On Modified Squared Givens Rotations for Sphere Decoder Preprocessing

Yun Wu, John McAllister

***Abstract*—Sphere Decoding for Multiple-Input Multiple-Output (MIMO) wireless systems is a complex operation, usually demanding custom accelerators in order to support real-time performance. The cost of these accelerators is disproportionately influenced by channel matrix preprocessing, which represents a relatively small fraction of the overall computational cost of detecting an OFDM MIMO frame in standards such as 802.11n, but consumes a very large amount of hardware resource. Modified Squared Givens' Rotations has been proposed to resolve this issue and shown to dramatically reduce accelerator cost. However, there is no analysis on the record of the complexity of this algorithm, nor its detection performance. This paper shows that, despite offering modest reductions in operational complexity, MFSD-SQRD enables dramatic cost reductions by explicitly addressing the overhead of matrix permutation steps. Further, it shows that for most SNR values of practical interest, the performance of MFSD-SQRD is not appreciably diminished relative to the standard SQRD approach to preprocessing. To the best of the authors' knowledge, the proposed modified SQRD preprocessing approach is the highest performance sub-optimal preprocessing approach on record.**

I. Introduction

Multiple-Input Multiple-Output (MIMO) communications topologies have been combined with Orthogonal Frequency-Division Multiplexing (OFDM) to great effect in modern wireless communication standards such as LTE-Advanced and IEEE 802.11 [2], [1]. However, achieving high data rates in such systems relies on highly accurate detection of received symbols via approaches such as Sphere Decoding (SD). All such SD approaches require a-priori manipulation of the Channel State Information (CSI) to order the received symbols for detection based on the estimated distortion experienced by each, in an operation known as preprocessing [7], [6].

In the most computationally efficient quasi-optimal SD - the Fixed-Complexity Sphere Decoder (FSD) [3] - preprocessing usually involves a process known as Sorted QRD (SQRD) [4]. This involves equalisation using either Zero-Forcing (ZF) or Minimum Mean Square Error (MMSE) techniques, followed by an iterative QR Decomposition (QRD) and matrix permutation procedure. The latter in particular makes preprocessing computationally demanding in absolute terms. Despite this, when computational complexity is accounted for an entire OFDM frame, FSD-SQRD is relatively inexpensive, constituting only a small fraction of the total computational cost of the FSD detector [8].

The only recorded real-time FSD SQRD for 4×4 802.11n MIMO on record is a Field Programmable Gate Array (FPGA) architecture presented in [4]; in this work it is noticeable that, despite the low complexity of the preprocessing step relative to the remainder of the FSD algorithm, the resource cost of the preprocessing components dominate the overall system resource cost; this resource inefficiency is expensive and will become increasingly so - the increase in the number of antennas in MIMO standards such as Long Term Evolution (LTE) and LTE-Advanced (LTE-A), coupled with the associated fourth-order increase in FSD-SQRD complexity means that preprocessing algorithms which are more amenable to low cost implementation are urgently required.

In [5] an alternative preprocessing approach, known as modified FSD-SQRD, was presented in an attempt to address the implementation cost issues associated with FSD preprocessing. Specifically, it was shown how FPGA-based for 802.11n could enable real-time throughput for 4×4 802.11n whilst incurring only 50% of the resource cost of the most direct comparable solution. However, there has been no report of either the complexity of this algorithm, nor any analysis of its detection performance.

This paper resolves those issues. It presents the first complexity and detection performance analysis of the MFSD-SQRD algorithm. The remainder of this paper is as follows. Section III describes the modified FSD-SQRD algorithm, before Sections IV and V respectively analyse the complexity and performance of the algorithm.

