Energy-Efficient Adaptive Modulated Fixed-Complexity Sphere Decoder

Energy-Efficient Adaptive Modulated Fixed-Complexity Sphere Decoder

Yun Wu
School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, UK
yw106@hw.ac.uk

John McAllister
School of Electronics, Electrical Engineering and Computer Science
Queen's University Belfast
Belfast, UK
jp.mcallister@qub.ac.uk

***Abstract*—Fixed-Complexity Sphere Decoder (FSD) is an quasi-optimal detector for Multiple-Input Multiple-Output (MIMO) system which is a hardware-friendly parallel tree-search customised to the modulation and antenna scheme employed. However, it is not able to adapt its behaviour for various modulation and antenna schemes, as demanded by modern wireless standard. This restricts its usage in modern adaptive MIMO systems. This paper proposes a solution to this problem. A configurable FSD structure in proposed where normalized higher order modulation schemes can accommodate lower ones. By exploiting clock-gating, FSD of all modulation schemes is equally trimmed to allow power savings of over 30% when implementing on Field Programmable Gate Array (FPGA). This architecture enables the facility to balance the power consumptions with compatible information rate in dynamic, adaptive MIMO communications environments.**

***Index Terms*—Adaptive Modulation, Fixed-complexity Sphere Decoder, Multiple-Input Multiple-Output, Field Programmable Gate Array**

I. Introduction

The maximum information transmission capacity of modern Multiple-Input Multiple-Output (MIMO) communication systems depend on the communication environment, specifically the Signal-to-Noise-Ratio (SNR) and the channel state [1]. Adapting to these conditions as they vary enable MIMO with very large capacity, and spatial diversity but which is highly reliant on Channel States Information (CSI) as well as SNR. As these vary, different numbers of antennas [2], modulation schemes [3], coding schemes [4], and transceiver power [5] and all be used in adaptive configurations; hence the communications environment is a key factor in the performance of fundamental signal processing operations, such as symbol detection, in MIMO transceivers.

In particular, Adaptive Modulation (AM) is widely adopted by MIMO detection algorithms such as linear equalization [6] or Maximum Likelihood (ML) [7]. However, while quasi-ML MIMO detection algorithms, Sphere Decoder (SD) [8] can achieve near-optimal detection accuracy for fixed configurations, it is a challenge to support AM.

The Fixed-Complexity Sphere Decoder (FSD) [9] is a deterministic SD algorithm which enables quasi-ML detection with complexity significantly reduced from the ideal ML detector. It exploits fixed antenna and modulation configurations to support efficient real-time accelerators for MIMO transceivers [10]. However, it cannot easily handle variation in modulation schemes as it necessitates full enumeration of modulation symbols; this change causes very significant variation in decoding algorithms structure, computational load, resource cost and power consumption. To the best of the authors' knowledge, there has been no real-time accelerator for FSD which can handle AM schemes of varying scales.

This paper addressing this shortcoming and proposes an adaptive FSD architecture supporting AM with significant power reduction by employing adaptive clock-gating on Field Programmable Gate Array (FPGA). By adapting the FSD detection data path according to the quantity of full enumeration, the unused data paths are pruned by gated clock to reduce the power consumption. Specifically, three contributions are made:

- The first AM FSD architecture is presented,
- An adaptive QAM demodulator is developed for AM FSD,
- A Clock-gate infrastructure for proposed AM architecture is developed enabling over 32% power savings.

The FSD with AM is introduced in Section II. In Section III, the proposed adaptive modulation and FSD architectures are illustrated after which the experiment results are demonstrated in Section IV with conclusion in Section V.

II. Background

A. Adaptive Modulated MIMO System

A MIMO system employs N_t transmit antennas and N_r receive antennas, assuming $N_t \leq N_r$ [11]. A data stream for transmission is modulated and multiplexed onto N_t transmit antennas as a symbol vector $\mathbf{s} \in \mathbb{C}^{N_t \times 1}$, which is distorted by both Additive White Gaussian Noise (AWGN) and multi-path fading as $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ at receiver. (1) shows the relation between $\mathbf{s}$ and $\mathbf{y}$:

$$\mathbf{y} = \sqrt{\frac{\rho}{N_r}} \cdot H \cdot \mathbf{s} + \mathbf{w} \tag{1}$$

where $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$ is the AWGN with power σ_w^2. Considering σ_s^2 as the signal power of $\mathbf{s}$, the Signal-to-Noise Ratio (SNR)

Authors undertook this work at the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfastk

ρ is given by $\frac{\sigma_s^2}{\sigma_w^2}$. Ideally, the Rayleigh-distributed multi-path channel is represented by $H \in \mathbb{C}^{N_r \times N_t}$ where $h_{i,j}$ denotes the path fading between the i^{th} receive and j^{th} transmit antennas [12].

