J Real-Time Image Proc (2019) 16:547–557
DOI 10.1007/s11554-016-0572-4

ORIGINAL RESEARCH PAPER

Design and implementation of an efficient hardware integer motion estimator for an HEVC video encoder

Estefania Alcocer[1] · Roberto Gutierrez[2] · Otoniel Lopez-Granado[1] · Manuel P. Malumbres[1]

Received: 30 September 2015 / Accepted: 26 February 2016 / Published online: 18 March 2016

Abstract High-Efficiency Video Coding (HEVC) was developed to improve its predecessor standard, H264/AVC, by doubling its compression efficiency. As in previous standards, Motion Estimation (ME) is one of the encoder critical blocks to achieve significant compression gains. However, it demands an overwhelming complexity cost to accurately remove video temporal redundancy, especially when encoding very high-resolution video sequences. To reduce the overall video encoding time, we propose the implementation of the HEVC ME block in hardware. The proposed architecture is based on (a) a new memory scan order, and (b) a new adder tree structure, which supports asymmetric partitioning modes in a fast and efficient way. The proposed system has been designed in VHDL (VHSIC Hardware Description Language), synthesized and implemented by means of the Xilinx FPGA, Virtex-7 XC7VX550T-3FFG1158. Our design achieves encoding frame rates up to 116 and 30 fps at 2 and 4K video formats, respectively.

Keywords HEVC · FPGA · Integer motion estimation · Inter-prediction · SAD architecture

✉ Estefania Alcocer
ealcocer@umh.es

[1] Physics and Computer Architecture department, Miguel Hernandez University of Elche, Alicante, Spain

[2] Communication Engineering department, Miguel Hernandez University of Elche, Alicante, Spain

1 Introduction

The High-Efficiency Video Coding (HEVC) standard is the most recent joint video project of the ITU-T VCEG and ISO/IEC MPEG standardization organizations, working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC) [1]. Previous video coding standards are currently used for many applications such as broadcast of High-Definition (HD) TV, video content acquisition, Internet and mobile video streaming, and real-time conversational applications. However, new video services with UltraHigh-Definition (UHD) formats are emerging, which need higher coding efficiency than previous standards. HEVC has been designed to deal with these demands, working with higher video resolutions and adapting its design to allow the use of parallel processing techniques. It can compress video about twice as much as its predecessor, H264/AVC, without sacrificing quality, providing video delivery with higher resolutions and frame rates, higher dynamic range, and a wider color gamut. Furthermore, HEVC contains new key features that are friendly with the use of parallel processing techniques [2].

As in previous video standards, Motion Estimation (ME) is one of the most critical modules in the video encoding process since it is able to efficiently remove the temporal redundancy between successive frames. However, the ME module is by far the most complex task of the encoder, requiring more than 90 % of the encoding time [3].

In HEVC, the complexity is even more critical due to several issues such as (a) a large set of Coding Tree Unit (CTU) partitioning modes, (b) the presence of multiple reference frames, and (c) the varying size of Coding Units (CU) in comparison with its predecessor H264/AVC. In addition, HEVC adopts Variable Block Size Motion Estimation (VBSME) to obtain advanced coding efficiency,

which comes at the expense of a huge increase of computational complexity.

For these reasons, several hardware architectures have been proposed to speed up the HEVC ME module, reducing the overall encoder complexity as much as possible. The Integer-pel Motion Estimation (IME) block is in charge of motion estimation and it is composed of (a) an integer motion search algorithm, and (b) a Rate/Distortion (R/D) optimization procedure that optimally reduces the temporal redundancy found at the video sequence. In most of the works found in the literature, the proposed IME hardware architectures are only focused on the motion search algorithm since it takes most of the computational complexity of the IME block. There are a lot of motion search algorithms that can be used to find the motion in video sequences. The most popular in hardware implementations is the Full Search (FS) algorithm. It follows greedy behavior by searching for motion at all points of the established search area of a reference frame, and, as a consequence, it is able to provide an optimal result (i.e., a motion vector that minimizes the residual error of the actual CTU).

