Turbo-Like Joint Data-and-Channel Estimation in Quantized Massive MIMO Systems

Fabian Steiner, Amine Mezghani, Lee Swindlehurst, Josef A. Nossek and Wolfgang Utschick

***Abstract*—We consider joint channel-and-data estimation for quantized massive MIMO systems. The estimation for both parts follows a turbo-like fashion, where the estimation error of one step is treated as additive Gaussian noise for the other. An approximate belief propagation algorithm is employed to obtain an approximate minimum mean square error estimate of both the data and channel. The performance of our scheme is compared to a Bayes optimal joint channel-and-data estimation approach by Wen *et al.* (2015). We observe that 10 turbo iterations are enough to achieve similar performance with lower complexity.**

I. Introduction

To address the explosively growing demand for faster data rates and increased energy efficiency, wireless communications faces a trend towards using a very large number (100 or more) antennas at the base station to simultaneously accommodate many users. This idea is commonly known as *Massive MIMO* [1]–[3]. However, there is still skepticism about the practicality of such systems; increasing both the number of antennas and the bandwidth by an order of magnitude for millimeter wave systems [4], [5] requires an enormous investment in hardware, particularly if a separate radio frequency chain is used for each antenna. The analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) must be extremely high-speed to handle the large bandwidths in the millimeter wave range. High-precision ADCs/DACs operating at gigasample-per-second rates have an extremely high fabrication cost and energy consumption which contradicts their employment in such situations. Low resolution ADCs and DACs, even down to a single bit of resolution, have therefore gained a lot of interest in the massive MIMO literature recently: Bounds on the capacity of quantized channels have been derived in [6]. In [7], the authors consider the problem of detection and channel estimation using linear methods based on least-squares, matched filtering and zero-forcing. They also formulate the maximum-likelihood (ML) detection problem, but do not solve it due to its claimed NP-hardness. The authors of [8] come up with a relaxed ML formulation that can be used for both data and channel detection. The work of [9] considers the possibility of using larger constellations and characterizes the tradeoff between the length of the pilot sequence and the channel coherence time for a pilot-only scheme. Joint channel-and-data (JCD) estimation for quantized massive MIMO systems is considered in [10], where the authors employ an adapted version of bilinear generalized approximate message passing (biGAMP) [11]. They characterize important system parameters (bit-error rate (BER), mean-squared error (MSE)) in the asymptotic limit of a large system analysis. In this work, we also investigate the joint channel-and-data estimation problem, but propose a turbo like detection algorithm that iterates between a channel and data estimation phase and provides improved estimates after each iteration. For that purpose, we employ generalized approximate message passing (GAMP) [12] that implements a loopy belief propagation (BP) approximating the minimum mean square error (MMSE) estimate. The suggested approach has smaller complexity than biGAMP, allows for a straightforward integration of channel decoding and a simpler state-evolution analysis.

The following paper is structured as follows: In Sect. II, we describe the employed system model. Sect. III reviews the general GAMP algorithm and points out implementation details. We propose our turbo-like joint channel-and-data detection approach in Sect. IV. Numerical validations are provided in Sect. V. We conclude in Sect. VI.

II. System Model

We consider a typical massive MIMO system in the uplink, where a flat block-fading channel model is assumed. The channel $\boldsymbol{H} \in \mathbb{C}^{M\times K}$ remains constant for a time of T_{ch} channel uses and its entries are iid. Gaussian with zero mean and unit variance, i.e., $h_{ij} \sim \mathcal{CN}(0,1)$. The K single antenna users transmit their symbols $x_{ij} \in \mathcal{X}$, which come from a power normalized QPSK constellation $\mathcal{X}$ and are collected in the matrix $\boldsymbol{X} \in \mathbb{C}^{K\times T_{\text{ch}}}$. The transmitted symbols are corrupted by additive white Gaussian noise $\boldsymbol{N} \in \mathbb{C}^{M\times T_{\text{ch}}}$, $n_{ij} \sim \mathcal{CN}(0,\sigma^2)$, resulting in the matrix $\boldsymbol{R} \in \mathbb{C}^{M\times T_{\text{ch}}}$:

$$\boldsymbol{R} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}. \tag{1}$$

