

High Throughput Modified MMSE Hardware Detector for High-Rate Spatial Modulation Systems

Xuan-Nghia Nguyen^{*†}, Van-Tu Nguyen^{*}, Ngoc-Nam Pham^{*}, Minh-Tuan Le^{†‡}, Xuan-Nam Tran[§], Vu-Duc Ngo^{*†}

^{*}Hanoi University of Science and Technology, Viet Nam

[‡]Ha Noi Department of Science and Technology, Viet Nam

[§]Le Quy Don Technical University, Viet Nam

[†]Mobifone Corporation, Viet Nam

Email:{vantu.bkhn, tuan.hdost}@gmail.com, nghianx@mobifone.vn, nam.phamngoc@hust.edu.vn, namtx@mta.edu.vn
duc.ngovu@hust.edu.vn

Abstract—A High-rate spatial modulation (HR-SM) transmission scheme was proposed in [1]. Compared to conventional Multiple-Input Multiple-Output (MIMO) systems, HR-SM has better bit-error-rate (BER) performance with trade off of spectral efficiency. In this paper, we propose a hardware architecture of HR-SM detector for 4 x 4 system using Modified-Minimum Mean Square Error - sorted QR decomposition (MSQRD) algorithm which was proposed in [2]. Implementation result shows that our design achieves higher throughput compared to other implementations in conventional MIMO systems despite the disadvantage of spectral utilization in HR-SM, while still keeps latency low and hardware usage acceptable.

Index Terms—MIMO, Detector, Spatial Modulation, FPGA, SQRD

I. INTRODUCTION

Spatial Modulation [3] is a promising MIMO system which provides better BER performance than conventional MIMO schemes (such as Spatial Multiplexing and Space-Time block codes) because it does not have Inter Carrier Interference (ICI) and Inter-Antenna Synchronization (IAS) problem [4]. In trade off, the original SM scheme has poor spectral utilization. A MIMO scheme - High Rate Spatial Modulation (HR-SM) was proposed in [1] based on the concept of SM to address this disadvantage. With a new design of Spatial Constellation (SC) codeword, an HR-SM system can achieve $2(n_T - 1) + \log_2(M)$ bpcu (bit per channel used), with n_T is the number of transmit antennas and M is the quadrature amplitude modulation deep. Compared to SM/GSM and SSK/GSSK [5–7], HR-SM not only has better spectral efficiency (although still lower than conventional MIMO systems), but its spectral utilization also grows linearly with the number of transmission antennas. In addition, Nguyen et al. [1] shows that HR-SM has better BER performance over SM/SSK system. The disadvantage of the HR-SM system compared to others SM systems is the power consumption, because all transmit antennas are activated at a time. In [2], a low complexity detection algorithm for HR-SM scheme has been presented, called MSQRD (Modified MMSE based on Sorted QR Decomposition).

To develop a hardware architecture for HR-SM system, we have studied a lot of implementations for conventional MIMO detector [8–12] and Sorted QR decomposition [13–15]. Al-

though Maximum Likelihood type detectors and V-BLAST [16] type detectors already have hardware implementations, the architecture becomes complex for MIMO systems which have high number of transmitting and receiving antennas. Thus, a class of MMSE-based detector has been focused, which bases on QR decomposition - a method that prevents detector from doing matrix inversion. In addition, QR decomposition has been used widely in many detection algorithms, including Zero Forcing (ZF) with Successful Interference Cancellation (SIC), Minimum Mean Square Error (MMSE) with Successful Interference Cancellation [16], or Sphere Decoding (SD) [9]. In [14], Huang and Tsai present an architecture of sorted Modified-Gram-Schmidt (MGS) QR decomposition (QRD), which has better BER compared to unsorted QRD, but without pipelining, this architecture provides poor throughput (only 1.27 million matrices per second, at a frequency of 200 MHz). In [12], Luethi et al. proposed another MGS-QRD architecture, which also has poor throughput (1.56 million matrices per second). In [8], Kim et al. proposed a pipeline architecture for QRD, which can provides matrix output every 8 clocks at a maximum frequency of 140 MHz in Xilinx Virtex 2. However, this architecture uses unsorted QRD, thus reduces BER performance compared to Sorted QRD [16]. Luethi et al. [12] also provide a pipeline architecture for QR Decomposition, but only for 2×2 systems. In [14], Huang and Tsai proposed a fully pipeline detector based on QR decomposition using Given Rotation algorithm, which is not as fast as MGS and also harder to implement in hardware. With the maximum clock frequency is up to 100 MHz, this architecture can reach a detection rate of 2.4 Gbps with 64-QAM. Finally, it is important to note that existing studies have focus on general MIMO systems, none of them is for HR-SM systems.

