A Low Complexity and High Throughput MIMO Detection VLSI Design for MIMO-OFDM Systems

Zhaohui Cai, Yu Hong Wang, Suttinan Chattong
Institute for Infocomm Research, 1 Fusionopolis Way, Singapore 138632

Abstract—This paper presents a linear Minimum Mean Square Error (MMSE) MIMO Detector design for MIMO-OFDM systems based on Application-Specific Instrument-set Processor (ASIP). As part of the IEEE 802.11ac-compliant PHY baseband transceiver, the proposed MIMO detector offers low latency, high throughput with efficient resource utilization. The design has been synthesized with TSMC 40 nm CMOS technology, the logic gate count for each QRD engine is about 245 K gates. It is able to support 20/40/80 MHz bandwidth and up to 4 spatial streams. Detection latency for 80 MHz VHT mode (234 data sub-carriers) is 750 ns.

I. INTRODUCTION

With the techniques such as advanced channel coding, multiple-input multiple output (MIMO) and orthogonal frequency-division multiplexing (OFDM), wireless local area networking (LAN) has experienced tremendous growth in the last ten years with the proliferation of IEEE 802.11 devices. MIMO systems employ multiple antennas at both sides of the wireless link and are able to improve the data throughput significantly by transmitting multiple data streams concurrently over the same frequency band. On the other hand, OFDM is well suited to wideband systems in frequency selective fading environments. OFDM is tolerant of time synchronization errors and drastically simplifies equalization design at the receiver. Now MIMO in combination with OFDM constitutes the basis for many wireless standards [1-2].

For MIMO-OFDM systems, the pre-processing circuit to obtain the filtering matrix from the channel estimation is a crucial component of the MIMO detector and an interesting research topic. Pre-processing algorithms perform all operations necessary to compute an estimate of the transmitted symbol, which do not depend on the data symbols [3].

Many MIMO detection algorithms have been reported in the literature, such as, linear MMSE detection [3-4], SIC detectors and tree-search based detection [5-6]. As the time available for PHY layer turn-around is short due to stringent latency constraints imposed by the IEEE 802.11n/ac standard: the combined latency has to be less than 16 µs short inter-frame space (SIFS) constraint, most 11n transceivers apply linear MMSE detection [3] due to its simplicity compared to SIC and sphere decoding.

There are two main architectures have been proposed to implement preprocessing circuits reported in the literature, namely, processor-like (ASIP-based) architectures [7-9] and pipelined architectures [3-4].

While the pipelined solutions divide the algorithm to be implemented into multiple-tasks and assign the individual tasks to dedicated modules that are separated by pipeline stages, an ASIP-based design computes all tasks required on the same piece of hardware with one or more customized processor cores associated with dedicated instruction ROMs. All the data processing is handled by some dedicated execution units.

Compared with the pipelined architectures, one advantage of the ASIP-based architecture is its ability to adapt to algorithm/interface changes by reprogramming the instruction ROM. For example, the ASIP engine proposed for MIMO detection in this paper can be reused for other tasks such as V compression required for the steering matrix computation in the beamforming (BF) mode [2] by adding some hardware resources, as there is no MIMO detection activities for the null date packet (NDP) processing.

In this paper, as part of our IEEE 802.11ac-compliant PHY baseband transceiver design, an ASIP based VLSI architecture is implemented as the QR decomposition (QRD) engine as part of the MIMO detector. Other tasks such as channel estimation, MMSE matrix generation, bias generation and data filtering are implemented with pipelined modules (This is different from our PHY baseband transceiver designed for IEEE 802.11n, where the channel estimation and the symbol detection are also part of the ASIP engine’s task, as the computation load for QRD in 11n is just roughly the half of that in 11ac).

In order to meet the 11ac latency constraint, two dedicated QRD engines are instantiated, one for odd-indexed subcarriers and one for even-indexed subcarriers, respectively. Our design offers low latency, high throughput MIMO detection with efficient resource utilization. With CMOS 40 nm processing technology, the equivalent gate-count (GE) for each QRD engine is about 245 KGE. The latency of 80 MHz VHT mode (234 data sub-carriers) is 750 ns.

The remainder of the paper is organized as follows. Section II introduces the IEEE 802.11ac system and presents the packet error rate (PER) performance result of an 11ac-compliant PHY that we have developed. The linear MMSE MIMO detection algorithm used in our design is also presented in this section. Section III discusses the VLSI architecture design for the MIMO detection in detail. The implementation result is compared with work of others in this section as well. The summary is given in section IV.

