Noncoherent Decision-Feedback Equalization in Massive MIMO Systems

Robert F. H. Fischer and Melanie Bense

1Institut für Nachrichtentechnik, Universität Ulm, Ulm, Germany, Email: robert.fischer@uni-ulm.de
2Lehrstuhl für Informationsübertragung, Universität Erlangen-Nürnberg, Erlangen, Germany, Email: bense@lnt.de

Abstract—In this paper, a noncoherent approach to decision-feedback equalization (DFE) in multi-user massive MIMO systems is presented. Thereby, the contradicting principles of DFE, where interference of already detected symbols is canceled using actual channel knowledge, and noncoherent reception, where the symbols are detected without any channel-state information, are combined. Based on an analysis of the statistics of the interference terms in autocorrelation-based noncoherent receivers, DFE is proposed and optimized. In combination with decision-feedback differential detection of the individual users, a low-complexity high-performance scheme is established.

I. INTRODUCTION

Multiple-input/multiple-output (MIMO) systems, where the base station is equipped with a very large number of receive antennas, so-called massive MIMO, gain more and more attention, e.g., [8], [9]. As the number of channel coefficients is extremely large, one of the main challenges is to acquire accurate channel estimates. One solution to overcome the problem of requiring a large amount of pilot symbols and the need to perform channel estimation is to resort to noncoherent reception based on the idea of decision-feedback equalization (DFE) in multi-user massive MIMO systems sorted DFE is known as noncoherent decision-feedback equalization (DFE) in multi-user massive MIMO systems. In this paper, we present a noncoherent approach to DFE over the users for use in massive MIMO systems. We analyze the statistics of the interference and noise terms when using autocorrelation receivers. Employing these insights, DFE based on statistical channel knowledge is proposed and optimized. As in H-BLAST [4], on the one hand, the users are successively detected; an optimum decision order is derived. On the other hand, the detection of each user is done block-wise over a temporal block (transmission burst) utilizing DFDD. Hence, two DFE procedures (temporal/spatial) are combined to establish low-complexity high-performance noncoherent multi-user detection schemes.

The paper is organized as follows. In Sec. II, the system model is introduced and noncoherent (single-user) detection is briefly reviewed. Sec. III studies the interference in autocorrelation-based detectors and derives noncoherent DFE. Numerical results on the performance of noncoherent DFE are presented in Sec. IV. Sec. V gives some conclusions.

II. SYSTEM MODEL AND NONCOHERENT DETECTION

Throughout the paper we consider the multi-user uplink scenario depicted in Fig. 1. $N_u$ users, each equipped with a single transmit antenna, transmit to a central base station equipped with a very large number, $N_{rx} \gg 1$, of receive antennas.

In each time step (discrete index $k$), user $u$ transmits symbols $b_{u,k}$ drawn from an $M$-ary PSK constellation $\mathcal{M} = \{e^{j2\pi i/M} | i = 0, 1, \ldots, M - 1\}$. In view of the noncoherent reception based on the autocorrelation principle, differential encoding is performed at the transmitter, i.e., the transmit symbols are computed via $b_{u,k} = a_{u,k}b_{u,k-1}, b_{u,0} = 1$, where the information symbols $a_{k,u}$ to be transmitted are drawn from the $M$-ary PSK constellation, too.

A. Massive MIMO Channel Model

We assume flat-fading channels—coefficients $h_{m,u}$ in complex baseband notation—from user $u$ to receive antenna $m$. They are characterized by the user-specific power-space profile (PSP), which describes the average receive power induced by user $u$ at receive antenna $m$. If the users are placed in front of a uniform linear antenna array (antenna spacing $d_a$) and a pure path loss model (exponent $\gamma$) is studied, the PSP is given by (for details see [11])

$$P_{m,u} \equiv E\{|h_{m,u}|^2\} = \text{const.} \cdot e^{-\frac{\lambda d_e^2}{4\pi^2}}, \tag{1}$$

1This work was supported by the DFG Priority Programme 1644, “Cooperative Communication Networks.”
where \( m_u \) is the antenna element closest (distance \( d_{mu} \)) to user \( u \) and \( \xi^2_u = d^2_u/(d^2_a \gamma) \). The channel coefficients \( h_{m,u} \) are then randomly drawn from a zero-mean, circular-symmetric complex Gaussian distribution, independently of each other, with a variance according to the power-space profile and are expected to be constant during a transmission burst.