In this paper, we propose a high throughput hardware detector for recently proposed 4×4 HR-SM system using Modified SQRD algorithm. Firstly, a pipelined Sorted QR decomposition architecture using Modified-Gram-Schmidt algorithm is presented. Secondly, a detector which was built upon SQRD architecture and uses MSQRD algorithm with hardware optimizations is shown. Although HR-SM has lower bpcu than conventional MIMO system, we still achieved a

promising throughput by fully pipelining our architecture and optimizing the design to reach the maximum frequency.

The rest of the paper is organized as follows. Section II describes an overview of HR-SM system and MSQRD detection. Section III shows the architecture of Sorted QRD. Section IV presents the architecture of modified detection to adapt with HR-SM schemes. Implementation result and comparison are presented in Section V. At last, conclusions are given in Section VI.

II. HR-SM SCHEMES AND DETECTION

A. HR-SM System

Considering a $n_T \times n_R$ HR-SM system in Rayleigh fading environment, with n_T transmit and n_R receive antennas (Fig. 1). In each transmission cycle, $l + m$ bits are fed into the transmitter, while the first l bits are conveyed into antenna index to construct SC codeword, and the last m bits are carried by M-PSK or M-QAM symbol ($M = 2^m$). The SC codeword is designed in such a way that the first component is fixed to 1, and other components are chosen from the set $\{\pm 1, \pm j\}$ depending on the l bits input. Signal going to the transmit antennas is constructed by multiplying the SC codeword with PSK/QAM symbol $c = s \times x$.

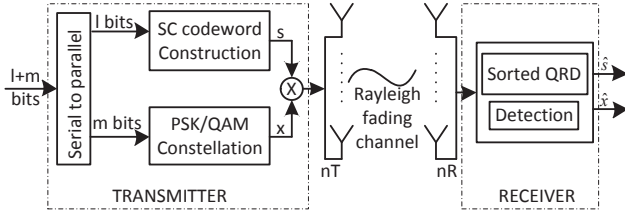


Fig. 1. HR-SM system with MSQRD Detector

The $n_R \times 1$ received signal vector \mathbf{y} at the receiver is given by:

$$\mathbf{y} = \sqrt{\frac{\gamma}{n_T E_s}} \mathbf{H} \mathbf{s} x + \mathbf{n}, \quad (1)$$

where \mathbf{H} is the channel matrix and \mathbf{n} is the $n_R \times 1$ noise vector.

B. MSQRD Detection

At the receiver, we place a HR-SM detector - a MSQRD detector, and assume that our system has $n_R = n_T$. MSQRD is a modified version of MMSE-SQRD in [16] that has been modified work with the HR-SM system. We define the $(n_T + n_R) \times n_T$ extended channel matrix \mathbf{D} and the extended received vector \mathbf{z} as follows:

$$\mathbf{D} = \begin{bmatrix} \mathbf{H} \\ \frac{1}{\sqrt{E_s}} \mathbf{I}_{n_T} \end{bmatrix}, \mathbf{z} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \quad (2)$$

We can apply MMSE-SQRD to decompose \mathbf{D} into \mathbf{Q} and \mathbf{R} matrix. Then, by multiplying \mathbf{z} with \mathbf{Q}^H , we have:

$$\begin{aligned} \mathbf{v} &= \mathbf{Q}^H \times \mathbf{z} \\ &= \mathbf{R} \times \mathbf{c} + \mathbf{w} \end{aligned} \quad (3)$$

