2 + N_h \lvert\mathbb{S}\rvert^{N_h})$	-
LE-RLS [23]	$\mathcal{O}(N_h^3 + N_f^3)$	-
APP-SD-KF [13]	$\mathcal{O}(N_{\mathsf{D}}^2 N_h^2 \lvert\mathbb{S}\rvert^{N_h})$	-

As can be seen from the discussion above, many approaches have been proposed for soft noncoherent equalization of DS channels. To help put these approaches into perspective, Table I[3] lists the complexity orders of several recently proposed algorithms. The goal, as we see it, is to minimize complexity while maintaining near-optimal performance.

[3]In constructing Table I, an elaboration of [32, Table II], we assumed that the equalization delay used in [32], [33] is proportional to N_h (as suggested in [32]). For single-carrier schemes, N_{D} corresponds to the BE model order as well as the AR model order (as suggested in [33]), and N is the BEM period. For multicarrier schemes, N_{D} corresponds to the ICI spread, and N is the number of subcarriers.

In accordance with this goal, we propose two novel methods of soft noncoherent equalization, both based on the combination of soft tree-search with generic BE channel modeling. Our "sequential noncoherent" (SNC) equalizer can be categorized as JCEE, and our "iterative noncoherent" (INC) equalizer can be categorized as EM-based ICEE. Our use of generic BE models facilitates a unified treatment of different channel types (e.g., time-variant ISI channels, frequency-variant ICI channels, and sparse versions of those channels), and our use of soft tree-search leverages recent ideas from the flat-fading multiple-input multiple-output (MIMO) literature (e.g., [25], [26]), facilitating an efficient tradeoff between performance and complexity. Our specific contributions are as follows.

1) We first derive the *optimal* soft noncoherent equalizer of BE-modeled doubly selective channels for a very general class of block modulation schemes that includes single- or multi-carrier schemes, cyclic- or zero-prefix, and rectangular or non-rectangular windowing. This optimal scheme involves the computation of a noncoherent metric for every possible bit sequence. Although the metrics do not explicitly involve channel estimates, we show that the metric can be recursively computed in a way that implicitly involves per-sequence MMSE estimates of the BE coefficients.
2) As an approximation of the optimal soft noncoherent equalizer, we propose a "sequential noncoherent" (SNC) equalizer that performs soft tree-search using the M-algorithm [35]. Our SNC scheme incurs a per-symbol complexity of $\mathcal{O}(N_{\mathsf{D}}^2 N_h^2)$, where N_h is the channel's discrete delay spread, and N_{D} is its discrete Doppler spread (i.e., the BE model order in the single-carrier case or the ICI spread in the multicarrier case). We note that SNC's complexity compares favorably[4] with the existing methods in Table I, given that N_h is often the dominant factor in practice.
3) Motivated by the possibility of *further* reducing the complexity dependence on N_h, we propose a novel "iterative noncoherent" (INC) scheme using the space-alternating generalized-EM (SAGE) framework from [36]. For the single-carrier case, our INC scheme performs soft channel estimation with per-symbol complexity $\mathcal{O}(N_{\mathsf{D}}^2 N_h)$ and, for the multicarrier case, with complexity $\mathcal{O}(N_{\mathsf{D}} N_h \log_2 N)$, where N is the number of subcarriers. In both cases, soft coherent equalization uses an $\mathcal{O}(N_{\mathsf{D}} N_h)$-complexity soft tree-search based on the M-algorithm. To our knowledge, complexity that depends *linearly* on N_h is unprecedented.
4) Finally, we discuss practical implementation details and numerically analyze the proposed methods in a turbo framework, demonstrating coded BER performance close to genie-aided bounds and robustness to BE choice and to Doppler-spread knowledge.

The system model is described in Section II, the optimal soft noncoherent equalizer and its sequential approximation

[4]With the exception of [21], all other approaches in Table I scale at least cubically in N_h. In comparing to [21], we note that $N > N_{\mathsf{D}} N_h$ for underspread channels, so that the complexity of our scheme becomes more favorable as the channel becomes more underspread.

in Section III, and the SAGE-based equalizer in Section IV. Implementational details are discussed in Section V, numerical results in Section VI, and conclusions in Section VII.

Notation: We use $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ to denote conjugate, transpose and Hermitian transpose, respectively. We write the k^{th} entry of vector $\boldsymbol{x}$ as $[\boldsymbol{x}]_k$, and the $(k,l)^{th}$ entry of matrix $\boldsymbol{A}$ as $[\boldsymbol{A}]_{k,l}$. The $N\times N$ identity matrix is denoted by $\boldsymbol{I}_N$, and the circular complex normal distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$ is denoted by $\mathcal{CN}(\boldsymbol{m},\boldsymbol{C})$. For vector norms, we use $\|\boldsymbol{x}\| \triangleq \sqrt{\boldsymbol{x}^H\boldsymbol{x}}$ and $\|\boldsymbol{x}\|_{\boldsymbol{A}} \triangleq \sqrt{\boldsymbol{x}^H\boldsymbol{A}\boldsymbol{x}}$, where $\boldsymbol{A}$ is positive semi-definite Hermitian. $\Re\{\boldsymbol{x}\}$ denotes taking the real part of a complex-valued vector $\boldsymbol{x}$. Finally, $\odot$ denotes the elementwise product of matrices, $\mathcal{D}_d(\boldsymbol{x})$ denotes the diagonal matrix constructed from the d^{th} cyclic down-shift of vector $\boldsymbol{x}$, and $\mathcal{D}(\boldsymbol{x})$ is shorthand for $\mathcal{D}_0(\boldsymbol{x})$.

II. System Model

At the transmitter, we assume that information bits $\{b_m^{(j)}\}$, are rate-R coded, interleaved, and mapped to 2^Q-ary QAM symbols. Groups of N_s information symbols are then combined with pilot and guard symbols to form symbol blocks of length $N \geq N_s$. We denote the j^{th} symbol block by $\boldsymbol{s}^{(j)} = [s_0^{(j)},\ldots,s_{N-1}^{(j)}]^T$, where $s_n^{(j)} \in \mathbb{S}$ for symbol alphabet $\mathbb{S}$, and the corresponding coded bit vector by $\boldsymbol{x}^{(j)} = [x_0^{(j)},\ldots, x_{N_sQ-1}^{(j)}]^T$, where $x_k^{(j)} \in \{0,1\}$. The symbols are then linearly block-modulated by either a single-carrier scheme or a multi-carrier scheme, represented by $\boldsymbol{G}\in\mathbb{C}^{N_t\times N}$ with $N_t \geq N$, to form the transmitted signal $\boldsymbol{t}^{(j)} \triangleq [t_0^{(j)},\ldots,t_{N_t-1}^{(j)}]^T = \boldsymbol{G}\boldsymbol{s}^{(j)}$. The construction of $\boldsymbol{G}$ will be described later.

At the channel output, the samples in the j^{th} received block $\boldsymbol{r}^{(j)} \triangleq [r_0^{(j)},\ldots,r_{N_r-1}^{(j)}]^T$ are assumed to take the form

$$r_n^{(j)} = \sum_{l=0}^{N_h-1} h_{n,l}^{(j)} t_{n-l}^{(j)} + \nu_n^{(j)}, \tag{1}$$

where $h_{n,l}^{(j)}$ is the time-n response of the channel to an impulse applied at time-$(n-l)$, where N_h is the discrete channel delay spread, and where $\{\nu_n^{(j)}\}$ is zero-mean circular white Gaussian noise (CWGN).

The received vector $\boldsymbol{r}^{(j)}$ is then linearly (single- or multi-carrier) demodulated via matrix $\boldsymbol{\Gamma}\in\mathbb{C}^{N\times N_r}$ to yield

$$\boldsymbol{y}^{(j)} = \underbrace{\boldsymbol{\Gamma}\boldsymbol{\mathcal{H}}^{(j)}\boldsymbol{G}}_{\triangleq\, \boldsymbol{H}^{(j)}}\boldsymbol{s}^{(j)} + \boldsymbol{w}^{(j)}. \tag{2}$$

In (2), $\boldsymbol{w}^{(j)} = \boldsymbol{\Gamma}\boldsymbol{\nu}^{(j)}$ and $\boldsymbol{\mathcal{H}}^{(j)} \in \mathbb{C}^{N_r\times N_t}$ is a convolution matrix constructed from the channel's time-varying impulse response according to $[\boldsymbol{\mathcal{H}}^{(j)}]_{n,n-l} = h_{n,l}^{(j)}$. Thus $N_r = N_t + N_h - 1$ and $\boldsymbol{\mathcal{H}}^{(j)}$ is banded with bandwidth N_h. Note that $\boldsymbol{H}^{(j)}$ represents the composite effect of modulation, channel propagation, and demodulation. When the single- or multicarrier scheme is appropriately designed, $\boldsymbol{H}^{(j)}$ can be closely approximated by a "circularly banded" matrix with bandwidth N_H, as illustrated in Fig. 2(a) [7]. For example,

- In single-carrier block zero-padded[5] schemes, $\boldsymbol{G} = \boldsymbol{I}_N$ (so that $N_t = N$) and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I}_{N_h-1} & \boldsymbol{0} & \boldsymbol{I}_{N_h-1} \\ \boldsymbol{0} & \boldsymbol{I}_{N-N_h+1} & \boldsymbol{0} \end{bmatrix}. \tag{3}$$

  Thus $\boldsymbol{H}^{(j)}$, with bandwidth $N_H = N_h$, contains the impulse response coefficients $\{h_{n,l}^{(j)}\}$.
- In cylic-prefixed[6] multicarrier modulation, $\boldsymbol{G} = \mathcal{D}(\boldsymbol{g})\boldsymbol{F}_t^H$, where $\boldsymbol{F}_t^H \in \mathbb{C}^{N_t\times N}$ is a period-N unitary IDFT matrix cyclically extended in the row dimension, and where $\mathcal{D}(\boldsymbol{g})$ is a diagonal matrix created from a time-domain transmission pulse $\boldsymbol{g}\in\mathbb{C}^{N_t}$. Then $\boldsymbol{\Gamma} = \boldsymbol{F}_r\mathcal{D}(\boldsymbol{\gamma}\odot\boldsymbol{m})$, where $\boldsymbol{F}_r \in \mathbb{C}^{N\times N_r}$ is a period-N unitary DFT matrix cyclically extended in the column dimension, $\boldsymbol{\gamma}\in\mathbb{C}^{N_r}$ is a time-domain reception pulse, and $[\boldsymbol{m}]_n = \exp(j\frac{2\pi}{N}\frac{N_{\mathsf{D}}-1}{2}n)$. With appropriate design of $\boldsymbol{g}$ and $\boldsymbol{\gamma}$ [39], the frequency-domain channel matrix $\boldsymbol{H}^{(j)}$ has bandwidth $N_H = N_{\mathsf{D}} \triangleq \lceil 2f_{\mathsf{D}}T_sN\rceil + \delta$ where f_{D} denotes the single-sided Doppler spread (in Hz), T_s denotes the channel-use interval (in sec), and δ is a (small) non-negative integer that controls out-of-band coefficient energy. The off-diagonal elements of $\boldsymbol{H}^{(j)}$ induce ICI.

We assume the last N_H-1 symbols in $\boldsymbol{s}^{(j)}$ are zero-valued guards, so $\boldsymbol{H}^{(j)}$ acts causally on the first $N-N_H+1$ symbols.

