

Linköping University Pre-Print

On Diversity Combining with Unknown Channel State Information and Unknown Noise Variance

Erik G. Larsson, Ragnar Thobaben and Gang Wang

N.B.: When citing this work, cite the original article.

©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Erik G. Larsson, Ragnar Thobaben and Gang Wang, On Diversity Combining with Unknown Channel State Information and Unknown Noise Variance, 2010, Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC).

<http://dx.doi.org/10.1109/WCNC.2010.5506707>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-52172>

On Diversity Combining with Unknown Channel State Information and Unknown Noise Variance

Erik G. Larsson¹, Ragnar Thobaben², and Gang Wang³

¹ Dept. of Electrical Engineering, Communication Systems, Linköping University, Linköping, Sweden

² School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden

³ KPMG Advisory Limited, Nanjing Branch, P.R. China (previously with KTH)

Abstract—We derive detection metrics for soft-output diversity combining for the case of imperfect channel state information at the receiver. We treat in particular the case when the noise variance at the receiver is unknown. We contrast conventional training-based methods to a detector based on the generalized likelihood-ratio (GLR) test paradigm. We study the performance of the detectors via EXIT chart analysis and via simulations of LDPC coded transmission over a fast Rayleigh fading channel. The results show that the GLR receivers can significantly outperform the conventional detectors.

I. INTRODUCTION

This paper deals with the problem of optimum diversity combining with imperfect channel state information (CSI) at the receiver. More specifically we consider an observed signal

$$\mathbf{y}_s = \mathbf{h}s + \mathbf{e}$$

where $s \in \mathbb{C}$ is a transmitted symbol comprising bits $\{b_k\}$, $\mathbf{h} \in \mathbb{C}^m$ is a channel vector, and $\mathbf{e} \in N(\mathbf{0}, \sigma^2 \mathbf{I})$ is white Gaussian noise. The variable m is the dimension of the problem, and it may represent the number of antennas in a multiple-antenna diversity-reception system, for example. The problem is to take soft decisions on s , more precisely to compute log-likelihood ratios of the form

$$\log \left(\frac{P(b_k = 1 | \mathbf{y}_s, \mathcal{H})}{P(b_k = 0 | \mathbf{y}_s, \mathcal{H})} \right)$$

where \mathcal{H} indicates the knowledge available about \mathbf{h} . This problem is nearly trivial if \mathbf{h} and σ^2 are perfectly known, since then $\mathbf{h}^H \mathbf{y}_s$ is a sufficient statistic, and the problem is reduced to demodulation for a unit-gain scalar AWGN channel. It is also well studied if σ^2 is known and \mathbf{h} has a Gaussian distribution with perfectly known parameters: for $m = 1$ (no diversity) this was treated in [1] and for general m in [2] (zero-mean \mathbf{h} ; Rayleigh fading) and [3] (nonzero-mean \mathbf{h} ; Rice fading). Essentially what must be done to obtain the optimal receiver is to marginalize the likelihood $p(\mathbf{y}_s | s, \mathbf{h})$ over the channel distribution (conditioned on the known information) $p(\mathbf{h} | \mathcal{H})$. If σ^2 is known, then only Gaussian densities are involved, and this is straightforward to do in closed form.

Contact email: erik.larsson@isy.liu.se

This work was supported in part by the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), and VINNOVA. E. Larsson is a Royal Swedish Academy of Sciences (KVA) Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

Here we are especially interested in the situation in which σ^2 is unknown. The motivation for this is that receivers often do not accurately know their noise floor. This noise floor typically depends on the carrier frequency, and it often varies with temperature and other external factors. More importantly, the “noise” term \mathbf{e} often includes other sources than thermal noise, such as co-channel interference. This co-channel interference may vary significantly with time and frequency (especially so in a frequency-hopping system) and its power can often be regarded as unknown by the receiver. With unknown σ^2 , a direct approach to the marginalization problem becomes numerically intricate. Additionally, it is not immediately clear what a *a priori* distribution for $p(\sigma^2)$ one should use. Hence a simpler, suboptimal solution is desirable.

The problem of soft detection with unknown noise statistics has received relatively little attention in the literature. Apart from our own previous work (see the next paragraph), we note reference [4], that considered interference-rejection combining by treating the interference as Gaussian noise, with an unknown covariance matrix. This interference covariance matrix was estimated from data; however, the white noise level was assumed to be known. The metrics can in principle deliver soft information but no evaluation of this was performed. The decision metrics in [4] require a numerical evaluation which appears hard to do in a real-time implementation.