Because of the structure of the upper triangular matrix \mathbf{R} , a Successive Interference Cancellation (SIC) procedure can be applied to get the output vector $\hat{\mathbf{c}}$: the last entry of \mathbf{v} (v_{n_T}) can be detected without interference from other antennas to have \hat{c}_{n_T} , by slicing $\frac{v_{n_T}}{R_{n_T, n_T}}$ (assuming that v_{n_T} is detected correctly). Next element of \mathbf{v} (v_{n_T-1}) can have interference canceled from \hat{c}_{n_T} by subtracting $R_{n_T-1, n_T} \times \hat{c}_{n_T}$. After that, result will be sliced with R_{n_T-1, n_T-1} to have \hat{c}_{n_T-1} . The loop is repeated until all entries of \mathbf{v} is detected. The SC codeword can be recovered by dividing all entries of $\hat{\mathbf{c}}$ to \hat{c}_{n_T} . A detailed description of the MSQRD algorithm is presented in [2]. Fig. 2 shows the BER performance of MSQRD Detector with $n_T = n_R = 4$ antennas and different modulation depths.

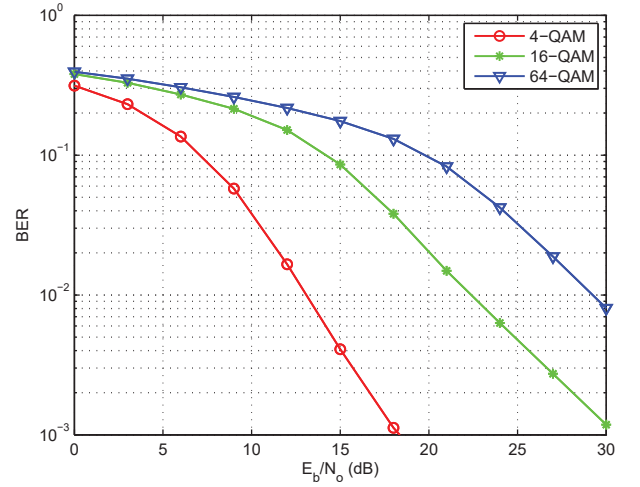


Fig. 2. BER performance

Algorithm 1: Sorted Modified Gram-Schmidt

Input: \mathbf{D}

Output: \mathbf{Q}, \mathbf{R}

- 1: $\mathbf{R} = \mathbf{0}; \mathbf{Q} = \mathbf{D}; \mathbf{p} = [0 : n_R - 1]$
- 2: for $n = 1$ to n_R begin
- 2: $norm(k) = \|\mathbf{Q}_k\|^2$
- 3: end
- 4: for $k = 0$ to $n_R - 1$ begin
- 5: $n_{min} = \min \{norm(k), \dots, norm(n_R - 1)\}$
- 6: $i = \text{index of } n_{min}$
- 7: swap column k and i of \mathbf{Q}, \mathbf{R} and \mathbf{p}
- 8: $\mathbf{R}(k, k) = \sqrt{n_{min}}, \mathbf{Q}_k = \frac{\mathbf{Q}_k}{\mathbf{R}(k, k)}$
- 9: for $k1 = k + 1$ to n_R begin
- 10: $\mathbf{R}(k, k1) = \mathbf{Q}_k^H \times \mathbf{Q}_{k1}$
- 11: $\mathbf{Q}_{k1} = \mathbf{Q}_{k1} - \mathbf{R}(k, k1) \times \mathbf{Q}_k$
- 12: $norm(k1) = norm(k1) - \|\mathbf{R}(k, k1)\|^2$
- 13: end
- 14: end