In a AM-MIMO system, the modulation order is updated at the transmitter on a per-antenna basis according to the SNR and channel state information during communication [3]. This allows for higher information rate and detection BER performance [13]. Fig. 1 shows the system diagram of AM-MIMO, where the transmitted signal $\mathbf{s}$ is of diverse modulation over N_t antennas. It is updated according to the feedback link at the receiver side based on the channel state information estimation with modulation order indicators.

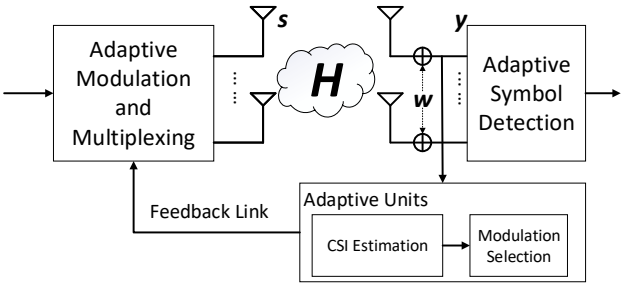


Fig. 1: Adaptive Modulated MIMO System Model

The received signal $\mathbf{y}$ is recovered into specific modulated constellation adaptively with MIMO detection techniques. Due to the diverse modulation in AM, it deeply affects on the performance of MIMO detector, especially those correspondent to the modulation order with fixed detection data path such as Fixed-Complexity Sphere Decoder (FSD) [9].

B. Fixed-Complexity Sphere Decoder

FSD employs breadth-first SD algorithm with a deterministic tree search behavior [9], which is of quasi-ML BER performance and highly parallel structure with much lower complexity than ML detection. It divides the tree search layers into two stages according to the number of enumerating symbols, which are Full Search (FS) with all possible modulated constellations and Single Search (SS) with only one quantized modulated constellation. The number of FS layers is defined by $NFS = \lceil \sqrt{N_t} - 1 \rceil$. By pre-processing the channel matrix H, the Vertical-Bell Laboratories Layered Space-Time (VBLAST) order is adopted for the SS stage while the reversed VBLAST order is employed by the FS stage [9].

By decomposing the ordered $\hat{H}$ with QR Decomposition (QRD), $(Q, R) = QR(\hat{H})$, the final detected symbols are determined by finding the minimum Accumulated Partial Euclidean Distance (APED) in (3) among all tree search branches,

$$P\tilde{E}D_{l,k} = \sum_{j=l}^{N_t} \tilde{r}_{j,j}^2 \left\| \tilde{y}_{j,k} - \tilde{x}_{j,k} \right\|^2, \; l \in [1, N_t] \qquad (2)$$

$$A\tilde{P}ED_{l,k} = \sum_{j=l}^{N_t} P\tilde{E}D_{j,k} \qquad (3)$$

where $\tilde{y}_{j,k}$ is the j^{th} element of Zero-Forcing equalized signal with successive noise cancellation at k^{th} APED branch; $\tilde{x}_{j,k}$ is the k^{th} symbol enumerated at the j^{th} layer; l is the index of current tree search layer.

Fig. 3 shows the AM FSD tree search structure with FS layer switching from QPSK to 16-QAM for 4×4 MIMO detection. As this shows, the topological differences in the algorithm structure are stark in terms of both the number of nodes used. By adapting FSD with different tree search structure for various modulation schemes, a good trade off among information rate, BER performance and detection complexity is obtained. However, AM FSD is only algorithmic achievable while no architecture supported.

III. AM Architecture

A. AM FSD

To maintain the deterministic tree search structure of FSD while adapting to various modulation schemes, a reconfigurable architecture with overlapped tree search structure is created while the APED branch across different modulation schemes are reused. Fig. 2 illustrates the concept of 'reuse' tree search structure designed to accommodate both QPSK and 16-QAM modulation schemes.