II. THE IEEE 802.11AC SYSTEM AND LINEAR MMSE MIMO DETECTION

A. The IEEE 802.11ac MIMO-OFDM System
IEEE 802.11ac, like its predecessor IEEE 802.11n, offers considerable advantages. 802.11ac (aka VHT, Very High Throughput) devices are required to support 20, 40, and 80 MHz channels and 1 spatial stream. Several optional features are also defined in 802.11ac, such as, 160/80 + 80 MHz bandwidth, 256 QAM, up to 8 spatial streams, Multi-User (MU) MIMO and low density parity check (LDPC) codes support. An 11ac device making use of only the mandatory parameters (80 MHz bandwidth, 1 spatial stream, and 64 QAM 5/6) will be capable of a data rate of ~293 Mbps while a device that implements all optional parameters (8 spatial streams and 256 QAM with rate 5/6 with a short guard interval) will be able to achieve almost 3.5 Gbps [2].

We have designed an IEEE802.11ac-compliant MIMO-OFDM PHY baseband transceiver with some optional features support, such as, up to 4 spatial streams support, 256 QAM, based on a systemC bit-trace platform. The baseband transceiver has been synthesized with TSMC CMOS 40 nm technology. Our design supports NON-HT, HT-MF, and VHT modes. The PHY baseband transceiver block diagram is sketched in Figure 1.

![Baseband Transceiver Block Diagram](image)

**Figure 1 IEEE 802.11ac PHY baseband transceiver block diagram**

With $N_{bs} = 4$, MCS = 9 (VHT 80 MHz mode), and 800 ns long guard interval (GI), our design is able to offer a maximum data rate of 1.56 Gbps. Using our SystemC platform, simulated PER performance for 802.11ac 4x4 VHT mode with convolutional coding and different number of spatial streams ($N_a$) is shown in Figure 2. In the simulation, the frequency offset of -13.75 ppm. The propagation channel is 80 MHz TGn channel B, and packet length is 1 Kbytes. The blue-solid curves and the red-dotted curves denote floating-point and fixed-point results, respectively.

One of the major parts of the PHY baseband transceiver is the MIMO detection, which usually contributes to a large percentage of the silicon area. The MIMO detection consists of data filtering and pre-processing tasks, i.e., channel estimate, post-equalization signal to interference and noise ratio (SINR) and $G_{MMSE}$ computation. Our MIMO detector is designed to work with up to 4 transmitter antennas and 4 receiver antennas, and to support up to 4 spatial streams.

B. The Linear MMSE MIMO Detection Algorithms

1) Channel Estimation

Channel estimation is performed using the L-LTF field and VHT-LTF (HT-LTF) field in frequency domain (after receiver FFT) for detection of legacy filed (such as L-SIG and VHT-SIGA) and VHT field, respectively. In the following, we describe the channel estimation algorithm using VHT-LTF filed. Suppose $N_{ss}$ transmit spatial stream and $N_{rx}$ receive antennas are used. Let $y_{l,k}^{(i)}$ denote the received signal of $k$-th subcarrier of $i$-th received antenna during $l$-th ($l=1,2,\ldots,N_{VHT}$) VHT-LTF symbol, and $d_{l,k}^{(j)}$ denote transmitted frequency domain signal from $j$-th ($j=1,2,\ldots,N_{s}$) spatial stream during $l$-th VHT-LTF symbol. Then received vector $y_{l,k} = [y_{l,k}^{(1)},\ldots,y_{l,k}^{(N_{rx})}]^T$ during $l$-th VHT-LTF ($l=1,2,\ldots,N_{VHT}$) symbol for $k$-th subcarrier can be written as

$$y_{l,k} = H_{k}d_{l,k} + n_{l,k},$$

where $H_{k}$ denotes the $N_{rx} \times N_{ss}$ channel matrix with the $(p,q)^{th}$ element being the random fading between $p^{th}$ receive antenna and $q^{th}$ transmit spatial stream. $d_{l,k} = [d_{l,k}^{(1)},\ldots,d_{l,k}^{(N_{ss})}]^T$ is the $N_{ss}$ x 1 transmitted data vector during $l$-th VHT-LTF symbol. $n_{l,k}$ is the $N_{rx}$ x 1 complex Gaussian noise vector with covariance matrix $N_{0}I$. Let $Y_{k} = [y_{l,k},\ldots,y_{N_{VHT},k}]$, and $D_{k} = [d_{l,k},\ldots,d_{N_{VHT},k}]$, then Least Square (LS) estimation of $H_{k}$ can be written as,

$$\hat{H}_{k} = (D_{k}H_{k}D_{k})^{-1}D_{k}^{H}Y_{k},$$

where $(\cdot)^{H}$ is conjugate-transpose operation. $N_{VHT}$ is number of VHT-LTF symbols in VHT-LTF field.