The paper in hand builds on our previous work [5] where we considered the detection of space-time block codes on a channel with additive noise of unknown spatial color, and proposed a GLRT-based detection metric that implicitly estimated both the channel matrix and the noise/interference covariance matrix. The formulas for diversity combining which is our interest here follow as special cases of those in [5]. However, [5] did not contrast the proposed GLRT technique with training-based estimation of the channel and noise parameters, nor with the optimal noncoherent receiver. Nor did it evaluate the performance with powerful channel coding. In this paper we perform such comparisons. Specifically we evaluate the performance using powerful LDPC channel coding and also visualize the capacity loss due to the imperfect CSI in an extrinsic information transfer (EXIT) characteristics analysis. The main conclusion is that the GLRT metric, while less widely used in the communications literature, is highly competitive in terms of performance. More conclusions are summarized in Section VII.

II. MODEL

We consider an idealized model of a block-fading channel. A large number of data bursts are transmitted, and each burst undergoes transmission over a quasi-static fading SIMO channel with gain vector $\mathbf{h} \triangleq [h_0, \dots, h_{m-1}]^T$. Each burst in turn consists of a known pilot sequence $\mathbf{s}_t \triangleq [s_t(0), \dots, s_t(n_t - 1)]^T$ of length n_t followed by a number of information symbols s drawn uniformly at random from a finite, complex-valued alphabet \mathcal{S} . That is, $P(s) = 1/|\mathcal{S}|$. A codeword is then constructed by taking one or more symbols from each burst and interleaving these. For the purpose of symbol detection, we can equivalently model each symbol in the codeword s as if it had been appended to a length- n_t pilot sequence, and the so-obtained length- $(n_t + 1)$ sequence had propagated through a channel with gain \mathbf{h} that stays constant for at least $n_t + 1$ time intervals. This equivalent signal model is formulated and used only for the purpose of modeling the detection problem. In practice, the same training sequence precedes many consecutively transmitted symbols.

For the demodulation of a specific symbol s , we consider the $m \times (n_t + 1)$ matrix \mathbf{Y} of channel observations at the receiver:

$$\begin{aligned} \mathbf{Y} &= [\mathbf{Y}_t \ \mathbf{y}_s] = \left[\underbrace{\mathbf{y}_t(0) \ \cdots \ \mathbf{y}_t(n_t - 1)}_{\text{RX pilots}} \quad \underbrace{\mathbf{y}_s}_{\text{RX data for } s} \right] \\ &= [\mathbf{h}\mathbf{s}_t^T \ \mathbf{h}s] + \mathbf{E} = \mathbf{h}\mathbf{s}^T + \mathbf{E} \end{aligned} \quad (1)$$

where we defined the $(n_t + 1) \times 1$ combined pilot and symbol-of-interest vector

$$\mathbf{s} \triangleq [\mathbf{s}_t^T \ s]^T$$

and the $m \times n_t$ matrix of received pilot symbols

$$\mathbf{Y}_t \triangleq [\mathbf{y}_t(0) \ \cdots \ \mathbf{y}_t(n_t - 1)],$$

and where $\mathbf{E} \in \mathbb{C}^{m \times (n_t + 1)}$ is a noise matrix with i.i.d. $CN(0, \sigma^2)$ elements. Note that $\|\mathbf{s}\|^2 = \|\mathbf{s}_t\|^2 + |s|^2$.

We are interested in the log-likelihood ratios (LLRs)

$$\text{LLR}(b_k | \mathbf{Y}) = \log \left(\frac{\sum_{s: b_k=1} p(\mathbf{Y}|s) \cdot P(s)}{\sum_{s: b_k=0} p(\mathbf{Y}|s) \cdot P(s)} \right) \quad (2)$$

where $P(s)$ is the *a priori* probability for the symbol s , which can be either given by the stationary probability distribution for the symbols or provided as additional *a priori* information from another decoding stage. The key to the calculation of (2) is to find (or to efficiently approximate) $p(\mathbf{Y}|s)$, or equivalently $\log(p(\mathbf{Y}|s))$. We will develop expressions for general signal-space constellations. We note in passing, however, that in the special case of binary (BPSK) signalling and training we have $|s| = 1$, $\|\mathbf{s}_t\|^2 = n_t$, and $\|\mathbf{s}\|^2 = n_t + 1$. Additionally, for BPSK, (2) simplifies to

$$\text{LLR}_{\text{BPSK}}(b_k) = \underbrace{\log \left(\frac{p(\mathbf{Y}|s=+1)}{p(\mathbf{Y}|s=-1)} \right)}_{\triangleq L_y} + \underbrace{\log \left(\frac{P(s=+1)}{P(s=-1)} \right)}_{\triangleq L_a}$$

where the LLRs L_y and L_a carry the information provided by the channel observations \mathbf{Y} and the *a priori* information, respectively.

