

An energy-aware system-on-chip architecture for intra prediction in HEVC standard

El Ansari Abdessamad¹, Anas Mansouri², Ali Ahaitouf³

^{1,2}Faculty of Sciences and Technology, LERSI, Laboratory, University of Sidi Mohammed Ben Abdellah Fez, Morocco

³National School of Applied Sciences, LERSI, Laboratory, University of Sidi Mohammed Ben Abdellah Fez, Morocco

Article Info

Article history:

Received May 1, 2018

Revised Jun 22, 2019

Accepted Jul 2, 2019

Keywords:

FPGA

Hardware

HEVC

Intra prediction

Video coding

ABSTRACT

High resolution 4K and 8K are becoming the more used in video applications. Those resolutions are well supported in the new HEVC standard. Thus, embedded solutions such as development of dedicated systems-On-Chips (SOC) to accelerate video processing on one chip instead of only software solutions are commendable. This paper proposes a novel parallel and high efficient hardware accelerator for the intra prediction block. This accelerator achieves a high-speed treatment due to pipelined processing units and parallel shaped architecture. The complexity of memory access is also reduced thanks to the proposed design with less increased power consumption. The implementation was performed on the 7 Series FPGA 28 nm technology resources on Zynq-7000 and results show, that the proposed architecture takes 16520 LUTs and can reach 143.65 MHz as a maximum frequency and it is able to support the throughput of 3840×2160 sequence at 90 frames per second.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

El Ansari Abdessamad,
Sidi Mohammed Ben Abdellah University,
Faculty of Sciences and Technology of Fez,
Laboratory of Renewable Energy and Smart Systems (LERSI),
P.O. Box 2202 Fez, 00212626243355, Morocco.
Email: abdessamad.elansari@usmba.ac.ma

1. INTRODUCTION

High efficiency video coding (HEVC) is the last video coding standard recently released and developed under the efforts of Joint Collaborative Team on Video Coding (JCT-VC) under the ISO/IEC MPEG (Moving Picture Expert Group) standardization organization and the ITU-T VCEG (Video Coding Expert Group). It integrated different algorithms of the video coding such as inter, intra prediction, entropy coding and filters. It achieves good performance improvement with an increase computational complexity compared with its predecessor H.264/AVC.

Among the important feature tools implemented in HEVC, the new proposed variable size block prediction unit (PU) up to 64×64 dedicated to support various high-resolution video in opposite to the fixed macroblock 16×16, existing in its predecessor H.264/AVC. Secondly, the intra prediction which is one of the important and more time consuming block of the HEVC video coding standard, used to compute the pixels predicted inside the PUs. In the new standard, it leads to achieve better efficiency, than defined in the previous standards video coding even if these improvements make the algorithm execution significantly complex [1, 2]. In fact, the intra prediction exploits the spatial redundancy existing only in a single frame, it has several prediction models (up to 35 modes) and it computes the large PU block (up to 64×64 pixels). This way, each intra prediction block is computed by all prediction modes and the final choice is based on the smallest prediction cost, computed for all modes. This prediction is performed for both Chroma and Luma mode decision.

One of the biggest benefits of using field programmable gate array (FPGA), to implement all or partial parts of the HEVC encoder, is to make available a time measuring unit inside the whole processing system using only a single FPGA. In fact the complexity analyses of HEVC hardware implementation needs a rigorous time consuming to be considered in the profiling due the high algorithms complexity of HEVC encoder, and can be helpful for the real time processing check. One of the last generation FPGA from Xilinx is Zynq-7000. It offered a software processing system on dual-core ARM cortex-A9 and a hardware processing system on Programmable Logic (PL) including on FPGA in a single device. It is thus particularly suitable to study the possibility for the implementation of the HEVC within this family.

Since the first software version of HEVC standard was proposed, many software and hardware implementations have been developed to offer statistical analyses concerning the execution time, power consumption and complexity analyses in order to reach real time processing for the higher resolutions like 4K and 8K. In This work, we focused on a hardware implementation of the Intra prediction block of the HEVC encoder on the Field-Programmable Gate Array (FPGA).

The intra prediction algorithm computes all predicted pixels using several modes. The intra prediction algorithm is inserted in both JM and HM reference software encoder for the standards H.264/AVC and HEVC respectively. Software profiling of all functions used for the intra mode prediction that allows for a comparison between the complexities of all functions in the whole encoder HEVC is presented in the work.

Most of portable devices supported applications that are dedicated for multimedia like streaming from online services. Lot of these devices are heterogeneous embedded systems and are designed in spirit of low power consumption. The Zynq System On Chip (SOC), from Xilinx, is one of these devices that combine heterogeneous embedded system including in the same time a Programmable Logic (PL) and Processing System (PS) on one Chip with various system of communications with them, consequently, lot of reported work in the literature give the computational complexity in this Chip. The authors Peng et al. [3] proposed and implemented a highly integrated accelerating solution for N - body MOND simulations on various processors like ARM and GPU, and on the FPGA Zynq-7020 SoC integrated using both simplified pipeline and bandwidth requirement techniques. Their proposed solution on the SOC has a good performance in power consumption up to 10 times and the best performance in terms of cost over 50% [4, 5].