The equalizer employs an N_b-term BE model for the variation of the composite channel over the block. In particular, it models the d^{th} "cyclic" diagonal of $\boldsymbol{H}^{(j)}$, i.e., $\boldsymbol{h}_d^{(j)} \triangleq \big[[\boldsymbol{H}^{(j)}]_{0,-d}, [\boldsymbol{H}^{(j)}]_{1,1-d},\ldots,[\boldsymbol{H}^{(j)}]_{N-1,N-1-d}\big]^T$, as

$$\boldsymbol{h}_d^{(j)} \approx \boldsymbol{B}\boldsymbol{\eta}_d^{(j)}, \quad d=0,\ldots,N_H-1, \tag{4}$$

where $\boldsymbol{B}\in\mathbb{C}^{N\times N_b}$ is a matrix of basis vectors and $\boldsymbol{\eta}_d^{(j)}\in\mathbb{C}^{N_b}$ is a vector of BE coefficients. Note that the approximation in (4) can be made arbitrarily accurate via large enough N_b. With single-carrier modulation, the BE models channel variation in the time domain, so that $N_b = N_{\mathsf{D}}$ suffices (with appropriate choice of $\boldsymbol{B}$ and δ). With multicarrier modulation, the BE models channel variation in the frequency domain, so that $N_b = N_h$ suffices, with $\boldsymbol{B}$ being a truncated[7] DFT matrix [9]. In either case, $N_bN_H = N_hN_{\mathsf{D}}$. Assuming an accurate BE model (4), the received vector $\boldsymbol{y}^{(j)}$ from (2) becomes

$$\boldsymbol{y}^{(j)} = \boldsymbol{A}^{(j)}\boldsymbol{\theta}^{(j)} + \boldsymbol{w}^{(j)}, \tag{5}$$

where $\boldsymbol{\theta}^{(j)} \triangleq [\boldsymbol{\eta}_0^{(j)T},\ldots,\boldsymbol{\eta}_{N_H-1}^{(j)T}]^T \in\mathbb{C}^{N_bN_H}$ and

$$\boldsymbol{A}^{(j)} \triangleq \big[\mathcal{D}_0(\boldsymbol{s}^{(j)})\boldsymbol{B},\ldots,\mathcal{D}_{N_H-1}(\boldsymbol{s}^{(j)})\boldsymbol{B}\big]. \tag{6}$$

The receiver infers the information bits $\{b_m^{(j)}\}$ using the "turbo" principle: "soft" information on the coded bits $\boldsymbol{x}^{(j)}$, in the form of log-likelihood ratios (LLRs), is iteratively refined through alternating soft-equalization and soft-decoding steps, as shown in Fig. 1. The equalizer's task is to produce extrinsic LLRs given the observation $\boldsymbol{y}^{(j)}$ and the prior LLRs provided by the decoder (or, in the first turbo iteration, from pilots).

[5]We note that the same $\boldsymbol{H}^{(j)}$ is obtained in the context of single-carrier cyclic-prefix modulation [37] through a different choice of $\boldsymbol{G}$ and $\boldsymbol{\Gamma}$ [7].

[6]We note that the same $\boldsymbol{H}^{(j)}$ is obtained in the context of zero-padded multicarrier modulation [38] through a different choice of $\boldsymbol{G}$ and $\boldsymbol{\Gamma}$ [7].

[7]If the channel impulse response is sparse with known support, then $\boldsymbol{B}$ contains only those columns of the DFT matrix indexed by the support [40].

The equalizers we propose are "noncoherent" in that they treat the channel realization $\boldsymbol{\theta}^{(j)}$ as unknown. They treat channel statistics as known, however, assuming that $\boldsymbol{w}^{(j)} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ and $\boldsymbol{\theta}^{(j)} \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}^{(j)}, \boldsymbol{R}_\theta)$ for full rank $\boldsymbol{R}_\theta$. The selection of $\bar{\boldsymbol{\theta}}^{(j)}$ and $\boldsymbol{R}_\theta$ is discussed in Section V-A.

In Section III, we describe the optimal noncoherent equalizer and a practical implementation based on tree-search, and in Section IV we describe equalization based on the Bayesian SAGE algorithm. Because the equalization procedure is invariant to block index j, we suppress the "$^{(j)}$" notation in the sequel.

III. Sequential Noncoherent Equalization

A. Optimum Soft Noncoherent Equalization

The log-likelihood ratio (LLR) of coded bit x_k given $\boldsymbol{y}$, i.e.,

$$L(x_k|\boldsymbol{y}) \triangleq \ln \frac{\Pr[x_k = 1|\boldsymbol{y}]}{\Pr[x_k = 0|\boldsymbol{y}]},\ k \in \{0, \ldots, N_s Q - 1\}, \quad (7)$$

can be written in the form [25]

$$L(x_k|\boldsymbol{y}) = \ln \frac{\sum_{\boldsymbol{x}:x_k=1} p(\boldsymbol{y}|\boldsymbol{x}) \exp \boldsymbol{l}^T \boldsymbol{x}}{\sum_{\boldsymbol{x}:x_k=0} p(\boldsymbol{y}|\boldsymbol{x}) \exp \boldsymbol{l}^T \boldsymbol{x}}, \quad (8)$$

where $\boldsymbol{l} \triangleq [L_a(x_0), \ldots, L_a(x_{N_s Q})]^T$ such that $L_a(x_k) \triangleq \ln(\Pr[x_k = 1]/\Pr[x_k = 0])$ is the *a priori* LLR of x_k. The "extrinsic" LLR $L_e(x_k|\boldsymbol{y}) \triangleq L(x_k|\boldsymbol{y}) - L_a(x_k)$ then becomes

$$L_e(x_k|\boldsymbol{y}) = \ln \frac{\sum_{\boldsymbol{x}:x_k=1} \exp \mu(\boldsymbol{x})}{\sum_{\boldsymbol{x}:x_k=0} \exp \mu(\boldsymbol{x})} - L_a(x_k) \quad (9)$$

using the noncoherent MAP sequence metric

$$\mu(\boldsymbol{x}) \triangleq \ln p(\boldsymbol{y}|\boldsymbol{x}) + \boldsymbol{l}^T \boldsymbol{x}. \quad (10)$$

Since $\boldsymbol{\theta}$ and $\boldsymbol{w}$ in (5) are both Gaussian distributed, we have

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{A}\bar{\boldsymbol{\theta}}, \boldsymbol{A}\boldsymbol{R}_\theta \boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_N), \quad (11)$$

where $\boldsymbol{A}$ depends on the coded bits $\boldsymbol{x}$ through the corresponding symbols $\boldsymbol{s}$. Thus, with $\boldsymbol{\Phi} \triangleq \boldsymbol{A}\boldsymbol{R}_\theta \boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_N$, we get

$$\mu(\boldsymbol{x}) = -\|\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{\theta}}\|^2_{\boldsymbol{\Phi}^{-1}} - \ln(\pi^N \det \boldsymbol{\Phi}) + \boldsymbol{l}^T \boldsymbol{x}. \quad (12)$$

The sequence metrics $\mu(\boldsymbol{x})$ can be evaluated using an N_s-stage 2^Q-ary tree, where, the partial metrics

$$\mu(\boldsymbol{x}_n) \triangleq \ln p(\boldsymbol{y}_n|\boldsymbol{x}_n) + \boldsymbol{l}_n^T \boldsymbol{x}_n \quad (13)$$

are evaluated recursively. In (13), $\boldsymbol{x}_n \triangleq [\underline{x}_0^T, \ldots, \underline{x}_n^T]^T$ with $\underline{x}_i \triangleq [x_{iQ}, \ldots, x_{iQ+Q-1}]^T$, $\boldsymbol{l}_n \triangleq [\underline{l}_0^T, \ldots, \underline{l}_n^T]^T$ with $\underline{l}_i \triangleq [L_a(x_{iQ}), \ldots, L_a(x_{iQ+Q-1})]^T$, and $\boldsymbol{y}_n \triangleq [y_0, \ldots, y_n]^T$. Note that $\underline{x}_i$ and $\underline{l}_i$ correspond to the i^{th} symbol. The recursion is derived in Appendix A and summarized in Table II, where $\boldsymbol{b}_n^H$ denotes the n^{th} row of $\boldsymbol{B}$. It is straightforward to show that each recursion consumes $N_b^2 N_H^2 + 3N_b N_H + 7$ multiplications.

The Table II quantity $\hat{\boldsymbol{\theta}}_n$ can be written (see Appendix A):

$$\hat{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\theta}} + \boldsymbol{R}_\theta \boldsymbol{A}_n^H \boldsymbol{\Phi}_n^{-1}(\boldsymbol{y}_n - \boldsymbol{A}_n \bar{\boldsymbol{\theta}}), \quad (14)$$

which can be recognized as the $\boldsymbol{x}_n$-conditional MMSE estimate of $\boldsymbol{\theta}_n$ from $\boldsymbol{y}_n$. Using this fact, Appendix B shows that

$$\mu(\boldsymbol{x}_n) = -\tfrac{1}{\sigma^2}\|\boldsymbol{y}_n - \boldsymbol{A}_n \hat{\boldsymbol{\theta}}_n\|^2 + \boldsymbol{l}_n^T \boldsymbol{x}_n - \ln(\pi^N \det \boldsymbol{\Phi}_n) - \|\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}\|^2_{\boldsymbol{R}_\theta^{-1}}. \quad (15)$$

TABLE II

Fast Recursion for Evaluating $\mu(\boldsymbol{x}_n)$

From the old quantities:
$\mu(\boldsymbol{x}_{n-1})$, $\hat{\boldsymbol{\theta}}_{n-1}$, $\boldsymbol{\Sigma}_{n-1}^{-1}$, $[s_{n-1}, \ldots, s_{n-N_H+1}]$,
and the inputs:
y_n, s_n, $\underline{l}_n$, $\underline{x}_n$,
calculate the new quantities:
$\mu(\boldsymbol{x}_n)$, $\hat{\boldsymbol{\theta}}_n$, $\boldsymbol{\Sigma}_n^{-1}$, $[s_n, \ldots, s_{n-N_H+2}]$,
using the recursion:
$\boldsymbol{a}_n = [s_n \boldsymbol{b}_n^H, \cdots, s_{n-N_H+1}\boldsymbol{b}_n^H]^H$
$\boldsymbol{d}_n = \boldsymbol{\Sigma}_{n-1}^{-1}\boldsymbol{a}_n$
$\zeta_n = (1 + \boldsymbol{a}_n^H \boldsymbol{d}_n)^{-1}$
$e_n = y_n - \boldsymbol{a}_n^H \hat{\boldsymbol{\theta}}_{n-1}$
$\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_{n-1}^{-1} - \zeta_n \boldsymbol{d}_n \boldsymbol{d}_n^H$
$\mu(\boldsymbol{x}_n) = \mu(\boldsymbol{x}_{n-1}) - \frac{\zeta_n}{\sigma^2}|e_n|^2 + \ln(\frac{\zeta_n}{\pi\sigma^2}) + \underline{l}_n^T \underline{x}_n$
$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} + \zeta_n e_n \boldsymbol{d}_n$,
initializing (iff $n = 0$) with:
$\mu(\boldsymbol{x}_{-1}) = 0$, $\hat{\boldsymbol{\theta}}_{-1} = \bar{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{-1}^{-1} = \sigma^{-2}\boldsymbol{R}_\theta$.

From (15), we see that the noncoherent MAP metric $\mu(\boldsymbol{x})$ is the sum of a "coherent MAP metric" $-\frac{1}{\sigma^2}\|\boldsymbol{y}_n - \boldsymbol{A}_n \hat{\boldsymbol{\theta}}_n\|^2 + \boldsymbol{l}_n^T \boldsymbol{x}_n$, a "bias term" $-\ln(\pi^N \det \boldsymbol{\Phi}_n)$, and a term $-\|\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}\|^2_{\boldsymbol{R}_\theta^{-1}}$ which penalizes the deviation of the conditional estimate $\hat{\boldsymbol{\theta}}$ from the prior statistics $\boldsymbol{\theta}_n \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}, \boldsymbol{R}_\theta)$. Thus, the recursive MAP sequence metric evaluation implicitly uses per-sequence processing [41].

It should be noted that, when the alphabet $\mathbb{S}$ is symmetric, sufficient asymmetry in the apriori LLR structure $\{L_a(x_k)\}$ is needed to circumvent the phase-ambiguity that results from both channel and symbols being unknown. For this purpose, it suffices to insert one pilot symbol per block.

