RAKE Receiver for Channels with a Sparse Impulse Response

Yngve Selén, *Student Member, IEEE*, and Erik G. Larsson, *Member, IEEE*

***Abstract*— We derive the optimal receiver for RAKE diversity combining on channels with a sparse impulse response. The receiver is based on the Bayesian philosophy and thus it requires the knowledge of certain *a priori* parameters. However, we also derive an empirical Bayesian version of our receiver, which does not require any *a priori* knowledge, nor the choice of any user parameters. We show that both versions of our detector can outperform a classical training-based maximum-ratio-combining detector.**

***Index Terms*— Demodulation, signal detection, spread spectrum communication, Bayes procedures, receivers, multipath channels.**

I. Introduction

DIVERSITY reception is a key component in many communication systems. To optimally exploit it, one must implement the receiver that determines the most likely transmitted message given the received data. This problem is complicated because normally the receiver has only imperfect (or partial) knowledge of the channel. In this context, a substantial body of work [1]–[10] quantifies the performance loss when using an estimated (or for other reasons imperfect) channel in lieu of the true one for symbol detection. Less work is available on the design of detectors that can optimally make use of an estimated channel. However, optimal decision metrics are known for some cases; see, e.g., [11] for single-channel (i.e., no diversity) reception and [12] for space-time codes.

In this paper we are concerned with diversity combining for *sparse* channels, i.e. channels where the gains for many of the diversity paths are zero. The main motivation for this is RAKE combining for spread-spectrum systems operating on channels where the signal bandwidth is so large that individual multipath components can be resolved, so that the chip-sampled impulse response is sparse. Basically, there are two types of RAKE receivers in the literature: either the RAKE fingers are placed at fixed locations (typically on a grid), or their locations are adaptively chosen via a finger search mechanism [13]. Fixed-grid RAKE usually performs well if there are enough many RAKE fingers [14]. Also, its corresponding channel model may be sparse. RAKE receivers with path searching, however, do generally not give sparse channel models. (If the finger search algorithm succeeds, it positions the fingers at the correlation lags where the impulse response has power. The resulting effective channel will then be non-sparse.) Therefore fixed-finger RAKE is the main motivation for our work. In this context the main goal of our work is to examine how much performance one can gain by knowing that the channel is sparse. By way of contrast, if such knowledge were available but not exploited, some of the channel taps included in the detection process would carry only noise (no signal). This would lead to unnecessary noise enhancement and reduced detection performance.

Receivers for sparse channels have previously been studied in some special cases. For example, [15] derived a generalized Akaike information criterion to estimate the channel structure for OFDM systems. However, the optimal diversity combiner for sparse channels, derived in this article, is novel to our knowledge.

II. Problem Formulation

We consider a setup where the transmitter first sends a training (pilot) sequence $s_t(k)$, $k = 0, \ldots, n_t - 1$, which is known to the receiver. Thereafter, the transmitter sends an information symbol s, to be detected by the receiver. For simplicity of the exposition we restrict the discussion to M-PSK modulation (so, $|s| = |s_t(k)| = 1, \forall k$). We assume that the transmitted data are spread with an ideal spreading sequence, and then matched-filtered and despread at the receiver. The data propagate over a frequency selective channel with the chip-sampled impulse response $\boldsymbol{h} = [h_0, ..., h_{n-1}]^T$. We further assume that the received data are disturbed by circular AWGN, but that there is no multiaccess interference. The received data can be written as

$$\begin{aligned}
\boldsymbol{y}_t(0) &= \boldsymbol{h}s_t(0) + \boldsymbol{e}_t(0)\ , & \boldsymbol{y}_t(0) \in \mathbb{C}^n \\
&\vdots \\
\boldsymbol{y}_t(n_t-1) &= \boldsymbol{h}s_t(n_t-1) + \boldsymbol{e}_t(n_t-1)\ , & \boldsymbol{y}_t(n_t-1) \in \mathbb{C}^n \\
\boldsymbol{y} &= \boldsymbol{h}s + \boldsymbol{e}\ , & \boldsymbol{y} \in \mathbb{C}^n
\end{aligned}$$

where the noise terms $\boldsymbol{e}_t(k)$ and $\boldsymbol{e}$ are i.i.d. $CN(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. (Throughout, the subscript $(\cdot)_t$ denotes quantities associated with the training.) The objective for the receiver is to detect s given $\{\boldsymbol{y}_t(k)\}$, $\{s_t(k)\}$ and $\boldsymbol{y}$.

