Channel Estimation for Very Low Power MIMO Envelope Detectors

Georgios K. Psaltopoulos and Armin Wittneben
Communication Technology Laboratory, ETH Zurich, CH-8092 Zurich, Switzerland
Email: {psaltopoulos, wittneben}@nari.ee.ethz.ch

***Abstract*— We consider a MIMO system for very low-power applications, e.g, sensor networks, where the receiver employs an envelope detector at each receive antenna. Properties of such systems have been studied in [1], [2] and [3], where the channel was considered to be perfectly known at the receiver, or a linear channel estimation model was used. In this paper we propose a very simple channel estimation scheme, which allows the non-coherent MIMO envelope detector to estimate all required channel knowledge necessary for maximum likelihood detection. In contrast to linear MIMO system, complete knowledge of the channel coefficients is not required.**

I. Introduction

We consider multiple-input multiple-output (MIMO) systems where the receiver has *access only to the amplitude* (envelope) of the complex-valued received signal. This implies that the receiver does not require the ubiquitous I/Q structure, but a simple *envelope detector* suffices instead. Power-intensive circuitry like a mixer and precise local oscillator reference are not necessary, and hence implementation of such a receiver is extremely low-power [4]. This type of design is especially suited for wireless sensor networks or alike systems, where complexity and low power-consumption have highest priority. The MIMO envelope detector has been studied in the framework of *nonlinear MIMO systems* in [1], [2] and [3]. The nonlinearity refers to the operation of extracting the envelope of the complex-valued signal.

The information theoretic limits of nonlinear MIMO systems were investigated in [1] for the case of perfect channel state information (CSI) at the receiver, and in [2] for noisy CSI. It was shown that an $N \times N$ nonlinear MIMO system achieves $N/2$ spatial multiplexing gain. The performance and diversity order of the maximum likelihood (ML) detector of MIMO systems with envelope detectors were studied in [3] for the case of perfect CSI. It was shown that the receive diversity of the MIMO envelope detector drops to one half, when the uncoded rate of the modulation alphabet is higher than 1 bit/s/Hz.

So far, we either assumed perfect CSI is available at the receiver (cf. [1], [3]), or used a linear minimum mean square error (MMSE) estimation model to include noisy CSI (cf. [2]). Meanwhile, it has not been clear how a MIMO envelope detector can estimate the channel coefficients. In fact, estimating the phases of the channel coefficients is indeed not possible, since the phase information is removed when taking the envelope of the received signal. However, as we will show, complete knowledge of the channel coefficient is not required. The MIMO envelope ML detector can operate using *only* knowledge of the norm of the channel coefficients, and the norm of linear combinations of the channel coefficients. We refer to this knowledge as *sufficient CSI*, and propose a simple channel estimation scheme that allows the MIMO envelope detector to estimate the necessary CSI without relying on other means, like obtaining CSI on a feedback channel. Our proposed scheme decomposes the estimation problem in several scalar estimations. We analyze different estimation techniques for the scalar estimations, such as maximum likelihood, Bayesian and method of moments estimation, following a similar approach to [5]. Finally we suggest an estimator based on the method of moments, which provides the best trade-off between performance and complexity.

The paper is structured as follows. Section II describes the system model. In Section III, we discuss the ML detector and describe the sufficient CSI. Section IV formulates the joint estimation problem. In Section V-A, we describe our proposed scheme, and in Section V-B we analyze various scalar estimation techniques. Finally, in Section VI we assess the performance of the scalar estimators, as well as the performance of the ML detector that is based on our proposed scheme.

Notation: Throughout the paper, bold-faced italic lower and upper case letters stand for vectors and matrices, respectively. b_i is the ith element of vector $\boldsymbol{b}$, and $[\boldsymbol{b}_i]_j$ is the jth element of vector $\boldsymbol{b}_i$. $\boldsymbol{I}_N$ is the $N \times N$ identity matrix and $\boldsymbol{e}_i$ is the ith column of $\boldsymbol{I}_N$. $\mathbf{0}$ and $\mathbf{1}$ are all-zero and all-ones vectors, respectively. The circularly symmetric complex Gaussian distributed random variable $\boldsymbol{x}$ with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{R}$ is denoted by $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{m}, \boldsymbol{R})$.

II. System Model

We consider the MIMO system depicted in Fig. 1, with N_{T} transmit and N_{R} receive antennas. The signal $\boldsymbol{s} \in \mathcal{S}$ is transmitted over the stationary memoryless flat fading channel $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$, with tap-gain h_{ij} from the jth transmit to the ith receive antenna. We assume a block fading model for $\boldsymbol{H}$. The channel remains constant during a burst, while a pilot is sent along with data, and changes to an independent realization in the next burst. The received vector $\boldsymbol{z} \in \mathbb{C}^{N_{\mathrm{R}}}$ is perturbed by a zero-mean circularly symmetric Gaussian noise vector $\boldsymbol{w} \in \mathbb{C}^{N_{\mathrm{R}}}$, with autocorrelation function $\mathcal{E}[\boldsymbol{w}_k \boldsymbol{w}_l^{\mathrm{H}}] = \sigma_w^2 \boldsymbol{I}_{N_{\mathrm{R}}} \delta(k-l)$, where k and l are symbol instants. The envelope of the received signal is extracted on