B. Practical Soft Sequential Noncoherent (SNC) Equalization

From (9), computation of exact soft outputs $L_e(x_k|\boldsymbol{y})$ is impractical because it requires evaluating and summing $\mu(\boldsymbol{x})$ for all $2^{N_s Q}$ hypotheses of $\boldsymbol{x}$. However, we expect the set $\{\exp \mu(\boldsymbol{x})\}$ to be dominated by a few "significant" bit vectors $\boldsymbol{x}$, which we collect into the set $\mathcal{S}$. Thus, we reason that near-optimal soft outputs will result from restricting the summations in (9) to $\boldsymbol{x} \in \mathcal{S}$, i.e.,

$$L_e(x_k|\boldsymbol{y}) \approx \ln \frac{\sum_{\boldsymbol{x}\in\mathcal{S}\cap\{\boldsymbol{x}:x_k=1\}} \exp \mu(\boldsymbol{x})}{\sum_{\boldsymbol{x}\in\mathcal{S}\cap\{\boldsymbol{x}:x_k=0\}} \exp \mu(\boldsymbol{x})} - L_a(x_k). \quad (16)$$

If desired, the "max-log" approximation $\sum_{\boldsymbol{x}:x_k=x} \exp \mu(\boldsymbol{x}) \approx \max_{\boldsymbol{x}:x_k=x} \mu(\boldsymbol{x})$ could be applied for further simplification:

$$L_e(x_k|\boldsymbol{y}) \approx \max_{\boldsymbol{x}\in\mathcal{S}\cap\{\boldsymbol{x}:x_k=1\}} \mu(\boldsymbol{x}) - \max_{\boldsymbol{x}\in\mathcal{S}\cap\{\boldsymbol{x}:x_k=0\}} \mu(\boldsymbol{x}) - L_a(x_k). \quad (17)$$

To find the significant bit vectors $\mathcal{S}$ and their metrics $\{\mu(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathcal{S}}$, we suggest a suboptimal breadth-first tree-search such as the M-algorithm or the T-algorithm [35]. The M-algorithm is particularly convenient because it yields a complexity that is invariant to channel realization and SNR. With search breadth M and the recursion in Table II, soft noncoherent equalization consumes only $\mathcal{O}(M 2^Q N_h^2 N_\mathsf{D}^2)$ operations per symbol (since $N_b N_H = N_h N_\mathsf{D}$ and $|\mathbb{S}| = 2^Q$). Furthermore, when the symbol constellation $\mathbb{S}$ satisfies a multi-level bit mapping, the complexity can be made nearly independent of Q, as discussed in [26], which is useful when Q is large.

Note that $\mathcal{S} \cap \{\boldsymbol{x} : x_k = 1\}$ or $\mathcal{S} \cap \{\boldsymbol{x} : x_k = 0\}$ may be empty for some k, which would make $L_e(x_k|\boldsymbol{y})$ infinite. For this, a simple solution is to clip $L_e(x_k|\boldsymbol{y})$ [26]. Note also that (arbitrarily placed) pilot symbols are easily incorporated by setting their apriori bit LLRs $\underline{l}_i$ to very large values.

IV. Noncoherent Equalization via SAGE

We now develop a soft iterative noncoherent (INC) equalizer based on the SAGE framework [36], a generalization of the EM framework [14] that allows updating of parameters in subsets that each use a different choice of hidden data. In our case, an estimate of $\boldsymbol{\theta}$ is updated one element at a time using $(\boldsymbol{z}_l, \boldsymbol{s})$ as the hidden data for θ_l, where $\boldsymbol{z}_l$ is defined as

$$\boldsymbol{z}_l \triangleq \boldsymbol{\alpha}_l \theta_l + \boldsymbol{w}, \tag{18}$$

with $\boldsymbol{\alpha}_l$ denoting the l^{th} column of $\boldsymbol{A}$. In the sequel, $\boldsymbol{\theta}_{\bar{l}}$ will be used to denote the vector $\boldsymbol{\theta}$ with the l^{th} term omitted, and $\boldsymbol{A}_{\bar{l}}$ used to denote $\boldsymbol{A}$ with the l^{th} column omitted, so that

$$\boldsymbol{y} = \boldsymbol{A}_{\bar{l}} \boldsymbol{\theta}_{\bar{l}} + \boldsymbol{z}_l. \tag{19}$$

We note that, when estimating θ_l, the hidden data $(\boldsymbol{z}_l, \boldsymbol{s})$ is "admissible" [36] because $p(\boldsymbol{y}|\boldsymbol{z}_l, \boldsymbol{s}, \boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{z}_l, \boldsymbol{s}, \boldsymbol{\theta}_{\bar{l}})$.

Our application of SAGE updates the estimate of θ_0 at iteration $i = 1$, then θ_1 at $i = 2$, and so on, until all $N_b N_H$ coefficients have been updated a total of K times. In particular, at the i^{th} iteration, we update index $l = (i \mod N_b N_H)$ as

$$\begin{aligned} \theta_l[i+1] = \arg\max_{\theta_l} \mathrm{E}\left\{\ln p(\boldsymbol{z}_l, \boldsymbol{s}|\theta_l, \boldsymbol{\theta}_{\bar{l}}[i])|\boldsymbol{y}, \boldsymbol{\theta}[i]\right\} \\ + \ln p(\theta_l, \boldsymbol{\theta}_{\bar{l}}[i]), \end{aligned} \tag{20}$$

while freezing the others (i.e., $\boldsymbol{\theta}_{\bar{l}}[i+1] = \boldsymbol{\theta}_{\bar{l}}[i]$). We adopt the Bayesian form of SAGE in (20), with $p(\boldsymbol{\theta}) \triangleq \mathcal{CN}(\boldsymbol{\theta}; \bar{\boldsymbol{\theta}}, \boldsymbol{R}_\theta)$. In Appendix C, we show that (20) reduces to

$$\begin{aligned} \theta_l[i+1] = \theta_l[i] + (\|\bar{\boldsymbol{\alpha}}_l\|^2 + c_{ll} + \sigma^2 \rho_{ll})^{-1} \\ \times \left(\bar{\boldsymbol{\alpha}}_l^H \boldsymbol{e} - \sigma^2 \boldsymbol{\rho}_l^H (\boldsymbol{\theta}[i] - \bar{\boldsymbol{\theta}}) - \boldsymbol{c}_l^H \boldsymbol{\theta}[i]\right), \end{aligned} \tag{21}$$

with $\boldsymbol{e} \triangleq \boldsymbol{y} - \bar{\boldsymbol{A}}\boldsymbol{\theta}[i]$, $\boldsymbol{\rho}_l \triangleq [\boldsymbol{R}_\theta^{-1}]_{:,l}$, $\rho_{ll} \triangleq [\boldsymbol{R}_\theta^{-1}]_{l,l}$, and with $\bar{\boldsymbol{\alpha}}_l \triangleq [\bar{\boldsymbol{A}}]_{:,l}$, $\boldsymbol{c}_l \triangleq [\boldsymbol{C}]_{:,l}$, and $c_{ll} \triangleq [\boldsymbol{C}]_{l,l}$ defined from

$$\bar{\boldsymbol{A}} = \left[\mathcal{D}_0(\boldsymbol{m})\boldsymbol{B} \cdots \mathcal{D}_{N_H-1}(\boldsymbol{m})\boldsymbol{B}\right] \tag{22}$$

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{B}^H \mathcal{D}_0(\boldsymbol{v})\boldsymbol{B} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{B}^H \mathcal{D}_{N_H-1}(\boldsymbol{v})\boldsymbol{B} \end{bmatrix}, \tag{23}$$

where $\boldsymbol{m} \triangleq [\bar{s}_0, \ldots, \bar{s}_{N-1}]^T$ and $\boldsymbol{v} \triangleq [v_0, \ldots, v_{N-1}]^T$ collect the $\boldsymbol{\theta}[i]$-conditional symbol means $\bar{s}_n \triangleq \mathrm{E}\{s_n|\boldsymbol{y}, \boldsymbol{\theta}[i]\}$ and variances $v_n \triangleq \mathrm{E}\{|s_n - \bar{s}_n|^2 \,|\boldsymbol{y}, \boldsymbol{\theta}[i]\}$.

Our SAGE-based soft INC equalization algorithm is summarized in Table III. Essentially, the algorithm alternates between channel (re)estimation (Step 1) and coherent soft equalization (Step 2). The input to Step 1 is the current channel estimate $\boldsymbol{\theta}[i]$ and the associated (coherent) posterior LLRs

$$L(x_k|\boldsymbol{y}, \boldsymbol{\theta}[i]) \triangleq \ln \frac{\Pr[x_k = 1|\boldsymbol{y}, \boldsymbol{\theta}[i]]}{\Pr[x_k = 0|\boldsymbol{y}, \boldsymbol{\theta}[i]]}, \tag{24}$$

from which $\Pr\left[s_n = s\middle|\boldsymbol{y}, \boldsymbol{\theta}[i]\right]$ is calculated and used to compute the symbol means and variances

$$\bar{s}_n = \sum_{s \in \mathbb{S}} s \Pr\left[s_n = s\middle|\boldsymbol{y}, \boldsymbol{\theta}[i]\right] \tag{25}$$

$$v_n = \sum_{s \in \mathbb{S}} |s - \bar{s}_n|^2 \Pr\left[s_n = s\middle|\boldsymbol{y}, \boldsymbol{\theta}[i]\right]. \tag{26}$$

Step 1 outputs an updated version of $\boldsymbol{\theta}[i]$, which Step 2 then uses to update the LLRs $L(x_k|\boldsymbol{y}, \boldsymbol{\theta}[i])$. To do this, we propose a tree-search based on the coherent MAP sequence metric

$$\begin{aligned} \mu(\boldsymbol{x}|\boldsymbol{\theta}[i]) &\triangleq \ln p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}[i]) p(\boldsymbol{x}) \qquad (27) \\ &= -\tfrac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}[i]\|^2 + \ln(\pi^N \sigma^{2N}) + \boldsymbol{l}^T \boldsymbol{x}, \qquad (28) \end{aligned}$$

which, when restricted to "significant" bit patterns $\boldsymbol{x} \in \mathcal{S}$, yields the approximated LLRs

$$L(x_k|\boldsymbol{y}, \boldsymbol{\theta}[i]) \approx \ln \frac{\sum_{\boldsymbol{x} \in \mathcal{S} \cap \{\boldsymbol{x}: x_k = 1\}} \exp \mu(\boldsymbol{x}|\boldsymbol{\theta}[i])}{\sum_{\boldsymbol{x} \in \mathcal{S} \cap \{\boldsymbol{x}: x_k = 0\}} \exp \mu(\boldsymbol{x}|\boldsymbol{\theta}[i])}. \tag{29}$$

For this tree-search, we propose to use breadth-first methods such as the M-algorithm, as previously suggested in the MIMO context [26].

TABLE III
SAGE-based Iterative Noncoherent (INC) Equalization

Initialize $i = 0$, $\boldsymbol{\theta}[0] = \bar{\boldsymbol{\theta}}$, and set LLRs according to pilots and (when available) previous decoder outputs.
Step 1. Update channel estimate $\boldsymbol{\theta}[i]$:
Compute soft symbol estimates $\boldsymbol{m}$ and $\boldsymbol{v}$.
Compute matrices $\bar{\boldsymbol{A}}$ and $\boldsymbol{C}$ (yielding $\bar{\boldsymbol{\alpha}}_l$, $\boldsymbol{c}_l$, and c_{ll} $\forall l$).
Set $\boldsymbol{e} = \boldsymbol{y} - \bar{\boldsymbol{A}}\boldsymbol{\theta}[i]$.
For $l = 0, \ldots, N_b N_H - 1$:
$\beta = (\|\bar{\boldsymbol{\alpha}}_l\|^2 + c_{ll} + \sigma^2 \rho_{ll})^{-1}$
$\theta_l[i+1] = \theta_l[i] + \beta\left[\bar{\boldsymbol{\alpha}}_l^H \boldsymbol{e} - \sigma^2 \boldsymbol{\rho}_l^H (\boldsymbol{\theta}[i] - \bar{\boldsymbol{\theta}}) - \boldsymbol{c}_l^H \boldsymbol{\theta}[i]\right]$
$\boldsymbol{e} \leftarrow \boldsymbol{e} + (\theta_l[i+1] - \theta_l[i])\bar{\boldsymbol{\alpha}}_l$
$\boldsymbol{\theta}_{\bar{l}}[i+1] = \boldsymbol{\theta}_{\bar{l}}[i]$
$i \leftarrow i + 1$
end.
Step 2. Update coded bit estimates:
Compute MAP metrics $\mu(\boldsymbol{x}|\boldsymbol{\theta}[i])$ for $\boldsymbol{x} \in \mathcal{S}$ via tree-search.
Compute posterior bit LLRs $\{L(x_k|\boldsymbol{y}, \boldsymbol{\theta}[i])\}$.
Repeat steps 1–2 a total of K times.
Output the final bit LLRs $\{L(x_k|\boldsymbol{y}, \boldsymbol{\theta}[K N_b N_H])\}$.