The architecture proposals in [3, 4, 6, 7, 9] present an IME hardware block using FS strategy. In [3], a Sum of Absolute Differences (SAD) unit on a Field-Programmable Gate Array (FPGA) is proposed that is able to test all partition modes of a CTU except the set of asymmetric partition modes. Authors fixed a search area size lower than the one established by the standard, being able to run as fast as 30 fps at 2k video resolutions. The work presented in [4] proposes a SAD unit that computes all CTU partitions, achieving the same frame rates as previous work at 4k video formats. In their proposal, the search area has the same size as the maximum CTU, being implemented on an Application-Specific Integrated Circuit (ASIC). In [6], the maximum CTU size is reduced to 32×32 with a search area size of ± 23 pixels. This architecture is implemented on an FPGA device and achieves 30 fps at 1080p video resolutions. Different configurable search areas are studied in [7], achieving a maximum frame rate of 57 fps for a 720p video resolution. Several SAD units implemented on FPGA are described in [9], with different levels of parallelization, but no data about search area size, memory management, or how they obtain the minimum SAD are included.

On the other hand, [5, 8, 15] have proposed architectures which increase the throughput by limiting the number of searches in the reference frame. In [15], a motion estimation system for the HEVC encoder is presented. This design includes both integer-pel and fractional-pel motion estimation, achieving video encoding speeds of 1080p@60fps and 2160p@30fps when implemented over FPGA and ASIC technologies, respectively. The process in [15] is interrupted when the number of motion searches arrives at a limit fixed for a given resolution.

In addition, in [5] and [8], different implementations of suboptimal motion search strategies called fast ME algorithms, such as new Diamond Search (DS) or new Three Step Search (TSS), are shown. Similar hardware ME architectures have also been studied for the previous H264/AVC standard in [10–14], which are of interest for our work due to the high similarity of the IME block architecture in both standards.

Therefore, our purpose is to design a new hardware architecture that may perform IME computation in a fast and accurate way to significantly reduce the computation cost of the overall encoder. We will use FPGA technology, since it encourages design reuse and can greatly enhance the upgradability of digital systems. The programmability of FPGAs is particularly useful for highly flexible encoding systems that can accommodate a multitude of existing standards as well as the emergence of new ones [12].

Regarding the novelty of the proposed architecture, we present both innovative techniques: (a) a new SAD adder tree structure, and (b) a new memory scan order.

Firstly, we designed a new SAD adder tree structure to perform the additions at the first level of the tree, starting from the maximum size of the CTU, and halving the amount of additions at the next tree levels. This approach is different from the rest of state-of-the-art works, which divide a CTU into smaller blocks for consecutive accumulations, keeping the same additions in each step and thus requiring a higher number of steps to acquire all SADs. With our proposal, we took advantage of the resources provided by the FPGA, obtaining the minimum possible latency when calculating SADs of all levels and partitions for a CTU. In this way, SADs corresponding to asymmetric partitions are obtained in a fast and efficient way.

Secondly, regarding the new memory scan order, a series of reconfigurable shift registers and processing elements are responsible for storing the necessary pixels of both reference and current frames, keeping them always available for calculating the SADs and MVs of a CTU. With our system, we avoid external memory accesses since available data are highly reused by reconfiguring the displacement in a more efficient way.

The rest of the paper is organized as follows. Section 2 describes the HEVC ME module. Section 3 presents the architecture design of the proposed ME system while in Sect. 4, implementation results are provided in terms of hardware resources, time encoding, and R/D performance. Finally, in Sect. 5 some conclusions are drawn.

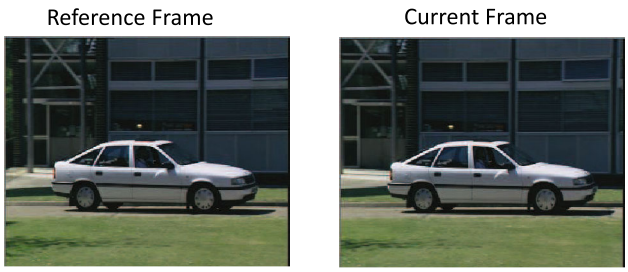
2 HEVC motion estimation

The motion estimation technique is based on the similarity between adjacent video frames, predicting the current frame based on a previous or subsequent reference frame in order of appearance.

