This paper investigates various MIMO detection methods for 3GPP LTE open-loop downlink multi-antenna transmission. Targeting VLSI implementation, these detection methods are evaluated with respect to complexity and detection performance. A realistic 3GPP LTE simulation chain is developed for the evaluation. The result shows that with the aid of Hybrid Automatic Repeat reQuest (H-ARQ), a recently proposed reduced complexity close-ML detector called MFCSO achieves a good tradeoff between achievable throughput and complexity. An adaptive transmission and detection scheme is also proposed based on user scenarios.

1. INTRODUCTION

Multi-antenna or multi-input and multi-output (MIMO) technologies have been widely adopted by latest wireless standards. 3GPP Long-Term Evolution (LTE) is the 4th generation radio access technology which incorporates Orthogonal Frequency Division Multiple Access (OFDMA) as the multiple access scheme in downlink. MIMO technologies are also mandatory in LTE to achieve the LTE bit-rate targets (e.g. 100 Mbit/s peak data rate for downlink). As part of the receiver chain depicted in Fig. 1, MIMO symbol detection is a significant challenge for VLSI implementation.

2. MULTI-ANTENNA TRANSMISSION IN LTE

As defined in 3GPP LTE standard [8], the procedure to map modulated symbols to different antennas is called antenna mapping which incorporates Orthogonal Frequency Division Multiple Access (OFDMA) as the multiple access scheme in downlink. MIMO technologies are also mandatory in LTE to achieve the LTE bit-rate targets (e.g. 100 Mbit/s peak data rate for downlink). As part of the receiver chain depicted in Fig. 1, MIMO symbol detection is a significant challenge for VLSI implementation.
In case \( H \) is badly conditioned (e.g. when line-of-site occurs), linear
detection based on the pseudo-inversion of \( H \) in (6) will perform
poorly. In other words, the gain of spatial multiplexing heavily de-
pends on the multipath fading. To allow close-loop beamforming
based on codebook, a pre-coding matrix \( W \) can be multiplied with
the layer mapped symbols at the transmitter side. For downlink, \( W \)
is usually computed at the baseestation based on the codebook and
UE feedback.

2.2 Space-Frequency Block Coding

Similar to Space-Time Block Coding (STBC), Space-Frequency
Block Coding (SFBC) [8] is a technique to transmit data for guaran-
teed diversity with a low complexity symbol detector on the receiver
side. Alamouti matrix [6] based orthogonal STBC has been widely
adopted in latest wireless standards for the reason that it is the only
full-rate linear STBC code with a diversity gain of 2. In other words,
the SFBC considered in this paper is an Alamouti schemes in space
and frequency domain. This assumes the channels of neighboring
subcarriers are identical, so that when a single codeword is mapped
to several neighboring subcarriers, frequency diversity is achieved.
The basic \( 4 \times 2 \) space-frequency channel matrix is defined as

\[
H = \begin{bmatrix}
    h_{11} & -h_{12} \\
h_{12} & h_{11} \\
h_{21} & -h_{22} \\
h_{22} & h_{21}
\end{bmatrix}
\]  

(3)

3. MIMO DETECTION ALGORITHMS

For MIMO systems, a major challenge is the symbol detection at
the receiver. As channel coding (e.g. Turbo) is used, soft-output, in
effect the likelihood ratio (LLR), must be computed. Maximum
Likelihood (ML) detection which is the optimum detector computes
the receiver. As channel coding (e.g. Turbo) is used, soft-output, in
For MIMO systems, a major challenge is the symbol detection at
the transmitter. For downlink, \( W \) is usually computed at the baseestation based on the codebook and
UE feedback.

3.1 Linear Detection

Linear detection schemes such as Zero-Forcing (ZF) and Minimum
Mean-Square-Error (MMSE) have very low complexity. The only
difference between ZF and LMMSE is the later one takes the noise
power \( \sigma^2 \) into consideration while the former does not. The ZF and
LMMSE detection is defined in the following

\[
ZF : \hat{s}_{ZF} = (H^H H)^{-1} H^H r
\]

(5)

\[
MMSE : \hat{s}_{MMSE} = (H^H H + \sigma^2 I)^{-1} H^H r
\]

(6)

The equation shows that matrix inversion is involved in the detection.
The low complexity of linear detection makes them attractive for
VLSI implementation, though they have relatively poor performance
especially when the channel is slow-fading [3]. Fortunately, the
“frequency hopping” of multiple users in OFDMA creates a fast
fading channel for each individual user, which will to some extent
improve the performance of linear detection.