2) Noise variance estimation

To perform the MMSE detection, we need to know noise variance $N_{0}$. Noise variance estimation algorithm makes use of 2 identical L-LTFs present in the preambles of an IEEE 802.11a/n/ac packet. Let $l_{L-LTF}(k)$ and $r_{L-LTF}(k+N_{FFT}) (k=0,1,2,\ldots,N_{FFT})$ represent the time domain samples (before
receiver FFT) at the \(i\)-th receiver antenna of first received L-LTF symbol and second L-LTF, respectively, and \(N_{\text{FFT}}\) is number of samples in one L-LTF symbol. Then an estimate of the SNR can be written as

\[
SNR = \frac{\sum_{i=1}^{N_{\text{FFT}}} \left| H_{i}\right|^2}{2\sum_{i=1}^{N_{\text{FFT}}} \left| I_{i}\right|^2} \left| H_{i}-L_{\text{FFT}}(k+i)\right|^2,
\]

(3)

and the estimate of \(N_0\) can be written as

\[
N_0 = \frac{1}{SNR}.
\]

(4)

3) Linear MMSE MIMO Detection

After the channel estimates \(H_k\) is obtained, the linear MMSE MIMO detector can be calculated. Without loss of generality, in the following section, we use \(H\) to denote the channel estimation of subcarrier \(k\). Then MMSE matrix \(G_{\text{MMSE}}\) can be written as,

\[
G_{\text{MMSE}} = (H^H + N_{\text{SS}}N_0^H)^{-1}H^H.
\]

(5)

The estimate of the transmitted symbol vector \(y\) is

\[
y = G_{\text{MMSE}}y.
\]

(6)

The per-stream post-equalization SINR is given by

\[
\text{SINR} = \frac{\text{bias}}{1-\text{bias}},
\]

(7)

where \(\text{bias} = \text{diag}\left(\text{abs}(\text{real}(G_{\text{MMSE}} \times H))\right)\).

Most MMSE detector implementations reported in the literature apply an explicit matrix inversion of the matrix \(A = H^H + N_{\text{SS}}N_0^H\). [10] employs direct matrix inversion with Strassen’s method by exploring the Hermitian property of the matrix \(A\), while [12-13] applies QRD on \(A\) as \(A = QR\), and subsequently, the matrix inversion problem can be solved as \(A^{-1} = R^{-1}Q^H\).

Another approach is to work on the augmented channel matrix defined as

\[
H_{\text{aug}} = \left[\frac{H}{\sqrt{N_{\text{SS}}N_0^H}}\right],
\]

(8)

and work on the QRD of \(H_{\text{aug}}\) as

\[
H_{\text{aug}} = QR = [Q_1 | Q_2] R,
\]

(9)

so that

\[
G_{\text{MMSE}} = \frac{Q_2 R^H}{\sqrt{N_{\text{SS}}N_0^H}}.
\]

(10)

This approach eliminates the need for matrix inversion. It also significantly reduces the precision requirements of the system [4].

Same as in [3] and [4], the solution based on QR Decomposition of \(H_{\text{aug}}\) is adopted in our MIMO detector design and the Modified Gram-Schmidt (MGS) algorithm is applied to decompose the \(H_{\text{aug}}\). The MGS QR decomposition has an advantage in numerical stability when compared to the original Gram-Schmidt algorithm. The MGS algorithm is summarized in Algorithm 1.

**Algorithm 1: Algorithm for QR decomposition**

\[
\begin{align*}
&[Q, R] = \text{MGS}(H_{\text{aug}}) \\
&m, n = \text{size}(H_{\text{aug}}) \\
&Q = [q_1, q_2, \ldots, q_m] \\
&R = 0_n \\
&\text{for } i = 1:n \\
&\quad r_{ii} = \sqrt{\text{real}(q_{i+1,i}^T \times q_{i+1,i})} \\
&\quad \text{inv}_{r_{ii}} = 1/r_{ii} \\
&\quad q_{i+1,j} = q_{i+1,j} \times \text{inv}_{r_{ii}} \\
&\quad \text{for } j = i+1:n \\
&\quad \quad r_{ij} = q_{i,j}^T q_{j} \\
&\quad \quad q_{j} = q_{j} - r_{ij} q_{i} \\
&\quad \text{end} \\
&\text{end} \\
\end{align*}
\]

**III. PROPOSED MIMO DETECTOR VLSI ARCHITECTURE**

**A. The Latency Requirement**

In 802.11n/ac the MIMO channel estimate is carried out with the HT/VHT long training filed (LTF) symbols. The LTF symbols are followed immediately by OFDM data symbols, as show in the Figure 3.

