Per-Antenna Maximum Likelihood Detector for Massive MIMO

Senjie Zhang[(✉)], Zhiqiang He, and Baoyu Tian

Key Laboratory of Universal Wireless Communications, Ministry of Education,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{senjie.zhang,hezq,tianbaoyu}@bupt.edu.cn

Abstract. Massive multiple-input multiple-output (MIMO) systems have attracted extensive attention recently due to their potentials to provide high system capacity. The uplink receiver of massive MIMO is with very high complexity due to the large number of antennas at base station, while the processing time budget reduces in order of magnitude due to the low latency requirement in next generation wireless systems like 5G. In this paper, a per-antenna maximum likelihood (PAML) detector is proposed to address this issue. The proposed PAML detector separates and distributes ML detection to a group of observation nodes (VNs). Each VN associates with a receiving antenna and extracts a fraction of soft information using ML detection. The soft information from all VNs is accumulated before being delivered to channel decoder. Thus the degree of parallelism scales up with antenna number. Furthermore, VN of PAML detector works independently each other. High localization benefits parallel processing a lot. Simulation results show that PAML detector approaches ML detector and outperforms MMSE detector.

Keywords: Massive MIMO · Maximum likelihood detection · MLD
Belief propagation · BP · Parallel processing · 5G

1 Introduction

Massive multiple-input multiple-output (MIMO) systems, also called as large-scale antenna systems, attracted extensive attention in the past few years as a promising key technology for next generation wireless systems like 5G [1].

In massive MIMO, dozens of or even hundreds of antennas are employed at base station (BS) [2,3]. The number of receiving antennas is much higher than the number of concurrent uplink user equipment (UE). Large antenna array provides high degrees of freedom and increase the system capacity significantly [4–6]. Although asymptotical analysis based on random matrix theory demonstrates that massive MIMO systems can achieve the capacity gain with simple signal processing methods, traditional MIMO detectors are still favorite for robust to different channel models. Due to the large number of antennas, traditional MIMO detectors become very complicated.

Another important point is the low latency requirements for next generation wireless systems like 5G [7]. A widely accepted end-to-end latency value is 1us. The time budget for signal processing may reduce to be lower than 500 ns. It makes the design of massive MIMO detector further challenging. Various algorithms were proposed to reduce complexity of massive MIMO detector, e.g. Neumann series expansion [5] and Gauss-Seidel method [8]. However these algorithms does not exploit the parallelism among receiving antennas so they are not efficient for low latency processing.

Belief propagation (BP) detector serves as an alternative solution to this problem [9–11]. In BP detector, the MIMO channel at a given subcarrier is represented by a factor graph where each receiving antenna is referred as an observation node (VN) and each UEs transmitted constellation symbol as a symbol node (SN). VN and SN is equivalent to variable node and check node in LDPC. Thus BP algorithm can be applied for detection in similar manner as its application in LDPC decoders. BP detector requires exchanging belief information between observation nodes and symbol nodes iteratively. The amount of information exchanged is very high. And iterative processing is less attractive for low latency processing, although BP detector is with high degree of parallelism.

Based on BP detector, in this paper we proposed the per-antenna maximum likelihood (PAML) detector for massive MIMO. PAML detector has similar framework with BP detector, while it replaces BP detectors SN with code block node (CN), i.e. PAML detector treats each UEs code block (cross multiple subcarriers) as the check node in LDPC.

In PAML detector, a group of VNs are employed. Each VN associates with a receiving antenna. From its associated antenna, VN extracts the log-likelihood ratio (LLR) for multiple UEs code blocks using maximum likelihood detection (MLD) described in [12]. The CN of PAML detector accumulates the LLR from different VNs and deliver the combined LLR to channel decoders. The enhanced LLR of each UE, extracted by channel decoder, can be used for a priori information to VNs. By utilizing the diversity provided by channel coding, this improvement reduces iteration number. At typical massive MIMO scenario, where the number of receiving antennas is much higher than the number of concurrent uplink UEs, PAML detector with two iterations outperforms MMSE detector and approaches traditional ML detector. Compared with MMSE and traditional ML detector, PAML detector is more suitable for low latency parallel processing.