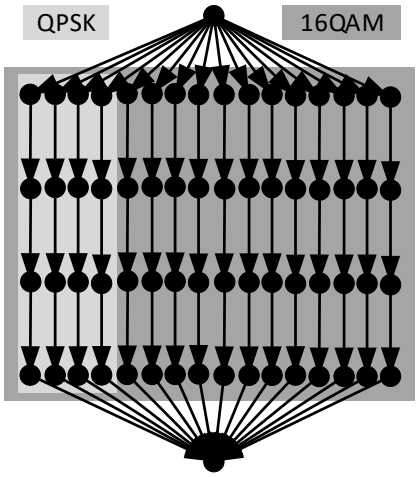
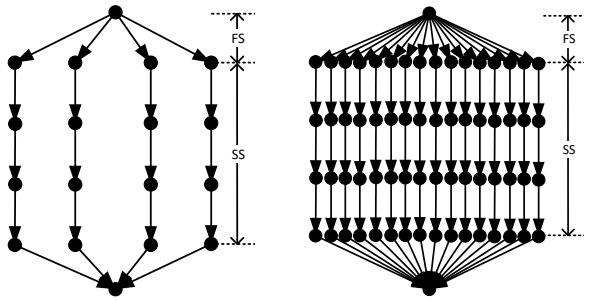


Fig. 2: Reused Tree Structure for AM FSD

It is clear that the tree search structure of QPSK can be accommodated by 16-QAM due to fewer APED branches. Therefore, the parallel 16-QAM FSD architecture can enact QPSK FSD by only activating the first 4 APED branches as shown in Fig. 2. The AM FSD using QPSK and 16-QAM in Section II-B can adopt the same deterministic architecture without switching between two different ones as shown in Fig. 3a and Fig. 3a.



(a) QPSK (b) 16-QAM

Fig. 3: AM FSD Tree Structure from QPSK to 16-QAM

As the parallel AM FSD structure is correspondent to the constellation sizes of QAM, all based on 2^M, the AM FSD with larger QAM architecture can be reused by those with lower QAM by closing $2^M - 2^{M/2}$ parallel APED branches, where M is the number of bits per symbol[1].

To switch between different modulation schemes and save the power consumptions, the active APED branches are controlled by using clock-gating technique [14]. Fig. 4 demonstrates the clock-gating architecture of AM FSD support modulation up to 256-QAM.

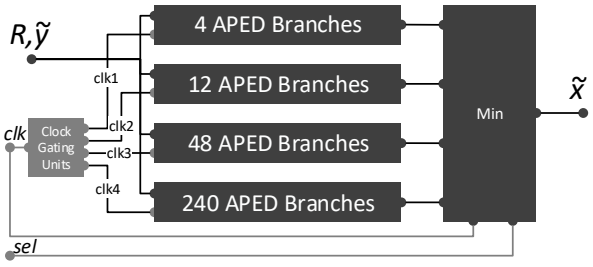


Fig. 4: AM FSD Architecture for Up to 256-QAM

As shown, the upper triangular matrix R and ZF equalized signal vector $\tilde{y}$ are input to all the APED branches while the final recovered modulated signal $\tilde{x}$ is output by judging the minimum APED of each branches according to different modulation scheme using 'sel' signal. The first 4 APED branch is separately clocked for QPSK while the increasingly differential number of APED branches for 16-QAM (12 APED branches), 64-QAM (48 APED branches), and 256-QAM (240 APED branches) are independently clocked.

The clock-gating units takes the system clock clk, as input which is split into four separate ones, $clk1, clk2, clk3, clk4$, with frequency tuning capability. By enabling the clock-gated combination from $clk1$ to $clk4$, a specific AM FSD tree structure is chosen. For instance, when $clk3$ and $clk4$ are clock-gated, the AM FSD is of tree structure 16-QAM.

Only one modulation scheme is enabled through architecture reusing and clock-gating. It performs conventional hard output FSD detection with chosen modulation scheme. With corresponding soft output extension, such as [15], it would perform the normal Adaptive Modulation and Coding (AMC) [16].