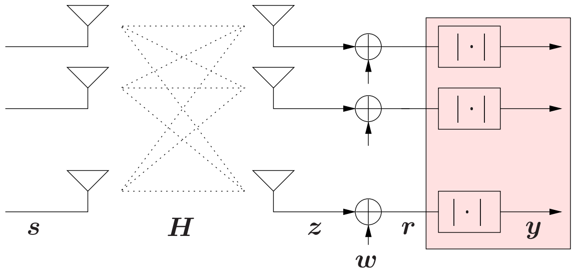

Fig. 1. MIMO System Reference Model

each antenna, yielding the observation $y_i = |z_i + w_i|$ on the ith antenna. The detector has access *only* to $\boldsymbol{y}$. The signal $\boldsymbol{s}$ at the transmitter is chosen from a set $\mathcal{S}$ of equiprobable symbols. One choice for $\mathcal{S}$ is the extension of OOK (*extended OOK*) to multiple antennas, where $\boldsymbol{s} \in \{0, \alpha\}^{N_\mathrm{T}}$, with $\alpha = \sqrt{2E_s/N_\mathrm{T}}$ [3, Appendix]. The average energy per symbol is E_s and the uncoded rate is N_T bits/s/Hz.

III. ML Detector and Sufficient CSI

The ML detector considered in [3] assumes that perfect CSI is available at the receiver, and chooses the transmit symbol $\tilde{\boldsymbol{s}}_\mathrm{ML}$ that satisfies

$$\tilde{\boldsymbol{s}}_\mathrm{ML} = \underset{\boldsymbol{s}\,\in\,\mathcal{S}}{\operatorname{argmax}}\, \mathrm{p}\left(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{H}\right) \quad (1)$$

$$= \underset{\boldsymbol{s}\,\in\,\mathcal{S}}{\operatorname{argmax}} \prod_{i=1}^{N_\mathrm{R}} \mathrm{p}\left(y_i|\boldsymbol{s},\boldsymbol{H}\right). \quad (2)$$

The second relation follows from the fact that the noise $\boldsymbol{w}$ is spatially white across different receive antennas. The distribution of y_i is Ricean:

$$\mathrm{p}\left(y_i|\boldsymbol{s},\boldsymbol{H}\right) = \frac{2y_i}{\sigma_w^2} e^{-\frac{y_i^2 + |\boldsymbol{e}_i^\mathrm{T}\boldsymbol{H}\boldsymbol{s}|^2}{\sigma_w^2}}\, \mathrm{I}_0\!\left(\frac{2y_i|\boldsymbol{e}_i^\mathrm{T}\boldsymbol{H}\boldsymbol{s}|}{\sigma_w^2}\right) \quad (3)$$

$$= \frac{2y_i}{\sigma_w^2} e^{-\frac{y_i^2 + x_i^2}{\sigma_w^2}}\, \mathrm{I}_0\!\left(\frac{2y_i x_i}{\sigma_w^2}\right), x_i = |\boldsymbol{e}_i^\mathrm{T}\boldsymbol{H}\boldsymbol{s}| \quad (4)$$

since y_i is the norm of a Gaussian random variable $r_i \sim \mathcal{CN}(\boldsymbol{e}_i^\mathrm{T}\boldsymbol{H}\boldsymbol{s}, \sigma_w^2)$ with non-zero mean [6]. x_i is the noiseless received signal at the ith antenna.

Observing (2) and (4), it is apparent that the ML detector does not require complete knowledge of all channel coefficients h_{ij}. Actually, it is sufficient to have knowledge of all elements in $\mathcal{X} = \{x_i | i = 1, \ldots, N_\mathrm{R}, \boldsymbol{s} \in \mathcal{S}\}$. We refer to knowledge of $\mathcal{X}$ as *sufficient CSI*, as opposed to perfect CSI, which implies complete knowledge of $\boldsymbol{H}$. For the considered nonlinear MIMO system, the ML detector with sufficient CSI performs exactly the same as the ML detector with perfect CSI.

The elements of $\mathcal{X}$ are the norms of linear combinations of the channel coefficients. Depending on $\boldsymbol{s}$, x_i is either the norm of one channel coefficient, or the norm of the sum of more coefficients. Furthermore, the size of $\mathcal{X}$ depends on the modulation alphabet $\mathcal{S}$. Without loss of generality, we will *consider extended OOK in the following*. For example, for those elements of $\mathcal{S}$ that have a single non-zero entry, i.e. $\boldsymbol{s}_i = [0, \ldots, 0, \alpha, 0, \ldots, 0]^\mathrm{T}$, the receiver requires knowledge of $x_j = \alpha|h_{ij}|,\ j = 1, \ldots, N_\mathrm{R}$, that is, knowledge of $|h_{ij}|,\ \ \forall i, j$. For the remaining symbols that have two or more non-zero entries, the receiver requires knowledge of the norm of combinations of the channel coefficients: $x_j = |\sum_{l=1}^{N_\mathrm{T}} h_{il}[\boldsymbol{s}_i]_l|,\ j = 1, \ldots, N_\mathrm{R}$. The cardinality of the sufficient CSI is $|\mathcal{X}| = (2^{N_\mathrm{T}} - 1)N_\mathrm{R}$. Note that $|\mathcal{X}|$ grows exponentially with N_T. Furthermore, the same information can be generated from the $N_\mathrm{T} \cdot N_\mathrm{R}$ complex-valued channel coefficients.