Naturally, if $\boldsymbol{h}$ were known to the receiver, then $\boldsymbol{h}^H\boldsymbol{y}/\|\boldsymbol{h}\|^2$ is a sufficient statistic for the detection of s, and the problem is trivial. The receiver then takes (assuming $P(s)$ is uniform for all s),

$$\hat{s} = \arg\min_s \left| s - \frac{\boldsymbol{h}^H\boldsymbol{y}}{\|\boldsymbol{h}\|^2} \right|^2 .$$

Manuscript received March 9, 2006; revised June 7, 2006; accepted June 20, 2006. The associate editor coordinating the review of this paper and approving it for publication was S. Kishore. Parts of this work were presented at the Asilomar Conference on Signals, Systems and Computers 2005. This work was supported in part by the Swedish Research Council (VR). E. Larsson is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Y. Selén is with the Division of Systems and Control, Dept. of Information Technology, Uppsala University, P.O. Box 337, SE-751 05 Uppsala, Sweden (email: yngve.selen@it.uu.se).

E. Larsson is with the Dept. of Communication Theory, School of Electrical Engineering, Royal Institute of Technology, Osquldas väg 10, SE-100 44 Stockholm, Sweden (email: erik.larsson@ee.kth.se).

Digital Object Identifier 10.1109/TWC.2007.06030033.

If $\boldsymbol{h}$ is unknown, however, the optimal receiver will in general depend on the specific assumptions made and on the philosophy used for the receiver design. Most commonly, an *a priori* density is assumed for $\boldsymbol{h}$, and the optimal receiver is derived under that assumption. However, it is also possible to take an "orthodox" approach (not making any assumptions on $\boldsymbol{h}$), for example using a generalized-likelihood ratio test. In this paper (like in most classical communication theory literature) we stick to the Bayesian paradigm, since it enables unique, well-defined inference. Additionally, it makes all assumptions explicit.

We shall assume for the purpose of receiver design that, with some probability, some elements of $\boldsymbol{h}$ are equal to zero. More precisely, we model $\{h_i\}$,[1] as i.i.d. random variables with the following mixture distribution:

$$h_i = \begin{cases} 0, & T_i = 0 \\ CN(0, \rho_i^2), & T_i = 1 \end{cases} \tag{1}$$

where $\{T_i\}_{i=0}^{n-1}$, are i.i.d. Bernoulli random variables with parameter p_i: $T_i \in \{0,1\}$ and $P(T_i = 0) = p_i = 1 - P(T_i = 1)$. This is not the only possible model, but it does seem to make sense from a physical perspective (for example, $|h_i|$ will have a Rayleigh distribution when $T_i = 1$), and also it leads to tractable mathematics.

III. The Optimal Receiver

The optimal receiver finds the s which has the largest probability of being transmitted, given all received data. Let us introduce the notation $\boldsymbol{Y} = [\boldsymbol{y}_t(0) \ \cdots \ \boldsymbol{y}_t(n_t-1) \ \boldsymbol{y}] \in \mathbb{C}^{n \times n_t+1}$. Then the receiver should find $\hat{s} = \operatorname{argmax}_s P(s|\boldsymbol{Y})$. (In what follows, the training sequence $\{s_t(k)\}$ is assumed to be known to the receiver.) To detect individual bits b_k in s, we should instead maximize $P(b_k = \beta|\boldsymbol{Y}) = \sum_{s:b_k=\beta} P(s|\boldsymbol{Y})$. To take soft decisions on a bit b_k, the quantity of interest would be $L_k = \log(P(b_k = 1|\boldsymbol{Y})/P(b_k = 0|\boldsymbol{Y}))$. Either way, what is important is to evaluate $P(s|\boldsymbol{Y})$.

For equiprobable symbols (an assumption that can be relaxed) we have that[2]

$$\begin{aligned} P(s|\boldsymbol{Y}) &\cong p(\boldsymbol{Y}|s) = \prod_{i=0}^{n-1} p(\boldsymbol{y}_i|s) \\ &= \prod_{i=0}^{n-1} \big(p_i p(\boldsymbol{y}_i|s, T_i = 0) + (1-p_i) p(\boldsymbol{y}_i|s, T_i = 1)\big) \end{aligned} \tag{2}$$

where $\boldsymbol{y}_i = [y_{ti}(0) \ \cdots \ y_{ti}(n_t-1) \ y_i]^T$, i.e., all data received for tap i. Note that (2) only holds if the tap gains $\{h_i\}$ are independent and the noise samples (elements of $\{\boldsymbol{e}_t(k)\}$ and $\boldsymbol{e}$) are independent, so that $\{\boldsymbol{y}_i\}$ are independent (conditioned on s). This was, however, an assumption that we did make. To compute (2) we need the conditional densities $p(\boldsymbol{y}_i|s, T_i)$. For $T_i = 0$ this follows directly, since in this case $\boldsymbol{y}_i \sim CN(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and

$$\begin{aligned} p(\boldsymbol{y}_i|s, T_i = 0) = &\frac{1}{\pi^{n_t+1}} \frac{1}{\sigma^{2(n_t+1)}} \\ &\exp\left(-\frac{1}{\sigma^2}\left(|y_i|^2 + \sum_{k=0}^{n_t-1} |y_{ti}(k)|^2\right)\right). \end{aligned} \tag{3}$$