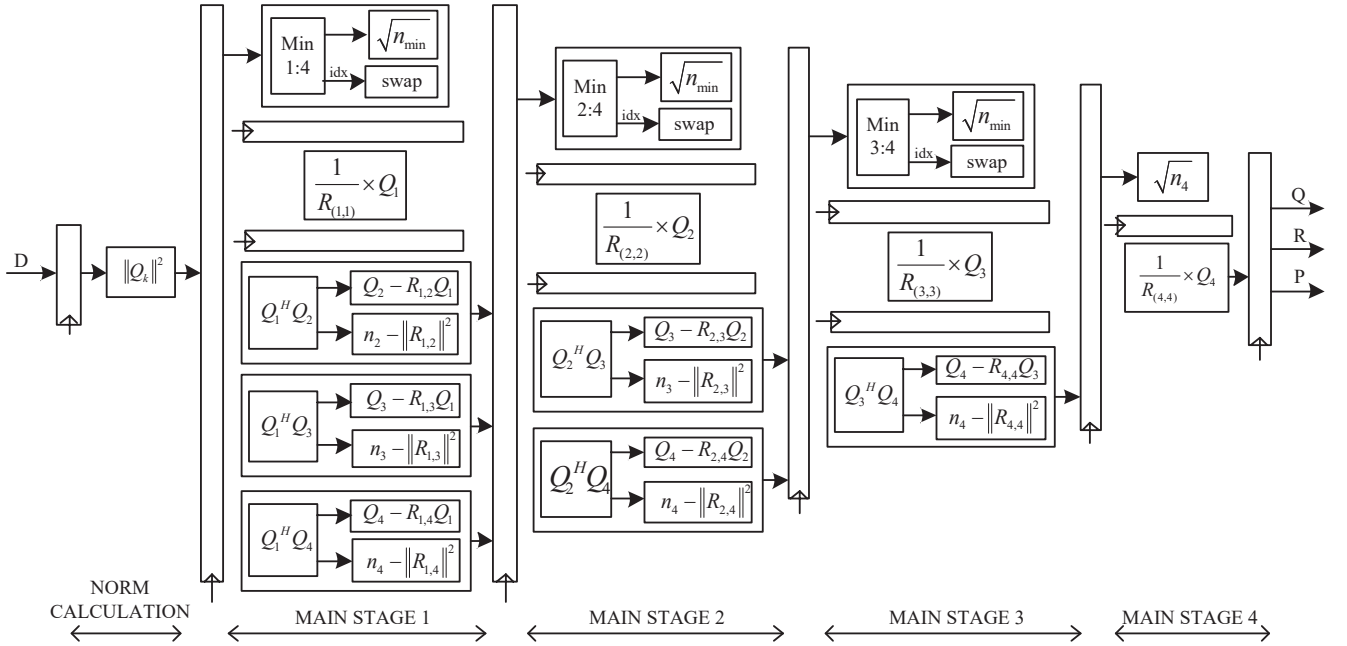


Fig. 3. Pipeline Architecture for Sorted MGS

III. SORTED QR DECOMPOSITION

There are two algorithms which have been used widely to process QR decomposition: Gram-Schmidt and Given Rotation. Since many analyses and implementations have been done for both of them, Gram-Schmidt is proven to be better to implement in FPGA for two reasons: 1) high speed multipliers is now available in FPGA, and 2) Gram-Schmidt is widely used in software using floating point [8, 12].

Because the original Gram-Schmidt algorithm has the problem at numerical stability [17], a Modified Gram-Schmidt (MGS) is used. In addition, a Sorted MGS will be applied to have better BER [16], although an increase in complexity of the algorithm will lead to an increase in hardware usage because of the extra steps required. In this design, values are represented in 12 bits fixed-point type. Higher-bit fixed point can be easily applied to the architecture to get a closer performance compared to floating-point, with some increase of slices and registers used (note that number of DSP usage will not increase, since a multiplier with less than 18 bits operators always uses one DSP).

A. Modified Gram-Schmidt Algorithm with Sorted procedure

Algorithm 1 describes the details of Sorted MGS, with Q_k denotes the column k of matrix Q . The idea of sorting comes to deal with the problem of SIC detection method: when detecting signal one after another and not in parallel, if the first detected signal is not the right one, other subsequent detected signals (in the same symbol duration) will be wrong. In order to archive the best BER performance, it is required to detect signal from the strongest to the weakest ones. The permutation

vector p is used to keep track of the sorting process. After all entries of vector y are detected, p will be used to re-sort the sequence of these entries as transmitted.

B. Pipelining Architecture

To maximize throughput, a pipelined architecture is used, as shown in (Fig. 3). The QR Decomposition is partitioned into five stages: NORM CALCULATION to calculate $\|Q_k\|^2$ with $k = \{1, 2, 3, 4\}$, four MAIN STAGE correspond to four column calculation in matrix Q . Each MAIN STAGE itself is divided into three smaller calculation stages (except the fourth one which is only divided into two smaller stages), following the Sorted MGS algorithm. The pipeline cycle is set to 8 clocks for the whole SQRD architecture. Data are processed as a continuous flows, and are backed up every stage, while local data inside each calculation stage is backed up every 8 clocks. We use built-in Xilinx ip-cores for Divider and Square Root blocks.