II. Background

MIMO communication systems adopt multiple antennas at both transmit and receive terminals, as shown in Fig. 1. Here, an M-element antenna array transmits a signal $\mathbf{s} \in \mathbb{C}^{M\times 1}$ via a multi-path fading channel $\mathbf{H} \in \mathbb{C}^{N\times M}$, resulting in a received symbol vector $\mathbf{r} \in \mathbb{C}^{N\times 1}$ at an N-element receiver. Assuming Additive White Gaussian Noise (AWGN), $\mathbf{w} \in \mathbb{C}^{N\times 1}$ at the receiver, $\mathbf{r}$ can be modelled as

$$\mathbf{r} = \mathbf{H} \cdot \mathbf{s} + \mathbf{w}. \tag{1}$$

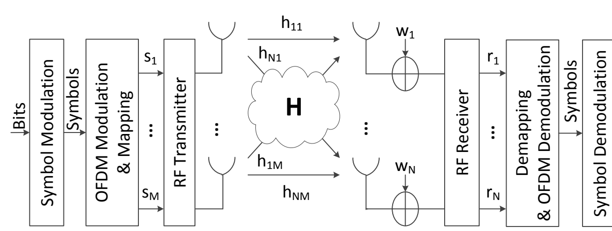


Fig. 1: MIMO OFDM System Model

A. FSD Preprocessing

Detection in MIMO systems involves deriving an estimate of the transmitted symbol vector $\mathbf{s}$ given $\mathbf{r}$. SD is one approach to performing this operation via two key steps:

1) *Preprocessing*: The symbols of $\mathbf{r}$ are ordered for detection and the centre of the decoding sphere is initialised.
2) *Metric Calculation & Sorting (MCS)*: An M-level decode tree forms an estimate of $\mathbf{s}$ based on cumulative Euclidean distance analysis.

The lowest complexity quasi-optimal SD approach, the Fixed-Complexity Sphere Decoder (FSD), exploits a process known as Sorted QRD for preprocessing. In SQRD, the received symbols are reordered based on the perceived distortion experienced by each. This process requires that $\hat{\mathbf{H}}$, an equalised version of $\mathbf{H}$ be derived according to the equalisation strategy employed:

$$\hat{\mathbf{H}} = \begin{cases} \mathbf{H} & \text{for ZF} \\ (\mathbf{H}; \sigma_w \cdot \mathbf{I}) & \text{for MMSE} \end{cases} \tag{2}$$

where σ_w is the normalized standard variance of white gaussian noise and $\mathbf{I}$ is a $M \times M$ identity matrix.

Decomposition of $\hat{\mathbf{H}}$ via QRD is then performed to yield

$$\hat{\mathbf{H}} = \mathbf{Q} \cdot \mathbf{R}, \tag{3}$$

where $\mathbf{R} \in \mathbb{C}^{\mathbf{M} \times \mathbf{N}}$ and $\mathbf{Q} \in \mathbb{C}^{\mathbf{M} \times \mathbf{M}}$ (ZF) or $\mathbf{Q} \in \mathbb{C}^{(\mathbf{2} \cdot \mathbf{M}) \times \mathbf{M}}$ (MMSE). SQRD-based derivation of $\mathbf{Q}$ and $\mathbf{R}$ commences by initialising $\mathbf{Q} = \hat{\mathbf{H}}$ and calculating the norm of each column:

$$norm_k = \|q_i\| \,. \tag{4}$$

Subsequently, an iterative sequence of permuations (5), orthogonalisation (6) and triangularisation (8) operations are then applied to order the elements of $\mathbf{r}$ according to the distortion experienced by each. In each iteration the i^{th} column of $\mathbf{Q}$, $\mathbf{norm}$ and $\mathbf{R}$ are permuted with the k^{th} according to

$$k = \begin{cases} \arg \max\limits_{l\, =\, i, \cdots, M} \mathbf{norm_l}, & i \leq nfs \\ \arg \min\limits_{l\, =\, i, \cdots, M} \mathbf{norm_l}, & i > nfs \end{cases} \tag{5}$$

where $nfs = \left\lceil \sqrt{M} - 1 \right\rceil$,. Subsequent to each permutation, $\mathbf{Q}$ is orthogonalised and R undergoes upper triangularisation according to (6) and (8) respectively.