III. REVIEW OF COHERENT DETECTOR

As a baseline, we consider the case when \mathbf{h} and σ^2 are perfectly known to the receiver. Here we have directly:

$$\log(p_{\text{coh}}(\mathbf{y}_s | s)) = -m \log(\pi \sigma^2) - \frac{\|\mathbf{y}_s\|^2 + \|\mathbf{h}\|^2 |s|^2 - 2\text{Re}\{s^* \mathbf{h}^H \mathbf{y}_s\}}{\sigma^2}. \quad (3)$$

For BPSK, the LLR takes on a very simple form:

$$L_y = \frac{4}{\sigma^2} \text{Re}\{\mathbf{h}^H \mathbf{y}_s\}. \quad (4)$$

This is simply the LLR associated with an AWGN channel after filtering the received data with the matched filter \mathbf{h}^* .

IV. DETECTORS FOR UNKNOWN \mathbf{h} , KNOWN σ^2

In practice, \mathbf{h} and σ^2 will be unknown or only partially known to the receiver. The remainder of this article deals with the strategies for handling this. We treat metrics for known σ^2 (this section) and unknown σ^2 (Section V) separately. While the formulas presented here can be obtained by appropriate specialization of the results in [5], for convenience we provide concise derivations from first principles. We also provide discussion that was not present in [5].

A. Optimal noncoherent detector

Here we consider the case when \mathbf{h} is unknown, but its distribution $p(\mathbf{h})$ is known. Specifically, we assume Rayleigh fading with a known power delay profile $\{\rho_1, \dots, \rho_m\}$. That is, h_i are i.i.d. $CN(0, \rho_i)$. We also assume that σ^2 is known. This receiver is an important baseline because it is the best one can do if the realizations of \mathbf{h} are unknown. Therefore we call it “optimal”.

To proceed, define $\mathbf{y}_i \triangleq [y_{ti}(0) \ \cdots \ y_{ti}(n_t - 1) \ y_{si}]^T$ for $i = 1, \dots, N$. Then (1) can be written as $\mathbf{Y} = [\mathbf{y}_0 \ \cdots \ \mathbf{y}_{m-1}]^T$ where $\mathbf{y}_i = h_i \mathbf{s} + \mathbf{e}_i$, $i = 0, 1, \dots, m-1$, and $\mathbf{e}_i \sim CN(0, \sigma^2 \mathbf{I})$. Since \mathbf{y}_i are independent conditioned on s , we have

$$\log(p_{\text{N-MAP}}(\mathbf{y} | s)) = \sum_{i=0}^{m-1} \log(p_{\text{N-MAP}}(\mathbf{y}_i | s)), \quad (5)$$

and hence it is sufficient to compute $p_{\text{N-MAP}}(\mathbf{y}_i | s)$. Conditioned on s , we have $\mathbf{y}_i \sim CN(\mathbf{0}, \mathbf{Q}_i)$, where $\mathbf{Q}_i \triangleq \rho_i^2 \mathbf{s} \mathbf{s}^H + \sigma^2 \mathbf{I}$. Hence

$$p_{\text{N-MAP}}(\mathbf{y}_i | s) = \frac{1}{\pi^{n_t+1} |\mathbf{Q}_i|} \exp(-\mathbf{y}_i^H \mathbf{Q}_i^{-1} \mathbf{y}_i). \quad (6)$$

By using the identity $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$, we find that

$$\begin{aligned} |\mathbf{Q}_i| &= |\sigma^2 \mathbf{I} + \rho_i^2 \mathbf{s} \mathbf{s}^H| \\ &= \sigma^{2(n_t+1)} \left(1 + \frac{\rho_i^2}{\sigma^2} \mathbf{s}^H \mathbf{s} \right) = \sigma^{2n_t} \left(\sigma^2 + \rho_i^2 \|\mathbf{s}\|^2 \right). \end{aligned}$$

Also, by using the matrix inversion lemma we find that

$$\mathbf{Q}_i^{-1} = \frac{1}{\sigma^2} \mathbf{I} - \frac{1}{\sigma^2} \frac{1}{\frac{\sigma^2}{\rho_i^2} + \|\mathbf{s}\|^2} \mathbf{s} \mathbf{s}^H.$$

Using these two identities in (6) and inserting the result in (5) gives:

$$\begin{aligned} \log(p_{\text{N-MAP}}(\mathbf{y}|s)) &= -mn_t \log(\pi\sigma^2) \\ &\quad - \sum_{i=0}^{m-1} \log\left(\pi\sigma^2 + \pi\rho_i^2\|\mathbf{s}\|^2\right) \\ &\quad + \frac{1}{\sigma^2} \left(\sum_{i=0}^{m-1} \frac{|\mathbf{y}_i^H \mathbf{s}|^2}{\frac{\sigma^2}{\rho_i^2} + \|\mathbf{s}\|^2} - \|\mathbf{Y}\|^2 \right). \end{aligned} \quad (7)$$

Note the metric in (7) is *not* a quadratic function of s , unless modulation is BPSK.