It is worth noting that the algorithm in Table III actually uses a modification of SAGE approach described in [36], in that the expectation in (20) is not recomputed at every $i = 0, 1, 2, \ldots$, but rather when i is a multiple of $N_b N_H$. In other words, it waits until *all* parameters in $\boldsymbol{\theta}$ have been updated before re-estimating the coded bits. This modification greatly reduces the overall computational complexity. Note that a direct implementation of the SAGE-based INC equalizer outlined in Table III requires $\mathcal{O}(N_b N_H)$ multiplications per symbol for Step 1 and $\mathcal{O}(N_b^2 N_H)$ multiplications per symbol for Step 2, for a grand total of $\mathcal{O}(K N_b^2 N_H)$ multiplications per symbol after K SAGE iterations.[8]

In the single-carrier case, $N_b = N_\mathsf{D}$ and $N_H = N_h$, implying an overall complexity of $\mathcal{O}(K N_\mathsf{D}^2 N_h)$, which is linear in discrete delay spread, N_h. Here, the quadratic dependence on

[8]We have verified numerically that the number of SAGE iterations K required for convergence does not scale with N, N_b, or N_H.

discrete Doppler spread, N_{D}, is not expected to be problematic because N_{D} is usually very small in practice.

In the multicarrier case, $N_b = N_h$ and $N_H = N_{\mathsf{D}}$, and so a direct implementation would require $\mathcal{O}(KN_h^2 N_{\mathsf{D}})$, which may be impractical when N_h is large (e.g., several hundred). However, we can now exploit the N-DFT structure of $\boldsymbol{B}$ and the $\mathcal{O}(N \log_2 N)$ complexity of the N-FFT to design an implementation with an overall per-symbol complexity of $\mathcal{O}(KN_h N_{\mathsf{D}} \log_2 N)$. To do this, we avoid explicit computation of $\boldsymbol{C}$ and instead notice that

$$c_{ll} = \left[\boldsymbol{B}^H \mathcal{D}_{l_1}(\boldsymbol{v})\boldsymbol{B}\right]_{l_2,l_2} = \frac{1}{N}\sum_{n=0}^{N-1} v_n \ \forall l \tag{30}$$

$$\boldsymbol{c}_l^H \boldsymbol{\theta}[i] = \left[\boldsymbol{C}^H \boldsymbol{\theta}[i]\right]_l = \left[\boldsymbol{B}^H \mathcal{D}_{l_1}(\boldsymbol{v})\boldsymbol{B}\boldsymbol{\theta}_{l_1}[i]\right]_{l_2} \tag{31}$$

for $l_1 \triangleq \lfloor l/N_b \rfloor$ and $l_2 \triangleq (l \mod N_b)$ and $\boldsymbol{\theta}_{l_1}[i] \triangleq \left[\boldsymbol{\theta}[i]\right]_{l_1 N_b:(l_1+1)N_b-1}$. Since $\boldsymbol{B}^H \mathcal{D}_{l_1}(\boldsymbol{v})\boldsymbol{B}\boldsymbol{\theta}_{l_1}[i]$ can be computed using i) an N-FFT, ii) N scalar multiples, and iii) an N-IFFT, each application of Step 1 requires only $\mathcal{O}(N_b N_H N \log_2 N)$ multiplies per symbol-block. The overall per-symbol complexity of this FFT-based multicarrier SAGE algorithm then becomes $\mathcal{O}(KN_b N_H \log_2 N)$, or equivalently $\mathcal{O}(KN_h N_{\mathsf{D}} \log_2 N)$ in the multicarrier case.

V. Implementational Considerations

A. *Choice of* $\bar{\boldsymbol{\theta}}^{(j)}$ *and* $\boldsymbol{R}_\theta$

Since both the sequential noncoherent (SNC) equalizer of Section III-B and the iterative noncoherent (INC) equalizer of Section IV employ the channel prior $\boldsymbol{\theta}^{(j)} \sim \mathcal{CN}(\bar{\boldsymbol{\theta}}^{(j)}, \boldsymbol{R}_\theta)$, it is worthwhile discussing the choice of $\bar{\boldsymbol{\theta}}^{(j)}$ and $\boldsymbol{R}_\theta$.

Under a Rayleigh fading assumption, one may be tempted to choose the non-informative prior $\bar{\boldsymbol{\theta}}^{(j)} = \boldsymbol{0}$. In doing so, however, the equalization of a symbols in the j^{th} block does not benefit from the knowledge of pilots (and, when available, previous decoder outputs) in *neighboring* blocks, whose BE coefficients $\{\boldsymbol{\theta}^{(j')}\}_{j' \neq j}$ may be strongly correlated with those of the current block. A simple way to exploit this knowledge is to set $\bar{\boldsymbol{\theta}}^{(j)}$ equal to the MMSE estimate of $\boldsymbol{\theta}^{(j)}$ based on out-of-block quantities. Note that, if only N_p out-of-block pilots are to be used, then the MMSE estimator for $\boldsymbol{\theta}^{(j)}$ can be computed in advance and implemented using only $\mathcal{O}(N_{\mathsf{D}} N_h N_p)$ operations, and reduced-rank techniques can further reduce the complexity [42]. If we want to incorporate out-of-block data-symbol estimates, then the MMSE estimator cannot be computed in advance. However, a procedure similar to Step 1 in Table III could be used to generate a near-MMSE estimate of $\boldsymbol{\theta}^{(j)}$, with per-symbol complexity $\mathcal{O}(N_{\mathsf{D}}^2 N_h)$ for single-carrier and $\mathcal{O}(N_{\mathsf{D}} N_h \log N)$ for multicarrier cases.

We recommend that the covariance $\boldsymbol{R}_\theta$ be constructed based on worst-case Doppler spread assumptions. In Section VI, a specific Doppler model is detailed and robustness to the assumed worst-case Doppler spread is investigated numerically.

B. *Pilot and Guard Patterns*

Recall that, in Section II, the last $N_H{-}1$ symbols in $\boldsymbol{s}$ were assumed to be zero-valued guards, so that $\boldsymbol{H}$ acts causally on the first $N{-}N_H{+}1$ symbols. This made the last $N_H{-}1$ columns of $\boldsymbol{H}$ inconsequential and allowed $\boldsymbol{H}$ to be treated as a lower-triangular N_H-banded matrix, a property that was exploited for both noncoherent and coherent tree-search. These guards notwithstanding, one may wonder whether the remaining $N-N_H$ symbols in each block should be data symbols, or whether a few should be dedicated as pilots or guards and—if so—how they should be arranged. Towards this aim, we review some related literature.

For communication over block-fading DS channels whose intra-block time-variation obeys a complex-exponential BE model, [43] derived the maximum achievable rate and showed that a pilot-aided system which places a cluster of $N_{\mathsf{D}} N_h$ pilots at the beginning of the block achieves this maximal rate. With a suboptimal receiver such as ours, however, there is no guarantee that this pilot pattern remains optimal, and in fact it is easy to show numerically that deviations from this pattern can yield improvements.

Other criteria have also been considered for pilot pattern design, such as minimizing the MSE attained during MMSE estimation of $\boldsymbol{\theta}^{(j)}$. For DS channels whose time-variation obeys a complex-exponential BE model, and for estimators which use only pilots within the current block, such “MMSE pilot patterns” were derived for single-carrier zero-padded schemes in [44], and, more generally, for the class of affine transmission schemes in [45]. Among the MMSE pilot patterns identified in [45] are single-carrier schemes with N_{D} “Kronecker-delta” pilot/guard clusters of length $2N_h - 1$ and multicarrier schemes with N_h “Kronecker-delta” pilot/guard clusters of length $2N_{\mathsf{D}} - 1$, recalling earlier heuristic designs [46]. While these MMSE patterns yield provably good channel estimates, they are rate-suboptimal in the sense that they do not allow full-rate transmission across the DS channel [43].

To conclude, the design of rate-maximizing pilot patterns for our *suboptimal* receivers remains an open problem. That said, [43]–[46] provide insights useful in constructing heuristic designs, as done in, e.g., [5], [47]. A particular family of patterns inspired by [43]–[46] is detailed in Section VI and examined numerically.

VI. Numerical Results

We now describe numerical experiments that compared the proposed equalizers to other approaches and performance bounds, for both single- and multicarrier cases.

1) Setup: Single- and multicarrier transmission schemes were then employed as described in Section II. In all experiments, the transmitter employed rate $R = \frac{1}{2}$ irregular low density parity check (LDPC) codes with average column-weight 3, generated by publicly available software [48]. The coded bits were block-interleaved by feeding them into an $8 \times (JQN_s/8)$ array column-wise, then reading them out row-wise. The interleaved bits were mapped to QPSK symbols (i.e., $Q = 2$) and partitioned into data blocks of length N_s, each of which was merged with N_p pilot/guards, as described below, to form a transmission block of length $N = N_s + N_p$. So that each codeword spanned $J = 32$ data blocks, $(JQN_s, RJQN_s)$-LDPC codes were employed. Unless otherwise noted, we used block length $N = 64$ with $N_p = 8$ pilot/guards per block.

The pilot/guard patterns illustrated in Fig. 2(b)-(c) were employed for single- and multicarrier cases, respectively. In the single-carrier case, each block contains $N_p - N_H + 1$ non-zero leading pilots and $N_H - 1$ zero-valued guards. In the multicarrier case, each block contains $\mathcal{K} \geq 1$ pilot/guard clusters, where each cluster contains $N_\mathsf{D} - 1$ leading[9] zero-valued guards and $N_p/\mathcal{K} - N_\mathsf{D} + 1$ trailing non-zero pilots. The cluster pattern repeats every $\mathcal{P} = N/N_p$ blocks, and the cluster locations are staggered[10] so that each subcarrier appears in a cluster exactly once every $\mathcal{P}$ blocks.

Jakes method [49] was used to generate realizations of a wide-sense stationary uncorrelated (WSSUS) Rayleigh fading channel, i.e., $\mathrm{E}\{h_{n,l} h^*_{n-m,l-\ell}\} = \xi_m \sigma_l^2 \delta_\ell$ with delay-power profile (DPP) σ_l^2 and temporal autocorrelation $\xi_m = J_0(2\pi f_\mathsf{D} T_s m)$. Here, $f_\mathsf{D} T_s$ denotes the normalized single-sided Doppler spread and $J_0(\cdot)$ the 0^{th}-order Bessel function of the first kind. To facilitate comparison of equalizers whose complexity grows rapidly in N_h, most experiments used the relatively short delay spread $N_h = 3$ with the uniform DPP $\sigma_l^2 = N_h^{-1}$. However, in some experiments that only tested the computationally efficient SAGE-based equalizer, $N_h = 64$ was used with an exponential DPP where the last tap was 20dB weaker than the first.

The case of $N_h = 3$ and $f_\mathsf{D} T_s = 0.002$ occurs, e.g., when a system with carrier frequency $f_c = 5.8$GHz and bandwidth 400kHz (i.e., $T_s = 2.5\mu$s) communicates over a channel with a maximum delay spread of 5μs and Doppler spread of $f_\mathsf{D} = v_{\max} f_c / c = 800$ Hz, where $v_{\max} = 149$km/h. On the other hand, the case of $N_h = 64$ and $f_\mathsf{D} T_c = 0.0005$ occurs, e.g., when a system with carrier frequency $f_c = 5.8$GHz and bandwidth 3.2MHz (i.e., $T_s = 0.3125\mu$s) communicates over a channel with a maximum delay spread of 20μs and Doppler spread of $f_\mathsf{D} = v_{\max} f_c / c = 1.6$kHz, where $v_{\max} = 298$km/h.

The BE models used by the equalizers were the following. In the single-carrier case, a Karhunen-Lóeve (KL) basis [50] was nominally used to model channel *time*-variation, i.e., $\boldsymbol{B}$ was constructed column-wise from the N_b principal eigenvectors of $\boldsymbol{R}_h \triangleq \mathrm{E}\{\boldsymbol{h}_d \boldsymbol{h}_d^H\}$ and diagonal $\boldsymbol{R}_\theta$ was constructed from the N_b principal eigenvalues of $\boldsymbol{R}_h$, where $N_b = 3$ was used in all cases. Robustness to the use of an oversampled complex-exponential (OCE) basis [51] is examined in Section VI-6. In the multicarrier case, a Fourier basis was used to model channel *frequency* variation, i.e., $\boldsymbol{B}$ was formed from N_h columns of the N-DFT matrix, and $\boldsymbol{R}_\theta$ was constructed to match the WSSUS statistics (as detailed in [52]). Unless otherwise specified, the mean $\bar{\boldsymbol{\theta}}^{(j)} = \mathbf{0}$ was assumed for the first turbo iteration.