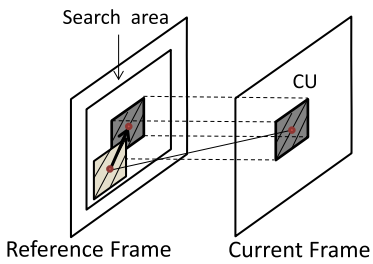
The Motion Vector (MV) represents the translational movement of a picture area in the current frame compared to its position in the reference frame. This movement is found inside a defined search area to bound the overall motion search complexity, as shown in Fig. 1.

In the ME process, each video frame is subdivided and partitioned into basic coding units called CTUs. The coding structure in HEVC consists of CUs with a maximum size of 64 × 64 pixels, as large as that of CTUs, which can be recursively divided into picture squares until achieving a block size of 8 × 8 pixels. Each CU consists of prediction units (Intra- or Inter-) and its size can vary from the maximum size of the CU to 4 × 4 pixels for Intra prediction, or to 4 × 8 or 8 × 4 for inter-prediction, supporting 8 partitioning modes as shown in Fig. 2. Prediction units of sizes 2N × 2N and N × N are called square motion partitions (square); 2N × N and N × 2N as Symmetric Motion Partitions (SMP); and 2N × nU, 2N × nD, nL × 2N, and nR × 2N as Asymmetric Motion Partitions (AMP). The total number of different partitions for a 64 × 64 CTU is more than 600, and for each of these partitions, the HEVC encoder performs one ME process to determine the best CTU partitions in terms of bit rate and video quality.

There are many kinds of algorithms for block-based IME. The most accurate strategy is the FS algorithm, which exhaustively finds motion for all prediction unit blocks at every single point of the established search area. Due to computational regularity and excellent video quality, FS motion estimation is commonly employed in hardware implementations [16]. Therefore, we will focus our work towards the design and hardware implementation of an FS



(a) Current and reference frames



(b) Obtaining motion vector

Fig. 1 Motion estimation

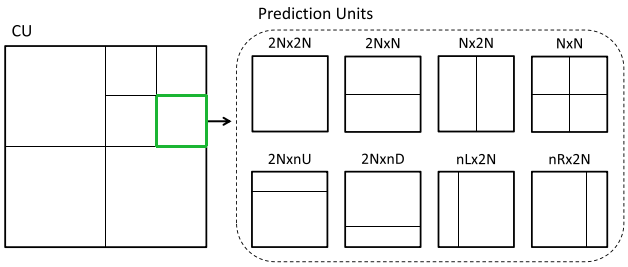


Fig. 2 Predictions units within a CU

algorithm that is able to significantly speed up the motion estimation process of the HEVC encoder without losing R/D performance.

3 Hardware architecture

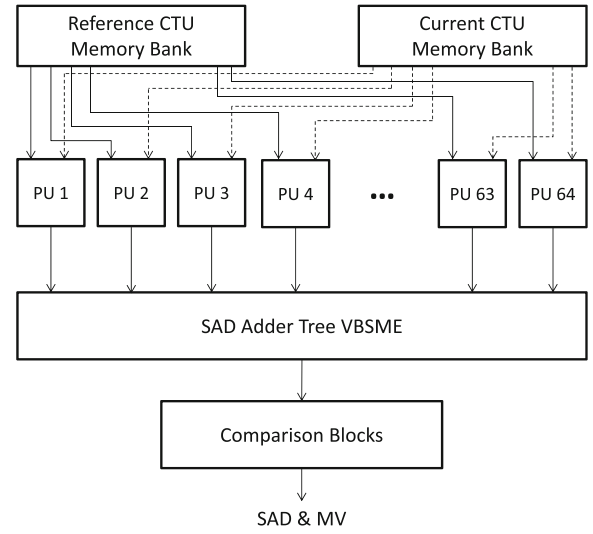
In this section, we present a high-performance IME hardware unit in HEVC that provides the minimum SADs and associated MVs of all possible partitions from a 64 × 64 CTU for inter-prediction, exploiting parallelism in an efficient way. The system is composed of memory areas for current CU and reference search area pixels, 64 Processing Units (PU), one SAD Adder Tree Block (SATB), and one comparison block that saves the minimum SAD values and their corresponding MVs for all CU partitions. Figure 3a shows the proposed hardware architecture.