Next, observe that, conditioned on s and $T_i = 1$, we have $\boldsymbol{y}_i \sim CN(\boldsymbol{0}, \boldsymbol{Q}_{s,i})$ where

$$\boldsymbol{Q}_{s,i} \triangleq \rho_i^2 \boldsymbol{s}\boldsymbol{s}^H + \sigma^2 \boldsymbol{I}$$

and $\boldsymbol{s} = [s_t(0) \ \cdots \ s_t(n_t-1) \ s]^T$ ($\boldsymbol{Q}_{s,i}$ is a function of s). So,

$$p(\boldsymbol{y}_i|s, T_i = 1) = \frac{1}{\pi^{n_t+1}} \frac{1}{|\boldsymbol{Q}_{s,i}|} \exp\left(-\boldsymbol{y}_i^H \boldsymbol{Q}_{s,i}^{-1} \boldsymbol{y}_i\right). \tag{4}$$

Equation (4) can be simplified as follows. By the matrix inversion lemma,

$$\boldsymbol{Q}_{s,i}^{-1} = \frac{1}{\sigma^2}\boldsymbol{I} - \frac{\rho_i^2}{\sigma^4 + \rho_i^2 \sigma^2 (n_t+1)} \boldsymbol{s}\boldsymbol{s}^H,$$

so

$$\begin{aligned} \boldsymbol{y}_i^H \boldsymbol{Q}_{s,i}^{-1} \boldsymbol{y}_i = &\frac{1}{\sigma^2}\left(\sum_{k=0}^{n_t-1} |y_{ti}(k)|^2 + |y_i|^2\right) \\ &- \frac{\rho_i^2}{\sigma^4 + \rho_i^2 \sigma^2 (n_t+1)} \left|\sum_{k=0}^{n_t-1} y_{ti}^*(k) s_t(k) + y_i^* s\right|^2. \end{aligned}$$

Next, by using the identity $|\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A}|$ we have

$$|\boldsymbol{Q}_{s,i}| = \sigma^{2(n_t+1)}\left(1 + (n_t+1)\frac{\rho_i^2}{\sigma^2}\right).$$

To summarize, we have found that the log-densities of the data are

$$\begin{aligned} \log p(\boldsymbol{y}_i|s, T_i = 0) \ \cong \ &-(n_t+1)\log(\sigma^2) \\ &-\frac{1}{\sigma^2}\left(\sum_{k=0}^{n_t-1} |y_{ti}(k)|^2 + |y_i|^2\right) \end{aligned}$$

and

$$\begin{aligned} \log p(\boldsymbol{y}_i|s, T_i = 1) \cong &-(n_t+1)\log(\sigma^2) \\ &-\frac{1}{\sigma^2}\left(\sum_{k=0}^{n_t-1} |y_{ti}(k)|^2 + |y_i|^2\right) - \log\left(1 + (n_t+1)\frac{\rho_i^2}{\sigma^2}\right) \\ &+\frac{\rho_i^2}{\sigma^4 + \rho_i^2 \sigma^2 (n_t+1)} \left|\sum_{k=0}^{n_t-1} y_{ti}^*(k) s_t(k) + y_i^* s\right|^2. \end{aligned}$$

Assembling things together and letting $J(a,b)$ stand for the Jacobian logarithm[3], $J(a,b) \triangleq \log(e^a + e^b)$, we have (omitting some irrelevant terms)

$$\begin{aligned} \log P(s|\boldsymbol{Y}) \cong &\sum_{i=0}^{n-1} \log\Big[p_i P(\boldsymbol{y}_i|s, T_i = 0) \\ &+(1-p_i) P(\boldsymbol{y}_i|s, T_i = 1)\Big] \cong \sum_{i=0}^{n-1} J(\alpha_i, \beta_i) \end{aligned} \tag{5}$$

where $\alpha_i = \log(p_i)$ and

$$\begin{aligned} \beta_i \ = \ &\log(1-p_i) - \log\left(1 + (n_t+1)\frac{\rho_i^2}{\sigma^2}\right) \\ &+\frac{\rho_i^2}{\sigma^4 + \rho_i^2 \sigma^2 (n_t+1)} \left|\sum_{k=0}^{n_t-1} y_{ti}^*(k) s_t(k) + y_i^* s\right|^2. \end{aligned}$$

[1]Generally, we let subscript $(\cdot)_i$ denote the ith element of a vector.