$$q_i = \frac{q_i}{r_{i,i}}, i \in [1, M]\,, \tag{6}$$

$$q_k = q_k - q_i \cdot r_{i,k}, k \in [i+1, M]\,, \tag{7}$$

$$r_{i,k} = \begin{cases} norm_k, & \text{if } i = k \\ \langle q_i, q_k \rangle\,, & \text{if } i < k \\ 0, & \text{if } i > k \end{cases} \tag{8}$$

The column norm of $\hat{\mathbf{H}}$ is updated during the decomposition as

$$norm_k = norm_k - \|r_{i,k}\|^2 \,. \tag{9}$$

The iterative ordering-update nature of SQRD makes it computationally demanding - the M iterations, each of which contains an $O\left(M^3\right)$ QRD operation, combine to result in $O\left(M^4\right)$ overall complexity.

This preprocessing step presents a significant challenge when attempting to realise SD-based detectors as custom accelerators on, for instance, FPGA. The work in [8] shows that it incurs a disproportionately high hardware cost, relative to its contribution to the overall complexity of decoding a symbol frame. Section III describes an alternative preprocessing approach which aims to resolve this issue.

III. Modified SGR for FSD PP

Using SQRD for FSD pre-processing avoids M iterations of an algorithm where QRD is a component. Instead, it employs a single iteration merging the ordering, orthogonalization and triangularization process. This has the effect of reducing computational complexity by an order of magnitude to $O\left(M^3\right)$.

However, it also has the effect of tightly coupling the permutation, orthogonalization and triangularization operations; when mapped to separate processors in the multi-core architecture presented in [4], a high inter-processor traffic load results; specifically a two way exchange of $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{norm}$ between pairs of processors is required M times per SQRD operation. This two-way communication has two consequences: it's bidirectional nature demands two FIFO queues, one per direction, between each pair of processors, whilst the communication process demands cycles, which would otherwise be dedicated to performing the algorithm computations.

By removing the integrated ordering before each iteration of the vector orthogonalization, a separated detection order rearrangement is performed based on the norm of each column before QRD. The pseudo-code of this Modified FSD-SQRD (MFSD-SQRD) is shown in Algorithm 1.

Adopting the same inputs, $\hat{\mathbf{H}}$ and M, $\mathbf{R}$ and $\mathbf{Q}$ are initialised as zero matrices and $\mathbf{O}$ as a zero vector. From this point, the modified preprocessing for FSD is specified as follows:

1) The $\mathbf{norm}$ is derived from the diagonal elements of $\hat{\mathbf{H}}^{\mathbf{H}} \cdot \hat{\mathbf{H}}$, in step 3), before the iterative SQRD process.
2) At the i^{th} iteration of a separated ordering process, the $(nfs+1)^{st}$ minimum element of $\mathbf{norm}$ is selected for $i \in [1, M - nfs]$ while the $(M-i+1)^{st}$ lowest element of $\mathbf{norm}$ is chosen for $i \in [M-i+1, M]$ in step 6) - 7).
3) The index of the selected element in $\mathbf{norm}$, j, is recorded in $\mathbf{O}$ and j^{th}. The i^{th} element is exchanged with j^{th} element in $\mathbf{norm}$ while the j^{th} column is recorded as i^{th} column of $\mathbf{Q}$ instead of swapping.
4) The processes 2) and 3) are performed iteratively until index i reaches M.
5) The determined detection order, $\mathbf{Q}$ is used for a iterative GM orthogonalization from step 10) to 17).