For all tree-searches, the M-algorithm was used with search breadth $M = 64$, and the LLR magnitudes were clipped to 2.3 in the noncoherent case and 8 in the coherent case. The LDPC decoder by MacKay and Neal [53] was used with a maximum of 60 LDPC iterations, and turbo equalization was used with a maximum of 8 turbo iterations, unless otherwise noted. We specify the *maximum* number of iterations because the receiver breaks out of both the LDPC and turbo loops as soon as the LDPC syndrome check indicates error-free decoding. For SAGE, $K = 3$ iterations were used unless otherwise noted.

In the sequel, we refer to the proposed equalizers as

- ncT-BE: the proposed sequential noncoherent (SNC) approach, which uses the M-algorithm to perform a tree-search according to a noncoherent BE-structured metric,
- $(\text{sBE+cT})^K$: the proposed SAGE-based iterative noncoherent (INC) approach, which iterates soft BE-channel estimation with soft coherent tree-search, K times per turbo iteration.

We also investigate the performance of

- sBE+cT: soft BE-channel estimation followed by coherent tree-search (once per turbo iteration),
- sAR+cT: soft AR-channel estimation followed by coherent tree-search (once per turbo iteration),
- sAR+cB: soft AR-channel estimation followed by coherent BCJR (once per turbo iteration), equivalent to the "SKTE" method proposed in [22],

as well as two genie-aided performance upper-bounds:

- pH+cT: coherent tree-search using *perfect* knowledge of the channel $\boldsymbol{H}$,
- pllrBE+cT: coherent tree-search based on a BE-channel estimate constructed using *perfect LLR feedback*.

As discussed in Section IV, *coherent tree-search* (cT) uses the M-algorithm to sequentially maximize the metric $\ln p(\boldsymbol{x}|\boldsymbol{y}, \hat{\boldsymbol{H}})$ for externally supplied $\hat{\boldsymbol{H}}$—a direct application of the MIMO technique [26]. Meanwhile, *coherent BCJR* (cB) refers to the use of the trellis-based BCJR (or "forward-backward") algorithm [15] to calculate bit posteriors with an externally provided channel estimate. *Soft BE-channel estimation* (sBE) uses Step 1 of the SAGE iteration in Table III. Finally, *soft AR-channel estimation* (sAR) uses the Kalman technique from [22], for which we employed a second-order (i.e., three coefficient) AR model.

2) Effect of Number of Pilot/Guard Symbols: We first examine the effect of N_p, the number of pilot/guard symbols per N-block. Although we report only a particular single-carrier ncT-BE experiment, we observed similar behaviors in other settings. Figure 3 shows coded BER versus E_b/N_o for various N_p. As can be seen, the performance increases with N_p until $N_p = 8$ and then remains constant through $N_p = 11$. As N_p increases further, to $N_p = 14$, the BER-vs-E_b/N_o actually degrades, because the penalty on E_b/N_o overwhelms the reduction in channel estimation error.

The case $N_p = 3$, corresponding to the use of 1 non-zero pilot and 2 guard symbols, demonstrates the ability of the noncoherent ncT-BE equalizer to function reliably with only a single (non-zero) pilot symbol. Recall that one pilot suffices to circumvent the inherent phase ambiguity of symmetric $\mathbb{S}$.

3) Algorithm Comparison—Single-carrier Transmission: For single-carrier transmission, Figs. 4 and 5 compare the proposed soft noncoherent equalizers ncT-BE and $(\text{sBE+cT})^K$ to the approaches sBE+cT, sAR+cT, and sAR+cB, as well as to the performance bounds pH+cT and pllrBE+cT.

[9] Here, we use only leading guards because the multicarrier channel is causal. When $N_p = N_h N_\mathsf{D}$ and $\mathcal{K} = N_h$, this pattern coincides with the "frequency domain Kronecker delta" pattern of [45], [46].

[10] Note that, by cyclically shifting the elements of both $\boldsymbol{y}$ and $\boldsymbol{s}$, it is possible to place $N_H - 1$ guards at the end of the block while maintaining the "circularly banded" structure of $\boldsymbol{H}$ illustrated in Fig. 2(a).

Figure 4 examines a Doppler spread of $f_{\mathsf{D}}T_s = 0.002$, where the proposed ncT-BE and (sBE+cT)^3 perform only 2dB from the perfect-CSI bound pH+cT and only 1.7dB from the perfect-LLR-feedback bound pllrBE+cT. Interestingly, both tree-search-based algorithms outperform the trellis-based method sAR+cB by 0.6dB. Figure 5 examines the larger[11] Doppler spread of $f_{\mathsf{D}}T_s = 0.005$. There, ncT-BE still performs very well, maintaining a 3dB gap from the pllrBE+cT bound and outperforming other schemes by about 2dB. Meanwhile, (sBE+cT)^6 performs within 1dB of the trellis-based sAR+cB, which is impressive given that the per-symbol complexity of (sBE+cT)^6 grows linearly in N_h while that of sAR+cB grows as $\mathcal{O}(N_h|\mathbb{S}|^{N_h})$.

A comparison of the sAR+cT and sAR+cB traces in Figs. 4–5 shows that, for soft coherent equalization, the use of M-algorithm tree-search in place of optimal BCJR yields only about 1dB SNR loss. A comparison of the sBE+cT and sAR+cT traces in Figs. 4–5 suggests that, when used to model channel *time*-variation, a 3-term BE model performs similarly to a 3-term AR model (i.e., one may slightly outperform the other depending on E_b/N_o). As we shall see in Section VI-4, the story is different when modeling channel *frequency*-variation. A comparison of the sBE+cT and (sBE+cT)^K traces in Figs. 4–5 suggests that the use of $K>1$ SAGE iterations yields about 1dB SNR gain in most cases. (No additional gains were observed for $K>3$ when $f_{\mathsf{D}}T_s=0.002$, and $K>6$ when $f_{\mathsf{D}}T_s=0.005$).

4) Algorithm Comparison—Multicarrier Transmission: In Fig. 6, the single-carrier experiment of Fig. 4 was repeated for multicarrier transmission with ICI span $N_{\mathsf{D}} = 3$. While the channel remains identical to that in Fig. 4 (i.e., $N_h=3$ and $f_{\mathsf{D}}T_c=0.002$), we used $N_p = 9$ pilots (in $\mathcal{K}=1$ cluster) per block, initialized $\bar{\boldsymbol{\theta}}^{(j)}$ using pilots from $\mathcal{P}=4$ neighboring blocks, and tuned SAGE somewhat differently: $K = 6$ and, in Table III, Step 1 was repeated four times for each Step 2.

Consistent with the single-carrier experiment in Fig. 4, the multicarrier experiment in Fig. 6 shows (sBE+cT)^K performing on par with ncT-BE and about 0.5dB better than both sBE+cT and sAR+cB. However, unlike the single-carrier experiment, where the 3-term AR model performed on par with the 3-term BE model under common tree-search decoding, the multicarrier experiment in Fig. 6 shows sAR+cT performing significantly (1.5dB) worse than sBE+cT. This suggests that the AR model is not as well suited to modeling channel frequency-variation as it is to modeling channel time-variation.

Looking back over Figs. 4–6, we see that ncT-BE performed consistently well regardless of Doppler spread and transmission scheme, and that (sBE+cT)^K performed as well as ncT-BE in all but the very-high-Doppler single-carrier experiment. Given that (sBE+cT)^K is computationally cheaper than ncT-BE, it is not surprising to see some sacrifice in performance.

5) Performance under Large Delay Spread: In Fig. 7, we examined multicarrier performance under the larger delay spread of $N_h = 64$ and $f_{\mathsf{D}}T_c = 0.0005$. To mimic the sparsity of typical wireless channels, all but 20 randomly chosen channel taps were zeroed. Here, we used a block of $N = 256$ with $N_p = 64$ pilots arranged so that $\mathcal{K} = 2$ and $\mathcal{P} = 4$. The receiver used ICI span $N_{\mathsf{D}} = 3$ with a maximum of 16 turbo iterations, and $K = 3$ SAGE iterations were used with four repetitions of Step 1 for every Step 2. Figure 7 demonstrates that a large delay-spread channel can be effectively equalized (e.g., a 2.5dB gap to the genie bound at 10^{-3} BER) by the proposed $\mathcal{O}(KN_hN_{\mathsf{D}}\log_2 N_h)$ complexity SAGE-based equalizer.

6) Robustness to Statistical Mismatch: As discussed in Section V-A, the proposed noncoherent equalizers rely on certain assumptions about the BE-coefficient covariance matrix $\boldsymbol{R}_\theta$. We first examine the robustness of these schemes to the mismatch in the assumed maximum (normalized) Doppler spread $f_{\mathsf{D}}T_s$. For this, Fig. 8 plots the "E_b/N_o required to attain 10^{-2} BER" as a function of (true) $f_{\mathsf{D}}T_s$, comparing equalizers that know the true $f_{\mathsf{D}}T_s$ to those that assume the fixed value 0.002. Figure 8 demonstrates that the proposed equalization schemes, ncT-BE and (sBE+cT)^3, are both robust to mismatch in Doppler-spread, in that the performance of the $f_{\mathsf{D}}T_s$-fixed scheme closely tracks the performance of the $f_{\mathsf{D}}T_s$-aware scheme over the full range of (true) $f_{\mathsf{D}}T_s$.

Next, we examine the robustness of the proposed noncoherent equalizers to the use of an OCE basis (instead of the optimal KL basis) *in conjunction with* a mismatch in the assumed maximum Doppler spread $f_{\mathsf{D}}T_s$. For this, we used the OCE basis $[\boldsymbol{B}]_{n,l} = \exp(-j\frac{2\pi}{PN}n(l-\frac{N_b-1}{2}))$ for $l=\{0,\dots,N_b-1\}$ with $P=5$ and $N_b=3$. Constructing the figure in the same way as Fig. 8, we obtained Fig. 9, which looks remarkably similar to Fig. 8. In particular, for the ncT-BE traces, there is very little difference between Fig. 9 and Fig. 8. For the (sBE+cT)^3 traces, we see essentially no loss in performance from the use of OCE when the correct $f_{\mathsf{D}}T_s$ is applied, and approximately 1dB in loss when a mismatched $f_{\mathsf{D}}T_s$ is applied.

Robustness aside, the "U shape" of the curves in Figs. 8–9 gives insight into equalizer performance as a function of $f_{\mathsf{D}}T_s$. First note that the required E_b/N_o increases as $f_{\mathsf{D}}T_s$ becomes small. We attribute this behavior to lack of Doppler diversity. Likewise, the required E_b/N_o increases as $f_{\mathsf{D}}T_s$ becomes large. We attribute this behavior to the limitations of the N_b-term BE models for channel time-variation, where $N_b \ll N$. Furthermore, when $f_{\mathsf{D}}T_s$ is large, we see that ncT-BE significantly outperforms (sBE+cT)^3, regardless of whether $f_{\mathsf{D}}T_s$ is fixed or known, and regardless of whether KL or OCE is assumed for the basis. This behavior is perhaps not surprising given the fact that ncT-BE is computationally more demanding than (sBE+cT)^3.

VII. Conclusion

In this paper, we proposed two soft noncoherent equalizers that are applicable to single- or multicarrier transmissions over unknown doubly selective channels and suited for use in a turbo-equalizing receiver. In all cases, we exploited basis expansion (BE) models for channel (time or frequency) variation. To design our sequential noncoherent (SNC) equalizer, we started with an expression for the optimal noncoherent

[11]For ncT-BE, pllrBE+cT, and pH+cT we used a maximum of 8 turbo iterations, while for (sBE+cT)^6, sBE+cT, sAR+cT, and sAR+cB we used 16.

metric, showed that it can be evaluated sequentially, and then proposed an implementation based on M-algorithm tree-search whose per-symbol complexity grows as $\mathcal{O}(N_{\mathsf{D}}^2 N_h^2)$, where N_h is the channel's discrete delay spread and N_{D} its discrete Doppler spread. Motivated by further reduction in complexity, we also proposed an iterative noncoherent (INC) equalizer using the SAGE framework, which iterates between soft channel estimation and soft coherent equalization, the latter implemented using an M-algorithm tree-search. For single-carrier transmission, the per-symbol complexity of this INC equalizer grows as $\mathcal{O}(KN_{\mathsf{D}}^2 N_h)$, where K is the number of SAGE iterations. Here, the quadratic dependence on N_{D} is deemed tolerable since, in practice, N_{D} is very small. In the multicarrier case, we presented an FFT-based implementation of the SAGE technique with per-symbol complexity $\mathcal{O}(KN_{\mathsf{D}}N_h \log_2 N)$, where N is the number of subcarriers, which becomes advantageous as N_h becomes large. Numerical experiments show that the SNC and INC equalizers both perform reasonably close to genie-aided performance bounds and are robust to lack of knowledge of the true Doppler spread, $f_{\mathsf{D}}T_s$, as well as to the choice of BE model.