The rest of the paper is organized as follows. Section 2 describes the system model. Section 3 presents details of PAML detector. Numerical simulation results are given in Sect. 4. Section 5 concludes the entire paper.

Notation: In this paper, lower-case and upper-case boldface letters are used to denote vectors and matrices, respectively. The operations $(\cdot)^H$ and $E\{\cdot\}$ stand for conjugate transpose and expectation, respectively. The entry in the i-th row and j-th column of $\boldsymbol{A}$ is a_{ij}; the k-th entry of $\boldsymbol{a}$ is a_k.

2 System Model

Consider a MIMO-OFDM system with M uplink UEs, each with a single antenna, and one BS having a large number of receiving antennas. Let N denote the number of BS antennas; N is in the range of tens to hundreds. In massive MIMO, $M \ll N$.

Given that the system has Q subcarriers, and the number of bits carried per constellation symbol is M_c. Each UEs channel encoder generates a code block with $Q \cdot M_c$ bits.

The code block generated by the j-th UE is separated in Q pieces: $\boldsymbol{x}_1^{<j>}$, $\boldsymbol{x}_2^{<j>}, \ldots, \boldsymbol{x}_Q^{<j>}$ and transmitted o Q subcarriers, where $\boldsymbol{x}_q^{<j>}$ is an $M_c \times 1$ vector of bits, $j = 1, 2, \ldots, M$ is the index of UE and $q = 1, 2, \ldots, Q$ is the index of subcarrier.

Given the q-th subcarrier, denote $\boldsymbol{s}_q = [s_{q,1}, s_{q,2}, \ldots, s_{q,M}]^T$ as an $M \times 1$ vector representing constellation symbols transmitted by all UEs. The entries of $\boldsymbol{s}_q$ are chosen from some complex constellation $\mathcal{C}$ with 2^{M_c} points, e.g. $M_c = 2$ for quaternary phase-shift keying (QPSK). Constellation $\mathcal{C}$ is normalized as $E\left\{\|s_{q,j}\|^2\right\} = \frac{1}{M}$ so the total transmitted power is one.

The entries of $\boldsymbol{s}_q$ is determined by using mapping function $s_{q,j} = \mathtt{map}(\boldsymbol{x}_q^{<j>})$ (e.g. gray mapping). The $M \cdot M_C \times 1$ vector $\boldsymbol{x}_q = [x_{q,1}, x_{q,2}, \ldots, x_{q,M \cdot M_c}]^T$ representing the transmitted bits from all UE is obtained by stacking $\boldsymbol{x}_q^{<1>}$, $\boldsymbol{x}_q^{<2>}, \ldots, \boldsymbol{x}_q^{<M>}$. The k-th bit of $\boldsymbol{x}_q$ equals the u-th bits transmitted by the j-th UE where

$$k = (j-1) \cdot M_c + u, u \in [1, M_c], k \in [1, M \cdot M_c] \tag{1}$$

In this paper, the logical zero for a bit is represented by amplitude level $x_{q,k} = -1$, and the logical one by $x_{q,k} = +1$, $k = 1, \ldots, M \cdot M_C$.

The vector of constellation symbols and the vector of transmitted bits is related as

$$\boldsymbol{s}_q = \mathtt{vmap}(\boldsymbol{x}_q) = [\mathtt{map}(\boldsymbol{x}_q^{<1>}), \mathtt{map}(\boldsymbol{x}_q^{<2>}), \ldots, \mathtt{map}(\boldsymbol{x}_q^{<M>})]^T \tag{2}$$

At the q-th subcarrier the vector of received signals can be expressed by

$$\boldsymbol{y}_q = \boldsymbol{H}_q \cdot \boldsymbol{s}_q + \boldsymbol{n}_q \tag{3}$$

where $\boldsymbol{H}_q$ is a $N \times M$ complex CSI matrix representing the MIMO channel and $\boldsymbol{n}_q$ is a $N \times 1$ vector of independent zero-mean complex Gaussian noise with variance σ^2 per real component.