Remainder of this paper is organized as follows. Section II gives a related work of intra prediction in HEVC, section III describes the intra prediction algorithm in video encoding HEVC standard. Section IV is dedicated to the proposed hardware architecture for every mode. The synthesis and performance results obtained in the FPGA and ASIC with comparison to other works and presented in section V. Finally, the conclusion and future works are given in Section VI.

2. RELATED WORK

Many research works have been published in the literature in order to propose an efficient hardware architecture only for intra prediction in the previous and last HEVC standard [6–14], however, a study of this algorithm on system on Chip has not yet been reported in our knowledge. For hardware implementations several works have proposed an accelerating design for intra prediction, implemented on FPGA-based technology in [6-11], or on an ASIC chips in [12], or in both ASIC and FPGA in [13, 14]. Some of these designs are proposed to support all mode decision and all PU block sizes.

In [6], the hardware implementation of HEVC decoder including the intra prediction on FPGA is presented, it can be used for decoding in real time 4K resolution with 30 frame per second (fps). The authors used optimization available on FPGA on the pipeline stages, therefore, they use a single cycle reference pixel processing in the intra prediction and another techniques for blocks transform and both filter inside HEVC decode, so that lead to a high throughput.

in [7], a multiple techniques are proposed so as to make hardware accelerator developed for intra prediction of HEVC functioning in full pipeline. The three techniques adopted are novel buffer structure for reference samples, mode dependent scanning order and an inverse method for reference samples extension, that allowed to process 4 pixels per clock cycle, therefore, the throughput of this architecture can support Quad Full HD (3840×2160) at 30 frames/s. and also the accelerator can compute all intra prediction modes.

In [8], a hardware for intra prediction is presented for processing all modes, it is based on two techniques to reduce the computation complexity and to accelerate treatment. The first is adopted for angular modes and it called Processing Element for Angular (PEA) modes and the second is called Processing element for planar (PER) mode, then the design is structured in repeated paths in order to compute in parallel way. Therefore, the architecture can process 1080p@100 or 4K@24 in real time.

In [9], the hardware accelerators for each size block from 4×4 to 32×32 and a design are presented with these synthesis results. The hardware is proposed on VHDL language and it implemented on the Xilinx Virtex-6 that is manufactured at the 40 nm technology node. The design takes 170K LUTs and 110K

registers without any block of memory from available resources. After The synthesis design, the maximum frequency up to 219 MHz and is capable to compute 24fps for 4K resolution.

Jiang et al. [10] developed their work in a pipelined way and implemented it on a Xilinx Virtex 5 device. It consumed 69k LUTs, 64 internal DSPs and 148k bit internal memories type BRAM. Kahn et al. in [11] proposed a pipelined design that contain the parallel processing Elements (PE) supporting all PU sizes and modes decision, it achieves 213 MHz and can work in real time constraints achieving 120 fps for 1080 resolution and 30fps for the 4k resolution.

In paper [12], a VLSI architecture is presented for intra prediction in HEVC standard. The design adopted three techniques to reduce systematically the complexity. The first is the integrated hierarchical memory instead of gates as registers used for storing the neighboring samples that can increase the throughput. The second is a mode-adaptive scheduling scheme that lead to provide at least 2 samples/cycles. The three is reducing the multipliers by sharing them in the proposed architecture. Using this technique can reduce the consumption area, but, these three techniques consume more power due to employed SRAM memory and the post layout simulation shows a power consumption up to 2.11 mW. The design is synthesized at 200 MHz using ASIC technology 40nm process and can supported the resolution 3840×2160@30 in real time processing.

In [13], both techniques pixel equality and pixel similarity are used in order to reduce the amount of the performed calculation by the intra prediction algorithm in HEVC. Consequently, the energy consumption is also reduced. The hardware designs performed only 4×4 and 8×8 angular prediction modes and it implemented on both Xilinx Virtex 6 FPGA with 150 MHz of frequency, and on ASIC technology using 90 nm standard cell library from Synopsys, therefore, they can process 30 frame per second for full HD (1920×1080) resolution for two cases.

In [14], a symmetric propriety of intra prediction equations of the horizontal and vertical directions is exploited, that lead to limite study one direction and it change the order of the selected references samples for the other direction. The hardware design is implemented on both Virtex 6 FPGA from Xilinx that works at 234MHz and TSMC 180 nm CMOS process technology with the frequency operation is 218 MHz. The authors propose an architecture for 4×4, 8×8, 16×16 and 32×32 angular prediction modes only on the HEVC standard, their proposed architecture is based on DSP block inside in FPGA (Xilinx XC6VLX75T FF1759 FPGA). their optimization, lead to 34.66% less energy consumption than the original FPGA implementation (i.e. without implementation on the DSP block) of HEVC intra prediction and this architecture, hardware can be processed 55 full hd 1080p (1920×1080) frame per second.