C. Resource Reuse

In order to achieve high frequency, all delay paths between two registers must not be greater than a serial combination of two 12-bit adders. In addition, with a 1-clock pipelining architecture, it could cost a large amount of hardware resources. Since the input matrix D has the size 8×4 , we use a 8-clock pipeline, so inside a stage, each column of matrix D can be completely processed by one unit block. With almost all blocks inside the Sorted QR Decomposition architecture are fully pipelined (except the divider, the reason will be discussed later in this section), we reuse all blocks in the Sorted MGS architecture eight times in eight cycles (Fig. 4)

to save hardware. Then, we can balance the trade-off between throughput and hardware utilization to a practical point.

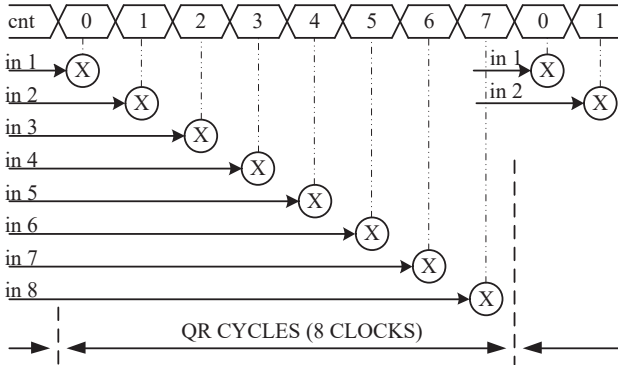


Fig. 4. Resource Reuse Technique

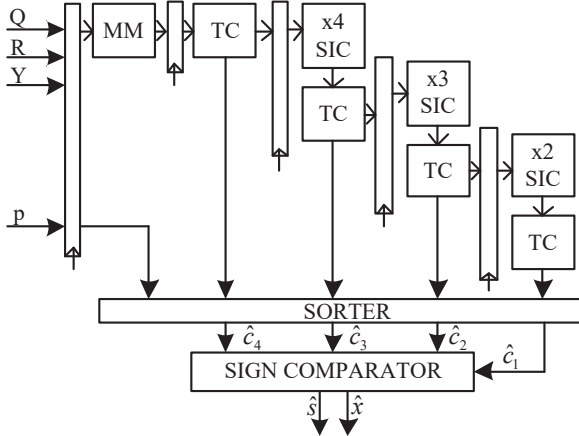


Fig. 5. Detector Architecture

Because of the structure of R matrix output (upper-triangular), $\frac{6}{16}$ or 37.5% registers for storing matrix R can be saved. In addition, values of R do not appear simultaneously at one time but in sequence, the early size of R need to be stored is small. On the other hand, although matrix Q has size $(n_T + n_R) \times n_T$, only the first n_R rows of Q will be used in detection, others are only used for computation inside Sorted QR block. Thus, we can reduce another significant amount of registers.

For the divider block, a direct vector-divider $Q_k = \frac{Q_k}{R_{(k,k)}}$ will cost extra hardware compared to the overall system. Thanks to the fact that $R_{k,k}$ is unchanged in a QR cycles (eight clocks), the inverse of $R_{k,k}$ will be calculated each QR cycle (initially by using eight clocks pipeline divider), then a pipeline multiplier will be used.

IV. MODIFIED DETECTION

In MSQRD [2], a SIC detection has been applied to detect signal in sequence, from the strongest to the weakest. Because

of the structure of SC codeword, each entry in the received vector y should correspond to a symbol in QAM constellation. Thus, according to [2], MSQRD can detect signal after applying SIC procedure by using threshold comparison. An illustration for 16-QAM is presented in (Fig. 6). To achieve high throughput, the architecture of Modified Detector is fully pipelined and can process each input vector y every clock cycle. After all entries of y are detected, a re-swap procedure will be applied to re-sort values into vector \hat{c} . Because first entries of SC codeword is always 1, the QAM modulated symbol should be the first entry of \hat{c} . Thanks to the structure of SC codeword, we can use a sign-comparator to recover others entries of SC codeword instead of using a complex divider, which lead to saving a remarkable amount of hardware. The Architecture of the Detector is shown in Fig. 5, with MM is matrix multiplication block(produce $Q \times y$), SIC is the Interference Cancellation block, and TC is the Threshold Comparison block.