This leaves very limited time for the pre-processing of the MIMO detection, i.e., channel estimation and MMSE weight computation.

**Figure 3 Target MIMO detection timing**

Ideally as long as the particular subcarrier MMSE matrix is ready once the corresponding data sample for this subcarrier is available from the phase compensation module, the latency of pre-processing (channel estimation and \(G_{\text{MMSE}}\) computation) added to the receiver latency is zero, i.e., there is no memory needed to buffer the data symbols; the MIMO detection latency just comes from the symbol detection.

For long GI systems with 800 ns cyclic prefix, this requires that the combined pre-processing should be less than 4 \(\mu\)s for the 4th data subcarrier, and should be less than 8 \(\mu\)s for the last data subcarrier.

Our design is very close to this target: the MIMO detection latency for 3rd OFDM data symbol onwards is just 750 ns (60 data samples for 80 MHz VHT mode).

**B. The Top level Architecture Overview**

The top level structure for our design is shown in Figure 4, where, the yellow color blocks are implemented using pipelined architecture, and light-blue color blocks are implemented using ASIP architecture.

It can be seen from the Algorithm 1, the MGS is an iterative process with computational intensive blocks such as complex vector dot product, square root and division. In case of 4x4 MIMO detector, the complex vector size is 8x1, and 8x1 complex vector dot product need 32 real multipliers. ASIP
The MIMO detector receives (i) vector \( y \) from phase compensation module (ii) estimated noise variance (from noise variance estimator module) (iii) \( N_{SS} \) and CBW (bandwidth) as its input. Based channel estimation, an 8x4 augmented channel matrix \( H_{aug} \) is build according to eqn (8).

\( H_{aug} \) is pushed to the QRD processors. To improve the throughput, two QRD engines to decompose \( H_{aug} \) and two MMSE weight matrix generation, two bias generation and two data filtering are instantiated, one for odd-indexed subcarriers and one for even-indexed subcarriers.

Once QRD is completed the MMSE matrix generator block reads Q matrices through memory interface from QRD processors and performs per-subcarrier MMSE matrix generation according to eqn (10) and filter the received signal vector \( y \) to generate \( \hat{y} \) according to eqn (6). In addition to this, it also performs per-subcarrier SINR calculation as in eqn (7).

Wherever possible, we try to overlap the processing of all these blocks to reduce latency, resulting in the scheduling diagram as shown in the figure below (using \( N_{SS}=4 \) as example).

This QRD engine comprises of:

1. Instruction ROM – this contains custom program (sequence of instructions) which implements the function of the QRD.
2. Opcode Modification based on \( N_{sa} \) and CBW - since our MIMO detector need to support different bandwidth and different \( N_{sa} \), the number of the combination of bandwidth and spatial stream is 12, and if we design a different instruction sequence for each of this combination, the instruction ROM would be very huge. This module is to online generate the modified instruction sequence based on number of spatial stream, bandwidth and basic instruction sequence.
3. Instruction Decoder – this module decodes each modified instruction into control signals required by the Arithmetic Unit.
4. Data Memory and read/write address generate –data memory stores \( inv \cdot f_{10} \), \( q_{r} \), and \( r_{ij} \) in memory (refer to Algorithm 1). The data is read by the Arithmetic Unit which performs processing, and then written back to the data memory.
5. Pipeline Balancer – this is a shift register which acts to balance data memory access delay with the decoded control signals, thus ensuring that these signals arrive at the Arithmetic Unit at the same time.
6. Arithmetic Unit – this block performs the mathematical computations as dictated by the opcodes within the instruction ROM. Refer to Algorithm 1, the mathematical computations includes vector dot product, square root, div and adder and subtractor.

The program flow of the system is linear without loop control. After reset is de-asserted, the QRD module starts in idle state. It starts operation in response to a trigger pulse from the MIMO channel estimation module. The program flow is linear, meaning the address output of program counter pulse when the last instruction is reached and will only be reset again when the module is disabled or reset.
The QRD module produces a start trigger pulse to the MMSE matrix generator module, telling it when to start MMSE matrix calculation. Both MMSE matrix and SINR values are calculated only once per frame. The data detection (filtering) process is done every symbol for all sub-carriers.