Note specifically the different output data initialization and the separation of the ordering process (lines 5 - 9) from the QRD (lines 10 - 17). Instead of permuting the column of $\mathbf{R}$, $\mathbf{Q}$,

input : N, n, m
output: $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{O}$

$\mathbf{R} = \mathbf{0}_{\mathbf{M}\times\mathbf{M}}$, $\mathbf{Q} = \mathbf{0}_{\mathbf{M}\times\mathbf{M}}$, $\mathbf{O} = \mathbf{0}_{\mathbf{1}\times\mathbf{M}}$;
for $i \leftarrow 1$ **to** M **do**
  $\mathbf{norm_i} = \left\|\hat{h}_i\right\|^2$;
end
for $i \leftarrow 1$ **to** M **do**
  $f = min\,(nfs + 1, M - i + 1)$;
  $j = \underset{l=i,\cdots,M}{\arg\min^{f}}\, \mathbf{norm}_{\boldsymbol{l}}$;
  $\mathbf{O_i} = \mathbf{j}$, $q_i = \hat{h}_j$, swap $\mathbf{norm_i}$ and $\mathbf{norm_j}$;
end
for $i \leftarrow 1$ **to** M **do**
  $r_{i,i} = \sqrt{\mathbf{norm_i}}$;
  $q_i = q_i / r_{i,i}$;
  for $k \leftarrow i + 1$ **to** M **do**
    $r_{i,k} = q_i^H \cdot q_k$;
    $q_k = q_k - r_{i,k} * q_i$;
    $\mathbf{norm_k} = \mathbf{norm_k} - \|\mathbf{r_{i,k}}\|^{\mathbf{2}}$;
  end
end

Algorithm 1: Modified FSD-SQRD

$\mathbf{O}$ and $\mathbf{norm}$ inside a SQRD process, the separated detection order is obtained iteratively before QRD. According to the chosen order, the $\mathbf{norm}$ is the only data for permutation while the selected column of $\hat{\mathbf{H}}$ is assigned to $\mathbf{Q}$ in sequence.

Given the selected ordering, the element assignment from $\hat{\mathbf{H}}$ to $\mathbf{Q}$ removes the need for iterative matrix permutation. Only the elements in $\mathbf{norm}$ are swapped, reducing the permutation to scalar in vector $\mathbf{norm}$ instead of vector in matrix $\mathbf{Q}$. In the next section, the effect of releasing such permuting operations on computational complexity is analyzed.

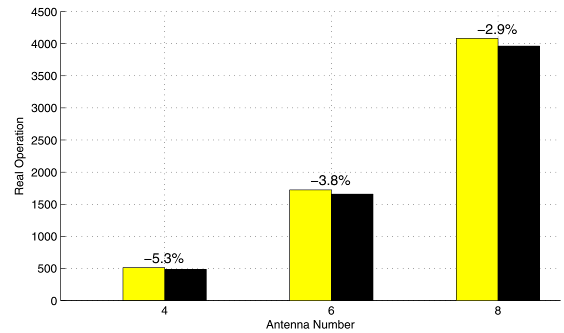
IV. SGR: Complexity

Performing the ordering in advance of QRD for FSD preprocessing, the computational complexity is reduced to only a single swap of elements of $\mathbf{norm}$ as described in Section III. Letting M denotes the number of transmit antennas and N the number of receive antennas, then the computational complexity (real data operations) for MFSD-SQRD is shown in Table I.

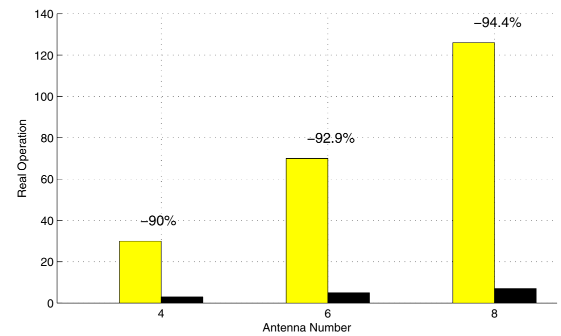
TABLE I: The Computational Complexity of MFSD-SQRD