[1]and only even bits per symbol or square QAM is considered in this work

B. AM Demodulator

Symbol demodulation is embedded in the APED branch computation to estimate the $\tilde{x}_{j,k}$ in (2) at SS stage of AM FSD. Assuming the scaled $\tilde{y}$ by $\sqrt{M_c/2}$, the estimated modulated symbol can be obtained by finding the symbol index based on pre-mapped constellation set with gray-code. The symbol index mappings with normalized symbol quantiztion are given by (4)-(7).

$$SI_{QPSK} = (BI_2 << 2) \mid BI_1, \quad when \begin{cases} BI_2 = 1, \Im(\tilde{y}_{j,k}) \geq 0 \\ BI_1 = 1, \Re(\tilde{y}_{j,k}) \geq 0 \end{cases} \tag{4}$$

$$SI_{16QAM} = (BI_4 << 3) \mid (BI_3 << 2) \mid (BI_2 << 2) \mid BI_1, \quad when \begin{cases} BI_4 = 1, \Im(\tilde{y}_{j,k}) \geq 0 \\ BI_3 = 1, \|\Im(\tilde{y}_{j,k})\|_1 < 2 \\ BI_2 = 1, \Re(\tilde{y}_{j,k}) \geq 0 \\ BI_1 = 1, \|\Re(\tilde{y}_{j,k})\|_1 < 2 \end{cases} \tag{5}$$

$$SI_{64QAM} = (BI_6 << 5) \mid (BI_5 << 4) \mid (BI_4 << 3) \mid (BI_3 << 2) \mid (BI_2 << 1) \mid BI_1, \quad when \begin{cases} BI_6 = 1, \Im(\tilde{y}_{j,k}) \geq 0 \\ BI_5 = 1, \|\Im(\tilde{y}_{j,k})\|_1 < 4 \\ BI_4 = 1, \|\Im(\tilde{y}_{j,k})\|_1 \geq 2 \ and \ |\Im(\tilde{y}_{j,k})|_1 < 6 \\ BI_3 = 1, \|\Re(\tilde{y}_{j,k})\|_1 \geq 2 \\ BI_2 = 1, \|\Re(\tilde{y}_{j,k})\|_1 < 4 \\ BI_1 = 1, \|\Re(\tilde{y}_{j,k})\|_1 \geq 2 \ and \ |\Re(\tilde{y}_{j,k})|_1 < 6 \end{cases} \tag{6}$$

$$SI_{256QAM} = (BI_8 << 7) \mid (BI_7 << 6) \mid (BI_6 << 5) \mid (BI_5 << 4) \mid (BI_4 << 3) \mid (BI_3 << 2) \mid (BI_2 << 1) \mid BI_1, \quad when \begin{cases} BI_8 = 1, \Im(\tilde{y}_{j,k}) \geq 0 \\ BI_7 = 1, \|\Im(\tilde{y}_{j,k})\|_1 < 8 \\ BI_6 = 1, \|\Im(\tilde{y}_{j,k})\|_1 \geq 4 \ and \ |\Im(\tilde{y}_{j,k})|_1 < 12 \\ BI_5 = 1, \big(\|\Im(\tilde{y}_{j,k})\|_1 \geq 2 \ and \ |\Im(\tilde{y}_{j,k})|_1 < 6\big) \ or \\ \qquad \big(\|\Im(\tilde{y}_{j,k})\|_1 \geq 10 \ and \ \|\Im(\tilde{y}_{j,k})\|_1 < 14\big) \\ BI_4 = 1, \Re(\tilde{y}_{j,k}) \geq 0 \\ BI_3 = 1, \|\Re(\tilde{y}_{j,k})\|_1 < 8 \\ BI_2 = 1, \|\Re(\tilde{y}_{j,k})\|_1 \geq 4 \ and \ |\Re(\tilde{y}_{j,k})|_1 < 12 \\ BI_1 = 1, \big(\|\Re(\tilde{y}_{j,k})\|_1 \geq 2 \ and \ \|\Re(\tilde{y}_{j,k})\|_1 < 6\big) \ or \\ \qquad \big(|\Re(\tilde{y}_{j,k})|_1 \geq 10 \ and \ \|\Re(\tilde{y}_{j,k})\|_1 < 14\big) \end{cases} \tag{7}$$

where $BI_1 - BI_8$ are the indexing offset corresponding to the gray-coded inphase and quadrature binary format of modulated symbol.