[2]Hereafter "$\cong$" stands for equality up to irrelevant constants.

[3]This can be implemented with good numerical stability and high computational efficiency. For example, one can show that $J(a,b) = \max(a,b) + \log(1 + e^{-|a-b|})$. The last term can be implemented via a table-lookup as a function of $a - b$.

(Note that the entire right hand side of (3) is independent of s and the same expression also appears in $p(\boldsymbol{y}_i|s, T_i = 1)$. Therefore it cancels in the above equation.)

In the limit when all p_i approach zero (i.e., there is no sparseness in the model), then the optimal receiver (5) becomes

$$\log P_{\text{non-sparse}}(s|\boldsymbol{Y}) \cong \sum_{i=0}^{n-1}\left[\frac{\rho_i^2}{\sigma^4+\rho_i^2\sigma^2(n_t+1)} \left|\sum_{k=0}^{n_t-1} y_{ti}^*(k)s_t(k)+y_i^*s\right|^2\right]. \tag{6}$$

If $\rho_i^2 = \rho^2,\ \forall i$, the receiver in (6) is equivalent to MRC (maximum-ratio combining) using the LS/ML (least squares/maximum likelihood) channel estimate, as far as hard decisions are concerned. To see this, consider the LS/ML estimate of $\boldsymbol{h}$:

$$\hat{\boldsymbol{h}}_{\text{LS}} = \frac{1}{n_t}\sum_{k=0}^{n_t-1}\boldsymbol{y}_t(k)s_t^*(k). \tag{7}$$

The MRC receiver using this estimated channel *minimizes*, with respect to s,

$$\begin{aligned}\left|s-\frac{\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}}{\|\hat{\boldsymbol{h}}_{\text{LS}}\|^2}\right|^2 &= \underbrace{|s|^2}_{=1}+\left|\frac{\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}}{\|\hat{\boldsymbol{h}}_{\text{LS}}\|^2}\right|^2-2\text{Re}\left(\frac{\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}}{\|\hat{\boldsymbol{h}}_{\text{LS}}\|^2}s^*\right)\\ &\cong -\text{Re}\left(\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}s^*\right).\end{aligned} \tag{8}$$

The receiver (6) *maximizes*, with respect to s (when $\rho_i^2 = \rho^2,\ \forall i$),

$$\begin{aligned}\sum_{i=0}^{n-1}\left|\sum_{k=0}^{n_t-1}y_{ti}^*(k)s_t(k)+y_i^*s\right|^2 &= \|\hat{\boldsymbol{h}}_{\text{LS}}n_t+\boldsymbol{y}s^*\|^2\\ &= n_t^2\|\hat{\boldsymbol{h}}_{\text{LS}}\|^2+\underbrace{|s|^2}_{=1}\|\boldsymbol{y}\|^2+2\text{Re}\left(n_t\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}s^*\right)\\ &\cong \text{Re}\left(\hat{\boldsymbol{h}}_{\text{LS}}^H\boldsymbol{y}s^*\right).\end{aligned} \tag{9}$$

Comparison of (8) with (9) proves the equivalence. The same conclusion holds if the receiver instead performs MMSE (minimum mean square error) channel estimation followed by MRC, but also only under the assumptions that $\rho_i^2 = \rho^2$ and $|s| = 1$.

Finally, note that the channel often is constant over many symbols and one may then want to detect several symbols using the same channel estimate. It is straightforward to extend the above optimal receivers to detect a vector of information symbols. However, the complexity of the resulting detector will be exponential in the number of information symbols. A more pragmatic approach is to maximize (5) with respect to each symbol separately. Strictly speaking, the so-obtained receiver is suboptimal, but its complexity is linear in the number of symbols.

IV. An Empirical Bayesian Receiver

Our receiver is Bayesian and it requires knowledge of $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2. How can we use this receiver if these parameters are unknown? There are several possible answers to this question. One could treat $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2 as random variables with certain *a priori* densities and then average $P(s|\boldsymbol{Y})$ over these. Alternatively, one could maximize (concentrate) $P(s|\boldsymbol{Y})$ with respect to $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2, treating them as hyperparameters. However, we choose another approach; namely to estimate $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2 from the data.

Estimating $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2 from the data is a pragmatic (but in a strict sense suboptimal) solution and the resulting receiver will be a so-called "empirical Bayesian" detector [16]. We will restrict the discussion to i.i.d. $\{h_i\}$, so that $\rho_i^2 = \rho^2$ and $p_i = p, \forall i$. This is not the most general approach, but it facilitates the use of simple moment-based estimators for $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$. Also, our Bayesian receiver turns out to be robust in the sense that it generally shows good performance even with mismatched $\{p_i\}_{i=0}^{n-1}$, $\{\rho_i^2\}_{i=0}^{n-1}$, σ^2 (see Section V). Thus, accurate estimation of these parameters does not seem crucial.