3. HEVC INTRA PREDICTION ALGORITHM

3.1. Overview of HEVC encoder

A description of the whole HEVC encoder block diagram is given in Figure 1. The main modules are the same into previous standards, but HEVC comes with flexibility coding block against existing fixed and less complexity modules proposed for HEVC encoder is known by high complexity than declared standards. Generally, it receives an input YUV frames types (or GOP [15]) and generates after treatment a bitstream data in its output sides. A brief description of these modules inside HEVC encoder is giving in this work, and for more details reader can return, for instance, to ref. [2].

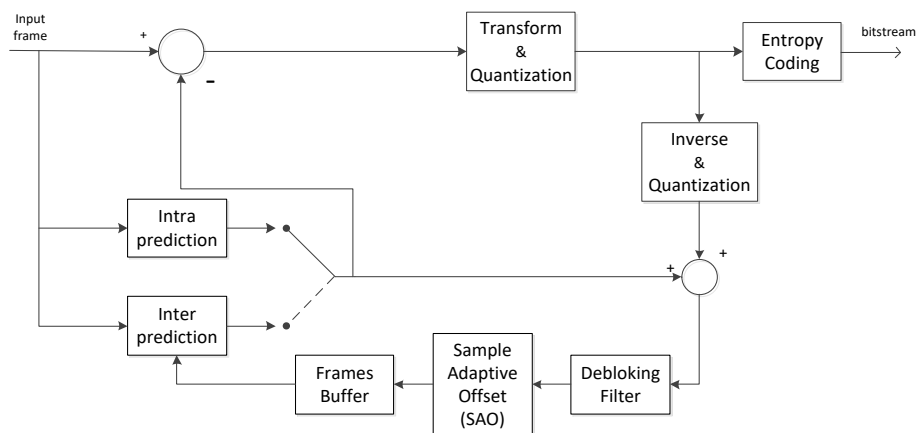


Figure 1. Block-diagram of the HEVC Encoder with the main modules

The several blocks are: 1) Inter prediction: the Inter prediction is used to reduce exchange between successive pictures in the YUV sequence by determination of the predicted block for the current picture available in the decoded picture buffer. It is processing in first by motion compensated prediction, based on motion vectors and it followed by sub-sample Interpolation filtering. 2) The Transform and inverse transform are referenced to TU to determine the size of matrix ranging from 4×4 to 32×32 to be used to treat the residual matrices after prediction process. 3) The quantization and inverse quantization are also similar to existing in the old standard H.264/AVC. This operation scales up to encode coefficients just after transform using by entropy coding and to decode them for filter block. 4) Entropy Encoding: The Entropy encoding used to decode the input bitstream Context Adaptive Binary Arithmetic Coding (CABAC), which is the only used technique for this decoding procedure, that is same as founded in the high profile into H.264/AVC [16]. 5) Deblocking filter & Sample-adaptive Offset: The Deblocking Filter and Sample-Adaptive Offset are two filters designed to improve coding efficiency after the HEVC processors and they remove the edge essentially affected by decision modes on one picture, which is stored in a buffer for use it in the calculation of the motion estimation. 6) The intra prediction modes in HEVC are fixed in 35 modes, which are DC, planar and 33 angular modes. The DC and planar are named mode 0 and 1 respectively. The prediction mode 2 to 18 are the first is the half-party horizontal angular prediction modes, and the second is vertical modes from 19 to 34. The intra prediction is applied to mode Luma and Chroma blocks, which are sized from 32×32 down to 4×4 pixels. Therefore, the equations perform the HEVC intra prediction algorithm that is implemented in HM (HEVC test model) of HEVC reference encoder is given below with each mode in [17].

3.2. Coding structure

Currently, the HEVC is working in highly flexible and efficient block coding portioning structure which is divided into four levels, giving as follows: CTU, CU, PU, and TU. First, each frame is divided into blocks named coding tree unit (CTUs) that have various sizes into 16×16 , 32×32 or 64×64 . The CTU includes one coding tree block (CTB) for Luma and two CTBs for component Chroma. The CTU is similar to fixed Macroblock 16×16 existing in the precedent standard H.264/AVC. Then, the coding unit (CU) is under CTU and it can be split into four depths as shown in Figure 2, it is sized a square region 8×8 , 16×16 , 32×32 and 64×64 pixels depending on resolution of sequences coding. Next, The CU itself can be divided in Prediction Unit (PU). This block can be used for the intra and inter block prediction with sizes ranging from 4×4 down to 64×64 according to mode decision. Finally, the transform Unit (TU) is determined from PU and it suitable for transform and quantification.

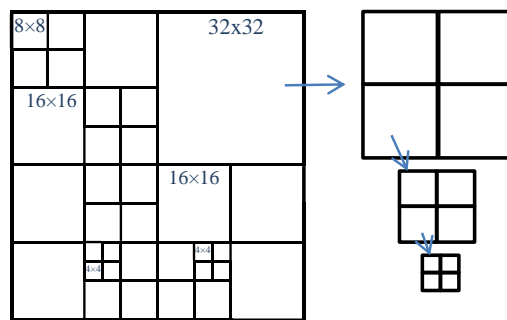


Figure 2. Example of CU in coding tree structure (CTU)