Appendix A
The Fast Recursive Update for $\mu(\boldsymbol{x}_n)$

First we write (13) as

$$\mu(\boldsymbol{x}_n) \;=\; -\|\boldsymbol{y}_n - \boldsymbol{A}_n\bar{\boldsymbol{\theta}}\|^2_{\boldsymbol{\Phi}_n^{-1}} - \ln(\pi^{n+1}\det\boldsymbol{\Phi}_n) + \boldsymbol{l}_n^T\boldsymbol{x}_n \quad (32)$$

$$\boldsymbol{\Phi}_n \;\triangleq\; \boldsymbol{A}_n\boldsymbol{R}_\theta\boldsymbol{A}_n^H + \sigma^2\boldsymbol{I}_{n+1}, \quad (33)$$

In the sequel, we use $\tilde{\boldsymbol{y}}_n \triangleq \boldsymbol{y}_n - \boldsymbol{A}_n\bar{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}_n \triangleq \hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}$. In the two sections below, we derive fast recursions for the first two terms in (32): $\mu_1(\boldsymbol{x}_n) \triangleq \tilde{\boldsymbol{y}}_n^H\boldsymbol{\Phi}_n^{-1}\tilde{\boldsymbol{y}}_n$ and $\mu_2(\boldsymbol{x}_n) \triangleq \ln\bigl(\pi^{n+1}\det\boldsymbol{\Phi}_n\bigr)$. Together, these recursions yield Table II.

A. Recursion for $\mu_1(\boldsymbol{x}_n)$

Rewriting $\boldsymbol{\Phi}_n$ with the aid of $\boldsymbol{A}_n = \left[\begin{smallmatrix}\boldsymbol{A}_{n-1}\\ \boldsymbol{a}_n^H\end{smallmatrix}\right]$, where $\boldsymbol{a}_n^H$ denotes the n^{th} row of $\boldsymbol{A}$, we have

$$\boldsymbol{\Phi}_n^{-1} \;=\; \begin{bmatrix}\boldsymbol{\Phi}_{n-1} & \boldsymbol{A}_{n-1}\boldsymbol{R}_\theta\boldsymbol{a}_n\\ \boldsymbol{a}_n^H\boldsymbol{R}_\theta\boldsymbol{A}_{n-1}^H & \boldsymbol{a}_n^H\boldsymbol{R}_\theta\boldsymbol{a}_n+\sigma^2\end{bmatrix}^{-1} = \begin{bmatrix}\boldsymbol{P}_n & \boldsymbol{p}_n\\ \boldsymbol{p}_n^H & p_n\end{bmatrix}, \quad (34)$$

for the block-inverse quantities

$$\boldsymbol{P}_n \;\triangleq\; \boldsymbol{\Phi}_{n-1}^{-1} + p_n^{-1}\boldsymbol{p}_n\boldsymbol{p}_n^H \quad (35)$$

$$\boldsymbol{p}_n \;\triangleq\; -\boldsymbol{\Phi}_{n-1}^{-1}\boldsymbol{A}_{n-1}\boldsymbol{R}_\theta\boldsymbol{a}_n p_n \quad (36)$$

$$p_n^{-1} \;\triangleq\; \sigma^2 + \boldsymbol{a}_n^H\left(\boldsymbol{R}_\theta - \boldsymbol{R}_\theta\boldsymbol{A}_{n-1}^H\boldsymbol{\Phi}_{n-1}^{-1}\boldsymbol{A}_{n-1}\boldsymbol{R}_\theta\right)\boldsymbol{a}_n. \quad (37)$$

Writing $\mu_1(\boldsymbol{x}_n)$ using (34) and $\tilde{\boldsymbol{y}}_n = \left[\begin{smallmatrix}\tilde{\boldsymbol{y}}_{n-1}\\ \tilde{y}_n\end{smallmatrix}\right]$, we get

$$\mu_1(\boldsymbol{x}_n) \;=\; \tilde{\boldsymbol{y}}_{n-1}^H\boldsymbol{P}_n\tilde{\boldsymbol{y}}_{n-1} + 2\Re\{\tilde{\boldsymbol{y}}_{n-1}^H\boldsymbol{p}_n\tilde{y}_n\} + p_n|\tilde{y}_n|^2. \quad (38)$$

Using the $\boldsymbol{s}_{n-1}$-conditional MMSE estimate of $\tilde{\boldsymbol{\theta}}$ from $\tilde{\boldsymbol{y}}_{n-1}$:

$$\tilde{\boldsymbol{\theta}}_{n-1} \;=\; \boldsymbol{R}_\theta\boldsymbol{A}_{n-1}^H\boldsymbol{\Phi}_{n-1}^{-1}\tilde{\boldsymbol{y}}_{n-1}, \quad (39)$$

we see that $\boldsymbol{r}_{n-1}^H\boldsymbol{p}_n = -\tilde{\boldsymbol{\theta}}_{n-1}\boldsymbol{a}_n p_n$. Applying this relationship to (35)-(37), we can rewrite (38) as

$$\begin{aligned}\mu_1(\boldsymbol{x}_n) \;&=\; \tilde{\boldsymbol{y}}_{n-1}^H\boldsymbol{\Phi}_{n-1}^{-1}\tilde{\boldsymbol{y}}_{n-1} + p_n\tilde{\boldsymbol{\theta}}_{n-1}^H\boldsymbol{a}_n\boldsymbol{a}_n^H\tilde{\boldsymbol{\theta}}_{n-1}\\ &\quad - 2p_n\Re\{\tilde{\boldsymbol{\theta}}_{n-1}^H\boldsymbol{a}_n\tilde{y}_n\} + p_n|\tilde{y}_n|^2 \quad (40)\\ &=\; \mu_1(\boldsymbol{x}_{n-1}) + p_n|\tilde{y}_n - \boldsymbol{a}_n^H\tilde{\boldsymbol{\theta}}_{n-1}|^2. \quad (41)\end{aligned}$$

Now we concentrate on p_n. Defining $\boldsymbol{\Sigma}_{n-1}$ and applying the matrix inversion lemma (MIL):

$$\boldsymbol{\Sigma}_{n-1} \;\triangleq\; \boldsymbol{A}_{n-1}^H\boldsymbol{A}_{n-1} + \sigma^2\boldsymbol{R}_\theta^{-1} \quad (42)$$

$$\sigma^2\boldsymbol{\Sigma}_{n-1}^{-1} \;=\; \boldsymbol{R}_\theta - \boldsymbol{R}_\theta\boldsymbol{A}_{n-1}^H\boldsymbol{\Phi}_{n-1}^{-1}\boldsymbol{A}_{n-1}\boldsymbol{R}_\theta, \quad (43)$$

we see from (37) that $p_n^{-1} = \sigma^2(1 + \boldsymbol{a}_n^H\boldsymbol{\Sigma}_{n-1}^{-1}\boldsymbol{a}_n)$. Using the fact that $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_{n-1} + \boldsymbol{a}_n\boldsymbol{a}_n^H$, a second application of the MIL yields $\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_{n-1}^{-1} - \zeta_n\boldsymbol{d}_n\boldsymbol{d}_n^H$ for

$$\boldsymbol{d}_n \;\triangleq\; \boldsymbol{\Sigma}_{n-1}^{-1}\boldsymbol{a}_n \quad (44)$$

$$\zeta_n \;\triangleq\; (1 + \boldsymbol{a}_n^H\boldsymbol{d}_n)^{-1} = p_n\sigma^2. \quad (45)$$

Together, this gives a fast update for $p_n = \zeta_n/\sigma^2$.

Finally, we tackle $\tilde{\boldsymbol{\theta}}_n$. Using the MIL again,

$$\boldsymbol{\Phi}_n^{-1} = \sigma^{-2}(\boldsymbol{I}_{n+1} - \boldsymbol{A}_n\boldsymbol{\Sigma}_n^{-1}\boldsymbol{A}_n^H), \quad (46)$$

which applied to (39) yields

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_n \;&=\; \tfrac{1}{\sigma^2}\boldsymbol{R}_\theta\bigl(\boldsymbol{\Sigma}_n - \boldsymbol{A}_n^H\boldsymbol{A}_n\bigr)\boldsymbol{\Sigma}_n^{-1}\boldsymbol{A}_n^H\tilde{\boldsymbol{y}}_n \quad (47)\\ &=\; \boldsymbol{\Sigma}_n^{-1}\boldsymbol{A}_n^H\tilde{\boldsymbol{y}}_n \quad (48)\\ &=\; \bigl(\boldsymbol{\Sigma}_{n-1}^{-1} - \zeta_n\boldsymbol{d}_n\boldsymbol{d}_n^H\bigr)\bigl(\boldsymbol{A}_{n-1}^H\tilde{\boldsymbol{y}}_{n-1} + \boldsymbol{a}_n\tilde{y}_n\bigr). \quad (49)\end{aligned}$$

Expanding (49) and applying $\boldsymbol{a}_n^H\boldsymbol{\Sigma}_n^{-1}\boldsymbol{a}_n = \zeta_n^{-1} - 1$, we get

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_n \;&=\; \tilde{\boldsymbol{\theta}}_{n-1} + \boldsymbol{d}_n\tilde{y}_n - \zeta_n\boldsymbol{d}_n\boldsymbol{a}_n^H\tilde{\boldsymbol{\theta}}_{n-1} - \zeta_n\boldsymbol{d}_n(\zeta_n^{-1}-1)\tilde{y}_n\\ &=\; \tilde{\boldsymbol{\theta}}_{n-1} + \zeta_n(\tilde{y}_n - \boldsymbol{a}_n^H\tilde{\boldsymbol{\theta}}_{n-1})\boldsymbol{d}_n. \quad (50)\end{aligned}$$

Notice that, in (41) and (50), $\tilde{y}_n - \boldsymbol{a}_n^H\tilde{\boldsymbol{\theta}}_{n-1} = y_n - \boldsymbol{a}_n^H\hat{\boldsymbol{\theta}}_{n-1}$.

B. Recursion for $\mu_2(\boldsymbol{x}_n)$

From (34), we can write

$$\boldsymbol{\Phi}_n = \begin{bmatrix}\boldsymbol{\Phi}_{n-1} & \boldsymbol{\phi}_n\\ \boldsymbol{\phi}_n^H & \phi_n\end{bmatrix}, \quad (51)$$

The Schur complement $\gamma_n \triangleq \phi_n - \boldsymbol{\phi}_n^H\boldsymbol{\Phi}_{n-1}^{-1}\boldsymbol{\phi}_n$ obeys [54]

$$\det(\boldsymbol{\Phi}_n) = \gamma_n\det(\boldsymbol{\Phi}_{n-1}). \quad (52)$$

Identifying $\boldsymbol{\phi}_n$ and ϕ_n from (34),

$$\begin{aligned}\gamma_n \;&=\; \sigma^2 + \boldsymbol{a}_n^H\boldsymbol{R}_\theta\boldsymbol{a}_n - \boldsymbol{a}_n^H\boldsymbol{R}_\theta\boldsymbol{A}_{n-1}^H\boldsymbol{\Phi}_{n-1}^{-1}\boldsymbol{A}_{n-1}\boldsymbol{R}_\theta\boldsymbol{a}_n\\ &=\; \sigma^2/\zeta_n \quad (53)\end{aligned}$$

using (43) and (45) for (53). Taking the logarithm of (52),

$$\mu_2(\boldsymbol{x}_n) = \mu_2(\boldsymbol{x}_{n-1}) + \ln(\pi\sigma^2/\zeta_n). \quad (54)$$

Appendix B
Derivation of (15)