After reception, the signal is quantized by the deterministic quantizer $\mathsf{Q}(\cdot)$:

$$\boldsymbol{Y} = \mathsf{Q}(\boldsymbol{R}). \tag{2}$$

The quantizer $\mathsf{Q}(\cdot)$ is implemented as a scalar quantizer $\mathsf{Q}_{\text{s}}(\cdot) : \mathbb{R} \to \mathcal{Q}$ that operates independently both on each component and the real $\Re(\cdot)$ and imaginary part $\Im(\cdot)$, i.e.,

$$[\mathsf{Q}(\boldsymbol{Y})]_{ij} = \mathsf{Q}_{\text{s}}(\Re(y_{ij})) + \mathrm{j}\,\mathsf{Q}_{\text{s}}(\Im(y_{ij})). \tag{3}$$

In the considered case, we investigate 1-bit quantizers such that $\mathsf{Q}_{\text{s}}(\cdot)$ can be replaced by the $\operatorname{sign}(\cdot)$ function and the quantization set is $\mathcal{Q} = \{\pm 1 \pm \mathrm{j}\,1\}$.

Since the channel matrix $\boldsymbol{H}$ needs to be estimated at the receiver, we allocate a slot of T_{T} time instances for training while another $T_{\text{D}} = T_{\text{ch}} - T_{\text{T}}$ channel uses are used for

transmitting the actual data. This allows a further refinement of the model as $\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}_\mathrm{T} & \boldsymbol{Y}_\mathrm{D}\end{bmatrix}$ and $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_\mathrm{T} & \boldsymbol{X}_\mathrm{D}\end{bmatrix}$:

$$\boldsymbol{Y}_\mathrm{T} = \mathsf{Q}(\boldsymbol{H}\boldsymbol{X}_\mathrm{T} + \boldsymbol{N}_\mathrm{T}) \in \mathbb{C}^{M\times T_\mathrm{T}}, \tag{4}$$

$$\boldsymbol{Y}_\mathrm{D} = \mathsf{Q}(\boldsymbol{H}\boldsymbol{X}_\mathrm{D} + \boldsymbol{N}_\mathrm{D}) \in \mathbb{C}^{M\times T_\mathrm{D}}. \tag{5}$$

The detection problem at the base station is to find an estimate for $\boldsymbol{X}_\mathrm{D}$ while only having access to the received and quantized symbols $\boldsymbol{Y}$ and the known pilots $\boldsymbol{X}_\mathrm{T}$. A Bayes optimal solution for the joint channel-and-data estimation problem is given by

$$\hat{\boldsymbol{X}}_\mathrm{D} = \mathrm{E}\left[\boldsymbol{X}_\mathrm{D}|\boldsymbol{Y}, \boldsymbol{X}_\mathrm{T}\right],$$

which is approximated by the scheme of [10] and is used as a reference in the following.

III. Generalized Approximate Message Passing

A. Summary of the Algorithm

Approximate message passing methods have been introduced in the context of compressive sensing in [13], [14] and build on the theory of BP algorithms that have been used for Bayesian inference problems for a long time. In [12], the notion of generalized approximate message passing (GAMP) is established and considers problems of the following form: A vector $\boldsymbol{x} = [x_1, \ldots, x_K]^\mathrm{T} \in \mathbb{C}^K$ with distribution $p_{\boldsymbol{x}} = \prod_{i=1}^K p_{x_i}$ is transformed by a mixing matrix $\boldsymbol{A} \in \mathbb{C}^{M\times K}$, $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$, before it is passed through a separable output channel $p_{\boldsymbol{y}|\boldsymbol{z}} = \prod_{i=1}^M p_{y_i|z_i}$. After observing $\boldsymbol{y}$, the vector $\boldsymbol{x}$ is to be estimated. The GAMP algorithm allows formulations for both approximating the maximum a posteriori (MAP) $\hat{\boldsymbol{x}} = \operatorname{argmax} p_{\boldsymbol{x}|\boldsymbol{y}}$ and MMSE estimate $\hat{\boldsymbol{x}} = \mathrm{E}\left[\boldsymbol{x}|\boldsymbol{y}\right]$. In the context of BP, these approaches are also known as max-sum and sum-product formulations. Algorithm 1 summarizes the sum-product version which is used in the following. For the considered system model of Sect. II, the abstract GAMP system model reads as

$$\boldsymbol{y} = \mathsf{Q}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}) = \mathsf{Q}(\boldsymbol{z} + \boldsymbol{n}),$$

which determines the output channel $p_{\boldsymbol{y}|\boldsymbol{z}}$ and the priors $p_{\boldsymbol{x}}$. In the next subsections, we instantiate those for our joint channel-and-data estimation problem.