3.2 Fixed-Complexity Soft-Output (FCSO) Detection

As a tradeoff between performance and complexity, sphere decod-
ing such as [1] have been proposed to reach close-ML performance with
lower complexity than ML. However, the complexity of sphere
decoding grows exponentially with the number of transmit antennas
and polynomially in the size of the signal constellation. More im-
portantly, the tree search used in sphere decoding is in principle a
sequential procedure which is difficult to parallelize. In [2], a fixed-
throughput sphere detector was proposed with fixed-complexity and
parallelism for hard-decision. A method namely layered orthogo-
nal lattice detector (LORD) is presented in [4] to compute the soft-
decision. Similarly, the FCSO detector [3] which computes soft-
output, achieves close-ML detection performance via fully enumer-
ating only one transmitted symbol and applying decision feedback
equalization (DFE) to the rest of the symbols. However, the com-
plexity of both FCSO and LORD will increase substantially as the
constellation grows (e.g. from 16-QAM to 64-QAM).

3.3 Modified FCSO Detection

In [5], a reduced complexity variant of FCSO [3] for high-order
modulation schemes is proposed called MFCSO for Modified
FCSO. This section essentially repeats the algorithm description
given in [5]. The approximation in MFCSO consists of only partially
enumerating the symbols selected for exact marginalization. Taking a \( 2 \times 2 \) MIMO system as an example, considering each complex-valued symbol as one layer, only one of them is ex-
actly marginalized with the other approximately marginalized (us-
ing DFE hard-decision). The channel rate processing of MFCSO
involves the QR decomposition (QRD) of two \( 2 \times 2 \) channel matrices
which are \( H_1 = H \) in (2) and

\[
H_2 = \begin{bmatrix}
h_{12} & h_{11} \\
h_{22} & h_{21}
\end{bmatrix}
\]

(7)

The QRD generates an upper triangular matrix \( R \), and a unitary ma-
trix \( Q \) so that

\[
H_1 = Q_1 R_1 \quad H_2 = Q_2 R_2
\]

(8)

Slightly different from the MFCSO presented in [5], the detec-
tion procedure for \( 2 \times 2 \) SM is in the following

1. Linear detection in (6) or (5) is carried out to estimate the \( 2 \times 1 \)
   initial symbol vector

   \[
   \hat{s}_{init} = \min_{s \in \mathcal{X}} \| H_1 s - r \|^2
   \]

   (9)

   Here \( s \) is the transmitted symbol vector, within which, \( s_k \) is the
   \( k \)th symbol.

   2. For each initially estimated symbol \( \hat{s}_{init,k} \), \( k \in \{1, 2\} \), a candidate
   set \( \mathcal{X}_k \) is created. \( \mathcal{X}_k \) contains \( N \) lattice points close to \( \hat{s}_{init,k} \).
   In this paper, it is decided that \( N = 16 \) for 64-QAM and \( N = 9 \) for
   16-QAM.

   3. First \( s_2 \) is chosen as the top-layer symbol. In order to perform
   DFE,

   \[
   \tilde{r} = Q_1^H r
   \]

   (10)

   needs to be computed. The same operation is needed once again
   when \( s_1 \) is chosen as the top layer later.

   4. For the \( n^{th} \) constellation point \( \zeta_n \in \mathcal{X}_2 \), its effect on \( \tilde{r}_1 \) will have to
   be canceled out.

   \[
   \tilde{r}_1 = \tilde{r}_1 - R_1 (1, 2) \zeta_n
   \]

   (11)

   Based on \( \zeta_n \), the partial Euclidean distance

   \[
   \delta_n = \| R_1 (2, 2) \zeta_n - \tilde{r}_2 \|^2
   \]

   (12)

   computed for the top-layer.