For convenience of describing per-antenna maximum likelihood detector, Eq. (3) can be reformulated as

$$\boldsymbol{y}_q = \begin{bmatrix} y_{q,1} \\ y_{q,2} \\ \vdots \\ y_{q,N} \end{bmatrix} = \begin{bmatrix} h_{q,11} & h_{q,12} & \cdots & h_{q,1M} \\ h_{q,21} & h_{q,22} & \cdots & h_{q,2M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{q,N1} & h_{q,N2} & \cdots & h_{q,NM} \end{bmatrix} \cdot \boldsymbol{s}_q + \boldsymbol{n}_q = \begin{bmatrix} \boldsymbol{h}_{q,1} \\ \boldsymbol{h}_{q,2} \\ \vdots \\ \boldsymbol{h}_{q,N} \end{bmatrix} \cdot \boldsymbol{s}_q + \boldsymbol{n}_q \tag{4}$$

where $\boldsymbol{h}_{q,i}$ is a $1 \times M$ complex CSI vector representing the MIMO channel between the i-th antenna and all M uplink UEs at the q-th subcarrier, and $i = 1, 2, \ldots, N$ is the index of the receiving antenna.

The CSI vector $\boldsymbol{h}_{q,i}$ is obtained using the received pilots at the i-th antenna. The pilot from different UE distinguishes each other in frequency domain or code domain, e.g. ZC sequences in LTE systems. So the VN associated with the i-th antenna can be assumed to know $\boldsymbol{h}_{q,i}$.

3   Per-Antenna Maximum Likelihood Detector

Figure 1 gives a flowchart of the proposed PAML detector.
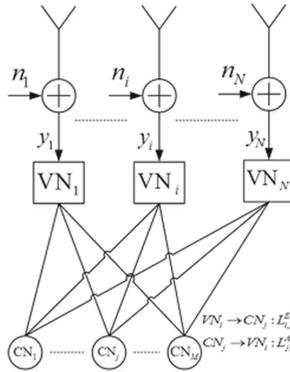
Fig. 1. The flowchart of PAML detector with factor graph based model

As Fig. 1 shown, PAML detector consists of N VNs and M CNs. Each VN associates with a receiving antenna. Each CN corresponds to an uplink UE. VNs and CNs are connected in fully meshed manner. Between VNs and CNs, LLR of each modulated bits are exchanged iteratively. VN sends extrinsic LLR to CN. CN sends *a priori* LLR to VN.

VN_i takes the received signals from the i-th antenna, and the *a priori* LLR from all CNs as its input. It extracts extrinsic LLR of bits transmitted by the j-th UE, denoted a s$\boldsymbol{L}_{i,j}^{E}$, and sends $\boldsymbol{L}_{i,j}^{E}$ to CN_j ($j = 1, 2, \ldots, M$). $\boldsymbol{L}_{i,j}^{E}$ is a $M_c \times Q$ matrix. The q-th column of $\boldsymbol{L}_{i,j}^{E}$ contains LLR of the M_C bits transmitted at the q-th subcarrier by the j-th UE.

CN_j takes $\boldsymbol{L}_{i,j}^{E}$, $i = 1, 2, \ldots, N$ as its input. It combines all these LLR matrices, de-interleaves them, and passes them through the soft in/out channel decoder for the j-th UE. Channel decoders extrinsic LLR output is interleaved and sent to VNs as *a priori* LLR. These *a priori* LLR from CN_j to VN_i (interleaved version of the j-th channel decoders extrinsic output), denoted as $\boldsymbol{L}_{j,i}^{A}$, has same dimension and correspondence as $\boldsymbol{L}_{i,j}^{E}$. At first iteration, $\boldsymbol{L}_{j,i}^{A}$ is an all-zero matrix since no channel decoder has been called yet.