The derivation is performed for full-block vectors rather than partial ones (e.g., $\boldsymbol{x}$ rather than $\boldsymbol{x}_n$), but applies to both. Applying the MIL to $\boldsymbol{\Phi}^{-1}$, the first term of (12) becomes

$$\tilde{\boldsymbol{y}}^H\boldsymbol{\Phi}^{-1}\tilde{\boldsymbol{y}} \;=\; \tfrac{1}{\sigma^2}\bigl(\tilde{\boldsymbol{y}}^H\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^H\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^H\tilde{\boldsymbol{y}}\bigr), \quad (55)$$

where $\boldsymbol{\Sigma} \triangleq \boldsymbol{A}^H\boldsymbol{A} + \sigma^2\boldsymbol{R}_\theta^{-1} = \boldsymbol{\Sigma}_{N-1}$ (via (42)) and the definition of $\tilde{\boldsymbol{y}}$ is from Appendix A. Writing

$$\tilde{\boldsymbol{y}}^H\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^H\tilde{\boldsymbol{y}} \;=\; 2\Re\{\tilde{\boldsymbol{y}}^H\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^H\tilde{\boldsymbol{y}}\} - \tilde{\boldsymbol{y}}^H\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^H\tilde{\boldsymbol{y}}$$

and plugging in $\tilde{\boldsymbol{\theta}} \triangleq \boldsymbol{\Sigma}^{-1}\boldsymbol{A}\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{\theta}}_{N-1}$ (via (48)), we find

$$\begin{aligned}\tilde{\boldsymbol{y}}^H\boldsymbol{\Phi}^{-1}\tilde{\boldsymbol{y}} \;&=\; \tfrac{1}{\sigma^2}\bigl(\tilde{\boldsymbol{y}}^H\tilde{\boldsymbol{y}} - 2\Re\{\tilde{\boldsymbol{y}}^H\boldsymbol{A}\tilde{\boldsymbol{\theta}}\} + \tilde{\boldsymbol{\theta}}^H\boldsymbol{\Sigma}\tilde{\boldsymbol{\theta}}\bigr) \quad (56)\\ &=\; \tfrac{1}{\sigma^2}\|\tilde{\boldsymbol{y}} - \boldsymbol{A}\tilde{\boldsymbol{\theta}}\|^2 + \|\tilde{\boldsymbol{\theta}}\|^2_{\boldsymbol{R}_\theta^{-1}}. \quad (57)\end{aligned}$$

The definitions of $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{\theta}}$ then yield (15).

Appendix C
Derivation of (21)

We analyze the two terms in (20) separately. For the first term, we recognize that $\ln p(\boldsymbol{z}_l, \boldsymbol{s}|\theta_l, \boldsymbol{\theta}_{\bar{l}}[i]) = \ln p(\boldsymbol{z}_l|\boldsymbol{s}, \theta_l, \boldsymbol{\theta}_{\bar{l}}[i]) + \ln p(\boldsymbol{s}|\theta_l, \boldsymbol{\theta}_{\bar{l}}[i])$, where $p(\boldsymbol{z}_l|\boldsymbol{s}, \theta_l, \boldsymbol{\theta}_{\bar{l}}[i]) = \mathcal{CN}(\boldsymbol{z}_l; \boldsymbol{\alpha}_l\theta_l, \sigma^2\boldsymbol{I})$ and where $p(\boldsymbol{s}|\theta_l, \boldsymbol{\theta}_{\bar{l}}[i])$ does not depend on θ_l, to write

$$\begin{aligned} &\mathrm{E}\left\{ \ln p(\boldsymbol{z}_l, \boldsymbol{s}|\theta_l, \boldsymbol{\theta}_{\bar{l}}[i]) \big| \boldsymbol{y}, \boldsymbol{\theta}[i] \right\} \\ &= c_1 - \sigma^{-2}\,\mathrm{E}\left\{ \|\boldsymbol{z}_l - \boldsymbol{\alpha}_l\theta_l\|^2 \big| \boldsymbol{y}, \boldsymbol{\theta}[i] \right\} \qquad (58) \\ &= c_2 + \sigma^{-2}\Big(2\Re\{\theta_l^*\,\mathrm{E}\{\boldsymbol{\alpha}_l^H\boldsymbol{z}_l|\boldsymbol{y}, \boldsymbol{\theta}[i]\}\} \\ &\quad - |\theta_l|^2\,\mathrm{E}\left\{\|\boldsymbol{\alpha}_l\|^2 \big| \boldsymbol{y}, \boldsymbol{\theta}[i]\right\}\Big), \qquad (59) \end{aligned}$$

where c_1 and c_2 do not depend on θ_l. To proceed, we write

$$\mathrm{E}\{\boldsymbol{\alpha}_l^H\boldsymbol{z}_l|\boldsymbol{y}, \boldsymbol{\theta}[i]\} = \mathrm{E}\left\{\boldsymbol{\alpha}_l^H\,\mathrm{E}\left\{\boldsymbol{z}_l|\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{\theta}[i]\right\} \big| \boldsymbol{y}, \boldsymbol{\theta}[i]\right\}, \quad (60)$$

taking the inner expectation w.r.t $\boldsymbol{z}_l$ and the outer expectation w.r.t $\boldsymbol{s}$. Since $\boldsymbol{z}_l$ and $\boldsymbol{y}$ are jointly Gaussian (given $\boldsymbol{s}$ and $\boldsymbol{\theta}[i]$),

$$\begin{aligned} &\mathrm{E}\left\{\boldsymbol{z}_l|\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{\theta}[i]\right\} \\ &= \mathrm{E}\left\{\boldsymbol{z}_l|\boldsymbol{s}, \boldsymbol{\theta}[i]\right\} + \boldsymbol{C}_{z_l y}\boldsymbol{C}_{yy}^{-1}(\boldsymbol{y} - \mathrm{E}\left\{\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta}[i]\right\}) \quad (61) \\ &= \boldsymbol{\alpha}_l\theta_l[i] + \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}[i], \qquad (62) \end{aligned}$$

$$\begin{aligned} \text{where} \quad \boldsymbol{C}_{yy} &\triangleq \mathrm{Cov}\{\boldsymbol{y}\boldsymbol{y}^H|\boldsymbol{s}, \boldsymbol{\theta}[i]\} = \sigma^2\boldsymbol{I}_N \qquad (63) \\ \boldsymbol{C}_{z_l y} &\triangleq \mathrm{Cov}\{\boldsymbol{z}_l\boldsymbol{y}^H|\boldsymbol{s}, \boldsymbol{\theta}[i]\} = \sigma^2\boldsymbol{I}_N. \qquad (64) \end{aligned}$$

Plugging (62) back into (60) yields

$$\begin{aligned} &\mathrm{E}\left\{\boldsymbol{\alpha}_l^H\boldsymbol{z}_l|\boldsymbol{y}, \boldsymbol{\theta}[i]\right\} \\ &= \mathrm{E}\left\{\|\boldsymbol{\alpha}_l\|^2 \big| \boldsymbol{y}, \boldsymbol{\theta}[i]\right\}\theta_l[i] + \bar{\boldsymbol{\alpha}}_l^H\boldsymbol{y} - \mathrm{E}\{\boldsymbol{\alpha}_l^H\boldsymbol{A}|\boldsymbol{y}, \boldsymbol{\theta}[i]\}\boldsymbol{\theta}[i] \\ &= (\|\bar{\boldsymbol{\alpha}}_l\|^2 + c_{ll})\theta_l[i] + \bar{\boldsymbol{\alpha}}_l^H\boldsymbol{y} - (\bar{\boldsymbol{\alpha}}_l^H\bar{\boldsymbol{A}} + \boldsymbol{c}_l^H)\boldsymbol{\theta}[i], \qquad (65) \end{aligned}$$

where $\bar{\boldsymbol{\alpha}}_l$, $\boldsymbol{c}_l$ and c_{ll} were defined after (21).

Expansion of $\ln p(\theta_l, \boldsymbol{\theta}_{\bar{l}}[i])$, the second term in (20), yields

$$\begin{aligned} \ln p(\theta_l, \boldsymbol{\theta}_{\bar{l}}[i]) &= c_3 - \rho_{ll}|\theta_l - \bar{\theta}_l|^2 \\ &\quad - 2\Re\{(\theta_l - \bar{\theta}_l)^*\boldsymbol{\rho}_{\bar{l}}^H(\boldsymbol{\theta}_{\bar{l}}[i] - \bar{\boldsymbol{\theta}}_{\bar{l}})\}, \quad (66) \end{aligned}$$

where ρ_{ll} was defined after (21), $\boldsymbol{\rho}_{\bar{l}}$ is defined as $\boldsymbol{\rho}_l$ with l^{th} entry omitted, and c_3 is irrelevant to the maximization.

Plugging (65)–(66) into (20), and zeroing the gradient of the resulting expression with respect to θ_l, yields (21).

References

[1] S.-J. Hwang and P. Schniter, "Fast noncoherent decoding of block transmissions over doubly dispersive channels," in *Proc. Asilomar Conf. Signals Syst. Comput.*, (Pacific Grove, CA), Nov. 2007.

[2] S.-J. Hwang and P. Schniter, "EM-based soft noncoherent equalization of doubly selective channels using tree search and basis expansion," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, (Perugia, Italy), June 2009.

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

[4] R. Koetter, A. C. Singer, and M. Tüchler, "Turbo equalization," *IEEE Signal Process. Mag.*, vol. 21, pp. 67–80, Jan. 2004.

[5] G. Kutz and D. Raphaeli, "Maximum-likelihood semiblind equalization of doubly selective channels using the EM algorithm," *EURASIP J. Appl. Signal Process.*, pp. Article ID 709143, 14 pages, 2010.

[6] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Trans. Commun.*, vol. 42, pp. 1017–1032, Feb./Mar./Apr. 1994.

[7] P. Schniter, S.-J. Hwang, S. Das, and A. P. Kannu, "Equalization of time-varying channels," in *Wireless Communications over Rapidly Time-Varying Channels* (F. Hlawatsch and G. Matz, eds.), pp. 321–352, Elsevier Academic Press, 2011.

[8] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adaptive Control & Signal Process.*, vol. 10, pp. 159–176, Mar. 1996.

[9] O. Edfors, M. Sandell, J.-J.van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, July 1998.

[10] J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, June 1990.

[11] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, Jul. 1994.

[12] C. Antón-Haro, J. A. R. Fonollosa, and J. R. Fonollosa, "Blind channel estimation and data detection using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 45, pp. 241–247, Jan. 1997.

[13] M. Nissilä and S. Pasupathy, "Adaptive Bayesian and EM-based detectors for frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 51, pp. 1325–1336, Aug. 2003.

[14] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, pp. 1–17, 1977.

[15] 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, pp. 284–287, Mar. 1974.

[16] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, pp. 754–767, May 2002.

[17] P. Schniter, "A low-complexity receiver for OFDM in doubly selective channels," in *Proc. IEEE Global Telecommun. Conf.*, (San Francisco, CA), Dec. 2003.

[18] F. Peng and W. E. Ryan, "A low-complexity soft demapper for OFDM fading channels with ICI," in *Proc. IEEE Wireless Commun. & Netw. Conf.*, (Las Vegas, NV), pp. 1549–1554, Apr. 2006.

[19] D. Liu and M. P. Fitz, "Reduced state iterative MAP equalization and decoding in wireless mobile coded OFDM," in *Proc. Allerton Conf. Commun. Control Comput.*, (Monticello, IL), Sep. 2007.

[20] K. Fang, L. Rugini, and G. Leus, "Low complexity block turbo equalization for OFDM systems in time-varying channels," *IEEE Trans. Signal Process.*, vol. 56, pp. 5555–5556, Nov. 2008.

[21] K. Fang, L. Rugini, and G. Leus, "Low complexity block turbo equalization for OFDM systems in time-varying channels," *EURASIP J. Appl. Signal Process.*, pp. Article ID 974652, 13 pages, 2010.

[22] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Signal Process.*, vol. 52, pp. 2885–2894, Oct. 2004.

[23] R. Otnes and M. Tüchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1918–1923, Nov. 2004.

[24] D. Liu and M. P. Fitz, "Joint turbo channel estimation and data recovery in fast fading mobile coded OFDM," in *Proc. IEEE Int. Symposium Pers. Indoor Mobile Radio Commun.*, (Cannes, France), pp. 1–6, Sep. 2008.

[25] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, pp. 389–399, Mar. 2003.

[26] Y. L. C. de Jong and T. J. Willink, "Iterative tree search detection for MIMO wireless systems," *IEEE Trans. Commun.*, vol. 53, pp. 930–935, June 2005.

[27] M. J. Gertsman and J. H. Lodge, "Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 788–799, Jul. 1997.