5. DFE is applied to detect the other layer. Using back-substitution
[7], \( s_1 \) can be estimated from

\[
\hat{s}_1 = \arg \min_{\hat{s}_1} \| R_1 (1, 1) \hat{s}_1 - \tilde{r}_1 \|^2
\]

(13)
6. The estimated $\delta_1$ with $\delta_2 = \zeta_2$ form a complete possible transmitted symbol vector $\hat{\delta}$, on which an accumulated full Euclidean distance

$$d_\text{Euclid} = d_\text{Euclid} + \|R_2(1,1)\delta_1 - \hat{r}\|^2$$

(14)
can be computed.

7. In total, there will be $N$ different $\delta_n$ computed when $\delta_1$ is chosen as the top layer. Then $\delta_1$ is chosen as the top-layer symbol as well. Based on $Q_2, R_2$ and $\delta_{n_{init}}$, the same procedure needs to be done once again to compute another $N$ different $\delta$. Hence for the $2 \times 2$ system, $2N$ different $\delta_n$ values need to be computed. They are used to update the LLR values in the end [5].

4. 3GPP LTE SIMULATION CHAIN

In order to carry out both fast prototyping and verification of the 3GPP LTE modems, a complete physical layer behavior model and simulation chain has been developed in Matlab and C. In combination to an LTE signal generator, it allows both quantitative performance evaluation and conformance test of the chip. The simulation chain includes a transmitter conforming to 3GPP technical spec [8][9] and [10], and a receiver which supports timing/frequency synchronization, channel estimation, subcarrier demapping, rate-matching, turbo decoding and Cyclic Redundancy Check (CRC). H-ARQ based on chase combining (CC) is included with up to three times retransmission allowed. The 3GPP SCME model [11] is used as the channel model. In the simulation done for this paper, 5000 subframes are simulated. Both $2 \times 2$ SM and $2 \times 2$ SFBC are chosen as the MIMO configuration. No close-loop precoding is assumed in this paper. Throughput is calculated based on the method in [12].

<table>
<thead>
<tr>
<th>Channel Quality Indicator (CQI)</th>
<th>Modulation</th>
<th>System bandwidth</th>
<th>Num of UE</th>
<th>Num of BS</th>
<th>Channel model</th>
<th>UE speed</th>
<th>Channel estimation</th>
<th>Turbo iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>16-QAM</td>
<td>5MHz</td>
<td>1</td>
<td>1</td>
<td>Urban Micro</td>
<td>3km/h</td>
<td>Chase Combining</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>64-QAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Simulation Parameters

<table>
<thead>
<tr>
<th>CQI</th>
<th>Modulation</th>
<th>Code rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>16-QAM</td>
<td>0.602</td>
</tr>
<tr>
<td>15</td>
<td>64-QAM</td>
<td>0.926</td>
</tr>
</tbody>
</table>

Table 2: CQI parameters in simulation [10]

5. PERFORMANCE ANALYSIS

Fig. 3, 4 and 5 show that in order to support CQI=15, relatively high SNR is required, which means the UE has to be close to the BS. Meanwhile, for $2 \times 2$ SM, FCSO achieves ML performance which is 7 dB better than MFSO when reaching FER=0.01 in Fig. 3 when the weakest code is used (0.926). MFSO is around 10 dB better than MMSE to reach FER=0.01 in the same criteria. Note that in wireless systems, compared to BER or FER, throughput is a more important performance factor (if not latency) which has direct effect on the user experience. Fig. 4 shows the gain in throughput brought by MFSO against MMSE is significant (up to 12Mbits/s, or 55% higher than the one achieved by MMSE). In comparison, the throughput gain brought by FCSD against MFSO is much smaller (up to 2.5Mbits/s, or 7% higher than that achieved by MFCSD). The much smaller gap in throughput in comparison to that of FER mainly owes to the H-ARQ retransmission with chase combining.

The result shows us that even with a sub-optimal detector (which also implies much lower complexity), a throughput that is close to the one achievable by ML detectors can be reached when H-ARQ is presented.