3.3. DC modes

The DC modes is dedicated to images that have a little change between the neighboring pixels, this mode is more convenient to generate a natural prediction block [1]. The DC mode is based on all neighboring vertical and horizontal adjacent pixels. The prediction is calculated using eq. (1) for the DC mode of pixel (x, y) and eq. (2) for the three pixels $(0,0)$, $(x,0)$ and $(0,y)$. Those equations are given as follows:

$$\begin{aligned}
 p_{\text{pred}}(x, y) & \text{ is the value of the prediction sample pixel with } x, y = 0 \dots N_c - 1 \\
 p_{\text{ref}}(x, y) & \text{ is the value of the neighbouring sample pixel with } x, y = 0 \dots N_c * 2 - 1 \\
 \text{dcVal} & = \left(\sum_{x=0}^{N_c-1} p_{\text{ref}}(x, -1) + \sum_{y=0}^{N_c-1} p_{\text{ref}}(-1, y) \right) \gg (\log_2(N_c) + 1) \\
 p_{\text{pred}}(x, y) & = \text{dcVal}
 \end{aligned}
 \tag{1}$$

With $N_c = nT_b$ s presented size of the transform block.

$$p_{\text{pred}}(0,0) = (p_{\text{ref}}(-1,0) + 2 * dcVal + p_{\text{ref}}(0, -1) + 2) \gg 2$$

$$p_{\text{pred}}(x, 0) = (p_{\text{ref}}(x, -1) + 3 * dcVal + 2) \gg 2$$

$$p_{\text{pred}}(0, y) = (p_{\text{ref}}(-1, y) + 3 * dcVal + 2) \gg 2 \quad (2)$$

3.4. Planar mode

The HEVC adopted also other mode named Planar in order to solve the problem caused by the angular prediction modes that have a bad results when the PUs are samples path not related to any direction. The planar mode is computed based on four sample reference pixels and giving by equation (3):

$$p_{\text{pred}}(x,y) = ((N_c - 1 - x) * p_{\text{ref}}(-1, y) * p_{\text{ref}}(N_c, -1) + (N_c - 1 - y) * p_{\text{ref}}(x, -1) + (y + 1) * p_{\text{ref}}(-1, N_c) + N_c) \gg (\log_2(N_c) + 1) \quad (3)$$

Where $\text{pred}(x,y)$ is the predicted sample corresponding to location (x,y) and N_c is the prediction unit size and the value of x and y are from 0 to N_c-1 .

3.5. The angular modes

The angular prediction modes are provided for complex texture and high frequency component, it is applied to 33 different directions that used the neighboring pixels, these directions are horizontal (2 to 18) including horizontal mode (10) and vertical (19 to 34) also including the vertical mode (26), These directions are given in Figure 3. Each sample is predicted using projection of its position on the set reference samples array, so the prediction samples is calculated by the following equations:

$$ildx = ((x + 1) * \text{IntraPredAngle}) \gg 5 \quad (5)$$

$$iFact = ((x + 1) * \text{IntraPredAngle}) \& 31 \quad (6)$$

$$\text{predSample}[x][y] = ((32 - iFact) * \text{ref}[y + \text{ildx} + 1] + iFact * \text{ref}[y + \text{ildx} + 2] + 16) \gg 5 \quad (7)$$

Where predsample is the predicted pixel attached to position (x,y) . And intrapredAngle is related to the intra prediction mode or direction, $\text{ref}[y + \text{ildx} + 2]$ corresponds to a reference simple array, this equation process the vertical modes, to compute the horizontal modes, we change the position of x by y and vice versa.

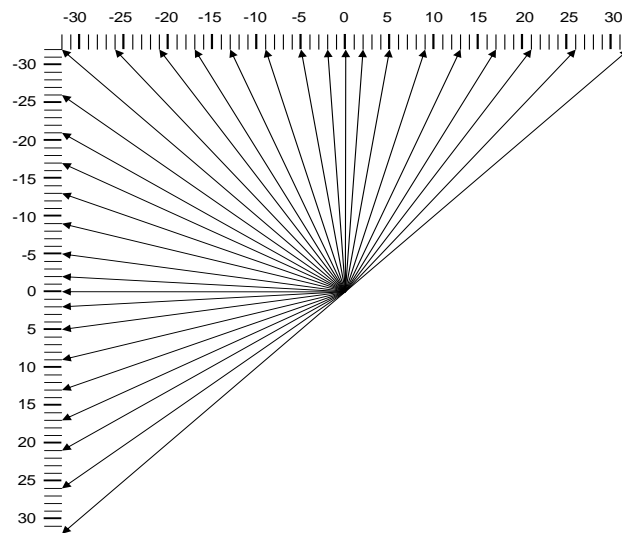


Figure 3. The different direction for 33 angular modes direction

4. PROPOSED ARCHITECTURE

The proposed design supports all sizes block defined for intra prediction in HEVC standard, as 4×4 to 32×32 coming from the size range of PU that are 64×64 down to 4×4. The overall view of the proposed architecture is presented in Figure 4. It is composed of 4 major blocks with a control unit. The first is the block dedicate for memory, this architecture leads to loading and storage of the neighboring pixel from one part, and to storage the pixel coming from the unit prediction process for from the other part.