Finally, $|\mathcal{X}|$ *also* depends on the rate of the symbol alphabet. If for instance we use a rate 1 bit/s/Hz alphabet, comprising the symbols $\boldsymbol{s} \in \{\boldsymbol{0}, \alpha \cdot \boldsymbol{1}\}$, the required channel knowledge is only $\alpha|\sum_{i=1}^{N_\mathrm{T}} h_{ij}|,\ j = 1, \ldots, N_\mathrm{R}$, that is, $|\mathcal{X}| = N_\mathrm{R}$.

IV. Nonlinear MIMO Channel Estimation

Considering again extended OOK, we gather the sufficient CSI quantities in a vector $\boldsymbol{\theta}$. We neglect α, since it is known to the receiver:

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^\mathrm{T} \quad (5)$$

$$= [|h_{11}|, \ldots, |h_{N_\mathrm{R}1}|, \ldots, |h_{1N_\mathrm{T}}|, \ldots, |h_{N_\mathrm{R}N_\mathrm{T}}|,$$

$$|h_{11} + h_{12}|, \ldots, |h_{N_\mathrm{R}1} + h_{N_\mathrm{R}2}|, \ldots \quad (6)$$

$$\ldots, |h_{N_\mathrm{R}1} + \ldots + h_{N_\mathrm{R}N_\mathrm{T}}|]^\mathrm{T} \quad (7)$$

The structure of $\boldsymbol{\theta}$ continues as shown above, including progressively all combinations of h_{ij}. Note that certain θ_i's are related. This implies that joint estimation of $\boldsymbol{\theta}$ is optimal, albeit complicated.

Consider for example the parameters $\theta_1 = |h_{11}|$, $\theta' := \theta_{N_\mathrm{R}+1} = |h_{12}|$ and $\theta'' := \theta_{N_\mathrm{T}N_\mathrm{R}+1} = |h_{11} + h_{12}|$. These three parameters are connected through the triangular inequality, as follows:

$$|\theta_1 - \theta'| \leqslant \theta'' \leqslant \theta_1 + \theta'. \quad (8)$$

Similar inequalities relate the rest of the parameters in $\boldsymbol{\theta}$. If we would consider ML estimation, the joint estimation problem would read as

$$\hat{\boldsymbol{\theta}}_\mathrm{MLE} = \underset{\boldsymbol{\theta}\,\in\,\mathbb{R}^+}{\operatorname{argmax}}\ \mathrm{p}(\boldsymbol{Y}; \boldsymbol{\theta}), \quad (9)$$

subject to a number of inequality constraints, in the form of (8). In this case, $\boldsymbol{Y}$ is the matrix of received vectors during the training phase of the channel. Performing such an estimation is complicated, and not practical for the complexity-reduced nature of nonlinear MIMO systems.

V. Reduced Complexity Nonlinear MIMO Channel Estimation

In this section we propose a very simple scheme that estimates each θ_i in the sufficient CSI separately. First we describe the scheme, that transforms the estimation into several scalar estimations. Then, we analyze various scalar estimations techniques.

Pilot				Data
s_2	s_3	$\cdots$	$s_{2^{N_\mathrm{T}}}$	

Fig. 2. Structure of a transmit block: Pilot consists of all $\boldsymbol{s} \in \mathcal{S}$ except $\boldsymbol{s}_1 = \mathbf{0}$. Here, $N = 1$.

A. Estimation Scheme

Since estimating $\boldsymbol{\theta}$ jointly is cumbersome, we propose a suboptimal estimator that estimates the entries of $\boldsymbol{\theta}$ separately. The frame structure is shown in Fig. 2. The pilot consists of transmitting all symbols $\boldsymbol{s} \in \mathcal{S}\backslash\{\mathbf{0}\}$, N times each. N is a parameter that controls the quality of each estimation. For every $\boldsymbol{s}_i$, the receiver can then estimate the corresponding x_j, for all $j = 1, \ldots, N_\mathrm{R}$. Since the noise at the receiver is spatially white, the estimation can be performed on each receive antenna separately. This implies that for every $\boldsymbol{s}_i$, N_R scalar estimation steps take place concurrently (on each receive antenna). In total, the pilot has a length of $N_\mathrm{tot} = N(2^{N_\mathrm{T}} - 1)$. During the channel estimation phase, the receiver performs $N_\mathrm{R}(2^{N_\mathrm{T}} - 1)$ scalar estimations. Note that the exclusion of the all-zero word from the pilot implies that we transmit slightly more power per symbol in the pilot phase. To compensate for this, we use the following value $\beta = \sqrt{(2^{N_\mathrm{T}} - 1)E_s/(N_\mathrm{T}2^{N_\mathrm{T}-1})}$ instead of α for the pilot symbols, such that the average energy per pilot symbol remains E_s.

A shortcoming of this scheme is the exponential growth of N_tot with the uncoded rate N_T (for extended OOK). There are two ways to circumvent this. One possibility is using multi-level modulation, like M–OOK. In this case, it suffices to transmit the first $2^{N_\mathrm{T}} - 1$ symbols. The additional terms in $\mathcal{X}$ can be computed using the already estimated terms and the cosine rule, while the rate increases to $N_\mathrm{T} \log_2 M$. For example, if we use three modulation levels $\{0, \gamma, \delta\}$, $\gamma, \delta \in \mathbb{R}^+$, we can compute combinations of the form $|\gamma h_1 + \delta h_2|$ as follows

$$|\gamma h_1 + \delta h_2| = \sqrt{\gamma^2|h_1|^2 + \delta^2|h_2|^2 - 2\gamma\delta|h_1||h_2|\cos\angle(h_1, h_2)}, \quad (10)$$

where the angle in $\angle(h_1, h_2)$ is computed from $|h_1|$, $|h_2|$, $|h_1 + h_2|$ and the cosine rule. The same holds when using more modulation levels. In essence, we use the property $\angle(h_1, h_2) = \angle(\gamma h_1, \delta h_2)$.