As shown in Fig. 3b, one PU consists of 64 Processing Elements (PEs), where each PE computes the difference of both the current and the reference pixel (see Fig. 3c). So each PU calculates the distortion values of a column of 64 pixels. At each clock cycle, current and reference pixel columns are delivered to the 64 PUs, being able to compute the pixel distortion values of a 64 × 64 block (maximum CU size) just in one clock cycle, that is, all distortions needed to obtain the SAD of a 64 × 64 CU in a particular position of the search area are calculated in just one clock cycle. The next block in our system, SATB, computes the SADs for all the possible prediction units (more than 600) by properly grouping the 64 × 64 pixel distortions obtained before.

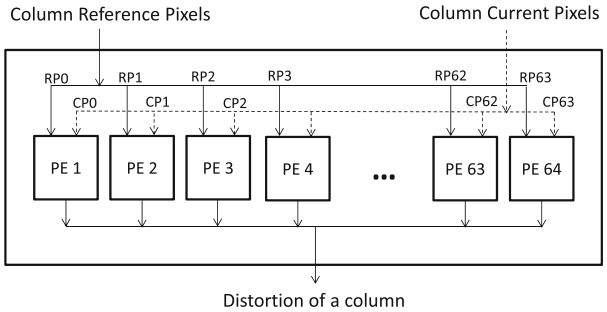
The process described above is performed for each of the positions of the search area, delivering the SADs to the comparison block, which is in charge of storing the minimum SADs with their corresponding MVs for each prediction unit of current CU. Table 1 lists the total number of different SAD partitions for a 64 × 64 CU.

3.1 Memory read controller block

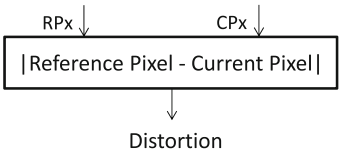
The memory read controller block is composed of a Block-RAM (BRAM) memory and a set of shift registers.



(a) Hardware block architecture



(b) Processing Unit (PU)



(c) Processing Element (PE)

Fig. 3 Proposed IME architecture

A BRAM consists on an embedded memory block within the FPGA. Pixels belonging to the search area of the reference frame are stored in the BRAM and current CTU pixels are saved in each PE. The reference pixels are spread from BRAM to the set of shift registers that are responsible for feeding PEs to calculate the distortion of the current CTU in a particular search area position.

The search area is just centered on the location of current CTU and the default search window spans ±64 pixels from the current CTU position, which defines a 128 × 128 search area as shown in Fig. 4, that is, the current CTU will be matched in 128 × 128 different pixel positions, being necessary to load on BRAM memory the pixels belonging to a reference frame area of 191 × 191 pixels.

To provide high data reuse, a snake scan order and a reconfigurable data path with 64 propagation registers are adopted. The snake scan order visits all positions of the search area following a Hamiltonian path composed by consecutive vertical scans with alternating directions (the first vertical scan begins from top to bottom, then moves one pixel to the right and starts the next vertical scan in a bottom to up direction, and so on) as illustrated in Fig. 4. So, there are three scanning directions U (upward), D (downward), and R (rightward).

The current 64 × 64 CTU pixels are stored in the PEs only once (at the beginning). The reference pixels will also be loaded to the PEs but instead of loading from BRAM will be loaded from the shift registers, since they will help us to perform the snake scan order and as a consequence a huge reduction of memory load operations will be achieved.

So, the memory controller will be the one that manages the shift registers set by loading rows of 64 pixels from the reference frame area (BRAM) and performing the shift register operations to cope with the snake scan order.

In Fig. 5, a diagram with the shift register set and the loading and shifting operations is shown. At the beginning, the register set is empty, so we have to perform several (64) load and shift operations before calculating the first SAD. As can be seen in Fig. 5, the first 64 clock cycles are dedicated to load the first 64-pixel rows starting from the left most upper position of the search area, following a downward (D) scan direction. In this figure, each 64-pixel row is labeled with the (x,y) pixel locations of the reference frame area. After loading the 64x64 reference frame block, all the pixels are sent to the PEs to compute the SAD in just one cycle (remember that the actual 64 × 64 CU pixels are already stored in the PEs waiting for this operation). At this point, the first SAD is computed. After that, we proceed in the D scan direction to compute the SAD of the next search area position. For this purpose, we only need to load an additional 64-pixel row in the D scan direction. So, in one clock cycle, (a) a right-shift operation takes place, discarding the first pixel row stored in the shift register 63, and (b) the new pixel row is loaded from BRAM in shift register 0. Then, the 64 × 64 pixels stored in the shift registers are sent to the PEs to compute a new SAD.