Shown in Figure 5, the second block *id* is the DC intra prediction mode. It consists of adders and shifter to perform calculation of (1) and (2). Then any reduction of the computational complexity coming with this architecture can be exploited for the huge size block treatment.

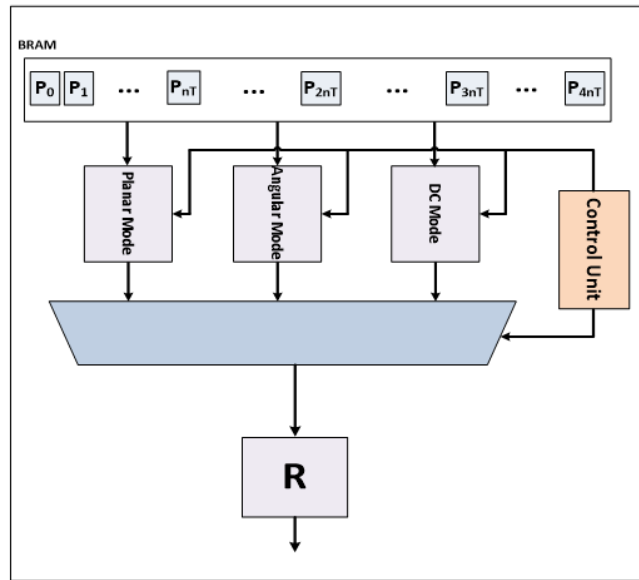


Figure 4. The global architecture for computing the intra prediction. Algorithm in HEVC encoder

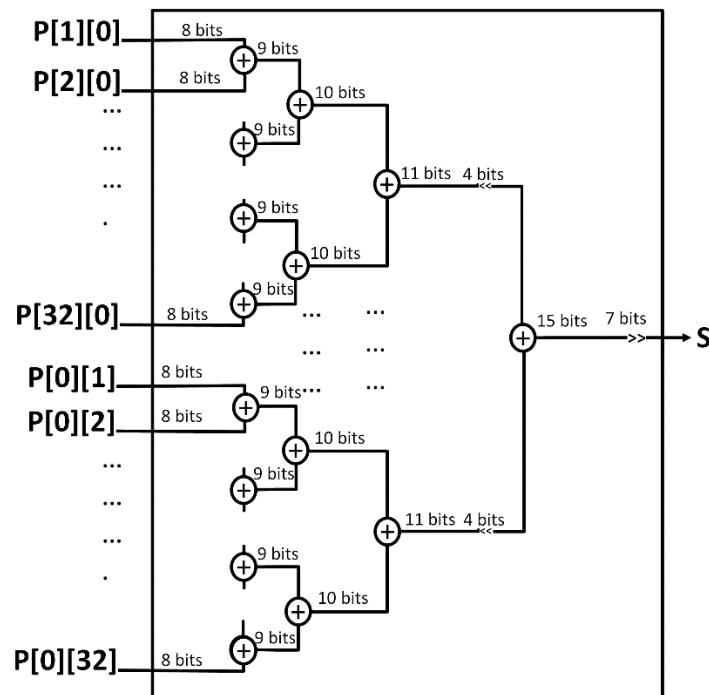


Figure 5. The proposed architecture for Intra prediction DC processing with 64 input samples

The two last blocks concern planar and angular mode prediction. For these two modes, we use the processing unit presented in Figure 6. This design performs calculation described by (3) to (7) and it consists of multiplexer, adders and two PE. The input data are received from the neighboring pixels and results S_1 and S_2 are supplied respectively for the planar and angular mode.

The control unit shown in Figure 4 is added in order to manage the global system of the whole design. It includes the organization of samples inputs and outputs for each case of sizes PUs blocks. also it applied in loop after clock cycles computed by a counter. It also used to control the design not shown in Figure 4 to select the best mode comparison between all predicted samples.

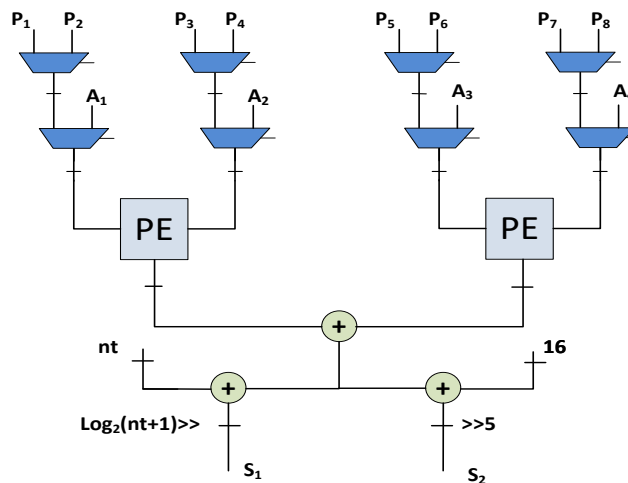


Figure 6. The proposed hardware accelerator processing element (PE) for computing the both Intra prediction planar and angular

As shown in Figure 7, the reference pixels values management for intra prediction processing, all reference samples are from the above and left of the existing PU, that are stored in memory Block RAM (BRAM) instead of registers, which can reduce the complexity associated to access a data, even if the power has little increased. The reference samples are organized in such a way that are available in one array, with the first available reference sample at the left are represented by the reference sample vertical and the second part available right by the horizontal reference samples.