Estimation of σ^2: An unbiased[4], consistent estimate of σ^2 can be obtained by taking

$$\hat{\sigma}^2 = \frac{1}{n_t n - n}\sum_{k=0}^{n_t-1}\|\boldsymbol{y}_t(k)-\hat{\boldsymbol{h}}_{\text{LS}}s_t(k)\|^2$$

where $\hat{\boldsymbol{h}}_{\text{LS}}$ is obtained from (7).

Estimation of ρ^2, p: Under the model (1) we have $E[|h_i|^2] = (1-p)\rho^2$ and $E[|h_i|^4] = 2(1-p)\rho^4$.[5] By combining these equations and using $E[|h_i|^2] \approx \sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^2/n$, $E[|h_i|^4] \approx \sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^4/n$, we obtain the following estimates:

$$\hat{\rho}^2 = \frac{\sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^4}{2\sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^2} \quad \text{and} \quad \hat{p} = 1-\frac{2}{n}\frac{\left(\sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^2\right)^2}{\sum_{i=0}^{n-1}|\hat{h}_{\text{LS},i}|^4}.$$

One must make sure that $\hat{p} \geq 0$, for instance, by setting it to a small positive value (or zero) if the above estimate turns out negative.

The estimates of σ^2, ρ^2, p proposed here are intuitively appealing and statistically sound, but not "optimal". Naturally, we expect the estimates to improve with increasing n_t. Also we expect that they will break down when $p \to 1$, and that they will not work very well for small n_t. In particular, they will break down completely for $n_t = 1$, since then $\hat{\sigma}^2$ is undefined.

V. Numerical Examples

We illustrate the performance of our receivers via Monte-Carlo simulation. We consider both receivers whose *a priori* parameters are matched to the channel, and mismatched receivers. Also, we consider both a simple "toy" channel and a more realistic channel model for indoor radio [18]. In all examples we use QPSK modulation and show the bit-error-rates (BERs) for (i) the coherent receiver, (ii) the non-sparse Bayesian receiver (6) (same as MRC using the

[4]Note that we normalize with $n_t n - n$ instead of just $n_t n$ to take into account the fact that the same pilot data are used to compute $\hat{\boldsymbol{h}}_{\text{LS}}$ and $\hat{\sigma}^2$.

[5]Using, e.g., that, when $x_i \in CN(0, \lambda_i^2)$, $E[x_1x_2x_3x_4] = E[x_1x_2]E[x_3x_4] + E[x_1x_3]E[x_2x_4] + E[x_1x_4]E[x_2x_3]$ [17].

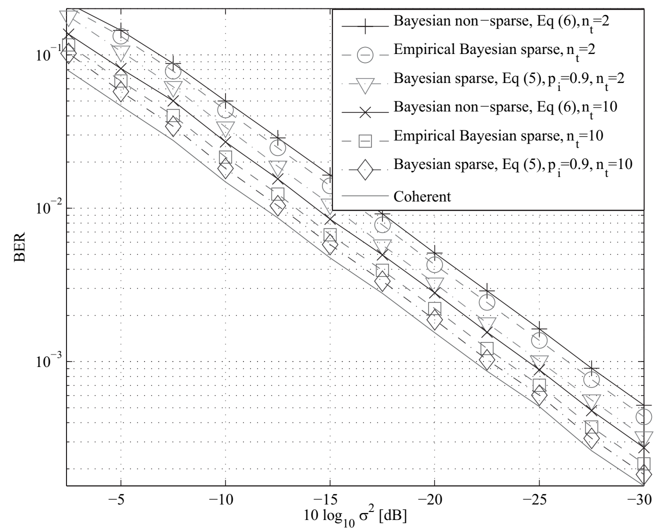

Fig. 1. Toy channel model, 20 i.i.d. taps. Each tap is $CN(0,1)$ with probability 10%, and zero with probability 90%. The receiver parameters $(p_i, \rho_i^2, \sigma^2)$ were the same as for the data generating process, i.e., $\rho_i^2 = 1$, $p_i = 0.9$.