For BPSK, the LLR value takes on a particularly simple form:

$$L_y = \frac{4}{\sigma^2} \sum_{i=0}^{m-1} \frac{1}{\frac{\sigma^2}{\rho_i^2} + n_t + 1} \text{Re}\{\hat{h}_i^* y_{si}\} \quad (8)$$

where $\hat{\mathbf{h}} = [\hat{h}_0, \dots, \hat{h}_{m-1}]^T$ is the ML estimate of \mathbf{h} from the training (see (9) in Section IV-B for details). Comparing (8) to (4) we can see that (8) penalizes diversity branches with a small SNR (small ρ_i^2/σ^2).

Note that though the derivation of the metric (7) involved explicit assumptions on $p(\mathbf{h})$ and $p(e)$, (7) is generally applicable regardless of the true distribution of the fading and noise. Yet, it is optimal *only* when \mathbf{h} and e have the assumed distributions. This observation limits the usefulness and generality of (7), as far as optimal receivers are desired. Also, even though the general form of the density $p(\mathbf{h})$ is known (for instance, one knows that the channel is Rayleigh fading), its parameters (e.g., ρ_i^2 above) may be unknown. These parameters could be estimated. However, if more elaborate assumptions on $p(\mathbf{h})$ would be made—say, for example, that \mathbf{h} is Gaussian with an arbitrary mean and arbitrary covariance matrix—then the approach requires the estimation of a large number of parameters which is unlikely to be viable in practice.

B. Training-based maximum-likelihood (TML) detector

In practice, the perfect CSI is unknown. Also, there may be reasons the optimal noncoherent detector (Section IV-A) cannot be used. For example, $p(\mathbf{h})$ may not be known, or complexity arguments disfavor the use of (7). In addition, the assumptions underlying (7)—specifically, that $p(\mathbf{h})$ is a Gaussian density—may be under question. It is then of interest to obtain a method that is simpler, and which does not make assumptions on the statistical distributions involved.

A common solution in practice is to first estimate \mathbf{h} from the training symbols, and insert the estimate into the coherent metric (3). The least-squares estimate (which coincides with maximum-likelihood under the assumptions made) of $\hat{\mathbf{h}}$ is

$$\begin{aligned} \hat{\mathbf{h}} &= \arg \min_{\mathbf{h}} \|\mathbf{Y}_t - \mathbf{h}\mathbf{s}_t^T\|^2 = \mathbf{Y}_t \mathbf{s}_t^* (\mathbf{s}_t^H \mathbf{s}_t)^{-1} \\ &= \frac{\mathbf{Y}_t \mathbf{s}_t^*}{\|\mathbf{s}_t\|^2} = \frac{1}{\|\mathbf{s}_t\|^2} \sum_{k=0}^{n_t-1} s_t^*(k) \mathbf{y}_t(k) \end{aligned} \quad (9)$$

where \mathbf{Y}_t denotes the noisy observations of the n_t pilots provided by the m diversity branches (see Section II). Inserting

(9) into (3), (4), we obtain the training-based maximum-likelihood (TML) detector.

We stress that obtaining an estimate of \mathbf{h} and inserting it into the coherent metric is strictly suboptimal, and the resulting performance should be worse than that of the optimal noncoherent metric (7), unless $p(\mathbf{h})$ deviates much from the distribution assumed in Section IV-A.

C. GLRT detector

The optimal noncoherent detector (Section IV-A) and the training-based detector (Section IV-B) have specific drawbacks, even disregarding possible complexity considerations. The noncoherent detector requires $p(\mathbf{h})$ to be known. Furthermore, for the so-obtained metric to be optimal, h_i must be independent and zero-mean Gaussian with variances ρ_i^2 . This is not necessarily so in practice. The training-based metric is suboptimal because inserting an estimate of \mathbf{h} , no matter how good it is, into the coherent metric is suboptimal.

In this section we derive a decision metric based on a GLR test. The GLR test is a well known tool for composite hypothesis testing (that is, hypotheses testing in the presence of unknown nuisance parameters). The GLR test is always applicable, even if a uniformly most powerful test does not exist. The main idea of GLR is to maximize the likelihood functions with respect to the unknown nuisance parameters, rather than to integrate them out (as done in the Bayesian framework). The chief advantage of the GLR paradigm over Bayesian tests is that no assumptions on the statistical distributions for the nuisance parameters need to be made. GLR tests are routinely used for various engineering problems, such as detection in radar signal processing and speech processing. They seem, however, to be much less known and used in communications applications.