A SIC detector requires extra complex subtractions and multiplications to be done in interference cancellation steps. Because one operand in multiplier is always a symbol in QAM constellation, a complex multiplier - which costs 4 DSPs with additional LUTs - can be replaced by a combination of shift and adder, for example:

$$\begin{aligned} & (a + bi)(1 + 3i) \\ &= (a - 3b) + (b + 3a)i \\ &= (a - b - b \ll 1) + (b + a + a \ll 1)i \end{aligned} \quad (4)$$

A multiplexer can be used to deal with different constellation symbol input, then extra DSP blocks can be saved at the cost of an acceptable increase of LUTs used.

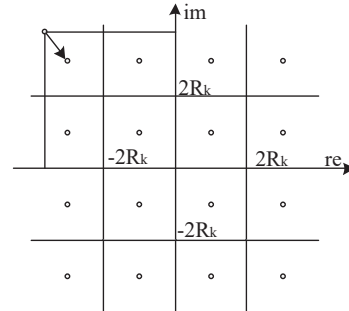


Fig. 6. Threshold Comparison

V. IMPLEMENTATION RESULT

The implementation result of our work is shown in Table I. The target FPGA chip is Virtex-6 VLX75T at speed grade -3. Our design is able to run at 429.9 MHz clock speed, producing output QR matrix every 8 clocks and detecting one vector y input every single clocks. Thus, the throughput of the design is 4.3 Gbps at 16-QAM, and even higher if we use 64-QAM.

For better look at the comparison, we re-synthesize our design in Virtex-5 VSX35T (speed grade -3) and Virtex-2 6000

TABLE I
IMPLEMENTATION COMPARISON

Section	In [13]	In [12]	In [14]	In [8]	This Work
Algorithm	Sorted MGS	Sorted MGS	Given Rotation	Un-sorted MGS	Sorted MGS
Target Chip	Virtex-5	UMC 0.18m 1P/6M CMOS	0.18um	Virtex-2 6000	Virtex-6 VLX75T
MaxFreq.	200 MHz / 349.3 MHz	162 MHz	100 MHz	140 MHz / 181.3 MHz	429.9 MHz
Hardware Usage				9003 LUTs 24 Blocks RAM 66 multipliers	14184 registers 10948 LUTs 141DSPs
QRD throughput (QRs/s)	1.27M / 43.7M	1.56M	25M	14.4M / 22.7M	53.73M
Detect throughput (instances/s)	1.27M / 349.3M		100M	14.4M / 181.3M	429.9M
Detect throughput			2.4 Gbps (64-QAM)	345.6 Mbps (64-QAM)	4.3 Gbps (16-QAM)
In-Out delay	157 clocks			388 clocks	269 clocks

⁽¹⁾ **Bold text** shows results when we re-synthesize our design in the same FPGA chip as the compared design.

⁽²⁾ **M** means million.

(speed grade -6). Synthesis results are shown in bold text in Table I. Notice that while our work is for HR-SM systems, all other works is for conventional 4×4 MIMO system with better spectral utilization [8, 14] ($n_T \times \log_2 M$ bpcu compared to $(2(n_T - 1) + \log_2 M)$ bpcu in HR-SM system, with M is the modulation depth). In [13], an architecture for sorted MGS has been implemented, providing a maximum frequency of only 200 MHz, compared to that of 349.3 MHz of our architecture. Furthermore, the design in [13] has poor throughput due to non-pipelining architecture. In [8], Kim et al. presented a pipeline architecture for Un-sorted MGS algorithm with a maximum frequency of 140 MHz in Virtex-2 6000, which is still smaller than that our architecture when synthesized in the same chip (181.3 MHz). In our design, we assume one that one input matrix D is used for 8 vectors y . If the number of equivalent vector y allowed is smaller (for example, 4 vectors), the throughput is reduced (to 2.15 Gbps at 16-QAM with 4 vectors y per input matrix D - still a high throughput), but the amount of hardware required for detector block will also decrease. Other related architectures ([12, 14]) are synthesized in ASIC, so a truly fair comparison is not possible. Overall, our work has higher frequency and throughput compared to others conventional system despite the spectral characteristic of HR-SM scheme, while maintaining an acceptable cost of hardware usage.