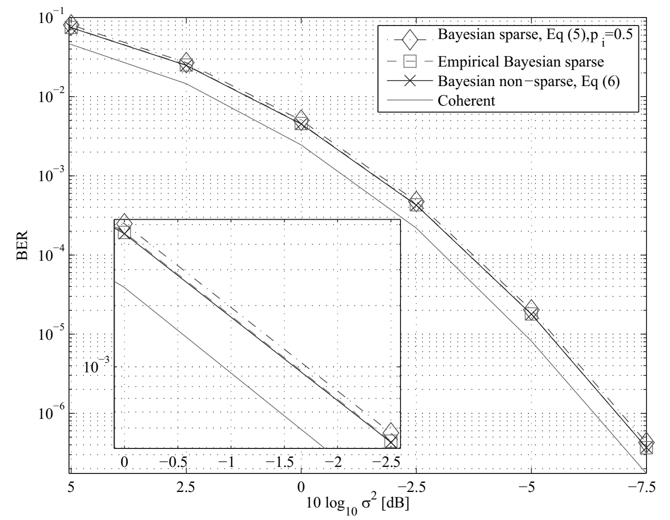

Fig. 2. Toy channel model, 10 i.i.d. taps. Here all taps are $CN(0,1)$, i.e., there is no sparsity in the channel impulse response. The "Bayesian sparse" receiver uses the true values of σ^2 and ρ_i^2, but a mismatched value of p_i (it takes $p_i = 0.5$). A training sequence of length $n_t = 10$ was used. The closeup shows the region $\sigma^2 \in [0, -2.5]$ dB.

LS/MMSE estimated non-sparse channel), (iii) our sparse Bayesian receiver (5) and (iv) its empirical version.

Sparse toy channel model: We first consider a sparse toy channel where the *a priori* parameters of the receiver are matched to the channel. This means that the channel impulse response is generated with the same parameters values as used in the receiver. The parameters are set to $p_i = 0.9$,[6] $\rho_i^2 = 1$, $n = 20$, and h_i are i.i.d. We consider two choices of the training sequence length: $n_t = 2$, and $n_t = 10$. The receiver performances are shown in Fig. 1. Our sparse Bayesian receiver outperforms the non-sparse Bayesian receiver with about 1.5 dB for both choices of n_t. This is expected, as the non-sparse Bayesian receiver uses mismatched channel knowledge ($p_i = 0$ instead of $p_i = 0.9$). The sparse Bayesian receiver with the long training sequence performs about 1 dB worse than the coherent receiver. The performance of our empirical Bayesian receiver beats that of the non-sparse Bayesian receiver by about 1 dB, even with only two training samples.

Non-sparse toy channel model—robustness to mismatch: Next, we consider a non-sparse toy channel example where the data are generated as above, but with $n_t = 10$, $n = 10$ and $p_i = 0$ (i.e., all taps are non-zero). We use a mismatched sparse Bayesian receiver with $p_i = 0.5$. The BERs are shown in Fig. 2. It is no surprise that the non-sparse Bayesian receiver performs slightly better than our (mismatched) sparse Bayesian receiver, as the former uses the true $p_i = 0$. However, the penalty for assuming sparseness is very small. Our empirical Bayesian estimator shows impressive performance—almost identical to that of the non-sparse Bayesian receiver, which is optimal in this scenario.

Indoor radio channel model: Finally, we consider the indoor radio channel model of Saleh-Valenzuela [18]. This model is based on an extensive set of measurements in the 1.5 GHz band. We use the channel model with the parameter values suggested in [18] ($1/\Lambda = 300$ ns, $1/\lambda = 5$ ns, $\Gamma = 60$ ns, $\gamma = 20$ ns). We assume a sampling interval of $T_s = 10$ ns, corresponding to a 100 MHz signal bandwidth, and root-raised-cosine filtering with a rolloff factor of $\beta = 0.5$ at the transmitter and at the receiver. We further assume that the receiver can acquire approximate timing, so that the first sample in the impulse responses is taken uniformly at random between $-1.5T_s$ and $-0.5T_s$ relative to the timing of the first multipath ray. We truncate the channel impulse responses after $n = 40$ taps and we use $n_t = 5$ training symbols.

One may ask whether, or how well, the channel model assumed in Section II corresponds to reality, or how well it matches a more realistic or specific model like that of [18]. Specifically, one might ask whether the distribution we assumed for the channel coefficients $\{h_i\}$ is reasonable. Under the model of [18] we can argue as follows. Owing to the random sampling timing, the pulse shaping, and since rays generally arrive in clusters according to [18] we cannot expect the independence assumption we made on $\{h_i\}$ to hold exactly. However, $\{h_i\}$ will be Rayleigh fading for the model of [18], because all channel taps in continuous time are complex Gaussian and $\{h_i\}$ are obtained via a linear operation on these. This is demonstrated in Figure 3, where we show "empirical pdfs" of the magnitudes of a few taps based on 10000 Saleh-Valenzuela channel realizations. This figure also illustrates how the average tap power varies as a function of tap delay.