Another way is transmitting only part of the alphabet $\mathcal{S}$, and computing the remaining components of $\mathcal{X}$ from the already estimated parts using trigonometric identities. E.g., it is possible to compute $|h_1+h_2+h_3|$ from $|h_1|, |h_2|, |h_3|, |h_1+h_2|, |h_1+h_3|$, and $|h_2+h_3|$. This means that in a 3×3 system it is not necessary to include the symbol $\beta\mathbf{1}$ in the pilot (extended OOK), hence reducing N_tot from $7N$ to $6N$. The gains are however low for small N_T, and become significant for high values of N_T. The added complexity is however not worth the effort for systems with moderate number of transmit and receive antennas, which are our main focus of interest.

B. Scalar Estimation

The proposed scheme breaks the estimation problem in several scalar estimations. In this subsection we present estimation techniques for the scalar estimation problem. The following description applies to estimation of any quantity of the sufficient CSI. The received symbol at the ith antenna can be written as

$$y_i = |h\beta + w_i| = ||h|\beta + w_i'| = |\theta\beta + w_i'|, \quad (11)$$

where h equals the sum of one or more channel coefficients. Let us assume that $h = \sum_{i=1}^{L} h_i$ — note that only the number L will be of interest. $\theta = |h| \geqslant 0$ is the parameter we wish to estimate. Collecting all N received symbols at the ith antenna, we obtain

$$\mathbf{y} = |\theta\mathbf{b} + \mathbf{w}'| \in \mathbb{R}^{N\times1}, \quad (12)$$

where $\mathbf{b} = \beta\mathbf{1}$, $\mathbf{w}' \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\boldsymbol{I}_N)$ and we omit the index i for the sake of brevity. The noise added at different time instances is independent, which implies that the receiver possesses N independent observations of the unknown parameter. Note that $\mathbf{y}$ in (12) refers to several instances of the received signal at *one* antenna, contrary to the previous section, where $\boldsymbol{y}$ described the received vector at *all* receive antennas at a given time instant.

Assuming θ is a non-random parameter, the distribution of $\mathbf{y}$ is given by:

$$\mathrm{p}(\mathbf{y}; \theta) = \prod_{i=1}^{N} \mathrm{p}(y_i; \theta) = \prod_{i=1}^{N} \frac{2y_i}{\sigma_w^2} e^{-\frac{y_i^2 + \beta^2\theta^2}{\sigma_w^2}} \mathrm{I}_0\!\left(\frac{2\beta y_i \theta}{\sigma_w^2}\right). \quad (13)$$

Hence, in essence we are interested in estimating the non-centrality parameter of a Rice distribution. A similar problem has been handled in [7], [5] and citations therein. There, the K-factor of a Ricean distribution is estimated with various methods. The scenario is that of a fading channel with a line-of-sight (LOS) component, where the Ricean factor describes the ratio of the power between the direct and non-direct paths. In our case, however, we already know the power of the noise (variance of non-LOS components) and we only need estimate the power of the LOS component. Our approach is similar to that in [5]. We consider the following estimation techniques:

1) Maximum Likelihood Estimation (MLE): We consider θ to be a parameter of unknown statistics. The MLE estimate follows from the maximization of the likelihood function $\ln \mathrm{p}(\boldsymbol{y}; \theta)$ with respect to θ. The estimate $\widehat{\theta}_\mathrm{MLE}$ is the root of the equation

$$\frac{\partial \ln \mathrm{p}(\mathbf{y}; \theta)}{\partial \theta} = \sum_{i=1}^{N} \left(\frac{\mathrm{I}_1\!\left(\frac{2\beta\theta y_i}{\sigma_w^2}\right)}{\mathrm{I}_0\!\left(\frac{2\beta\theta y_i}{\sigma_w^2}\right)} y_i - \beta\theta \right) = 0 \quad (14)$$

that maximizes the likelihood function (there might be several roots). Eq. (14) is nonlinear with no closed form solution. We will numerically compute the MLE and compare it with our complexity reduced estimators in Section VI.

Cramér-Rao Lower Bound (CRLB): The lower bound of the variance for estimating θ using N independent observations of the receiver output $\mathbf{y}$, is given by

$$\text{CRLB}(\theta) = \frac{1}{N \cdot i(\theta)} \tag{15}$$

where $i(\theta)$ is the Fischer information for one observation

$$i(\theta) = \mathcal{E}\left[\left(\frac{\partial \ln \mathrm{p}(y;\theta)}{\partial\theta}\right)^2\right] \tag{16}$$

$$= -\frac{4\beta^4\theta^2}{\sigma_w^4} + \int_0^\infty \frac{8\beta^2 y^3\, \mathrm{I}_1^2\left(\frac{2\beta\theta y}{\sigma_w^2}\right)}{\sigma_w^6\, \mathrm{I}_0\left(\frac{2\beta\theta y}{\sigma_w^2}\right)} e^{-\frac{y^2+\beta^2\theta^2}{\sigma_w^2}} \mathrm{d}y. \tag{17}$$

We compute the integral in (17) numerically in Section VI. The CRLB provides the minimum variance for an *unbiased* estimator. Since our estimators are only asymptotically unbiased [8], comparison with the bound is only meaningful as the sample size increases (high N).