B. Output Nonlinear Step

The expectation and variance in line 12 and 13 are calculated according to $p_{\tilde{z}_j|y_j}(\hat{z}_j|y_j)$, where $p_{\tilde{z}_j|y_j}(\hat{z}_j|y_j) \propto P_{y_j|z_j}(y_j|\hat{z}_j) \cdot p_{\tilde{z}_j}(\hat{z}_j)$ and $\tilde{z}_j \sim \mathcal{CN}(\hat{p}_j, \tau_j^p)$. The latter assumption is based on the central limit theorem and the definition of $\hat{z}_j$ in line 8. Its derivation is detailed in [12]. The output distribution $P_{y_j|z_j}(y_j|\hat{z}_j)$ is given as [7], [15]

$$P_{y_j|z_j}(y_j|\hat{z}_j) = \Phi\left(\frac{\Re(y_j)\cdot\Re(\hat{z}_j)}{\sqrt{\sigma^2/2}}\right)\cdot\Phi\left(\frac{\Im(y_j)\cdot\Im(\hat{z}_j)}{\sqrt{\sigma^2/2}}\right). \tag{6}$$

The output channel (6) is the same for both the channel (8) and data (16) estimation, as both are only observed after the quantization operation $\mathsf{Q}(\cdot)$. The involved expressions for the expectation and variance can be solved in closed form as shown in [16, Ch. 3.9].

Algorithm 1 Summary of the GAMP algorithm.

INPUT: observation $\boldsymbol{y}$, distributions p_{x_i} and $p_{y_i|z_i}$, N_gamp

1: **while** $t \leq N_\mathrm{gamp}$ **do**
2: **for** $i = 1$ to K **do** ▷ Initialization
3: $\hat{x}_i = \mathrm{E}_{p_{x_i}}\left[x_i\right]$
4: $\tau_i^x = \mathrm{Var}_{p_{x_i}}\left[x_i\right]$
5: **end for**
6: **for** $j = 1$ to M **do** ▷ Output Linear step
7: $\tau_j^p = \sum_i |a_{ji}|^2 \tau_j^x$
8: $\hat{z}_j = \sum_i a_{ji}\hat{x}_i$
9: $\hat{p}_j = \hat{z}_j - \tau_j^p \hat{s}_j$
10: **end for**
11: **for** $j = 1$ to M **do** ▷ Output nonlinear step
12: $\hat{s}_j = \mathrm{E}_{p_{\tilde{z}_j|y_j}}\left[\tilde{z}_j|y_j\right]$
13: $\tau_j^s = (1/\tau_j^p)\cdot\left(1 - \mathrm{Var}_{p_{\tilde{z}_j|y_j}}\left[\tilde{z}_j|y_i\right]/\tau_j^p\right)$
14: **end for**
15: **for** $i = 1$ to K **do** ▷ Input linear step
16: $\tau_i^r = \left(\sum_j |a_{ji}|^2 \tau_j^s\right)^{-1}$
17: $\hat{r}_i = \hat{x}_i + \tau_i^r \sum_j a_{ji}^* \hat{s}_j$
18: **end for**
19: **for** $j = 1$ to K **do** ▷ Input nonlinear step
20: $\hat{x}_i = \mathrm{E}_{p_{x_i|\tilde{r}_i}}\left[x_i|\tilde{r}_i\right]$
21: $\tau_i^x = \mathrm{Var}_{p_{x_i|\tilde{r}_i}}\left[x_i|\tilde{r}_i\right]$
22: **end for**
23: **end while**

C. Input Nonlinear Step

The expectation and variance in line 20 and 21 are calculated according to $p_{x_i|\tilde{r}_i}(x_i|\hat{r}_i) \propto p_{\tilde{r}_i|x_i}(\hat{r}_i|x_i)\cdot p_{x_i}(x_i)$, where it is assumed $\tilde{r}_i|x_i \sim \mathcal{CN}(x_i, \tau_i^r)$. For the channel estimation part, the priors p_{x_i} are iid. Gaussian with zero mean and unit variance as stated for the assumption on h_{ij} in Sect. II. For the data estimation, we have a uniform distribution over the QPSK set $\mathcal{X}$. In this special case, a closed form solution can be given as

$$\mathrm{E}\left[x_i|\tilde{r}_i\right] = \frac{1}{\sqrt{2}}\left(\tanh\left(\frac{2\Re(\tilde{r}_i)}{\sqrt{2}\tau_i^r}\right) + \mathrm{j}\tanh\left(\frac{2\Im(\tilde{r}_i)}{\sqrt{2}\tau_i^r}\right)\right)$$

$$\mathrm{Var}\left[x_i|\tilde{r}_i\right] = 1 - \left|\mathrm{E}\left[x_i|\tilde{r}_i\right]\right|^2$$