VI. CONCLUSION

In this paper, we have presented a pipelined and optimized hardware detector architecture for High-rate Spatial Modulation system using MSQRD algorithm proposed in [2]. Although HR-SM does not have spectral efficiency advantage compared to conventional MIMO schemes, by fully pipelining and optimizing hardware architecture, our design can achieve a comparable and even higher throughput (QR/s and instances/s) compared to other existing systems. With the advantages of HR-SM, our proposed detector can be used for high-speed wireless systems such as WLAN IEEE 802.11ac or 5G.

ACKNOWLEDGEMENT

This work is sponsored by National Foundation for Science and Technology Development (Nafosted) under project number 102.02-2015.23.

REFERENCES

- [1] Thu-Phuong Nguyen, Minh-Tuan Le, Vu-Duc Ngo, Xuan-Nam Tran, and Hae-Wook Choi, "Spatial modulation for high-rate transmission systems," *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, pp. 1–5, May 2014.
- [2] Tien-Dong Nguyen, Xuan-Nam Tran, Trung-Minh Do, Vu-Duc Ngo, and Minh-Tuan Le, "Low-complexity detectors for high-rate spatial modulation," *Advanced Technologies for Communications (ATC), 2014 International Conference on*, pp. 652–656, 2014.
- [3] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 4, pp. 2228–2241, July 2008.
- [4] M. Di Renzo, H. Haas, A. Ghayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized mimo: Challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, Jan 2014.
- [5] J. Fu, C. Hou, W. Xiang, L. Yan, and Y. Hou, "Generalised spatial modulation with multiple active transmit antennas," *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pp. 839–844, Dec 2010.
- [6] J. Jeganathan, A. Ghayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for mimo channels," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 7, pp. 3692–3703, July 2009.
- [7] J. Jeganathan, A. Ghayeb, and L. Szczecinski, "Generalized space shift keying modulation for mimo channels," *Personal, Indoor and Mobile Radio Communications*,

2008. *PIMRC 2008. IEEE 19th International Symposium on*, pp. 1–5, Sept 2008.
- [8] H. Kim, W. Zhu, J. Bhatia, K. Mohammed, A. Shah, and B. Daneshrad, “A practical, hardware friendly mmse detector for mimo-ofdm-based systems,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 267460, 2008.
- [9] M. Khairy, M. Abdallah, and S.-D. Habib, “Efficient fpga implementation of mimo decoder for mobile wimax system,” *Communications, 2009. ICC '09. IEEE International Conference on*, pp. 1–5, June 2009.
- [10] J. Soler-Garrido, D. Milford, M. Sandell, and H. Vetter, “Implementation and evaluation of a high-performance mimo detector for wireless lan systems,” *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 4, pp. 1519–1527, November 2011.
- [11] A. Salari, S. Fakhraie, and A. Abbasfar, “Algorithm and fpga implementation of interpolation-based soft output mmse mimo detector for 3gpp lte,” *Communications, IET*, vol. 8, no. 4, pp. 492–499, March 2014.
- [12] P. Luethi, C. Studer, S. Duetsch, E. Zraggen, H. Kaeslin, N. Felber, and W. Fichtner, “Gram-schmidt-based qr decomposition for mimo detection: Vlsi implementation and comparison,” *Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on*, pp. 830–833, Nov 2008.
- [13] G. L. Nazar, C. Gimpler, and N. Wehn, “Implementation comparisons of the qr decomposition for mimo detection,” *Proceedings of the 23rd Symposium on Integrated Circuits and System Design*, pp. 210–214, 2010.
- [14] Z.-Y. Huang and P.-Y. Tsai, “Efficient implementation of qr decomposition for gigabit mimo-ofdm systems,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 10, pp. 2531–2542, Oct 2011.
- [15] M. Abels, T. Wiegand, and S. Paul, “Efficient fpga implementation of a high throughput systolic array qr-decomposition algorithm,” *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pp. 904–908, Nov 2011.
- [16] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, “Mmse extension of v-blast based on sorted qr decomposition,” *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 1, pp. 508–512 Vol.1, Oct 2003.
- [17] . Bjrkck, “Numerics of gram-schmidt orthogonalization,” *Linear Algebra and its Applications*, vol. 197198, no. 0, pp. 297 – 316, 1994.