3.1   Per-antenna Maximum Likelihood Detection

VN_i stacks all of its *a priori* LLR inputs to be a global $M \cdot M_c \times Q$ *a priori* LLR matrix as

$$\boldsymbol{L}^A = \begin{bmatrix} \boldsymbol{l}_1^A \ \boldsymbol{l}_2^A \ \dots \ \boldsymbol{l}_Q^A \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{1,i}^A{}^T \ \boldsymbol{L}_{2,i}^A{}^T \ \dots \ \boldsymbol{L}_{M,i}^A{}^T \end{bmatrix}^T \tag{5}$$

where $\boldsymbol{l}_q^A$ is the q-th column of $\boldsymbol{L}^A$. The k-th entry of $\boldsymbol{l}_q^A$ gives the *a priori* information of $x_{q,k}$ as

$$l_{q,k}^A = \log \frac{P(x_k = +1)}{P(x_k = -1)} \tag{6}$$

As described in [12], at the q-th subcarrier, extrinsic LLR of bit $x_{q,k}$ can be obtained using a scalar input $y_{q,i}$ as

$$l_{q,k}^E = \log \frac{\sum_{\boldsymbol{x}_q \in \mathbb{X}_{+1}^k} p(y_{q,i} \mid \boldsymbol{x}_q) \cdot \exp(0.5 \cdot \boldsymbol{x}_{[k]}^T \cdot \boldsymbol{l}_{q,[k]}^A)}{\sum_{\boldsymbol{x}_q \in \mathbb{X}_{-1}^k} p(y_{q,i} \mid \boldsymbol{x}_q) \cdot \exp(0.5 \cdot \boldsymbol{x}_{[k]}^T \cdot \boldsymbol{l}_{q,[k]}^A)} \tag{7}$$

where $\mathbb{X}_{+1}^k$ is the set of $2^{M \cdot M_c - 1}$ possible values of $\boldsymbol{x}_q$ having $x_{q,k} = +1$; $\mathbb{X}_{-1}^k$ is the set of $2^{M \cdot M_c - 1}$ possible values of $\boldsymbol{x}_q$ having $x_{q,k} = -1$; $\boldsymbol{x}_{[k]}$ denotes the subvector of $\boldsymbol{x}_q$ obtained by omitting its k-th entry x_k; $\boldsymbol{l}_{q,[k]}^A$ denotes the subvector of $\boldsymbol{l}_q^A$ also omitting its k-th entry $l_{q,k}^A$.

From (3) and (4) it is easily to found

$$p(y_{q,i} \mid \boldsymbol{x}_q) = \frac{\exp(-\frac{1}{2\sigma^2} \cdot \|y_{q,i} - \boldsymbol{h}_{q,i} \cdot \boldsymbol{s}_q\|^2)}{2\pi\sigma^2} \tag{8}$$

where $\boldsymbol{s}_q = \mathtt{vmap}(\boldsymbol{x}_q)$.

With approximation $\log(e^{a_1} + e^{a_2}) \approx \mathtt{max}(a_1, a_2)$, the extrinsic LLR of bit $x_{q,k}$ becomes

$$\begin{aligned} l_{q,k}^E = &\frac{1}{2} \cdot \max_{\boldsymbol{x} \in \mathbb{X}_{+1}^k} \left\{ -\frac{1}{\sigma^2} \cdot \|y_{q,i} - \boldsymbol{h}_{q,i} \cdot \boldsymbol{s}_q\|^2) + \boldsymbol{x}_{[k]}^T \cdot \boldsymbol{l}_{q,[k]}^A \right\} - \\ &\frac{1}{2} \cdot \max_{\boldsymbol{x} \in \mathbb{X}_{-1}^k} \left\{ -\frac{1}{\sigma^2} \cdot \|y_{q,i} - \boldsymbol{h}_{q,i} \cdot \boldsymbol{s}_q\|^2) + \boldsymbol{x}_{[k]}^T \cdot \boldsymbol{l}_{q,[k]}^A \right\} \end{aligned} \tag{9}$$

where $l_{q,k}^E$ is the u-th entry of the q-th column in VN_is output $\boldsymbol{L}_{i,j}^E$. The indexes u and k obey the relation in (1).