IV. Turbo-Like, Joint Data-and-Channel Detection

Our proposed scheme can be divided into two phases that repeat for each iteration run. In the first iteration, we obtain an initial MMSE channel estimate using GAMP. At the same time, the algorithm also provides the variance of the channel estimation error. Using the channel estimate, we then compute the MMSE estimate of the data symbols, while taking the knowledge about the channel estimation error as an additive Gaussian distributed noise component into account. The statistical properties of this noise term are derived in Sect. IV-A. As before, we model the error in the data estimation as an additional additive Gaussian noise term in the next channel estimation phase (cf. Sect. IV-B). Abiding the

paradigm of joint channel-and-data estimation, the subsequent channel estimation phases have access to both the training data and the estimate of the payload data to improve upon the previous iterations (cf. Sect. IV-C).

A. Initial Channel Estimation Phase

Using the identity $\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{B})$ the system model for the training symbols (4) can be rewritten as

$$\underline{\boldsymbol{y}}_{\mathrm{T}} = \mathsf{Q}\left((\boldsymbol{X}_{\mathrm{T}}^{\mathrm{T}} \otimes \boldsymbol{I})\underline{\boldsymbol{h}} + \underline{\boldsymbol{n}}_{\mathrm{T}}\right) \in \mathbb{C}^{M \times T_{\mathrm{T}}}, \tag{7}$$

where the underlined vectors represent the stacking operation performed by the $\mathrm{vec}(\cdot)$ operator. Rewriting the model in this form allows to use GAMP to obtain the MMSE estimate of $\underline{\boldsymbol{h}}$:

$$\hat{\underline{\boldsymbol{h}}} = \mathrm{E}\left[\underline{\boldsymbol{h}} | \underline{\boldsymbol{y}}_{\mathrm{T}}\right]. \tag{8}$$

The original channel matrix $\boldsymbol{H}$ can be expressed as

$$\boldsymbol{H} = \hat{\boldsymbol{H}} + \tilde{\boldsymbol{H}}, \tag{9}$$

where $\tilde{\boldsymbol{H}}$ denotes the zero mean channel estimation error. Because of this property, the variance of the entries $\tilde{h}_{ij}$ is given by

$$\sigma^2_{\tilde{h}_{ij}} = \mathrm{Var}\left[\tilde{h}_{ij}\right] = \mathrm{Var}\left[\underline{x}_{(j-1)M+i} | \underline{\tilde{r}}_{(j-1)M+i}\right], \tag{10}$$

which is calculated by GAMP in the nonlinear output step (cf. line 13, Algorithm 1).

B. Data Estimation Phase

Eventually, the adapted system model (5) reads as

$$\boldsymbol{Y}_{\mathrm{D}} = \mathsf{Q}((\hat{\boldsymbol{H}} + \tilde{\boldsymbol{H}})\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{N}_{\mathrm{D}}) = \mathsf{Q}(\hat{\boldsymbol{H}}\boldsymbol{X}_{\mathrm{D}} + \underbrace{\tilde{\boldsymbol{H}}\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{N}_{\mathrm{D}}}_{\tilde{\boldsymbol{N}}_{\mathrm{D}}^{\mathrm{ch}}})$$

$$= \mathsf{Q}(\hat{\boldsymbol{H}}\boldsymbol{X}_{\mathrm{D}} + \tilde{\boldsymbol{N}}_{\mathrm{D}}^{\mathrm{ch}}). \tag{11}$$

The columns $\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}, j \in \{1, \ldots, T_{\mathrm{D}}\}$ exhibit the statistical properties

$$\mathrm{E}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}\right] = \boldsymbol{0} \tag{12}$$

$$\mathrm{Cov}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}\right] = \mathrm{diag}(\sigma^2_{\tilde{n}^{\mathrm{ch}}_{\mathrm{D},1}}, \ldots, \sigma^2_{\tilde{n}^{\mathrm{ch}}_{\mathrm{D},M}}), \tag{13}$$

where the the diagonal elements $\sigma^2_{\tilde{n}^{\mathrm{ch}}_{\mathrm{D},i}}$ can be calculated as (see Appendix A))

$$\sigma^2_{\tilde{n}^{\mathrm{ch}}_{\mathrm{D},i}} = \sum_{k=1}^{K} \sigma^2_{\tilde{h}_{ik}} + \sigma^2. \tag{14}$$