2) Bayesian (MMSE) Estimation: The Bayesian approach takes the distribution of θ into account. In our scenario, $h \sim \mathcal{CN}(0, L\sigma_h^2)$ and $\theta = |h|$ is Rayleigh distributed, like

$$\mathrm{p}(\theta) = \frac{2\theta}{L} e^{-\theta^2/L},\ \theta \geqslant 0, \tag{18}$$

where we assumed $\sigma_h^2 = 1$. The Bayesian estimate can be computed in closed form only for $N = 1$ sample. The conditional distribution $\mathrm{p}(y|\theta)$ is identical to (13). The distribution of y is Rayleigh, computed as

$$\mathrm{p}(y) = \int_0^\infty \mathrm{p}(y|\theta)\,\mathrm{p}(\theta)\mathrm{d}\theta = \frac{2y}{L\beta^2+\sigma_w^2} e^{-\frac{y^2}{L\beta^2+\sigma_w^2}}. \tag{19}$$

The posterior distribution of θ follows from $\mathrm{p}(\theta)$, (13), (19) and Bayes rule

$$\mathrm{p}(\theta|y) = \frac{\theta}{\sigma^2} e^{-\frac{\theta^2+\nu^2}{2\sigma^2}} \cdot \mathrm{I}_0\left(\frac{\theta\nu}{\sigma^2}\right) \tag{20}$$

and is Ricean distributed with parameters

$$\{\nu, \sigma^2\} = \left\{\frac{Ly\beta}{L\beta^2+\sigma_w^2}, \frac{L\sigma_w^2}{2(L\beta^2+\sigma_w^2)}\right\}. \tag{21}$$

The Bayesian MMSE estimate is then given by

$$\begin{aligned}\widehat{\theta}_{\text{MMSE}} &= \mathcal{E}[\theta|y] = \sqrt{\frac{\pi\sigma^2}{2}}\, {}_1\mathrm{F}_1\left(-1/2; 1; -\frac{\nu^2}{2\sigma^2}\right) \\ &= \frac{\sigma_w\sqrt{L\pi}}{2\sqrt{L\beta^2+\sigma_w^2}}\, {}_1\mathrm{F}_1\left(-1/2; 1; -\frac{L\beta^2 y^2}{\sigma_w^2(L\beta^2+\sigma_w^2)}\right),\end{aligned}$$

using the mean of a Rice distribution. ${}_1\mathrm{F}_1$ is the confluent hypergeometric function. The conditional mean is an ascending function of the observation y. The minimum value is attained for $y = 0$ end equals $\widehat{\theta}_{\text{MMSE}}^{\min} = \sigma_w\sqrt{L\pi}/(2\sqrt{L\beta^2+\sigma_w^2})$, which means that the MMSE estimator can not estimate realizations of the channel below $\widehat{\theta}_{\text{MMSE}}^{\min}$. This shortcoming is prevalent especially at low SNR, where the estimator relies heavily on the prior distribution of the parameter. This limitation on the value of θ is imposed, since small values of the parameter are less likely. At high SNR, $\widehat{\theta}_{\text{MMSE}}^{\min}$ tends to zero. For $N > 1$, no closed form solution exists. We choose to average the individual Bayesian estimates, although this is not optimum as we will see in Section VI.

3) Method of Moments (MoM) Estimation: A popular estimation method which often leads to simple and consistent estimators uses the moments of the given distribution. In our case, we will consider the first two raw moments of the Ricean distribution [6]

$$\mu_1 = \mathcal{E}[y] = \frac{\sqrt{\pi}\sigma_w}{2}\, {}_1\mathrm{F}_1\left(-\frac{1}{2}; 1; -\frac{\beta^2\theta^2}{\sigma_w^2}\right), \tag{22}$$

$$\mu_2 = \mathcal{E}[y^2] = \sigma_w^2 + \beta^2\theta^2. \tag{23}$$

In general $\mu_n = f_n(\theta)$ is a function of θ. We only consider the first two moments since estimating higher order moments requires increasing the sample size accordingly. It is evident from (22) and (23) that the second moment is related to θ in a less complicated way. Let $\widehat{\mu}_n = \frac{1}{N}\sum_{i=1}^N y_i^n$ be an estimate of the nth moment using N samples. We obtain the MoM estimators

$$\widehat{\theta}_{\mu_n} = f_n^{-1}(\widehat{\mu}_n) = g_n(\widehat{\mu}_n), \tag{24}$$

where we assume the inverse $f_n^{-1}(\cdot)$ exists. In case of $\widehat{\theta}_{\mu_1}$, this inverse must be computed numerically. However, the second moment yields readily a very simple estimator:

$$\widehat{\theta}_{\mu_2} = g_2(\widehat{\mu}_2) = \frac{1}{\beta}\sqrt{\widehat{\mu}_2 - \sigma_w^2}, \qquad \widehat{\mu}_2 - \sigma_w^2 \geqslant 0. \tag{25}$$