To proceed, consider (1). The GLR detector computes

$$\begin{aligned} p_{\text{GLR}}(\mathbf{Y}|s, \sigma^2) &= \max_{\mathbf{h}} p(\mathbf{Y}|s, \sigma^2, \mathbf{h}) \\ &= \frac{1}{(\pi\sigma^2)^{m(n_t+1)}} \exp\left(-\frac{\|\mathbf{Y}\|^2 - \frac{\|\mathbf{Y}\mathbf{s}^*\|^2}{\|\mathbf{s}\|^2}}{\sigma^2}\right) \end{aligned} \quad (10)$$

and the probability density function in log-domain is given by

$$\begin{aligned} \log(p_{\text{GLR}}(\mathbf{Y}|s, \sigma^2)) &= -m(n_t+1) \log(\pi\sigma^2) \\ &\quad - \frac{1}{\sigma^2} \left(\|\mathbf{Y}\|^2 - \frac{\|\mathbf{Y}_t \mathbf{s}_t^* + \mathbf{y}_s s^*\|^2}{\|\mathbf{s}\|^2} \right). \end{aligned} \quad (11)$$

For BPSK, the LLR simplifies to

$$L_y = \frac{n_t}{n_t+1} \frac{4}{\sigma^2} \text{Re}\{\hat{\mathbf{h}}^H \mathbf{y}_s\} \quad (12)$$

A comparison of (12) to the corresponding training-based detector (obtained by replacing \mathbf{h} by $\hat{\mathbf{h}}$ in (4)) shows that the GLR detector takes the power allocation between the training sequence and the data symbol into account. This is especially

relevant if $\|\mathbf{s}_t\|^2$ is in the same order of magnitude as $|s|^2$, and it becomes less significant if $\|\mathbf{s}_t\|^2 \gg |s|^2$.¹

V. DETECTORS FOR UNKNOWN \mathbf{h} , UNKNOWN σ^2

We now move on to the case of unknown σ^2 . Conceptually, the noncoherent detector (Section IV-A) is easy to extend (choose a prior $p(\sigma^2)$ and marginalize the conditional density over this prior). However this tends to result in computationally very intractable problems. Since our focus is on metrics that can be easily computed in real time, we will not pursue this direction.

A. Training-based maximum-likelihood (TML) detector

If σ^2 is unknown, we can estimate it by ML from the training too:

$$\hat{\sigma}^2 = \frac{1}{mn_t} \left\| \mathbf{Y}_t - \frac{\mathbf{Y}_t \mathbf{s}_t^*}{\|\mathbf{s}_t\|^2} \right\|^2 = \frac{1}{mn_t} \sum_{k=0}^{n_t-1} \|\mathbf{y}_t(k) - \hat{\mathbf{h}}_{s_t}(k)\|^2. \quad (13)$$

The training-based detector is obtained by inserting the estimates in (9) and (13) into (3), (4).

The estimate $\hat{\sigma}^2$ provided by (13) will generally be poor if the amount of training is small. It will also be poor at low SNR, because then $\hat{\mathbf{h}}$ is inaccurate and consequently the residual $\|\mathbf{y}_t(k) - \hat{\mathbf{h}}_{s_t}(k)\|$ will deviate significantly from its “true” value $\|\mathbf{y}_t(k) - \mathbf{h}_{s_t}(k)\|$. In particular, the estimate, and hence the resulting detector, is only well defined for $n_t \geq 2$. (For $n_t = 1$ the residual $\|\mathbf{y}_t(k) - \hat{\mathbf{h}}_{s_t}(k)\| = 0$, so the noise variance estimate is zero.)

B. GLRT detector

The GLRT paradigm provides an entirely systematic framework for the derivation of decision metrics also in the case when σ^2 is unknown. Specifically, it provides a way of avoiding explicit estimation of σ^2 , which is hard as explained in Section V-A.

To compute the GLRT we are interested in the maximum of the concentrated likelihood function in (11) with respect to σ^2 :

$$\log(p_{\text{GLR, unknown } \sigma^2}(\mathbf{Y}|s)) = \max_{\sigma^2} \log(p_{\text{GLR}}(\mathbf{Y}|s, \sigma^2)) \quad (14)$$

The maximization can be easily performed in closed form. For example, start with (11), take the derivative with respect to σ^2 and set it equal to zero. This gives

$$\log(p_{\text{GLR, unknown } \sigma^2}(\mathbf{Y}|s)) = m(n_t + 1) \left(\log(m(n_t + 1)/\pi) - 1 - \log \left(\|\mathbf{Y}\|^2 - \frac{\|\mathbf{Y}_t \mathbf{s}_t^* + \mathbf{y}_s s^*\|^2}{\|\mathbf{s}\|^2} \right) \right). \quad (15)$$

Note that this GLR metric is applicable for one pilot ($n_t = 1$) even if σ^2 is unknown. This stands in contrast to the training-based detector (V-A). It is worth stressing again, that not only

¹Remember that for BPSK we have $|s| = 1$, $\|\mathbf{s}_t\|^2 = n_t$, and $\|\mathbf{s}\|^2 = n_t + 1$.

the GLRT metrics have a simple analytical form, none of them need any side information in terms of statistical distributions of the channel gain vector or noise variance.