For each time instance of the data phase $j \in \{1, \ldots, T_{\mathrm{D}}\}$, we have the system model

$$\boldsymbol{y}_{\mathrm{D},j} = \mathsf{Q}(\hat{\boldsymbol{H}}\boldsymbol{x}_{\mathrm{D},j} + \tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}), \quad \tilde{\boldsymbol{n}}_{\mathrm{D}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_{\tilde{\boldsymbol{n}}^{\mathrm{ch}}_{\mathrm{D},j}}), \tag{15}$$

where $\boldsymbol{C}_{\tilde{\boldsymbol{n}}^{\mathrm{ch}}_{\mathrm{D},j}}$ is the diagonal covariance matrix of (13). As in (7), we can now employ GAMP to compute the MMSE estimate for the data symbols:

$$\hat{\boldsymbol{x}}_{\mathrm{D},j} = \mathrm{E}\left[\boldsymbol{x}_{\mathrm{D},j} | \boldsymbol{y}_{\mathrm{D},j}\right]. \tag{16}$$

As before, we investigate the statistical properties of the estimation error $\tilde{\boldsymbol{X}}_{\mathrm{D}} = \boldsymbol{X}_{\mathrm{D}} - \hat{\boldsymbol{X}}_{\mathrm{D}}$, which can be calculated as

$$\sigma^2_{\tilde{x}_{\mathrm{D},ij}} = \mathrm{Var}\left[\tilde{x}_{\mathrm{D},ij}\right] = \mathrm{Var}\left[\underline{x}_{(j-1)M+i} | \underline{\tilde{r}}_{(j-1)M+i}\right] \tag{17}$$

C. Subsequent Channel Estimation Phases

At this point, we start over and improve our channel estimation by incorporating the additional knowledge about the data estimates. Hence we obtain,

$$\begin{aligned}\boldsymbol{Y} &= \mathsf{Q}(\boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}) = \\ &= \mathsf{Q}(\boldsymbol{H}\left[\boldsymbol{X}_{\mathrm{T}} \quad \hat{\boldsymbol{X}}_{\mathrm{D}} + \tilde{\boldsymbol{X}}_{\mathrm{D}}\right] + \left[\boldsymbol{N}_{\mathrm{T}} \quad \boldsymbol{N}_{\mathrm{D}}\right]) \\ &= \mathsf{Q}(\boldsymbol{H}\left[\boldsymbol{X}_{\mathrm{T}} \quad \hat{\boldsymbol{X}}_{\mathrm{D}}\right] + [\boldsymbol{N}_{\mathrm{T}} \quad \underbrace{\boldsymbol{N}_{\mathrm{D}} + \boldsymbol{H}\tilde{\boldsymbol{X}}_{\mathrm{D}}}_{\tilde{\boldsymbol{N}}_{\mathrm{D}}^{\mathrm{D}}}]) \\ &= \mathsf{Q}(\boldsymbol{H}\tilde{\boldsymbol{X}} + \tilde{\boldsymbol{N}}).\end{aligned} \tag{18}$$

The derivation of the statistical properties of $\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{D}}$ follow the lines of (12), (13) and are detailed in Appendix B.

$$\mathrm{E}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{D}}\right] = \boldsymbol{0}, \tag{19}$$

$$\mathrm{Cov}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{D}}\right] = \left(\sum_{k=1}^{K} \sigma^2_{\tilde{x}_{kj}} + \sigma^2\right)\boldsymbol{I}. \tag{20}$$

We rewrite (18) using the vectorization operator and obtain the stacked model as

$$\underline{\boldsymbol{y}} = \mathsf{Q}((\tilde{\boldsymbol{X}}^{\mathrm{T}} \otimes \boldsymbol{I})\underline{\boldsymbol{h}} + \underline{\tilde{\boldsymbol{n}}}) \in \mathbb{C}^{MT_{\mathrm{ch}}}. \tag{21}$$

Therefore, the new MMSE estimate for $\boldsymbol{H}$ is given by

$$\hat{\underline{\boldsymbol{h}}} = \mathrm{E}\left[\underline{\boldsymbol{h}} | \underline{\boldsymbol{y}}\right]. \tag{22}$$

The iterative procedure is summarized in Algorithm 2.