In uplink MIMO, inter-UE interference and noise constrain system performance. Inter-UE interference is mitigated by ML detection described in (9). Averaging multiple observations, as the essential anti-noise method, cannot be carried out in VN since each VN has only one observation from the i-th antenna. To boost performance against noise, multi-antenna combination is introduced in CN.

3.2   Multi-antenna Combination and Iterative Detection

CN_j associates with the j-th UE. For the $M_C \times Q$ bits in j-th UE's code block, CN_j has N sets of LLR from N VNs, denoted as $\boldsymbol{L}_{i,j}^E$, $i = 1, 2, \ldots, N$. Using typical BP iteration, CN_j outputs the *a priori* LLR of these bits in code block to VN_i as

$$\boldsymbol{L}_{j,i}^A = \mathtt{dec}(\sum_{1 \leq i' \leq N, i' \neq i} \boldsymbol{L}_{i',j}^E) \tag{10}$$

where $\mathtt{dec}(\cdot)$ represents the procedure of de-interleaving, soft-in/out decoding and re-interleaving. Any soft-in soft-out channel decoding scheme is applicable, e.g. BahlCCockeCJelinekCRaviv (BCJR) algorithm.

Introduction of channel decoding in CN utilizes the diversity provided by channel code. Compared with the typical operations in CN, where simple synthesis of VN output is employed, the iteration number reduces significantly. The typical iteration number of iterative MIMO receiver is less than 10. On the other hand, we can see therere lots of girth-4 ring in PAML detectors factor graph from Fig. 1. With this issue no much improvement should be expected using typical BP algorithm with a huge iteration number.

With small iteration number, the accumulation of *a priori* information has negligible impact. So a simple procedure, described as

$$\boldsymbol{L}_{j,i}^A = \boldsymbol{L}_j^A = \mathtt{dec}(\sum_{1 \leq i' \leq N} \boldsymbol{L}_{i',j}^E) \tag{11}$$

can be used in CN. In this simple procedure, CN does not generate specific output for different VN. It combines LLR from all VN for channel decoding, and broadcasts a single version output to all VN.

3.3   Benefit to Parallel Processing

Because of per-antenna detection, a VN in PAML detector does not requires any information from other VNs. All VNs work independently so a full degree of parallelism, up to the huge number of receiving antenna N in massive MIMO, is obtained. High degree of parallelism conduces low latency processing. Similar independent processing is also obtained among CNs at per-UE level.

Introduction of channel decoding in CN converts the iteration between VN and CN into the iteration inside the channel decoder in CN. The significantly reduced iteration number between VN and CN benefits PAML detector with low latency detection and low intra-node communication load. Also a low complexity CN is obtained since the number of calling channel decoder reduces from N (once per VN) to 1 (once for the associated UE).

4   Simulation Results

In numerical simulations, two configurations are used: $N = 32$, $M = 2$ and $N = 64$, $M = 4$. Independent and identically distributed (i.i.d.) Rayleigh fading