VI. EMPIRICAL PERFORMANCE EVALUATION

We present numerical examples to illustrate the performance of the proposed detectors. In all examples we consider reception with $m = 5$ diversity branches. The (average) signal-to-noise ratio (SNR) is defined as $\text{SNR} \triangleq \text{E}[|s|^2]/\sigma^2$ where $\sigma^2 = \text{E}[|e_k|^2]$ is the variance (per complex dimension) of the noise at each antenna.

A. Loss in achievable rate due to imperfect CSI

The use of extrinsic information transfer characteristics (EXIT charts, see e.g., [8], [9]) allows us to quantify the fundamental loss in achievable rate incurred by imperfect CSI and by the use of specific, suboptimal decision metrics. To facilitate this, we consider an iterative detection and decoding setup where (similar to bit-interleaved coded modulation, BICM) an outer encoder/decoder pair is present and the detection metrics from Sections IV and V are used by the inner soft-input/soft-output SIMO detector. See, for example, (2) in Section II and Eq. (1) in [7].

In general, EXIT functions characterize the input-output behavior of a soft decoder or detector by describing the mutual information I_{extr} between the extrinsic output of the decoder and the corresponding quantity at the encoder as a function of the mutual information I_{apri} that the *a priori* input provides. By drawing the EXIT functions of the component decoders into one diagram while taking into account that the *a priori* information of the one decoder becomes the extrinsic information of the other, the exchange of extrinsic information during iterative decoding can be analyzed. Figure 1 shows an example for the soft-input/soft-output SIMO detector considering the different detection metrics discussed in this paper and different bit mappings to a 16-QAM signal space constellation. We consider both Gray mapping, and the 16-QAM mapping proposed in [7] which is optimized such that for perfect *a priori* information, the extrinsic information is maximized. The modulation is combined with an outer rate-1/2 recursive systematic convolutional (RSC) code (7, 5)₈.

One feature of the EXIT functions which is of special importance for quantifying the loss in achievable rate (which is encountered when using suboptimal detection metrics) is given by the area property² [9]: in our setup the area $A^{(\text{SIMO})}$ under the EXIT function of the SIMO detector equals the ratio $C/R^{(i)}$ of the channel capacity C and the rate $R^{(i)}$ of the signal space mapping (i.e., the inner code), given that $R^{(i)} \geq 1$ and that an optimal detector is used [9]. Furthermore, the area under the (inverse) EXIT function of the outer code (as shown in Figure 1) satisfies $A^{(\text{RSC})} = R^{(o)}$. Note that a necessary condition for convergence is given by $A^{(\text{SIMO})} \geq A^{(\text{RSC})}$

²To our knowledge the area properties of the EXIT charts have only been proven for the case of binary erasure channels. However it was observed by different authors that they hold to good approximation for other channels as well.

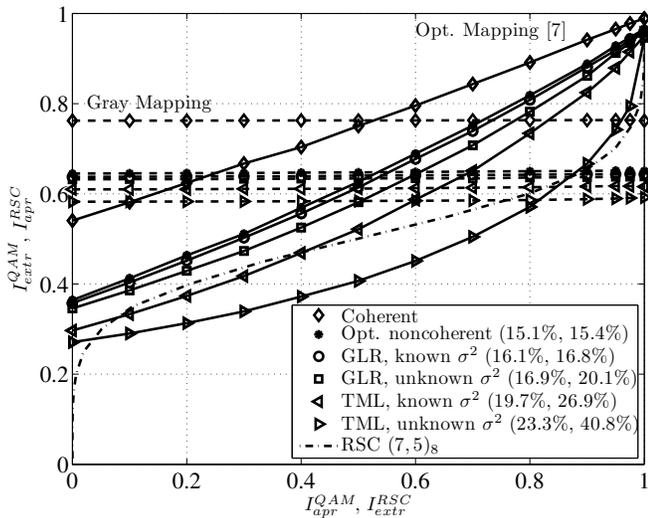


Fig. 1. EXIT charts of soft-input/soft-output detectors using the detection metrics from Sections IV and V applied to a 1×5 system, fast Rayleigh fading, uniform power delay profile, uniform noise level, $n_t = 2$ pilot symbols, and SNR = 3 dB: 16 QAM with Gray mapping (dashed curves) and optimized mapping from [7] (solid curves).

while at the same time intersections of the EXIT charts for $I_{apri}, I_{extr} < 1$ have to be avoided.