After computing the last SAD in the downward scanning direction, we have to change the scan direction from D to R, following the snake pattern described before. Moving the search area position one pixel to the right could be easy if we simply shift to the left one pixel in all shift registers (see Fig. 5 at the R scan direction). So, shift registers will

Table 1 Total number of SADs for each partition in a 64 × 64 CU

Block size	No. of SADs	Block size	No. of SADs
64 × 64 (2N × 2N)	1	32 × 32 (2N × nU)	8
64 × 64 (2N × N)	2	32 × 32 (2N × nD)	8
64 × 64 (N × 2N)	2	16 × 16 (2N × 2N)	16
64 × 64 (N × N)	4	16 × 16 (2N × N)	32
64 × 64 (nL × 2N)	2	16 × 16 (N × 2N)	32
64 × 64 (nR × 2N)	2	16 × 16 (N × N)	64
64 × 64 (2N × nU)	2	16 × 16 (nL × 2N)	32
64 × 64 (2N × nD)	2	16 × 16 (nR × 2N)	32
32 × 32 (2N × 2N)	4	16 × 16 (2N × nU)	32
32 × 32 (2N × N)	8	16 × 16 (2N × nD)	32
32 × 32 (N × 2N)	8	8 × 8 (2N × 2N)	64
32 × 32 (N × N)	16	8 × 8 (2N × N)	128
32 × 32 (nL × 2N)	8	8 × 8 (N ×2 N)	128
32 × 32 (nR × 2N)	8	Total	677

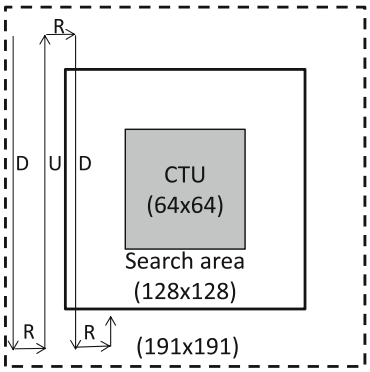


Fig. 4 Scan order of the search area

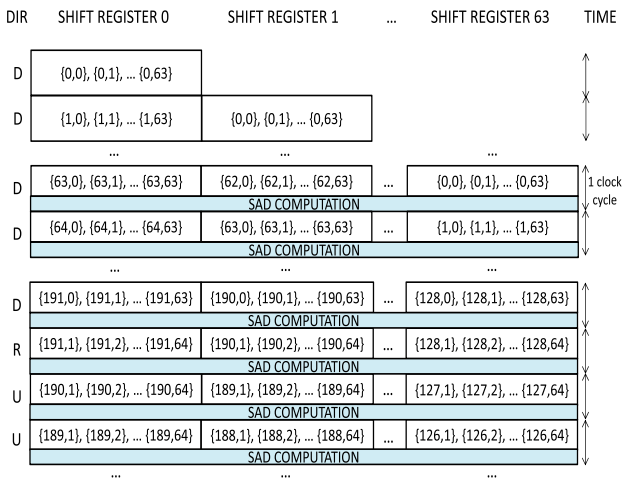


Fig. 5 Shift registers set: loading and shifting operations

contain the 64 × 64 search area block corresponding to the new position, and ready for the corresponding SAD computation.

After computing this SAD, we again change the scan direction from R to U, so we need to load a new 64-pixel row from BRAM, but now the loading is performed in the last shift register (63) and the register shift operation will be set to the left, discarding the contents of the first shift register (0).

The new SAD may now be computed, and as the scan direction is upwards, loading and shifting operations will be performed in the same way until a new change in scan direction is found.

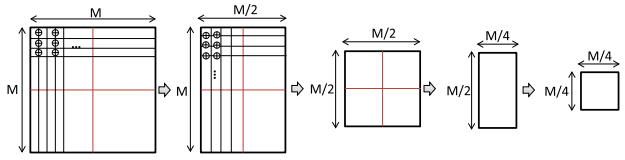
This procedure will iterate until all searching area positions have been processed, providing one SAD at every clock cycle to the next module in the proposed architecture, the SATB.