By shifting $BI_1 - BI_8$ according to the position of bit, the quantized symbol index (SI) is aggregated using bit-wise or operation, which is then used to address the pre-memorized modulated constellations for estimate $\tilde{x}_{j,k}$ of the APED computations.

Fig. 5 describes AM demodulators for different APED branches. The input is $\tilde{y}_{j,k}$ while the output is the quantized symbol constellation $\tilde{x}_{j,k}$. Each type of demodulator is connected to the memory of pre-stored modulation constellations, where the indexing is multiplexed to output specific modulated scheme through the input 'sel' signal.

As shown in Fig. 2, since QPSK APED branch is reused in all four modulation schemes, it equips the most demodulator schemes. As the modulation order increases from 16-QAM to 256-QAM, the equipped demodulator decreases as APED branch of higher modulation schemes does not need to support lower ones, where 256-QAM APED branch only needs its own demodulator in this case.

By providing separated clocking source to different demodulation modules, each demodulating data path can be clock-gated, where only the selected modulate scheme is enabled for specific FSD APED computations. Notice that the clock-gating of various demodulator is not standalone but embedded in the APED branches as show in Figure 4. The detailed cost and power consumption are given in next section.

IV. Evaluation

Fixed-point implementations, using a total of 16 bits (8 integer bits), for both AM demodulator and FSD are evaluated on Zynq ZC702 evaluation and development kit using Xilinx HLS 2016.4, which are packaged as customized peripherals Intellectual Property (IP) on FPGA and data streaming from ARM core using AMBA bus. 12 IPs are created as a batch corresponding to the sub-carrier number in a single Resource Blocks (RB) of OFDM symbol [17].

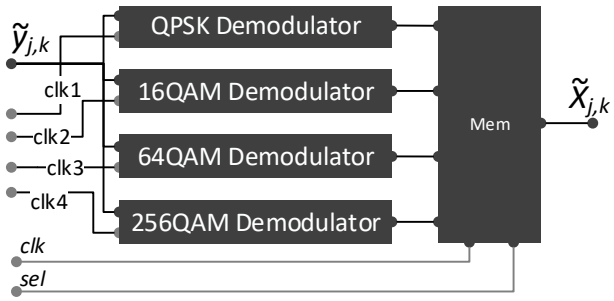
A. Cost

TABLE I: AM Demodulator of RB

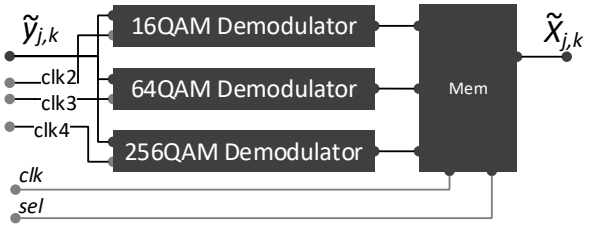
APED Scheme	QPSK	16-QAM	64-QAM	256-QAM
LUT ($\times 10^3$)	5.8	8.8	6.7	3.9
DSP48E1	0	0	0	0
BRAM	12	0	0	0
Clock (MHz)	256	265	294	308

TABLE II: AM FSD of RB, 4×4

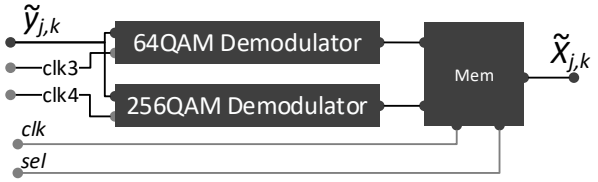
Scheme	AM FSD
LUT ($\times 10^3$)	22.7
DSP48E1	144
BRAM	30
Clock (MHz)	132
Throughput (Mbits/s)	10.95 / 21.90 / 32.85 / 43.80



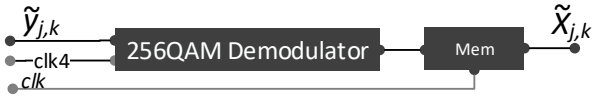
(a) QPSK APED



(b) 16-QAM APED



(c) 64-QAM APED



(d) 256-QAM APED

Fig. 5: AM Demodulator of APED Branch

Table I shows the IP resource utilization using all four different demodulator combinations in Fig. 5 for a single RB. The AM demodulator for QPSK APED consumes the most costs with both Look-Up-Tables (LUT) and Block RAM (BRAM) while the costs reduce as the supporting modulation sizes grows up. The frequency is increasing from AM demodulator of QPSK to 256-QAM APED branch. This corresponds as shown in Fig. 5 that the higher the supported modulation density, the less complex the demodulator.