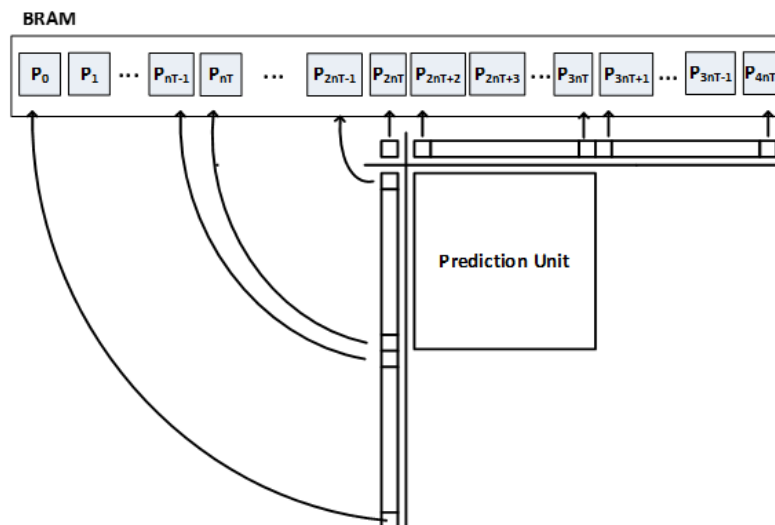


Figure 7. Placement the neighbouring pixels in the RAM memory. P_x indicates the pixel with $x = 0, \dots, nTbS * 2 - 1$ and $nTbS$ is equal to the length times the width of the block

The Processing Part (PP) in Figure 5 is a 32 unsigned bits multiplier based on the Carry Save Adder proposed by Singh et al. in work [18]. Then we avoid using DSPs known by their high power consumption, even with some increase of the consumed area.

The global design in Figure 4 is designed in a way to have a higher throughput by exploiting the available pipeline and employed deep parallelism to achieve real time processing. Regarding only the rate requirements, it can be anticipated that both pipeline and deeper parallelism included in the intra prediction design attain the highest rate. Nevertheless, the stage or parallelism is only beneficial if many logic output paths in hardware architecture, and the new inputs samples are available after several clock cycles, however, the power consumption is lower when one consider the power requirement for the parallel architecture instead of pipeline that needs more resources area. Because of those abovementioned properties, it necessary to include the pipeline and parallelism stage on whole intra prediction design. The advantage of the pipeline is more detailed in work [19], which the authors are used a fast VLSI pipelined multiplierless fixed-point 8x8 DCT. These architectures can be improved when applied 2-stage pipeline than non-pipeline.

The solution used in this work is as follows: first stage is exploited four PEs, that is processed in the parallelism (has four input samples for each case of PE in the same time) for the purpose of predicted four output samples values, then pipeline stage is applied to compute all remaining predicted outputs samples on each sizes PUs. Our hardware architecture works in parallel with four level samples and takes only eight sample clock cycle, including the consumed time with loading and storing data between the external memories in order to have real time processing.

The proposed parallel processing for the small 4x4 intra prediction block as shown in Figure 5 and 6 is exploited in a pipelined procedure to process bigger block 8x8, 16x16 and 32x32. This allows an increase of the bit stream with a drawback of some increase in hardware resources by loading and processing data at the same time. The mode DC predicted all samples based on the design in Figure 5, which have as input the reference sample neighboring available in BRAM $p[i][0]$ and $p[0][j]$ with x_i and y_j , respectively. It proposed on tree to compute 32x32 size block and its small one.

5. EXPERIMENTAL RESULTS

The proposed design block for the intra prediction in HEVC is firstly coded in a high-level language C++ in order to validate and reduce the computational complexity. Secondly, this design is implemented in VHDL language, then the hardware architecture prototyped on Artex-7 FPGA, that is integrated on ZYNQ-7000 processor that is xc7z020-1clg400 from Xilinx and acceleration hardware was verified on Microzed. Table 1 shows the synthesis results of the global architecture proposed in Figure 4. The maximum frequency attained is 143 MHz. One block RAM was used for storage the neighboring pixels and also for the storage of all possible output intra prediction pixels results. The pixels generated after the proposed hardware intra prediction architecture processing are first compared to predicted pixels coming from the reference software HEVC HM.15.0 version. Then they also compared with software implemented on language C++ to validate the proposed architecture. All these comparisons show a good argument.

5.1. Resources analysis

The proposed architecture was synthesized using recent design environment (Xilinx Vivado 2016.3). The results are presented in Table 1. This architecture works up to 143.65 MHz maximum clock frequency and it consumes 16520 LUTs that represents 31.05 % of available resources in our device (53200 LUTs) and 1345 registers bits corresponding to 1.27% of 106400 registers resources, finally, it doesn't exploit any DSPs from the 220 DSPs existing in our FPGA. More details of the steps of the implementation in platform Vivado Integrated Design Environment (IDE) are given in the work [20].