A shortcoming is the fact that the estimate is not defined when $\widehat{\mu}_2 - \sigma_w^2 < 0$. This can happen at low SNR, in which case we set the estimate to zero. Alternatively, we could choose $\widehat{\theta}_{\text{MMSE}}^{\min}$ as the most probable channel value in this cases. However, simulations show that system performance is not affected by such a modification. It can be shown that $\widehat{\theta}_{\mu_2}$ is a consistent estimator. Using Jensen's inequality, we obtain

$$\mathcal{E}\left[\widehat{\theta}_{\mu_2}\right] = \mathcal{E}\left[\frac{1}{\beta}\sqrt{\widehat{\mu}_2 - \sigma_w^2}\right] \tag{26}$$

$$\leqslant \frac{1}{\beta}\sqrt{\mathcal{E}[\widehat{\mu}_2] - \sigma_w^2} = \theta \tag{27}$$

meaning that $\widehat{\theta}_{\mu_2}$ is biased. However, using a first-order Taylor expansion about $\mathcal{E}[\widehat{\mu}_2] = \mu_2$ for high N, we get [8]

$$\widehat{\theta}_{\mu_2} \simeq g_2(\mu_2) + \left.\frac{\mathrm{d}g_2(\widehat{\mu}_2)}{\mathrm{d}\widehat{\mu}_2}\right|_{\widehat{\mu}_2=\mu_2} \cdot (\widehat{\mu}_2 - \mu_2) \tag{28}$$

Taking the mean in (28) we find that $\widehat{\theta}_{\mu_2}$ is asymptotically unbiased,

$$\mathcal{E}\left[\widehat{\theta}_{\mu_2}\right] \simeq \theta, \qquad N \to \infty. \tag{29}$$

Using the same approximation for the variance of $\widehat{\theta}_{\mu_2}$, we obtain

$$\begin{aligned}\mathrm{var}\left(\widehat{\theta}_{\mu_2}\right) &= \mathcal{E}\left[\left(\widehat{\theta}_{\mu_2} - \mathcal{E}\left[\widehat{\theta}_{\mu_2}\right]\right)^2\right] \\ &\simeq \mathcal{E}\left[\left(g_2(\mu_2) + \left.\frac{\mathrm{d}g_2(\widehat{\mu}_2)}{\mathrm{d}\widehat{\mu}_2}\right|_{\widehat{\mu}_2=\mu_2} \cdot (\widehat{\mu}_2 - \mu_2) - \theta\right)^2\right] \\ &= \left(\left.\frac{\mathrm{d}g_2(\widehat{\mu}_2)}{\mathrm{d}\widehat{\mu}_2}\right|_{\widehat{\mu}_2=\mu_2}\right)^2 \cdot \mathrm{var}(\widehat{\mu}_2) \Rightarrow\end{aligned}$$

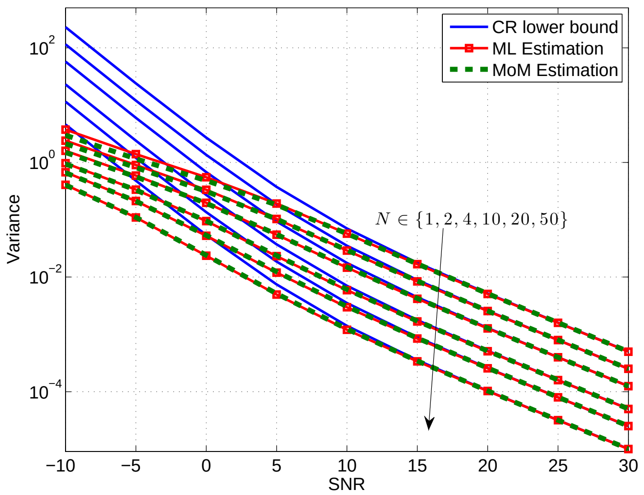

Fig. 3. Variance of MLE and MoM Estimator. The CRLB is valid above 15 dB. ML and MoM curves overlap.

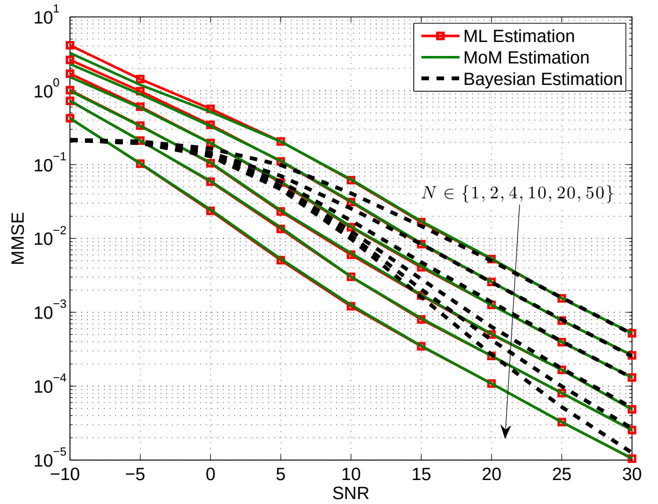

Fig. 4. MMSE comparison of MLE and MoM and Bayesian estimators. ML and MoM curves overlap almost everywhere.

$$\mathrm{var}\left(\widehat{\theta}_{\mu_2}\right) \simeq \frac{1}{4\beta^4\theta^2}\frac{\mathrm{var}\left(y^2\right)}{N} = \frac{1}{4\beta^4\theta^2 N}\left(\mathcal{E}\left[y^4\right] - \mathcal{E}\left[y^2\right]^2\right)$$
$$= \frac{\sigma_w^4 + 2\beta^2\theta^2\sigma_w^2}{4\beta^4\theta^2 N}. \quad (30)$$

which means that the variance becomes asymptotically zero, and thus $\widehat{\theta}_{\mu_2}$ is consistent.