	Norm	Ordering	QRD
$\pm$	$M(2N+3)$	0	$2MN2 + \frac{N(N+1)}{2}(8M-2)$
$\times$	$4MN$	0	$2MN + \frac{N(N+1)}{2}8m$
$\div$	0	0	$2MN$
$\sqrt{\cdot}$	0	0	N
Compare	0	$\frac{M(M+1)}{2}$	0
Permute	0	$M-1$	0

As described in Section III, all of $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{O}$ and $\mathbf{norm}$ are permuted iteratively. As $\mathbf{R}$ and $\mathbf{Q}$ are matrics, the complexity of the permutation operation is of $O(M^2)$ while the complexity of permutation operation on $\mathbf{O}$ and $\mathbf{norm}$ is of $O(M)$. Fig. 2 describes the relative complexities of the original FSD-SQRD approach (yellow) and the MFSD-SQRD (black) alternative.



(a) Operational Complexity



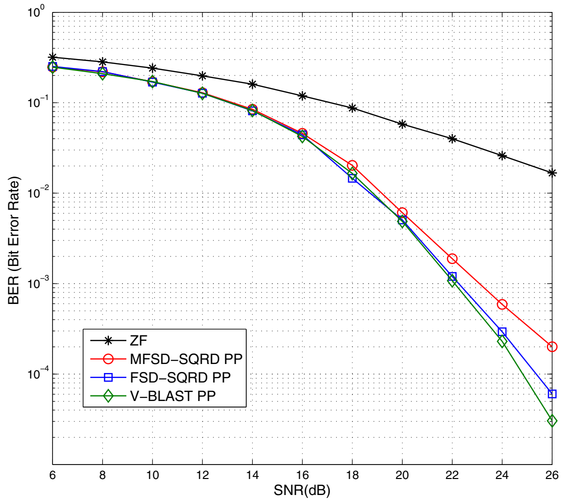
(b) Permute Complexity

Fig. 2: Complexity Reduction of FSD-SQRD and MFSD-SQRD

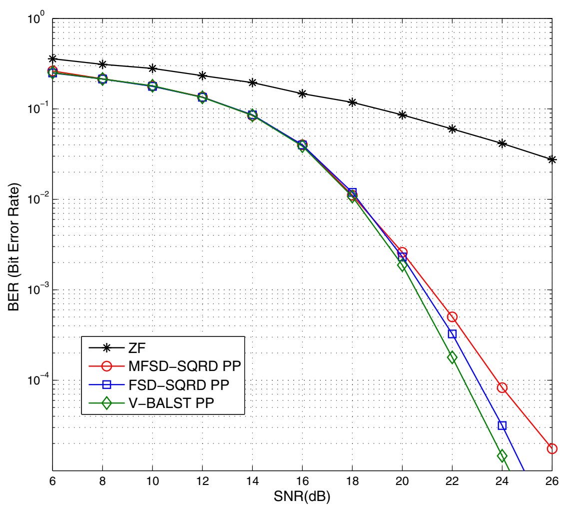
It is instructive to consider the trends in Fig. 2. According to Fig. 2a, the complexity of MFSD-SQRD (measured in terms of the number of real operations) is not significantly reduced relative to the original SQRD approach. The largest relative reduction in the number of operations is just in excess of 5.3%. This reduction decreases as the number of antenna grows. However, Fig. 2b provides an alternative interpretation. Specifically, when the application is distributed across a number of processing elements (as is required for real-time throughput), the localization of the permutation and QRD operations lead to a considerable reduction in the amount of data transferred between processors. As Fig. 2b describes, more than 90% of the permutation operations are removed which significantly reduces the amount of data swapping.

V. SGR: Performance

In this section, the detection accuracy of an FSD detector employing modified SQRD preprocessing is evaluated and compared to the equivalent detector employing SQRD. The perceived BER resulting from the FSD detector for both configurations is measured for both in 4×4 and 6×6 MIMO systems employing 16-QAM modulation. The evaluation is performed using Monte Carlo simulation over 100,000 transmitted symbols, accompanied with corresponding channel state information, as SNR is varied from 6 to 26 dB. Fig. 3a and Fig. 3b illustrate the relative detection performance when the modified SQRD (MFSD-SQRD), standard SQRD (FSD-SQRD), Zero-forcing (ZF) and V-BLAST ordered preprocessing are employed.