Algorithm 2 Iterative Data Detection and Channel Estimation

INPUT: $\boldsymbol{Y}$, $\boldsymbol{X}_{\mathrm{T}}, \sigma^2$, N_{turbo}

1: $t \leftarrow 1$
2: **while** $t \leq N_{\text{turbo}}$ **do**
3: **if** $t = 1$ **then**
4: Estimate $\hat{\boldsymbol{H}}^{(1)}$ using (8)
5: **else**
6: Estimate $\hat{\boldsymbol{H}}^{(t)}$ using (22)
7: **end if**
8: Calculate $\boldsymbol{C}^{(t)}_{\tilde{\boldsymbol{n}}^{\mathrm{ch}}_{\mathrm{D}}}$ using (13)
9: Estimate $\boldsymbol{x}^{(t)}_{\mathrm{D},j}, \forall j \in \{1, \ldots, T_{\mathrm{D}}\}$ using (16)
10: Calculate $\boldsymbol{C}^{(t)}_{\tilde{\boldsymbol{n}}^{\mathrm{D}}_{\mathrm{D},j}}, \forall j \in \{1, \ldots, T_{\mathrm{D}}\}$ using (20)
11: $t \leftarrow t + 1$
12: **end while**

D. Complexity Analysis

The complexity of the JCD approach [10] is dominated by 10 matrix multiplications per iteration as highlighted in [11, Sect. 2-H], i.e., we have an overall complexity of $\mathcal{O}(10MKT_{\mathrm{ch}}N_{\mathrm{JCD}})$, where N_{JCD} denotes the number of required iterations to converge.

As Algorithm 1 reveals, we need four matrix-vector multiplications per iteration (cf. lines 7, 8, 16, 17). For the channel estimation part of the turbo iteration the complexity is therefore $\mathcal{O}(4MKT_{\mathrm{T}}N_{\mathrm{gamp}})$ for the initial iteration and $\mathcal{O}(4MKT_{\mathrm{ch}}N_{\mathrm{gamp}})$ for the subsequent ones, when the data estimates can be used as well. We note that this is because of the

sparse mixing matrix of (7) and (21) which have only MKT_{T} and MKT_{ch} non-zero entries, respectively. For the data estimation phase, the complexity is $\mathcal{O}(4MKT_{\mathrm{D}}N_{\mathrm{gamp}})$. Hence, the overall complexity for a number of N_{turbo} turbo iterations is $\mathcal{O}((4MKT_{\mathrm{T}}+4MKT_{\mathrm{ch}}(N_{\mathrm{turbo}}-1)+4MKT_{\mathrm{D}}N_{\mathrm{turbo}})N_{\mathrm{gamp}})$.

Comparing both complexity expressions, we observe that the turbo-like scheme is clearly advantageous if $N_{\mathrm{gamp}}N_{\mathrm{turbo}} \leq N_{\mathrm{JCD}}$.

V. Simulation Results

In the following, we consider a quantized MIMO system with $M = 200$ antennas at the base station and $K = 50$ users. In order to allow comparisons with [10], we set the total channel coherence time to $T = 500$ channel uses. The signal-to-noise (SNR) ratio is defined as SNR $= 1/\sigma^2$. Fig. 1 illustrates the need for JCD estimation by considering a pilot-only scheme. Even if the number of channel uses for pilot transmission is increased from the minimum number of $T_{\mathrm{T}} = 50$ to $T_{\mathrm{T}} = 200$, we observe that the uncoded BER performance saturates which results in a large gap to the perfect CSI performance. In all three cases, the channel estimate was obtained by (8) without any further iterations.

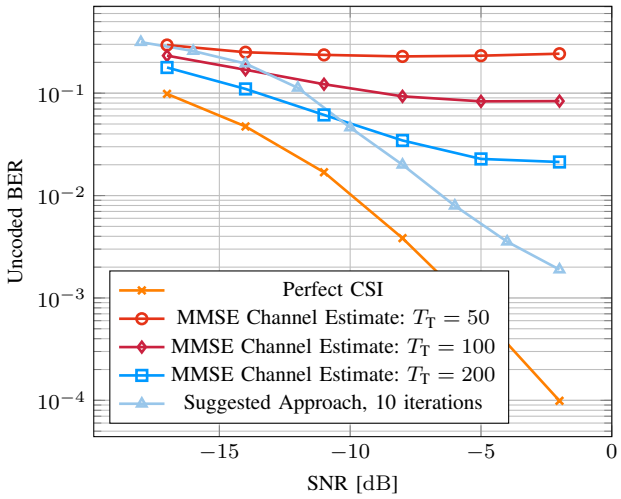

Fig. 1. Comparison of a pilot-only scheme to the perfect CSI performance.