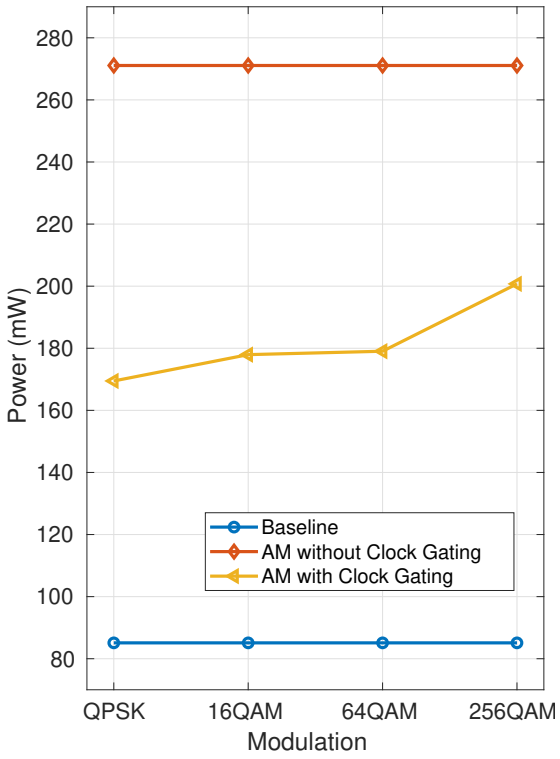
Table II shows the IP resource utilization of AM FSD using the AM demodulator modules. The AM FSD does consumes the most cost since it includes all four SM FSD IP while the SM FSD costs increase as supporting modulation sizes grows up. The frequency of AM FSD is the lowest one as it is the most complex while the frequency increase as the modulation order goes up. The corresponding throughput for QPSK-256QAM are various from 10.95 Mbits/s to 43.80 Mbits/s.

B. Power

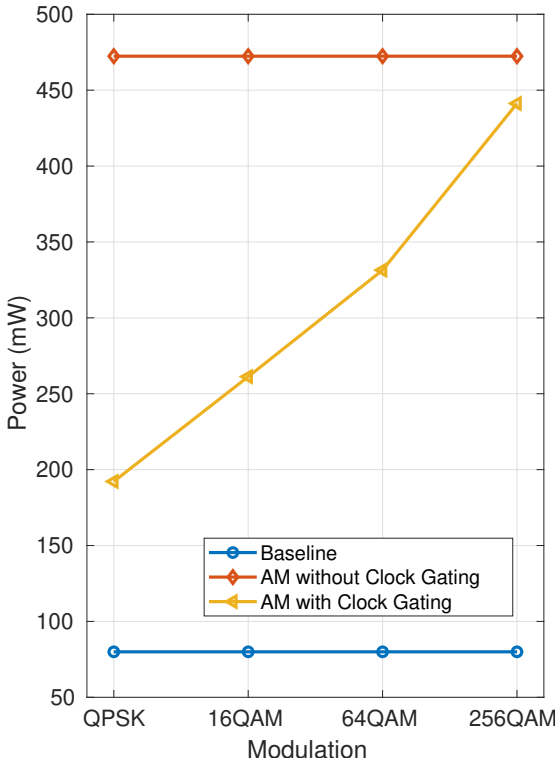
The power are measured through onboard power rail of Programmable Logic (PL) based on PMBus [18], where the average power is derived with 1000 random sampling measurements over 5 seconds for each scenario. The clock-gating is implemented using four fabric clock source from Zynq 7000 processing system, $FCLK\_CLK\{0, 2, 3, 4\}$, by unlocking the System Level Control Register (SLCR). Both power monitor and clock-gating are controlled through software, which are tunable at run-time.

Figure 6 demonstrates the power consumption of example AM demodulator, Figure 6a, and AM FSD, Figure 6b. The measured PL power without configuration is considered as the baseline static power, while the active power is measured against the baseline to show the dynamic power. The AM without clock-gate refers refers to fully active APED branches while AM with clock-gate indicates adaptive active APED branches.