(a) 4×4



(b) 6×6

Fig. 3: BER Performance Comparison of FSD with Various Preprocessing Methods

A number of noticeable trends are apparent in the detection performance. In all cases, ZF offers the worst detection performance, with noticeable performance degradation at all SNR values. In the range 0 dB - 16 dB for both 4×4 and 6×6 topologies the detection performance for MFSD-SRD, FSD-SQRD and V-BLAST are all similar, with no notable performance degradation in any within this range. For SNR values exceeding 16 dB, V-BLAST preprocessing offers the best absolute BER performance, followed by FSD-SQRD and MFSD-SQRD. At the highest levels of SNR, MFSD-SQRD suffers from reductions in detection performance of less than 2 dB. Whilst this reduced performance is clearly undesirable, the high quality of the environment which would support such a large SNR renders this occurrence very rare. For most scenarios of practical interest - where SNR is much lower - the degradation of BER performance is quite mild, if even tangible at all.

VI. Summary

Pre-processing of channel matrices is an operation of critical importance to the effectiveness of Sphere Decoders for detection of symbols received from multi-antenna MIMO receivers. Over the course of a typical OFDM frame, it contributes little to the overall complexity of the SD process, yet consumes a major proportion of the hardware cost for SD accelerators. This paper has described the complexity and performance of a sub-optimal pre-processing algorithm inspired by SQRD-based QR decomposition. It offers little operational complexity reduction, yet by dramatically reducing the number of matrix permutation steps by over 90% for 4×4 MIMO topologies, effects a major reduction of close to 50% on the cost of the accelerator in [4].

Furthermore, this paper shows that, in representative 4×4 and 6×6 MIMO operating scenarios, the detection performance of this sub-optimal detector is barely reduced, relative to the alternatives. Specifically, for SNR values less than 16 dB, i.e. those of practical interest, the reduction in detection performance is almost indistinguishable. To the best of the authors' knowledge, this makes the proposed modified SQRD approach unique in offering both the most efficient FPGA-based acceleration on record, alongside quasi-optimal detection accuracy.

References

[1] *OFDM: Concepts for Future Communications Systems*. Springer, 2001.
[2] *MIMO: From Theory to Implementation*. Academic Press, Oxford, 2011.
[3] L. G. Barbero and J. S. Thompson. Fixing the Complexity of the Sphere Decoder for MIMO Detection. *IEEE Transactions on Wireless Communications*, 7(6):2131–2142, June 2008.
[4] X. Chu and J. McAllister. Software-Defined Sphere Decoding for FPGA-Based MIMO Detection. *IEEE Transactions on Signal Processing*, 60(11):6017–6026, Nov 2012.
[5] Y. Wu, J. McAllister, and P. Wang. High Performance Real-time Pre-Processing for Fixed-Complexity Sphere Decoder. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 1250–1253, Dec 2013.
[6] D. Wubben, R. Bohnke, V. Kuhn, and K. . Kammeyer. MMSE Extension of V-BLAST Based on Sorted QR Decomposition. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, pages 508–512 Vol.1, Oct 2003.
[7] D. Wubben, R. Bohnke, J. Rinas, V. Kuhn, and K. D. Kammeyer. Efficient Algorithm for Decoding Layered Space-time Codes. *Electronics Letters*, 37(22):1348–1350, Oct 2001.
[8] C. Zheng, X. Chu, J. McAllister, and R. Woods. Real-Valued Fixed-Complexity Sphere Decoder for High Dimensional QAM-MIMO Systems. *IEEE Transactions on Signal Processing*, 59(9):4493–4499, Sept 2011.