It is interesting to study the extreme cases when the receiver is supplied much *a priori* information, and no such information, respectively. To simulate a scenario with relatively much *a priori* knowledge, we consider our sparse Bayesian receiver where p_i and ρ_i^2 are set to the values given in Table I. These values were obtained by averaging over many realizations of the Saleh-Valenzuela channel. (Since no taps are exactly zero-valued in the Saleh-Valenzuela model, a tap was considered to be zero when its magnitude was less than 0.05 times the

[6] To avoid an error floor, we make sure that the channels $\boldsymbol{h}$ have at least one non-zero tap (when all taps are zero, no signal reaches the receiver and detection will be pure guesswork).

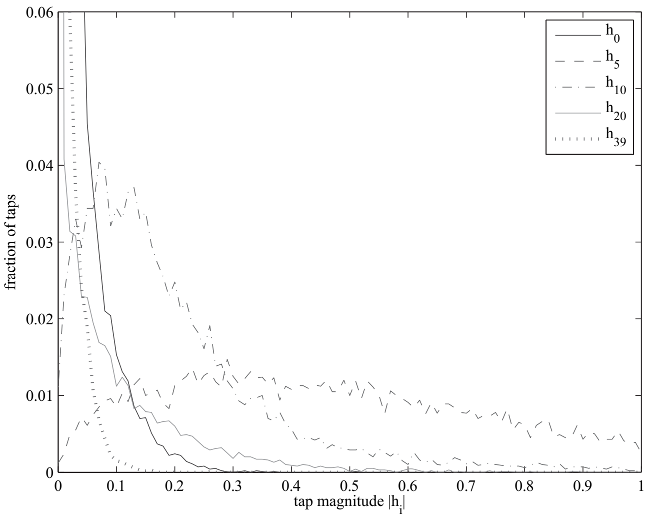

Fig. 3. "Empirical pdfs" of some $|h_i|$ for the Saleh-Valenzuela model.

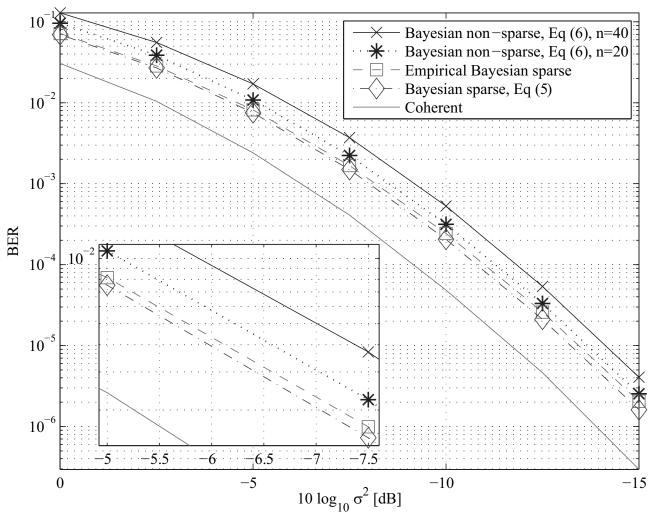

Fig. 4. Saleh-Valenzuela channel model. The *a priori* knowledge from Table I was used by "Bayesian sparse". A training sequence of length $n_t = 5$ was used. The closeup shows the region $\sigma^2 \in [-5, -7.5]$ dB.

magnitude of the largest tap in the realization.) If a typical channel is available for measurement prior to the selection of the *a priori* parameters, this is the type of information that can possibly be obtained. To study the opposite case, i.e. when no *a priori* knowledge about the channel is available, we use our empirical Bayesian receiver. Note that this detector uses $p_i = \hat{p}$ and $\rho_i^2 = \hat{\rho}^2$ (i.e., p and ρ^2 are the same for all taps). This results in a severely mismatched receiver (cf. Table I where the values of p_i, ρ_i^2 vary rather heavily with i).

TABLE I
THE MEASURED VALUES OF $\{\rho_i^2\}$ AND $\{p_i\}$ USED FOR THE "BAYESIAN SPARSE" CURVE IN FIGURE 4

$\boldsymbol{\rho}^2 =$	[0.014124, 0.09741, 1.4791, 1.1895, 0.77271, 0.51005, 0.33896, 0.23924, 0.17412, 0.12844, 0.10363, 0.086537, 0.078362, 0.072019, 0.071404, 0.07419, 0.067682, 0.066193, 0.061858, 0.056244, 0.046561, 0.044409, 0.03711, 0.032066, 0.030457, 0.025896, 0.022456, 0.021203, 0.020268, 0.017297, 0.014361, 0.01396, 0.012081, 0.010978, 0.010211, 0.0093643, 0.0081263, 0.007429, 0.0067383, 0.0066416]
$\boldsymbol{p} =$	[0.835, 0.1767, 0.0044, 0.0119, 0.0222, 0.035, 0.0519, 0.0698, 0.1036, 0.1399, 0.2016, 0.286, 0.3888, 0.5171, 0.6101, 0.6956, 0.7405, 0.7623, 0.7787, 0.7858, 0.7946, 0.7998, 0.8102, 0.8195, 0.8324, 0.8467, 0.8543, 0.8666, 0.879, 0.8923, 0.8989, 0.9124, 0.9139, 0.9266, 0.9327, 0.9372, 0.9488, 0.9563, 0.9642, 0.9682]