Fig. 2 shows the performance of our proposed turbo-like scheme and compares it with the JCD approach of [10]. For both the channel and data estimation we set $N_{\mathrm{gamp}} = 10$ (cf. Algorithm 1) and vary the number of turbo iterations N_{turbo} from 3 to 20 (cf. Algorithm 2). The Bayes optimal JCD approach needs $N_{\mathrm{JCD}} = 250$ iterations to converge. We use $T_{\mathrm{T}} = 50$ channels uses for training and $T_{\mathrm{D}} = 450$ for data transmission. Increasing the number of turbo iterations to more than 10 does not improve the performance significantly. This is also reflected in Fig. 3 which shows the corresponding MSE for the channel estimation. The Bayes optimal JCD saturates at around $-14\,\mathrm{dB}$ and we obtain the same performance in the medium and high SNR regime with 10 turbo iterations.

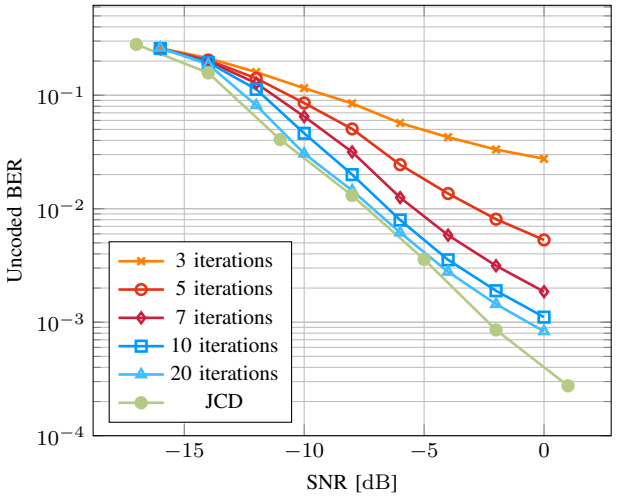

Fig. 2. Performance comparison of the Bayes optimal JCD scheme and the proposed turbo-like approach for different number of iterations.

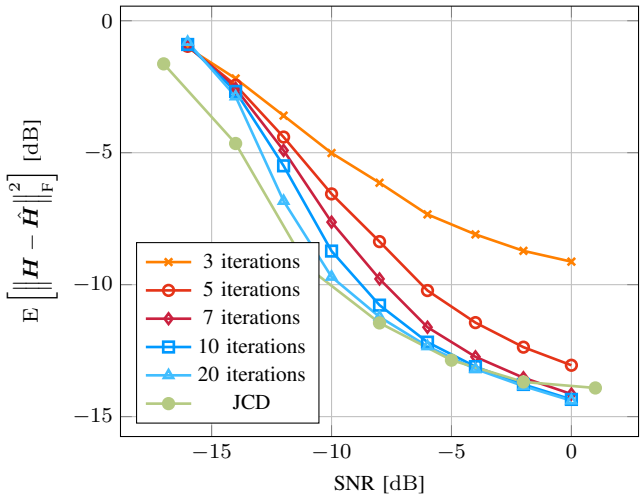

Fig. 3. Comparison of the quality of the channel estimation for both Bayes optimal JCD and the proposed scheme in terms of the MSE.

VI. Conclusion

We presented a novel approach for JCD estimation in a quantized massive MIMO system. The channel and data are estimated subsequently and the estimation error of each phase is modeled as additional additive noise for the following one. The simulation results show that the performance of a Bayes optimal JCD scheme based on biGAMP can be achieved with lower complexity. For the future, we aim at combining this approach with channel coding and perform an analysis in the large system limit that allows for tractable expressions to optimize the number of pilots.