The simulation in ModelSim demonstrates that our architecture can compute 32x32 block in 467 clock cycles and takes 6 cycles to process one Processing Element (PE). The proposed architecture can compute 30 frames per second (fps) for 4K resolution and 120 fps for the Full HD resolution. In fact, for the 4 K (3840 x 2160 pixels) resolutions, to achieve 30 frames per second with 4:2:0 sampling one requires (3840x2160x1.5x30) i.e 373.24 MSample/s as throughput, so our throughput is higher than this value. Moreover, this proposed hardware architecture can be improved to process more frames per second by exploiting the other existing more resources on the used device.

The processing time of hardware accelerate was computed by counting the total number of clock cycles taken to determine all possible output pixels for basketball sequence that takes into consideration the number of frames, resolution and all possible size block. The whole time is in relation to a clock frequency that is equal to 143.65 MHz.

5.2. Comparison of intra prediction modes algorithm implementations on FPGA.

Comparison our work on FPGA implementation of HEVC intra prediction with the proposed hardware on the FPGA in [21], it has less area and high frequency. however the performance the its architecture up to 55 frames per second only for resolution full HD against our up to real time processing for resolution 4K. Whole hardware architecture for decoder is presented by Abeydeera et al. [6]. In order to compare our performances for intra prediction, we consider only their intra prediction results presented in Table 2. The hardware accelerator is implemented on Xilinx Zynq 7045 with 28 nm technology process. The consumed resources area is 43K of LUTs and 22K Registers. The 94 BRAMs (BLOCK RAM) are occupied, in totally 18Kbits from the entire memory. The maximum frequency is up to 150 MHz. The design can be produced in the average 2.601 samples pixels per clock cycle, consequently, it can be up real time processing (30 fps) for the sequence 4K.

Table 2. Comparison of Synthesis Results for FPGA Implementation

	Our work	Abaydeer [6]	Biao [7]	Amish [8]	Angelhart [9]	Jiang [10]	Khan [11]	Hasan [21]
Technology (FPGA)	28 nm Artix-7	28 nm Artix-7	65 nm	40 nm Virtex-6	65 nm	65 nm Virtex-5	40 nm Virtex-6	40 nm Virtex-6
LUT/ALM ^a	16.52K	43K	14K	170K	8.8K	69k	140k	4.4k
Registers	1.34K	22K	5.5K	110K	13.K			1.1K
Max frequency	143Mhz	150MHz	110Mhz	219 MHz	110MHz	204Mhz	213Mhz	227Mhz
Frame rate	4K@30fps	4K@30fps	4K@30fps	4K@24fps	2160@30fps	--	4K@30fps	1080@55fps
memory	36Kb	18Kb	6K	--	96Kb	148kb	150k REG	--
PUs	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

^aLook-up Table (LUT) for Xilinx FPGA and (Adaptive Logic Module (ALM) for devices from ALTERA INC.

The related references in state of art are shown in Table 2. That includes lists key performance metrics of circuit architectures for intra prediction as FPGA technology, consumption of LUTs or ALMs for Xilinx devices or for ALETRA devices respectively, utilization of registers, the number resolution frames can be processing per second, and also the PUs size is reached. Our work and all references are the same, since all variable PUs size for intra prediction are supported. Due to different fabrication technology that related the maximum frequency reached for each hardware design on different FPGA platforms should be not compared, therefore, the throughput is included in calculation. For the throughput, all hardware can be process in real time 4K video frames except in [8], which can be process 24 frames/s. Both this work and [6] have the same memory occupied in the design, but [7] achieved minimum size memory (6 kbits) in its circuit, however, in [9–11] have a higher resource of memory, consequently, they design will consume more energy than others. Our design reaches a lower frequency thanks to the technology used and also due to the propose architectural design, that can lead to higher bit rate. So, our proposed architecture can achieve high throughput up to 3 times compared with the best results in [11] and can compute all bocks sizes like in [11], with a less increase in the needed area and memory caused by the reference sample data.

6. CONCLUSION

In this paper, a novel efficient accelerator for intra prediction of HEVC is presented, our hardware design supported all modes for intra prediction in HEVC, and it includes both parallel and pipeline techniques in order to reduce the computational complexity and to have high efficient throughput. The architecture is implemented in the Xilinx Zynq-7000 Artix-7. The FPGA Synthesis results show that the proposed hardware can reach 143 MHz of the maximum clock frequency and can process in real time the high resolution video 4K.

In future work, we plan to add the blocks more complex such as inter prediction and entropy encoding in or order to achieve real time processing for co-design on both system software and hardware instead of co-simulation with data store already in memory. Furthermore, details about power estimation based on system outside device and inside so as to estimate the consumption energy more than estimate by Xilinx tool.