As we can observe from Figure 1, for the suboptimal detectors the area $A^{(SIMO)}$ under the EXIT function is reduced compared to the coherent detector. They suffer hence from a capacity loss (since the rate $R^{(i)}$ is fixed). Taking into account the necessary condition for convergence, the maximum achievable rate of the system in this case is given by $R_a^{(\max)} = A^{(SIMO)} \cdot R^{(i)}$ assuming that we can find an outer code that perfectly matches the characteristics of the SIMO detector. Therefore, the area $A^{(SIMO)}$ under the EXIT function of the detector provides an appropriate measure to quantify the loss in achievable rate due to imperfect CSI. The loss in achievable rate for the different detection metrics, relative to the optimal coherent detector, is given in the legend of Figure 1 for both the Gray mapping and for the optimized mapping of [7]. These values are based on measurements of the areas under the EXIT functions. The example in Figure 1 provides a good illustration of this: assuming that the RSC code $(7, 5)_8$ is used as the outer code in connection with the optimized mapping [7], convergence becomes impossible for the TML detectors while it would be maintained by using the GLR detectors. Accordingly, in the case where detection is based on the TML metrics, a lower-rate outer code needs to be employed in order to maintain reliable communication, and data rate is lost.

From Figure 1 we can draw two conclusions. First, for the proposed GLR detectors we can expect a good performance which will be close to the optimal noncoherent detector, while the conventional training-based detectors yield in part a significant capacity loss (up to 23.3% for Gray-mapped 16 QAM and 40.8% for the optimized mapping compared to 16.9% and 20.1%, respectively, for the corresponding GLR

detector). Second, the rate loss depends on the mapping from the bits into the signal space. A BICM scheme using the optimized mapping from [7] will suffer more from suboptimal detectors with imperfect CSI, than a scheme which uses Gray mapping and a LDPC code with separate detection and decoding. Especially the conventional training-based detector with unknown noise variance performs very poorly if the optimized mapping [7] is used.

B. Performance for LDPC-encoded transmissions

In order to show how the loss in achievable rate translates into a loss in SNR we show examples for a LDPC-coded system using Gray mapping and separate detection and decoding. Here, we can expect the LDPC code to converge at an SNR for which the area under the EXIT chart exceeds the amount of information which is required for the LDPC code to converge.

In the following, the codewords have 10000 bits, and the degree distributions from Table 2 of [6] (maximum variable node degree: 20). This code is optimized for the AWGN channel, but it is expected to work well for a fast Rayleigh fading channel too, especially with diversity reception. In fact, when more and more diversity paths are being combined, the channel will behave more and more like the AWGN channel. The parity check matrix was randomly constructed, and loop-removal was performed so that the resulting code matrix had no cycles of length 6 or smaller. The variable nodes were randomly interleaved.

Example 1 (Fig. 2): Here we consider Gray-mapped QPSK modulation, and an i.i.d. Rayleigh fading channel ($\rho_i = \rho$). Two pilot symbols ($n_t = 2$) precede each data burst.³ Figure 2 shows the result. As predicted in Section VI-A, the GLR detector with known noise variance performs close to the optimal noncoherent detector. The gain of the GLR detector with unknown noise variance compared to its TML counterpart is clearly visible.

Example 2 (Fig. 3): Same as Example 1, but for Gray-mapped 16-QAM. Here the difference between the methods is larger. In particular, the gain of using the GLR based metric over the TML method is increased. We believe that the reason is that by contrast to QPSK demodulation, detection of 16-QAM signals requires channel amplitude information. Hence a demodulator that inserts estimates of the channel parameters into the coherent metric will be more sensitive to errors in σ^2 and \mathbf{h} .

Example 3 (Fig. 4): Same as Example 2, but with a channel that has an exponentially decaying power profile $\rho_i = e^{-0.3i}$, $i = 0, \dots, 4$. In this example we also vary the noise variance within a codeword according to $\sigma^2 = \sigma_{\text{nominal}}^2 \cdot 10^{\xi/10}$ where ξ is uniformly distributed between -10 and +10 dB. That is, the noise variance fluctuates ± 10 dB. This models the situation on a frequency-hopping channel where each data burst is subject to different interference level. It is also a rough model for the variations in intercell downlink interference level that can be

³We choose $n_t = 2$ in this example instead of $n_t = 1$, because for $n_t = 1$ the GLR detector with unknown σ^2 is undefined.

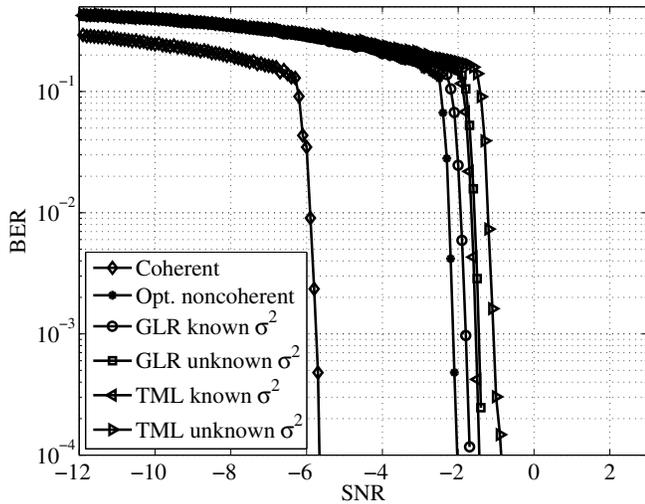


Fig. 2. 1×5 system, fast Rayleigh fading, uniform power delay profile, QPSK modulation, uniform noise level, $n_t = 2$ pilot symbols.