is used, and CSI is assumed to be known in receiver perfectly. For channel coding, we use the rate-1/2 LDPC code of MacKay in [13], which is based on a random regular factor graph with $d_s = 3$, $d_c = 6$. The length of code block is 4000. The BP iteration number for LDPC decoding in CN is 50. Four types of receivers are compared: MMSE, MLD, and the proposed PAML detector with one or two iterations between VN and CN. The simulation results are shown in Figs. 2 and 3.
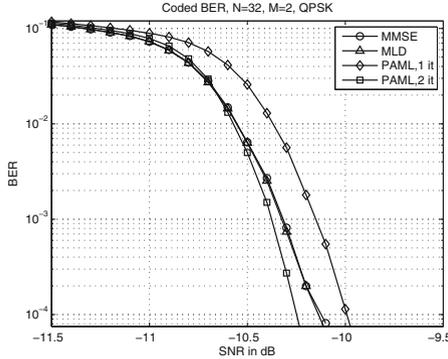
Fig. 2. Coded BER curves of massive MIMO with 32 receiving antennas and 2 concurrent uplink UEs using MMSE, MLD and PAML detector, QPSK
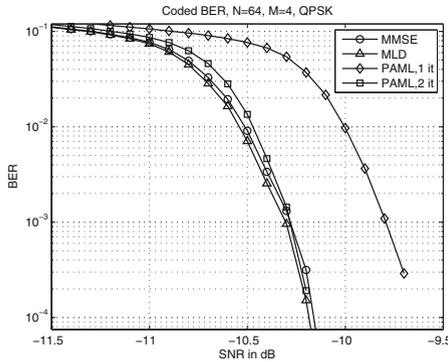
Fig. 3. Coded BER curves of massive MIMO with 64 receiving antennas and 4 concurrent uplink UEs using MMSE, MLD and PAML detector, QPSK

Figures 2 and 3 show that MMSE and MLD detector has almost same performance in massive MIMO, while PAML detector with one iteration approaches them with 0.2 0.5 dB gap. With two iterations, PAML detector slightly outperforms these two classical receivers.

5   Conclusion

In this paper, the per-antenna maximum likelihood (PAML) detector is proposed. PAML detector benefits low latency parallel processing for massive MIMO uplink receiver in terms of high degree of parallelism, low iteration number and low intra-node communication load. Simulation results show PAML detector slightly outperforms the classical MMSE and MLD detectors with two iterations. Further work will be directed towards BP-like detection for MIMO systems with short ring avoidance utilizing CSI feedback and its cooperation with channel codes.

References

1. Larsson, E.G., Tufvesson, F., Edfors, O., Marzetta, T.L.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**(2), 186–195 (2014)
2. 3GPP TR 36.873 V12.2.0 (2014-09), Study on 3D channel model for LTE
3. 3GPP TR 36.897 V13.0.0 (2015-12), Study on Elevation Beamforming/ Full-Dimension (FD) MIMO for LTE
4. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**, 3590–3600 (2010)
5. Rusek, F., Persson, D., Lau, B.K., Larsson, E.G., Marzetta, T.L., Edfors, O., Tufvesson, F.: Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Sig. Process. Mag. **30**, 40–46 (2013)
6. Lu, L., Li, G.Y., Swinlehurst, A.L., Ashikhmin, A., Zhang, R.: An overview of massive MIMO: benefits and challenges. IEEE J. Sel. Topics Sig. Process. **8**(5), 742–758 (2014)
7. Recommendation ITU-R M.2083 (2015-09), IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond
8. Dai, L., Gao, X., Su, X., Han, S., Chih-Lin, I., Wang, Z.: Low-complexity soft-output signal detection based on Gauss-Seidel method for uplink multiuser large-scale MIMO systems. IEEE Trans. Veh. Technol. **64**(99), 4839–4845 (2014)
9. Haroun, A., Nour, C.A., Arzel, M., Jego, C.: Symbolbased BP detection for MIMO systems associated with non-binary LDPC codes. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Istanbul, Turkey, pp. 212–217, April 2014
10. Yang, J., Zhang, C., Liang, X., Xu, S., You, X.: Improved symbol-based belief propagationdetection for large-scale MIMO. In: IEEE Workshop on Signal Processing Systems (SiPS), October 2015
11. Lakshmi Narasimhan, T., Chockalingam, A.: Detection in large-scale multiuser SM-MIMO systems: algorithms and performance. In: IEEE 79th Vehicular Technology Conference (VTC Spring) (2014)
12. Hochwald, B.M., ten Brink, S.: Achieving near-capacity on a multiple-antenna channel. IEEE Trans. Commun. **51**(3), 389–399 (2003)
13. MacKay, D.J.C.: Online database of low-density parity-check codes. http://wol.ra.phy.cam.uk/mackay/codes/data.html