Since the application of multiple-symbol differential detection (MSDD) [2] improves the performance of noncoherent receivers significantly, we consider the received signal over the entire burst of \( N_t \) time steps. In vector/matrix notation, the block of all \( N_t \) receive signals over \( N_t \) time steps (\( N_t \times N \) matrix) can be written as

\[
R = \sum_{u=1}^{N_t} h_u b_u + N,
\]

where \( h_u \triangleq [h_{1,u}, \ldots, h_{N_t,u}]^T \) collects the channel coefficients for user \( u \) and \( b_u \triangleq [b_{1,u}, b_{2,u}, \ldots, b_{N_t,u}] \) contains the transmit symbols of user \( u \). The noise matrix \( N \triangleq \sigma_n^2 I \) gathers the circular-symmetric (zero-mean) complex Gaussian noise \( n_{m,k} \) with variance \( \sigma_n^2 \).

### B. Noncoherent Detection

It is well known, e.g., [5], [10], that the differential detection of the symbols of user \( u \) can be based on the \( N \times N \) correlation matrix

\[
Z_u \triangleq R^H W_u R,
\]

with a user-specific diagonal weighting matrix \( W_u \). In view of the PSP (1), a suited choice is [11]

\[
W_u \triangleq \text{diag}(w_{1,u}, \ldots, w_{N_t,u}), \quad w_{m,u} = e^{-\frac{|m-m_u|^2}{\zeta_u}},
\]

where the parameter \( \zeta_{u,m} \) may be optimized for each user \( u \) individually.

Based on the correlation matrix, the optimum (block-wise, MSDD) detection scheme for user \( u \) calculates [6], [11]

\[
\hat{b}_{m,u}^{\text{MSDD}} = \arg \max_{b_c \in \mathbb{C}^N} b_c Z_u^H \hat{b}_u^H.
\]

For \( N = 2 \) the simplest noncoherent detection scheme—symbol-wise differential detection—results.

The high computational complexity of MSDD can significantly be reduced with only a marginal loss by applying the principle of decision-feedback differential detection [1], [3], [12]. In [11] this strategy has been adapted to the massive MIMO setting and an optimum decision order within the temporal block (based on the actual receive symbols) has been derived. A pseudo-code description of sorted DFDD (operating over the temporal dimension) is given in [11, Fig. 3].

In the final step, estimates for the information-carrying symbols are calculated via \( \hat{a}_k = \hat{b}_k \cdot \hat{b}_{k-1}^H, k = 1, \ldots, N_t \).

### III. NONCOHERENT DECISION-FEEDBACK EQUALIZATION

Employing DFDD, the detection is done for each user individually (in parallel) over a temporal block of \( N \) symbols. Thereby, the decision-feedback principle is utilized in temporal direction. Since a central receiver is assumed, the feedback of decisions may also be applied over the users, cf. Fig. 2.

However, classical DFE requires channel knowledge for the coherent subtraction of the interference of the already decided symbols. Subsequently, we discuss how known symbols of already detected users may be used in the present correlation-based differential detection of not yet detected users. An optimized sorting over the users for this noncoherent DFE (nDFE) is derived.