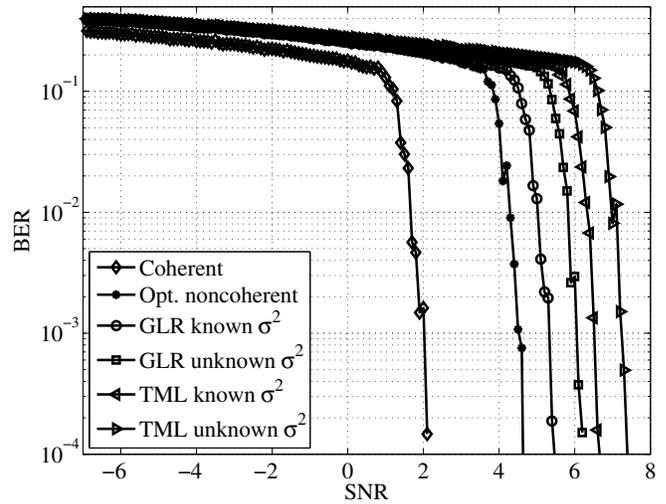


Fig. 4. Same as Figure 3, but with exponential power delay profile and fluctuating noise level.

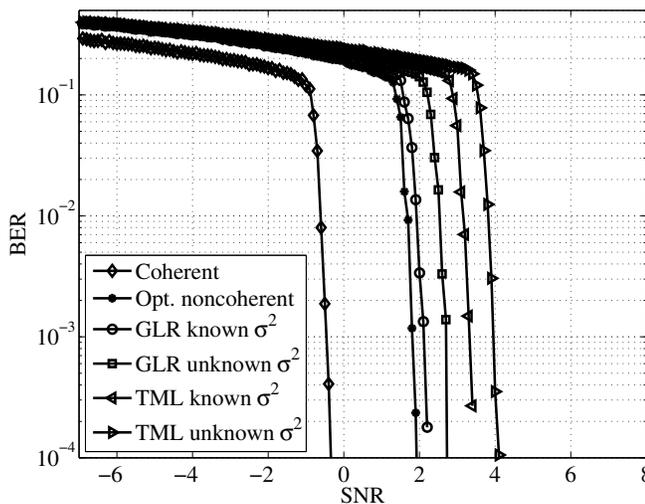


Fig. 3. Same as Figure 2, but with 16-QAM modulation.

observed in systems where neighboring base stations change their transmit powers in a rapid but uncoordinated way. In Figure 4, σ^2 is randomly drawn for each channel realization and the SNR is defined as $E[|s|^2]/E[\sigma^2]$. In this example, the noncoherent detector benefits from its knowledge of the ρ_i 's, while the GLR and TML detectors show similar gains relative to each other.

To obtain simulation code that can be used to reproduce the performance results in this section, visit www.commsys.isy.liu.se/~egl/rr.

VII. CONCLUSIONS

We have presented a range of decision metrics for diversity combining with unknown noise variance at the receiver. EXIT chart analysis and simulation results for an LDPC-coded 1×5 system show that the GLR detectors can have significant advantages compared to the training-based detectors. Specifically, the GLR metrics provide better performance than the

training-based metrics and also they do not need any *a priori* assumptions on the statistical distributions of the channel coefficients or the noise variance. In addition, in contrast to the training-based metrics, the GLR detectors work even in the case of a single pilot symbol and unknown noise variance (this is because the training-based detector is unable to estimate the noise variance in this case). The implementation complexity of the GLR metrics is about the same as for the training-based detector as none of them involves any search algorithm (the metric computation only involves single evaluations of elementary functions).

REFERENCES

- [1] P. Frenger, "Turbo decoding for wireless systems with imperfect channel estimation," *IEEE Transactions on Communications*, vol. 48, pp. 1437–1440, Sept. 2000.
- [2] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1874–1888, July 2005.
- [3] G. Taricco and G. Coluccia, "Optimum receiver design for correlated Rician fading MIMO channels with pilot-aided detection," *IEEE Journal on Selected Areas of Communications*, vol. 25, pp. 1311–1321, Sept. 2007.
- [4] R. K. Mallik, "Optimized diversity combining with imperfect channel estimation", *IEEE Trans. Information Theory*, pp. 1176–1184, Mar. 2006.
- [5] E. G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Transactions on Signal Processing*, vol. 50, pp. 937–944, Apr. 2002.
- [6] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, pp. 619–637, Feb. 2001.
- [7] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Communication Letters*, vol. 7, pp. 593–595, Dec. 2003.
- [8] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, pp. 1727–1737, Oct. 2001.
- [9] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, pp. 2657–2673, Nov. 2004.