VI. Simulation Results

Finally we assess the performance of the various scalar estimation methods presented in Section V-B. Subsequently, we apply our proposed scheme on a system and evaluate the performance of the ML detector in terms of BER.

A. Scalar Estimation

Fig. 3 depicts the variance of the MLE and MoM estimator together with the CRLB for estimating the norm a Gaussian fading channel coefficient. The SNR is defined as $\mathrm{SNR} = \frac{E_s}{\sigma_w^2}$. The two estimators have practically the same variance, which means that although $\widehat{\theta}_{\mu_2}$ is a very simple estimator, it is as powerful as the MLE. The CRLB is violated at low SNR, since the two estimators are biased. However, at high SNR and for high number of samples the bound is tight.

Fig. 4 depicts the MMSE of the MLE, the MoM and the Bayesian ($L = 1$) estimators for different block lengths. Like in Fig. 3, the performance of MLE and MoM is practically identical. For $N = 1$, the Bayesian estimator yields the best estimate for all SNR, since it is MMSE optimal. For $N > 1$, we average the individual Bayesian estimates, although this is not optimum. As a result, the MoM and ML estimates progressively improve and outperform the Bayesian estimate. At high SNR the estimators perform similarly. The Bayesian estimator performs very good with respect to the MMSE for small N and at low SNR. In the next subsection, we will see how this gain reflects to the BER performance.

Fig. 5 depicts the asymptotic variance $N \cdot \mathrm{var}(\widehat{\theta}_{\mu_2})$ of $\widehat{\theta}_{\mu_2}$ computed in (30), and $N \cdot \mathrm{CRLB}(\theta) = \frac{1}{i(\theta)}$ computed in

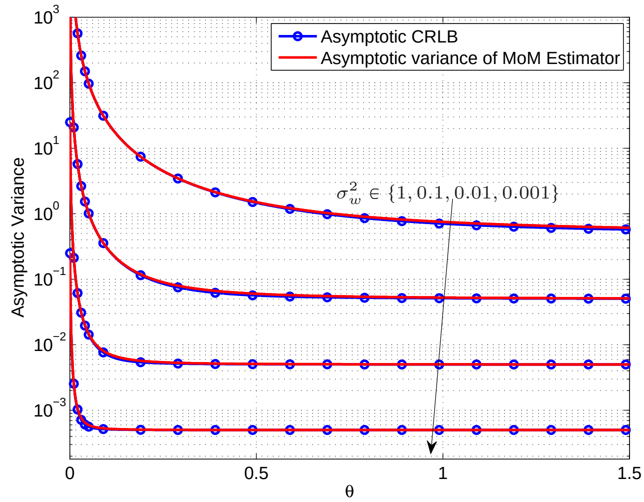

Fig. 5. Asymptotic variance of MoM estimate and CLRB.

(17), as a function of the estimation parameter, for different SNR. The MoM Estimate achieves asymptotically the CRLB. The estimation variance grows unbounded when θ goes to zero, since in this case the receiver observes solely noise, and saturates at $\sigma_w^2/2$ as θ grows.

B. System Performance

Fig. 6 depicts the BER performance of a 2×2 system, with perfect CSI and with estimated CSI. We compare the behavior of a nonlinear MIMO system that performs channel estimation with a linear MIMO reference system, that uses BPSK and spatial multiplexing, yielding the same uncoded rate of $R = 2$ bits/s/Hz. Note that the nonlinear MIMO system is not an alternative to a linear MIMO system, since its performance is clearly inferior. It is rather intended as an enhancement to single-input single-output non-coherent OOK systems, that are widely used in wireless sensor networks [4]. The linear MIMO system performs MMSE channel estimation using a properly scaled submatrix $\boldsymbol{P} \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{tot}}$ of the DFT matrix as the training sequence. The estimated channel is given by (cf. [9])

$$\widehat{\boldsymbol{H}}_\mathrm{LMMSE} = N_\mathrm{R} \boldsymbol{Y} \left(N_\mathrm{R} \boldsymbol{P}^\mathrm{H} \boldsymbol{P} + \sigma_w^2 \boldsymbol{I} \right)^{-1} \boldsymbol{P}^\mathrm{H}. \tag{31}$$

For a fair comparison, we have allocated the same resources for the pilot in both systems. Recall that for $N = 1$, the nonlinear MIMO system requires totally $(2^{N_\mathrm{T}} - 1)N = 3$ training symbols. In Fig. 6 we use $N_\mathrm{tot} = 3$ for both systems. This translates to $N = 1$ for the nonlinear MIMO system. The SNR per bit is given by $E_b/N_0 = \mathrm{SNR}/R$. We see that both systems exhibit a similar SNR penalty when the channel is estimated. The Bayesian estimator is insignificantly better than the MoM estimator at low SNR. Hence, the superiority of the Bayesian estimator at low SNR, as seen in Fig. 4, is lost in the BER performance. This indicates that at low SNR the performance degrades mainly due to large noise, and the channel estimation quality has a much smaller impact. Hence, it is meaningful to allocate more power to data transmission than to channel estimation at low SNR.