### A. Correlation Matrix and Interference Terms

Assuming a system with \( N_u \) users, the receive matrix is given in (2). Since \( h_{m,u}^H W_u h_u \) is scalar, the correlation matrix (3) for the detection of a particular user \( u \) calculates to

\[
Z_u = (\sum_{v=1}^{N_u} b^*_v h^*_v + N^H) W_u (\sum_{v=1}^{N_u} b_v h_v + N) \quad (6)
\]

\[
= h_{m,u}^H W_u h_m b^*_m b_m + \sum_{v \neq m} h_{m,u}^H W_u h_v b^*_v b_v + \sum_{v \neq m} b^*_v b_v + h_{m,u}^H W_u h_y b^*_m b_y + \sum_{v \neq m} h_{m,u}^H W_u h_v b^*_v b_v + \sum_{v \neq m} b^*_v b_v + N^H W_u \sum_{v=1}^{N_u} b^*_v b_v + N + N^H W_u N. \quad (7)
\]

Term (i) contains the desired correlation coefficients for the detection of user \( u \); (ii) and (iii) are interferences due to the other users, and (iv) are “noise × signal” and “noise × noise” terms.

For the further analysis, the statistics of the interference and noise terms have to be known (cf. also [10]). The quadratic form in (i), (ii) calculates to

\[
\eta_{u,v} \triangleq E(\xi_{u,v}) = \sum_{m=1}^{N_u} w_{m,u} h_{m,v}^2 + \sum_{m \neq v} w_{m,v}^2 \quad (8)
\]

\[
\sigma^2_{u,v} \triangleq E((\xi_{u,v} - \mu_{u,v})^2) = \sum_{m=1}^{N_u} w_{m,u}^2 h_{m,v}^2 + \sum_{m \neq v} w_{m,v}^2 \quad (9)
\]

The terms in (iii) are given by \( \Xi_{u,v,m} \triangleq E(\xi_{u,v,m}) \), where the definition

\[
\Xi_{u,v,m} \triangleq h_{m,u}^H W_u h_m + \sum_{m \neq v} w_{m,u} h_{m,u}^H h_{m,v} + \sum_{m \neq v} w_{m,v} h_{m,v}^H h_{m,u} \quad (10)
\]

\(1\)W.l.o.g. we consider the block with time indices \( k = 0, \ldots, N - 1 \).

\(2\)Note: \( h_u \) is a column vector over the receive antennas, whereas \( b_u \) is a row vector over the time.
has been used. \( \xi_{u,\nu,\mu} \) is the sum over products of independent, zero-mean complex Gaussians and is well approximated by a zero-mean complex Gaussian distribution. Moreover, since the elements of \( b_{\nu}^\dagger b_{\mu} \) are drawn from the \( M \)-PSK set \( \mathcal{M} \), the entries of the hermitian matrix \( \Xi_{u,\nu,\mu} \) have the form \( \text{e}^{	ext{i} \theta u,\nu,\mu} [\xi + \xi^* \text{e}^{	ext{i} \theta u,\nu,\mu}] \). The term in brackets lies on a line in the complex plane with direction \( \text{e}^{	ext{i} \theta u,\nu,\mu} \), hence the elements of \( \Xi_{u,\nu,\mu} \) are real-valued Gaussian distributed along the lines with direction \( \text{e}^{	ext{i} \theta u,\nu,\mu} \), \( l = 0, \ldots, M - 1 \), and variance

\[
\sigma^2_{u,\nu,\mu} \approx 2 \sum_{m=1}^{N_u} w^2_{m,u} P_{u,m} P_{m,u} .
\]

Finally, the elements of the sum of noise terms (iv) are zero-mean complex Gaussian distributed with variance

\[
\sigma^2_{u,\nu} \approx 2 \sigma_n^2 \sum_{m=1}^{N_u} w^2_{m,u} + \sum_{\nu \neq \mu} \sum_{m=1}^{N_u} \eta_{\nu,\mu} P_{m,u} + \sigma_n^4 \sum_{m=1}^{N_u} w^2_{m,u} .
\]

B. Decision-Feedback Equalization

The main principle of decision-feedback equalization (successive interference cancellation) is the subtraction of the interference caused by already detected users. To this end, both, the data symbols of the users and the channel coefficients, via which the respective data symbols interfere, have to be known.