In Figure 6a, by clock-gating the QAM demodulator corresponding to inactive APED branch, the power consumption is arising from QPSK to 256-QAM. By assuming equal occurring probability of various QAM schemes, up to 35% power reduce is achieved with clock-gating compared to fully active version. In Figure 6b, similar situation happens for the AM FSD compared to SM FSD. By assuming equal occurring probability of various QAM schemes are selected at FS stage, over 32% is achieved compared to its fully active AM FSD.



(a) AM Demodulator



(b) AM FSD

Fig. 6: AM Power Consumption with Clock-Gate

V. Conclusion

In this work, AM FSD with corresponding AM demodulator are presented for AM MIMO system. It enables various FSD tree search structure by activating specific modulation at FS stage and closing the rests using clock-gating. The ability of adaptation is at a cost of power consumption than its corresponding single modulated version, however, significant power saving is achieved compared to fully clocked version. To the best of authors' knowledge, this is the first AM FSD prototype so far while further optimization, such as soft AM FSD joint with channel coding, is ongoing as future works.

References

[1] A. J. Goldsmith and P. P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[2] H. Zhang, H. Dai, Q. Zhou, and B. L. Hughes, "On the Diversity Order of Spatial Multiplexing Systems With Transmit Antenna Selection: A Geometrical Approach," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5297–5311, 2006.

[3] P. Sebastian, H. Sampath, and A. Paulraj, "Adaptive Modulation for Multiple Antenna Systems," in *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers (Cat. No.00CH37154)*, 2000, vol. 1, pp. 506–510 vol.1.

[4] M. D. Dorrance and I. D. Marsland, "Adaptive Discrete-Rate MIMO Communications with Rate-Compatible LDPC Codes," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3115–3125, 2010.

[5] Taesang Yoo and A. Goldsmith, "Capacity and Power Allocation for Fading MIMO Channels with Channel Estimation Error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.

[6] R. W. Heath and D. J. Love, "Multimode Antenna Selection for Spatial Multiplexing Systems with Linear Receivers," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3042–3056, 2005.

[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[8] C. P. Schnorr and M. Euchner, "Lattice Basis Reduction: Improved Practical Algorithms and Solving Subset Sum Problems," *Math. Program.*, vol. 66, no. 2, pp. 181–199, Sept. 1994.

[9] L. G. Barbero and J. S. Thompson, "Fixing the Complexity of the Sphere Decoder for MIMO Detection," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2131–2142, 2008.

[10] X. Chu and J. McAllister, "Software-Defined Sphere Decoding for FPGA-Based MIMO Detection," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6017–6026, 2012.

[11] B. M. Hochwald and S. ten Brink, "Achieving Near-Capacity on A Multiple-Antenna Channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.

[12] V. Bhojak and A. Sharma, "MIMO Wireless Systems: V-BLAST Architecture," in *2013 Third International Conference on Advanced Computing and Communication Technologies (ACCT)*, April 2013, pp. 215–220.

[13] F. Kharrat-Kammoun, S. Fontenelle, and J. J. Boutros, "Accurate approximation of qam error probability on quasi-static mimo channels and its application to adaptive modulation," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1151–1160, 2007.

[14] M. Hosseinabady and J. L. Nunez-Yanez, "Run-time power gating in hybrid arm-fpga devices," in *2014 24th International Conference on Field Programmable Logic and Applications (FPL)*, 2014, pp. 1–6.

[15] X. Chen, J. Li, J. Ma, J. Wang, and G. He, "A low complexity soft-input soft-output fixed-complexity sphere decoding algorithm," in *2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*, 2012, pp. 1–4.

[16] K. Manolakis, M. A. Gutierrez-Estevez, and V. Jungnickel, "Adaptive modulation and turbo coding for 3gpp lte systems with limited feedback," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, 2014, pp. 1–5.

[17] Afif Osseiran, Jose F. Monserrat, and Patrick Marsch, *5G Mobile and Wireless Communications Technology*, Cambridge University Press, USA, 1st edition, 2016.

[18] Y. Wu, J. Nunez-Yanez, R. Woods, and D. S. Nikolopoulos, "Power modelling and capping for heterogeneous arm/fpga socs," in *2014 International Conference on Field-Programmable Technology (FPT)*, 2014, pp. 231–234.