Simulation result of CQI=9 are depicted in Fig. 6, 7 and 8. The result shows that 16-QAM only requires moderate SNR which will be available in most part of the cell range. It also shows that MFSO ($N = 9$) achieve the same performance as FCSO and ML detectors. It has a throughput that up to 68% higher than the one achieved by MMSE.

![Figure 3: Frame-Error-Ratio (2 × 2 SM, CQI=15)](image)

![Figure 4: BLock-Error-Ratio (2 × 2 SM, CQI=15), red curves are the BLER of the 1st retransmission of H-ARQ)](image)

![Figure 5: Minimum SNR to reach FER=0.01)](image)

![Figure 6: Throughput gain brought by FCSO against MMSE is much smaller (up to 12Mbits/s, or 55% higher than the one achieved by MMSE).](image)

Fig. 11 depicts the achievable throughput using two-level adaptive modulation and coding (AMC). The result shows that when SNR is worse than 10 dB, SFBC achieves both higher throughput and lower BLER than SM even if ML detector is used.
6. IMPLEMENTATION CONSIDERATIONS

In LTE [8], taking a 5 MHz bandwidth LTE system as an example, up to 7 OFDM symbols need to be processed within one slot (0.5 ms) which contain 1900 data subcarriers. This means that there will be no more than 0.26 μs to finish the detection of each subcarrier in average. Therefore, proper detection methods have to be chosen in order to maximize the data rate at reasonable implementation cost.

As depicted in Eq. (6), for 2 × 2 SM, the MMSE detector needs to compute the inverse of a 2 × 2 matrix. It has been presented in [13] that the inversion of small matrices can be done using direct inversion which supplies sufficient precision for most of the channels. The FCSO and MFCSO detector involves the search of a number of trellis nodes as depicted in Tab. 4. The FCSO detector always visits the complete constellation (e.g. 16 for 16-QAM and 64 for 64-QAM) while MFCSO only visits a subset of it (e.g. 9 for 16-QAM and 16 for 64-QAM). Note that MFCSO requires MMSE detection to compute the initial estimate (9) which is an extra cost.
In practice, the hardware is usually implemented taking both the cost and performance issues into consideration. Based on the complexity analysis in Tab. 4 and the performance analysis in Sec. 5, MFCFSO falls into the favor of the authors to be chosen as the target algorithm for ASIC implementation. Using ST 65nm CMOS process, while meeting the 0.26μm constraint, the implemented detector supporting both MMSE and MFCFSO for 2×2 SM and up to 64-QAM modulation occupies less than 0.25 mm².

7. ADAPTIVE TRANSMISSION AND DETECTION

As depicted in Tab. 4, a detector supporting MFCFSO/MMSE consumes 2.5 times the area of the one only supporting MMSE. Hence the former one is assumed to target high-end users willing to pay more in area and power for performance (e.g. laptops). The MMSE single-mode detector is in favor of low-end users for connectivity with minimum cost (e.g. smartphones). Note that the user cares about latency as well as throughput, and latency is partly determined by the number of retransmissions. Hence it is also important to keep the retransmissions to a minimum (which requires low FER).

Fig. 11 shows that with AMC, SM using MFCFSO detector always brings higher throughput when SNR is greater than 10 dB. For both types of users, when SNR is worse than 10 dB (a in Fig. 11), SFBC is preferred instead of SM. For low-end users, SFBC is still preferable (due to the low FER) since the algorithm has sufficiently low complexity [5], it is chosen over FCFSO [3] and other close-ML detection schemes for VLSI implementation. Based on the adaptive scheme proposed in Sec. 7, a good performance and cost tradeoff can be achieved. The result also emphasizes the need of a configurable detector to enable the adaptive scheme in real-time.

9. ACKNOWLEDGEMENT

The work of D. Wu, J. Eilert and D. Liu was supported in part by the European Commission through the EU-FP7 Multi-base project with Ericsson AB, Infineon Austria AG, IMEC, Lund University and KU-Leuven. The authors would like to thank Christian Mehlführer and the Christian Doppler Laboratory for Design Methodology of Signal Processing Algorithms at Vienna University of Technology, for contributions on the LTE simulation chain.

REFERENCES