In the present situation of noncoherent detection, the channel coefficients are not known—only their statistics is available via the PSP. However, with this knowledge and assuming a given receive weighting \( W_u \), the statistics of the interference terms (ii) and (iii) can be calculated. All terms in (iii) are zero-mean Gaussian and no knowledge can be exploited. However, the means \( \eta_{u,\nu} \) of the terms in (ii) contributed by the already detected users can be utilized.

When \( D \) denotes the index set of the already detected users and \( \bar{b}_\nu \) is the estimated vector of differentially encoded symbols \( \nu \), the correlation matrix for detection of user \( u \) hence should be calculated according to

\[
Z'_{u} = Z_u - \sum_{\nu \in \bar{D}} \eta_{u,\nu} \cdot \bar{b}_{\nu}^\dagger \bar{b}_\nu ,
\]

where \( Z_u \) is the conventional correlation matrix (3).

C. Optimization and Sorting

We assume that the data vectors \( b_{\nu} \) of users \( \nu \in D \) are already (error-free) detected and that the receiver knows the PSPs of all users.\(^4\) Then it is easy to calculate the signal-to-noise-plus-interference ratio (SINR) of user \( u \notin D \) for correlation-based detection, when the mean interference in the correlation matrix is calculated according to (13). Combining the above results it reads

\[
\text{SINR}_u = \frac{\eta_{u}^2 + \sigma^2_{u,u}}{\sum_{\nu \neq u} \sum_{\nu \neq \mu} \eta_{\nu,\mu}^2 + \sum_{\nu \neq \mu} \sum_{\nu \neq \mu} \sigma^2_{\nu,\mu,u} + \sigma^2_{u,u}} .
\]

Based on this analytic solutions two optimization tasks are carried out: On the one hand, as in the BLAST sorting strategy [4], in each step the user with the highest SINR should be detected (greedy approach). In a successive way—calculating the SINR of all not yet detected users, thereby taking the interference reduction due to the already detected ones into account and deciding for the best one—the optimum decision order is found.

Please note, as in H-BLAST [4], the users are successively detected; the sorting is based on the SINR (14) which is derived from statistical channel knowledge. For each user, block-wise detection over the temporal dimension is used (the entire block is decided). When, particularly, applying DFDD, for each user the detection order over this temporal block is optimized individually (based on the actual receive symbols). Hence, two DFDE procedures (temporal DFDD; nDFE over the users) with two different (and separated) sorting strategies are present.

On the other hand, as all interference terms in (14) depend on the weighting matrix \( W_{u} \), this matrix can be optimized for each user and each detection step individually. Choosing \( W_u \) according to (4), the only free parameter is \( \varsigma_{u,\nu} \). Via numerical optimization the SINR-maximizing value can be found.

In summary, based on (13), noncoherent DFE can be performed. The optimum receive weighting and the optimum decision order can be determined based on the SINRs (14).

Noteworthy, this optimization depends only on the PSPs of the users (statistical channel knowledge) and has only to be carried out when the configuration changes. Given the interference-reduced correlation matrix \( Z'_{u} \) (13), estimates on the data symbols of user \( u \) are obtained by DFDD. However, in principle, any noncoherent detection scheme utilizing the correlation matrix can be employed at this step.

Alg. 1 summarizes this procedure—\( P \) is a matrix containing all PSPs; \( \hat{B} \) row-wise contains the estimated transmit symbol vectors \( \hat{b}_u \). In Line 5 the subroutine for detecting the respective user is denoted as \( \hat{b}_u = \text{DFDD}(Z'_u) \). A pseudocode description of this DFDD algorithm (with inherent optimal sorting over the temporal dimension, cf. Fig 2) can be found in [11, Fig. 3]. In Line 4, based on the SINRs of the not yet detected users, the optimum decision order over the users (vertical dimension in Fig 2) is determined.