In Fig. 6 we used the smallest value for N_tot. As N_tot increases, the performance gradually improves. This is captured in Fig. 7, where the SNR penalty with respect to perfect CSI is plotted as a function of N_tot, at a BER of 10^{-2}. The performance gap of the nonlinear MIMO system with channel estimation is bigger than that of a linear MIMO system. Note that the linear MIMO system benefits from the proper design of the pilot sequence, which exhibits good correlation properties. This leads to an additional SNR gain, as seen in Fig. 7. Unfortunately, using such trainings sequences with the nonlinear MIMO system is not possible. Furthermore, we observe that the Bayesian estimator offers practically no advantage at small values of N_tot, and slowly performs worse than the MoM estimator as the length of the pilot increases.

Concluding, we can say that the simple MoM scalar estimator is the preferable choice for the complexity-reduced, low-power MIMO envelope detector. The resulting BER performance is within 1 dB from the ML detector with perfect CSI, for moderate pilot lengths.

VII. Conclusions

We presented a channel estimation scheme which enables a MIMO envelope detector to estimate all required channel information in order to perform ML detection. With the ability to estimate the channel on its own, the MIMO envelope detector can function autonomously, without relying on other means to obtain CSI. MIMO envelope detector can function autonomously, without relying on other means to obtain CSI.

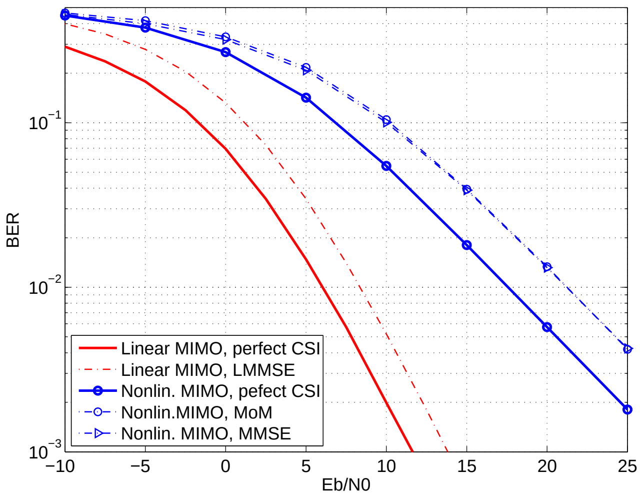

Fig. 6. BER of a 2×2 MIMO system. $N_\mathrm{tot} = 3$.

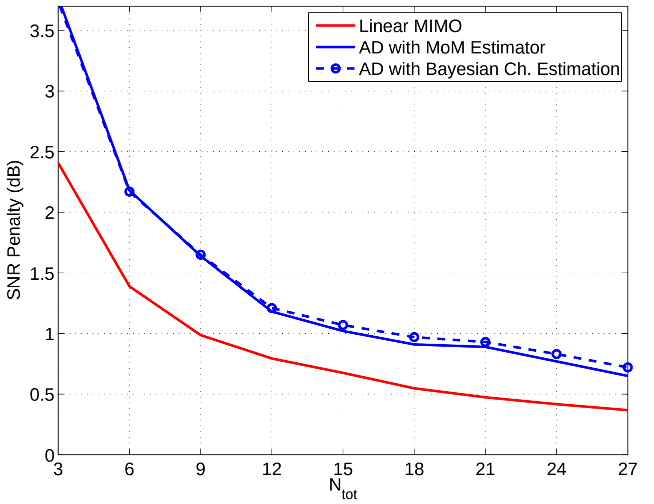

Fig. 7. 2×2 system, BER=10^{-2}. SNR penalty of various estimation methods with respect to perfect CSI, as a function of $N_\mathrm{tot} = 3N$.

References

[1] G. K. Psaltopoulos, F. Trösch, and A. Wittneben, "On Achievable Rates of MIMO Systems with Nonlinear Receivers," in *IEEE International Symposium on Information Theory*, June 2007, pp. 1071–1075.

[2] G. K. Psaltopoulos and A. Wittneben, "Achievable rates of nonlinear MIMO systems with noisy channel state information," in *IEEE International Symposium on Information Theory*, July 2008, pp. 2659–2662.

[3] ——, "Diversity and Spatial Multiplexing of MIMO Amplitude Detection Receivers," in *IEEE Personal, Indoor and Mobile Radio Communications Symposium*, September 2009.

[4] D. C. Daly and A. P. Chandrakasan, "An Energy-Efficient OOK Transceiver for Wireless Sensor Networks," *IEEE Journal of Solid State Circuits*, vol. 42, pp. 1003–1011, May 2007.

[5] C. Tepedelenlioğlu, A. Abdi, and G. Giannakis, "The Ricean K Factor: Estimation and Performance Analysis," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 799–810, July 2003.

[6] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2001.

[7] K. K. Talukdar and W. D. Lawing, "Estimation of the parameters of the Rice distribution," *The Journal of the Acoustical Society of America*, pp. 1193–1197, March 1991.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[9] M. Biguesh and A. Gershman, "Training-Based MIMO Channel Estimation: A Study of Estimator Tradeoffs and Optimal Training Signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, March 2006.