The receiver performances are shown in Figure 4. We also include the non-sparse Bayesian receiver where the channel has been truncated after $n = 20$ taps in this figure: this is a relevant comparison since $p_i \approx 1$ for $i \geq 20$ (see Table I). The Bayesian sparse receiver exploiting the full *a priori* knowledge from Table I is denoted "Bayesian sparse." This receiver performs about 1 dB better than the non-sparse Bayesian receiver with $n = 40$, and about 0.5 dB better than the non-sparse Bayesian receiver with $n = 20$. The empirical Bayesian receiver performs rather similarly to the sparse Bayesian receiver. One contributing reason for this is that the Bayesian receiver is very robust to mismatching *a priori* knowledge (as indicated by Figures 1 and 2).

VI. Concluding Remarks

We have derived the optimal Bayesian RAKE receiver under the assumption of a sparse channel impulse response. We also derived an empirical Bayesian version of our receiver, which does not require any *a priori* knowledge, nor the choice of any user parameters. The sparsity assumption was motivated both by physical considerations and by the fact that some existing channel models (e.g., for indoor radio) tend to generate sparse impulse responses. Via numerical simulations we demonstrated that our new receiver, and its empirical version, can significantly outperform the non-sparse Bayesian receiver when the channel is sparse. Our receiver has about the same complexity as MRC with channel estimation, which is the conventional solution.

References

[1] J. G. Proakis, "Probabilities of error for adaptive reception of M-phase signals," *IEEE Trans. Commun. Technol.*, vol. 16, pp. 71–81, Feb. 1968.

[2] M. J. Gans, "The effect of Gaussian error in maximal ratio combiners," *IEEE Trans. Commun.*, vol. 19, pp. 492–500, Aug. 1971.

[3] B. R. Tomiuk, N. C. Beaulieu, and A. A. Abu-Dayya, "General forms for maximal ratio diversity with weighting errors," *IEEE Trans. Commun.*, vol. 47, pp. 488–492, Apr. 1999.

[4] M. K. Simon and M.-S. Alouini, *Digital Communications over Fading Channels: A Unified Approach to Performance Analysis*. New York: John Wiley and Sons, 2000.

[5] A. Annamalai and C. Tellambura, "Analysis of hybrid selection/maximal-ratio diversity combiners with Gaussian errors," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 498–512, July 2002.

[6] W. Gifford, M. Win, and M. Chiani, "On the SNR penalty for realistic diversity systems," in *Proc. Conference on Information and System Sciences (CISS)*, Princeton, NJ, Mar. 2004.

[7] D. Singh, A. Kumari, R. K. Mallik, and S. S. Jamuar, "Analysis of RAKE reception with MRC and imperfect weight estimation for binary coherent orthogonal signaling," *IEEE Commun. Lett.*, vol. 6, pp. 245–247, June 2002.

[8] J. P. Fonseka, Y. Yin, and I. Korn, "Sensitivity to fading estimates in maximal ratio combining," *IEE Proc. Commun.*, vol. 150, pp. 269–274, Aug. 2003.

[9] F. A. Dietrich and W. Utschick, "Maximum ratio combining of correlated Rayleigh fading channels with imperfect channel knowledge," *IEEE Commun. Lett.*, vol. 7, pp. 419–421, Sept. 2003.

[10] Y. Akyildiz and B. D. Rao, "Maximum ratio combining performance with imperfect channel estimates," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2485–2488, 2002.

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

[12] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1874–1888, July 2005.

[13] X. Chu and R. D. Murch, "Generalized selection combining with log-likelihood ratio threshold test per path for RAKE reception in ultra-wideband communications," in *Proc. IEEE VTC 2005*, vol. 2, pp. 1360–1364.

[14] C. Cozzo, G. E. Bottomley, and A. S. Khayrallah, "Rake receiver placement for realistic channels," in *Proc. IEEE WCNC*, March 2004, vol. 1, pp. 316–321.

[15] M. R. Raghavendra and K. Giridhar, "Improving channel estimation in OFDM systems for sparse multipath channels," *IEEE Signal Processing Lett.*, vol. 12, no. 1, pp. 52–55, Jan. 2005.

[16] E. I. George and D. P. Foster, "Calibration and empirical Bayes variable selection," *Biometrika*, vol. 87, no. 4, pp. 731–747, Dec. 2000.

[17] P. H. M. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Autom. Control*, vol. 9, pp. 867–870, Sept. 1988.

[18] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.