IV. Numerical Results

The performance of the proposed scheme is now assessed by means of numerical simulations. To this end, a uniform linear array with \( N_{\text{rx}} = 100 \) antennas is expected, e.g., covering a

---

\(^4\) At the receiver side, only knowledge of the PSP is expected—not actual channel knowledge, \( h_{m,u} \), but only statistical one is assumed. The estimation of the PSP is much easier than that of the actual channel coefficients.
Fig. 3. Symbol error rate vs. $E_s/N_0$ (in dB). $N_u = 3$; $m_1 = 20$ ( ), $m_2 = 50$ ( ), $m_3 = 85$ ( ); colors corresponding to users. Uniform linear array with $N_a = 100$ antenna elements. Power-space profile according to (1), $\zeta_u = 20$, $\forall u$. Burst length $N = 200$. $M = 4$-ary DPSK.

Fig. 4. Top: symbol error rate vs. the position of user 2 ( ). Bottom: optimized receive window width $\zeta_{w,u}$. The decision order is indicated. $E_s/N_0 = 14$ dB. $N_u = 3$; $m_1 = 20$ ( ), $m_3 = 85$ ( ); colors corresponding to users. Uniform linear array with $N_a = 100$ antenna elements. Power-space profile according to (1), $\zeta_u = 20$, $\forall u$. Burst length $N = 200$. $M = 4$-ary DPSK.

long hallway, Transmission bursts of length $N = 200$ (equal to the DFDD block) are used; the block-fading channel changes after each burst. The results are averaged over 250 000 channel realizations. $N_u = 3$ users are active—one is located at $m_1 = 20$ (marked ), the next at $m_2 = 50$ ( ), and the last at $m_3 = 85$ ( ). The power-space profile (1) with $\zeta_u = 20$, $\forall u$, is used; it is normalized to $\sum_{\nu=1}^{N_u} P_{m,u} = 1$, $\forall u$. All users employ quaternary $(M = 4)$ DPSK.

In Fig. 3, the symbol error rates (uncoded symbols $a_{k,u}$) of the three users (via the respective color) are depicted over the signal-to-noise ratio $E_s/N_0 = 1/\sigma_n^2$; the darker curves correspond to individual detection of the users (fixed receive windowing with $\zeta_{w,u} = 15$, $\forall u$). Conventional DD, which does not perform well, is shown for comparison. Using nDFE, for the “middle” user ( ) significant gains can be achieved. The other users ( ) slightly profit from the optimization of $\zeta_{w,u}$.

For reference, the performance of coherent BLAST detection with perfect channel knowledge is shown; here all users perform nearly the same. When taking some degradation due to non-perfect actual channel knowledge into account, noncoherent reception with nDFE (over the users) and DFDD (over the temporal dimension) will perform almost comparable.

The variation of the optimum parameters in nDFE when user 2 ( ) moves along the hallway is shown in the top plot of Fig. 4 ($E_s/N_0 = 14$ dB). In the bottom plot, the optimized receive window width $\zeta_{w,u}$ is depicted. For $m_2 \leq 53$ user 3 ( ) is detected first, then user 1 ( ), and finally user 2 ( ). For $m_2 > 53$ users 1 and 3 exchange their role. If the users are close to each other, they cannot be separated based only on the PSPs. Here, coherent detection, where the phase of the channel coefficients is utilized, too, is clearly superior. However, if a sufficient spacing of the users with moderate overlap of the PSPs is present, nDFE/DFDD is attractive and is able to lower the SER of the middle user significantly.

V. CONCLUSIONS

In conclusion, it can be stated that noncoherent detection in massive MIMO systems is an attractive alternative to coherent schemes. By combining two DFE procedures low-complexity high-performance noncoherent multi-user detection schemes are enabled. First, DFE over the users utilizing statistical channel knowledge is present. Optimized parameters (sorting, weighting) can be derived purely from the statistical knowledge (PSP). Second, DFE over the temporal direction (block-wise processing) employing DFDD for each user individually is active. Here, an optimized processing order is obtained from the actual receive signal [11]. A joint spatial/temporal DFE with optimum global ordering would be possible, too.

REFERENCES