D. The Implementation Result

The MIMO detection design, together with other blocks of an 11ac PHY baseband transceiver, has been synthesized with Synopsys® Design Compiler using 40nm CMOS library, with a clock frequency of 320 MHz (This clock frequency constrain is imposed by the fact that free slow SRAM is used in our PHY baseband transceiver design). The logic gate count breakdown for the proposed MIMO detector is listed in Table 1 (excluding the slow SRAM). Comparison of our design with other MIMO detection architectures is presented in Table 2.

Table 1 Logic Gate count breakdown for the proposed MIMO detector

<table>
<thead>
<tr>
<th>Module</th>
<th>Logic Gate (kGE)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel Estimator</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>QRD processor 0</td>
<td>245</td>
<td>42</td>
</tr>
<tr>
<td>QRD processor 1</td>
<td>245</td>
<td>42</td>
</tr>
<tr>
<td>$W_{MSSS}$ and filtering</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>580</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2 Comparison of the MIMO detection implementations

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MGS</th>
<th>MGS</th>
<th>Givens</th>
<th>Givens</th>
<th>MCMC</th>
<th>MGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>Pipelined</td>
<td>Pipelined</td>
<td>Givens</td>
<td>Givens</td>
<td>Pipelined</td>
<td>Pipelined</td>
</tr>
<tr>
<td>Dimension</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
<td>4x4</td>
</tr>
<tr>
<td>Modulation</td>
<td>64QAM</td>
<td>64QAM</td>
<td>64QAM</td>
<td>64QAM</td>
<td>16QAM</td>
<td>64QAM</td>
</tr>
<tr>
<td>Clock Freq</td>
<td>360 MHz</td>
<td>140 MHz</td>
<td>250 MHz</td>
<td>100 MHz</td>
<td>272 MHz</td>
<td>400 MHz</td>
</tr>
<tr>
<td>Gate count</td>
<td>473 K</td>
<td>-</td>
<td>66 K</td>
<td>152 K</td>
<td>0.41mm</td>
<td>482 K</td>
</tr>
<tr>
<td>Technology</td>
<td>90 nm</td>
<td>FPGA</td>
<td>130 nm</td>
<td>180 nm</td>
<td>180 nm</td>
<td>65 nm</td>
</tr>
<tr>
<td>Spec.</td>
<td>11n</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Latency</td>
<td>0.95 µs</td>
<td>2.77 µs</td>
<td>0.032 µs</td>
<td>0.041 µs</td>
<td>1.41 µs</td>
<td>0.75 µs</td>
</tr>
<tr>
<td>Throughput</td>
<td>20 M</td>
<td>14.4 M</td>
<td>1000/32n</td>
<td>1000/40n</td>
<td>1/1.41n</td>
<td>-</td>
</tr>
</tbody>
</table>

The last row in the Table 2 is for the QRD processing throughput in Million Matrices per second, which is different from the one listed in [9]. Our design is able to process 234 sub-carriers within 8 µs time, i.e., the throughput > $234/8 = 29.25$ M matrices/s. Noted that the entries marked with ‘N.A.’ in ‘Spec.’ indicate the values (Gate count, latency and throughput) are for QRD only. Others are for entire MIMO detection and follow the corresponding specifications. To authors’ knowledge, our design is the first complete 11ac MIMO detection solution presented in public.

IV. SUMMARY

In this paper we have presented a VLSI architecture for MIMO detector which adopts hybrid pipelined and ASIP architecture. Our MIMO detector can support 12 different bandwidth and $N_s$ combinations. Instead of designing a different instruction sequence for each of this combination, we use a novel method which only uses one basic instruction ROM and another Opcode modification module to online modify instruction based on $N_s$ and bandwidth. And total ASIC gates count is 580 kGE. The latency of 80 MHz VHT mode (234 data sub-carriers) is 750 ns.

By adding Coordinate Rotation Digital Computer (CORDIC) engine into the ASIP Arithmetic Unit, our ASIP is able to support the compression of the V matrix in the beamforming mode of the 11ac, as the V-compression essentially consists of a series of Givens rotations of the orthogonal matrix V derived from the singular value decomposition (SVD) of the channel H.

This MIMO detector is part of an IEEE 802.11ac-compliant PHY baseband transceiver. Our PHY baseband transceiver design can support up to $N_s = 4$ and 256 QAM, with a maximum data rate of 1.56 G bps.

REFERENCES