APPENDIX

A. Derivation of the noise covariance matrix for the data estimation phase

We have

$$\mathrm{Cov}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}\right] = \mathrm{Cov}\left[\tilde{\boldsymbol{H}}\boldsymbol{x}_{\mathrm{D},j} + \boldsymbol{n}_{\mathrm{D},j}\right] = \mathrm{Cov}\left[\tilde{\boldsymbol{H}}\boldsymbol{x}_{\mathrm{D},j}\right] + \sigma^2\boldsymbol{I}.$$

We define $\boldsymbol{p} = \tilde{\boldsymbol{H}}\boldsymbol{x}_{\mathrm{D},j}$ such that $p_i = \sum_{k=1}^{K} \tilde{h}_{ik} x_{\mathrm{D},kj}$. Since both the data symbols $x_{\mathrm{D},kj}$ and the entries of the estimation error matrix $\tilde{h}_{ik}$ are stochastically independent, we calculate

$$\mathrm{E}\left[p_i p_j^*\right] = 0, \forall i \neq j,$$
$$\mathrm{E}\left[|p_i|^2\right] = \sum_{k=1}^{K} \sigma_{\tilde{h}_{ik}}^2.$$

Hence, we arrive at

$$\mathrm{Cov}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{ch}}\right] = \mathrm{diag}\left(\sum_{k=1}^{K} \sigma_{\tilde{h}_{1k}}^2 + \sigma^2, \ldots, \sum_{k=1}^{K} \sigma_{\tilde{h}_{Mk}}^2 + \sigma^2\right).$$

We note that the noise covariance matrix is the same of each time index $j \in \{1, \ldots, T_{\mathrm{D}}\}$ so that we drop its dependance in the following.

B. Derivation of the noise covariance matrix for the channel estimation phase

We have

$$\mathrm{Cov}\left[\boldsymbol{H}\tilde{\boldsymbol{x}}_{\mathrm{D},j} + \boldsymbol{n}_{\mathrm{D},j}\right] = \mathrm{Cov}\left[\boldsymbol{H}\tilde{\boldsymbol{x}}_{\mathrm{D},j}\right] + \sigma^2\boldsymbol{I}.$$

We define $\boldsymbol{p} = \boldsymbol{H}\tilde{\boldsymbol{x}}_{\mathrm{D},j}$ such that $p_i = \sum_{k=1}^{K} h_{ik}\tilde{x}_{kj}$. As before, we exploit the stochastic independence of both $\tilde{x}_{\mathrm{D},kj}$ and h_{ik}, yielding

$$\mathrm{E}\left[p_i p_j^*\right] = 0, \forall i \neq j,$$
$$\mathrm{E}\left[|p_i|^2\right] = \sum_{k=1}^{K} \sigma_{\tilde{x}_{kj}}^2.$$

In this case, the adapted noise covariance matrix turns out to be a scaled identity that may be different for each $j \in \{1, \ldots, T_{\mathrm{D}}\}$:

$$\boldsymbol{C}_{\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{D}}} = \mathrm{Cov}\left[\tilde{\boldsymbol{n}}_{\mathrm{D},j}^{\mathrm{D}}\right] = \left(\sum_{k=1}^{K} \sigma_{\tilde{x}_{kj}}^2 + \sigma^2\right)\boldsymbol{I}.$$

REFERENCES

[1] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] A. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[5] S. Talwar, D. Choudhury, K. Dimou, E. Aryafar, B. Bangerter, and K. Stewart, "Enabling technologies and architectures for 5G wireless," in *Microwave Symposium (IMS), 2014 IEEE MTT-S International*, Jun. 2014, pp. 1–4.

[6] J. Mo and R. W. Heath, Jr., "Capacity Analysis of One-Bit Quantized MIMO Systems With Transmitter Channel State Information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[7] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv:1404.7736*, Apr. 2014.

[8] J. Choi, J. Mo, and R. W. Heath, Jr., "Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems with One-Bit ADCs," *arXiv:1507.04452*, Jul. 2015.

[9] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *Proc. Int. Conf. on Commun. Workshop. (ICCW)*, Jun. 2015, pp. 1304–1309.

[10] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-Optimal Joint Channel-and-Data Estimation for Massive MIMO with Low-Precision ADCs," *IEEE Trans. Signal Process.*, to appear.

[11] J. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Approximate Message Passing – Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[12] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 2168–2172.

[13] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[14] A. Montanari, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.

[15] C. Studer and G. Durisi, "Quantized Massive MU-MIMO-OFDM Uplink," *arXiv:1509.07928*, Sep. 2015.

[16] C. E. Rasmussen, *Gaussian processes for machine learning*. MIT Press, 2006.