[28] L. M. Davis, I. B. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 2106–2114, Dec. 2001.

[29] B. D. Hart and S. Pasupathy, "Innovations-based MAP detection for time-varying frequency-selective channels," *IEEE Trans. Veh. Tech.*, vol. 48, pp. 1507–1519, Sept. 2000.

[30] A. Anastasopoulos and K. M. Chugg, "Iterative detection for channels with memory," *IEEE Trans. Commun.*, vol. 48, pp. 1638–1649, Oct. 2000.

[31] Y. Zhang, M. P. Fitz, and S. B. Gelfand, "Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models," in *Proc. IEEE Global Telecommun. Conf.*, (Phoenix, AZ), pp. 720–724, Oct. 1997.

[32] X. Li and T. F. Wong, "Turbo equalization with nonlinear Kalman filtering for time-varying frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 691–700, Feb. 2007.

[33] H. Kim and J. K. Tugnait, "Turbo equalization for doubly-selective fading channels using nonlinear Kalman filtering and basis expansion models," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2076–2087, June 2010.

[34] T. Cui and C. Tellambura, "Blind receiver design for OFDM systems over doubly selective channels," *IEEE Trans. Commun.*, vol. 55, pp. 906–917, May 2007.

[35] J. B. Anderson and S. Mohan, "Sequential decoding algorithms: A survey and cost analysis," *IEEE Trans. Commun.*, vol. 32, pp. 169–172, 1984.

[36] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, pp. 2664–2677, Oct. 1994.

[37] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, pp. 58–66, Apr. 2002.

[38] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel, "Cyclic-prefixed or zero-padded multicarrier transmissions?," *IEEE Trans. Commun.*, vol. 50, Dec. 2002.

[39] S. Das and P. Schniter, "Max-SINR ISI/ICI-shaping multi-carrier communication over the doubly dispersive channel," *IEEE Trans. Signal Process.*, vol. 55, pp. 5782–5795, Dec. 2007.

[40] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Process. Mag.*, vol. 12, pp. 52–55, Jan. 2005.

[41] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[42] P. Schniter, "On doubly dispersive channel estimation for pilot-aided pulse-shaped multi-carrier modulation," in *Proc. Conf. Inform. Science & Syst.*, (Princeton, NJ), Mar. 2006.

[43] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2829–2844, June 2010.

[44] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1351–1366, May 2003.

[45] A. P. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, pp. 1148–1160, Mar. 2008.

[46] B. D. Hart and D. P. Taylor, "Extended MLSD diversity receiver for the time and frequency selective channel," *IEEE Trans. Commun.*, vol. 45, pp. 322–333, Mar. 1997.

[47] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1490–1498, Mar. 2009.

[48] I. Kozintsev, "Matlab programs for encoding and decoding of LDPC codes over GF(2^m)." http://www.kozintsev.net/soft.html.

[49] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974. [reprinted by IEEE Press].

[50] D. K. Borah and B. D. Hart, "Receiver structures for time-varying frequency-selective fading channels," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1863–1875, Nov. 1999.

[51] T. A. Thomas and F. W. Vook, "Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications," in *Proc. IEEE Sensor Array & Multichannel Signal Process. Workshop*, (Boston, MA), pp. 444–448, Mar. 2000.

[52] S.-J. Hwang, *Noncoherent Communication over Time and Frequency Selective Channels*. PhD thesis, The Ohio State University, Columbus, Ohio, Dec. 2009.

[53] J. D. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 33, pp. 457–458, Mar. 1997.

[54] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1983.

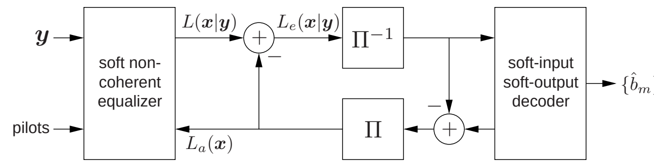

Fig. 1. Turbo receiver with soft noncoherent equalizer.

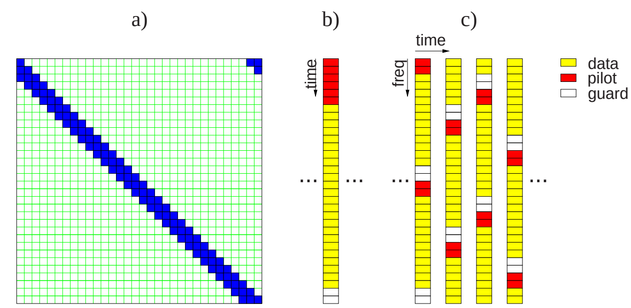

Fig. 2. For $N = 32$, $N_H = 3$, and $N_p = 8$, illustration of a) $\boldsymbol{H}^{(j)}$ support, b) single-carrier pilot pattern, and c) multicarrier pilot pattern with $\mathcal{P} = 4$ and $\mathcal{K} = 2$.

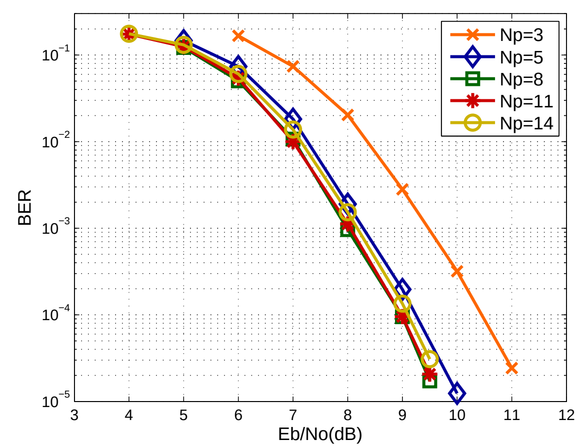

Fig. 3. Coded BER versus E_b/N_o for various pilot/guard numbers N_p. Single-carrier ncT-BE was used with $f_{\mathsf{D}}T_s = 0.002$, $N_h = 3$, and $N = 64$.

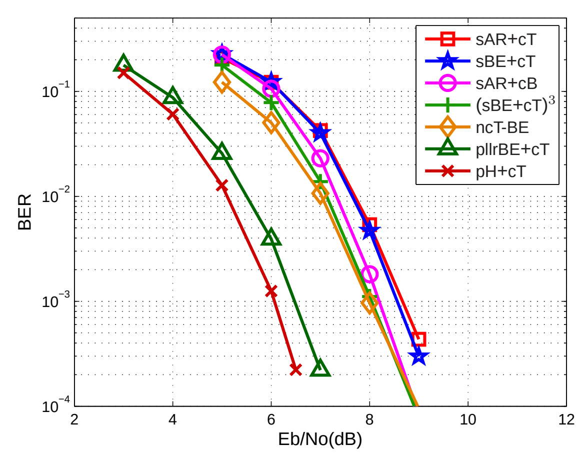

Fig. 4. BER vs. E_b/N_o for various equalization schemes under single-carrier transmission, $f_{\mathsf{D}}T_s = 0.002$, $N_h = 3$, and $N = 64$.

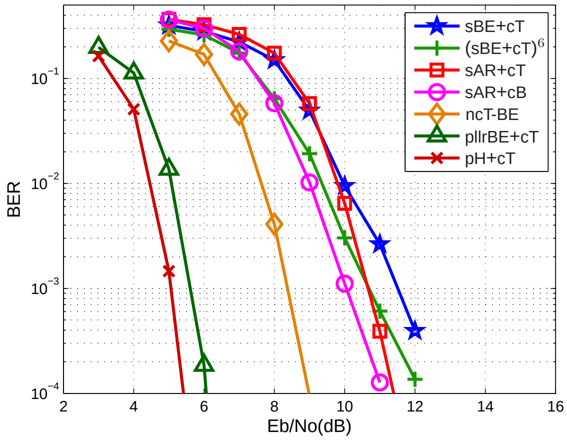

Fig. 5. BER vs. E_b/N_o for various equalization schemes under single-carrier transmission, $f_\mathsf{D}T_s = 0.005$, $N_h = 3$, and $N = 64$.

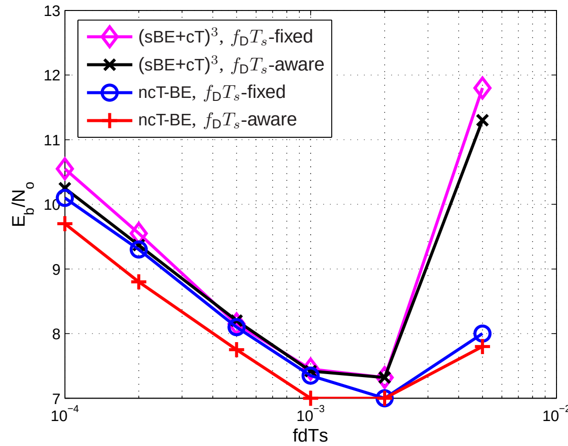

Fig. 8. E_b/N_o required to attain 10^{-2} BER versus $f_\mathsf{D}T_s$, for $f_\mathsf{D}T_s$-aware and $f_\mathsf{D}T_s$-fixed-at-0.002 schemes, under single-carrier transmission, $N_h = 3$, $N = 64$, and an $N_b = 3$-term KL-BE model.

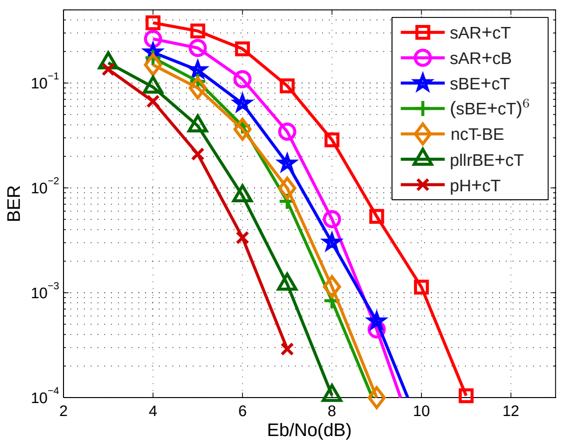

Fig. 6. BER vs. E_b/N_o for various equalization schemes under multicarrier transmission, $f_\mathsf{D}T_s = 0.002$, $N_h = 3$, and $N = 64$.

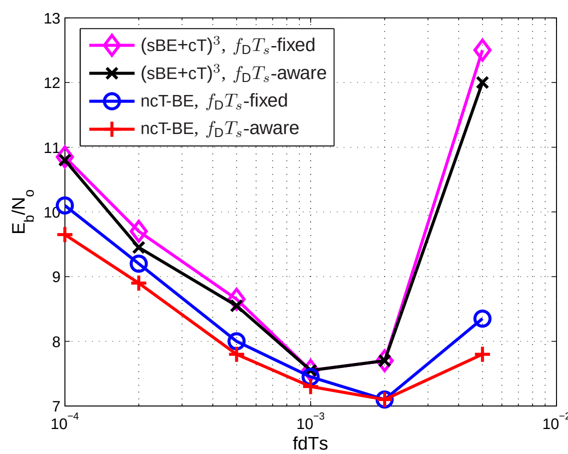

Fig. 9. E_b/N_o required to attain 10^{-2} BER versus $f_\mathsf{D}T_s$, for $f_\mathsf{D}T_s$-aware and $f_\mathsf{D}T_s$-fixed-at-0.002 schemes, under single-carrier transmission, $N_h = 3$, $N = 64$, and an OCE-BE model.

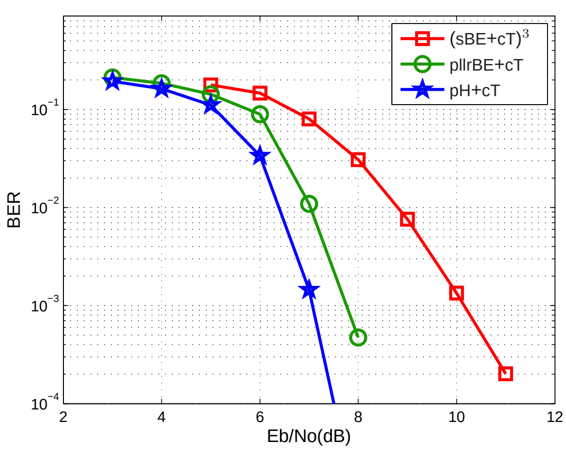

Fig. 7. BER vs. E_b/N_o for various equalization schemes under multicarrier transmission, $f_\mathsf{D}T_s = 0.0005$, $N_h = 64$, and $N = 256$.