REFERENCES

- [1] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra coding of the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1792–1801, 2012.
- [2] F. Bossen, B. Bross, K. Suhling, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1685–1696, 2012.
- [3] B. Peng, T. Wang, X. Jin, and C. Wang, "An Accelerating Solution for-Body MOND Simulation with FPGA-SoC," *Int. J. Reconfigurable Comput.*, vol. 2016, 2016.
- [4] "Zynq-7000 All Programmable SoC," 30-Jan-2017. [Online]. Available: <http://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html>. [Accessed: 30-Jan-2017].
- [5] L. H. Crockett, R. A. Elliot, M. A. Enderwitz, and R. W. Stewart, *The Zynq Book: Embedded Processing with the Arm Cortex-A9 on the Xilinx Zynq-7000 All Programmable Soc*. Strathclyde Academic Media, 2014.
- [6] M. Abeydeera, M. Karunaratne, G. Karunaratne, K. De Silva, and A. Pasqual, "4K real-time HEVC decoder on an FPGA," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 236–249, 2016.
- [7] B. Min, Z. Xu, and R. C. Cheung, "A Fully Pipelined Hardware Architecture for Intra Prediction of HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, 2016.
- [8] F. Amish and E.-B. Bourenane, "Fully pipelined real time hardware solution for High Efficiency Video Coding (HEVC) intra prediction," *J. Syst. Archit.*, vol. 64, pp. 133–147, 2016.
- [9] D. Engelhardt, J. Moller, J. Hahlbeck, and B. Stabernack, "FPGA implementation of a full HD real-time HEVC main profile decoder," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 476–484, 2014.
- [10] W. Jiang, H. Ma, and Y. Chen, "Gradient based fast mode decision algorithm for intra prediction in HEVC," in *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*, pp. 1836–1840, 2012.
- [11] M. U. K. Khan, M. Shafique, M. Grellert, and J. Henkel, "Hardware-software collaborative complexity reduction scheme for the emerging HEVC intra encoder," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013*, pp. 125–128, 2013.
- [12] C.-T. Huang, M. Tikekar, and A. P. Chandrakasan, "Memory-hierarchical and mode-adaptive HEVC intra prediction architecture for quad full HD video decoding," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 22, no. 7, pp. 1515–1525, 2014.
- [13] E. Kalali, Y. Adibelli, and I. Hamzaoglu, "A low energy intra prediction hardware for high efficiency video coding," *J. Real-Time Image Process.*, pp. 1–14, 2014.
- [14] M. Kammoun, A. B. Atitallah, and N. Masmoudi, "An optimized hardware architecture for intra prediction for HEVC," in *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*, pp. 1–5, 2014.
- [15] M. Jeon and B.-D. Lee, "Toward Content-Aware Video Partitioning Methods for Distributed HEVC Video Encoding," *Int. J. Electr. Comput. Eng. IJECE*, vol. 5, no. 3, pp. 569–578, 2015.
- [16] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [17] I.-K. Kim, K. McCann, K. Sugimoto, B. Bross, and W.-J. Han, "Hm9: High efficiency video coding (HEVC) test model 9 encoder description," in *Proc. 9th JCT-VC Meeting*, pp. 6–11, 2012.
- [18] R. P. P. Singh, P. Kumar, and B. Singh, "Performance analysis of 32-bit array multiplier with a carry save adder and with a carry-look-ahead adder," *Int. J. Recent Trends Eng.*, vol. 2, no. 6, pp. 83–86, 2009.
- [19] N. M. Zabidi and A. A.-H. Ab Rahman, "VLSI Design of a Fast Pipelined 8x8 Discrete Cosine Transform," *Int. J. Electr. Comput. Eng. IJECE*, vol. 7, no. 3, pp. 1430–1435, 2017.
- [20] A. Rani and N. Grover, "An Enhanced FPGA Based Asynchronous Microprocessor Design Using VIVADO and ISIM," *Bull. Electr. Eng. Inform.*, vol. 7, no. 2, pp. 199–208, 2018.
- [21] H. Azgin, A. C. Mert, E. Kalali, and I. Hamzaoglu, "An efficient FPGA implementation of HEVC intra prediction," in *Consumer Electronics (ICCE), 2018 IEEE International Conference on*, pp. 1–5, 2018.

BIOGRAPHIES OF AUTHORS



El Ansari Abdessamad was born in Morocco in . In 1987. he received his master from the faculty of science and technology (FST) of Fez Morocco, in 2012, he joined the team microelectronics and embedeed systems of the laboratory of energies and smart system. His research interests include video coding and implementation of video standard on embedded processor and VLSI architecture.



Anas Mansouri received M.S. and Ph.D degrees in Microelectronics and Telecommunication from Faculty of sciences \& technology, Fes, MOROCCO, in 2005 and 2009, respectively. He is a Assistant Professor in National School of Applied Sciences, Fes. His major research interests include VLSI and embedded architectures design, video and image Processing.



Pr. Ali Ahaitouf, teacher and researcher at the Faculty of Science and Technology of the University of Sidi Mohammed Ben Abdellah of the University of Fes (Morocco). He obtained his doctorate degree in electronics in Fès in 1998 and the Ph.D. in Metz University (France) in 1999. He is director of the Renewable Energies and Intelligent Systems Laboratory and head of the research team microelectronics and embedded systems. His research domain include microelectronics, digital and analog design of integrated circuits, image and data compression, and solar cells. He has led numerous multilateral research projects related to the optimization of analog design, characterization and optimization of electronic components, and concentrated photovoltaic (CPV). He has published a hundred or so articles in journals and conferences and supervised a dozen doctoral theses.