3.2 SAD adder tree block

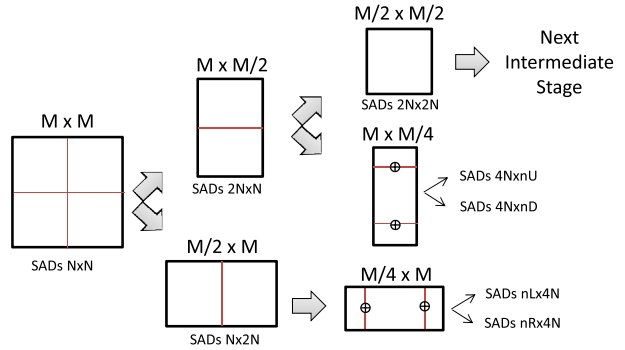
The SATB block is in charge of computing the SAD values for all partitions of each 64 × 64 CTU at every clock cycle. For inter-prediction, the HEVC standard proposes a partition size that ranges from 64 × 64 (maximum CU size) to 4 × 8/8 × 4 with different shapes—square, symmetric, and asymmetric partitions. After receiving the 64 × 64 distortions associated to the current search area position, a succession of aggregation stages are performed in this block to compute the corresponding SAD values for all the CTU partitions (a total number of 677), as shown in Fig. 6.

At the first stage, Fig. 6a, all pairs of consecutive distortion columns/rows of the input 64 × 64 SAD block (M = 64) are added, reducing the width/height of the resulting partition by one-half, until the block size of these added distortions is reduced to 16 × 16, from which the first SADs are obtained.

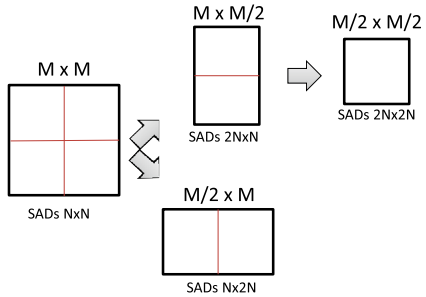
At the next three intermediate stages, a similar process to the one described above is followed. The successive



(a) First stage: M=64



(b) Three intermediate stages: M=16 and N=4; M=8 and N=8; M=4 and N=16



(c) Last stage: M=2 and N=32

Fig. 6 Structure of the SAD adder tree block

sums of different configurations (row–column, column–column, row–row) are performed to get the SADs of all partitions of a 64 × 64 CTU. For instance, in the first intermediate stage, starting with a 16 × 16 block of intermediate values (M = 16), all pairs of consecutive values for columns/rows are added as shown in Fig. 6b. So, both the 16 × 8 (M × M/2) and the 8 × 16 (M/2 × M) intermediate blocks, each one with 128 SADs, correspond to 2N × N and N × 2N symmetric partitions of all possible 8 × 8 CUs contained in the current 64 × 64 CTU. This SAD aggregation process is followed until the last partition size is reached (1 × 1), i.e., the SAD in the last stage corresponding to the 2N × 2N partition of 64 × 64 CU (see Fig. 6c).

A particular case is the way asymmetric partitions are obtained from SADs corresponding to symmetric partitions. The idea is to repeat the same type of aggregation as the last one performed. If the start block has been obtained by the sum of consecutive columns, then the resulting consecutive columns are added again. The obtained values are SADs corresponding to asymmetric partitioning (left, right, up, and down) of the next size of CUs. For instance, in the last intermediate stage ($M = 4, N = 16$), after a sum of consecutive columns, we start with a 4 × 2 block of 8 SADs values corresponding to the 2N × N symmetric partition of the 4 32 × 32 CUs contained in the current 64 × 64 CTU. Then, all pairs of consecutive columns are added again as shown in Fig. 6b. Thus, a 4 × 1 block of SAD values are obtained corresponding to 2N × nU and 2N × nD asymmetric partitions of the current 64 × 64 CTU.

Thus, in the proposed architecture, the SATB module delivers 677 SADs of the current CTU block every single clock cycle to the next module, the comparison block.

3.3 Comparison block

The comparison block should keep the minimum SAD values for each CU partition with their corresponding motion vectors (search area positions). So, it will compare all incoming SADs from the SATB with the minimum SADs previously found. In a clock cycle, the comparison block receives 677 SADs corresponding to all partitions of all CUs contained in the current CTU, which is located in a particular position of the search area. So, in the next cycle, this module again receives 677 SADs corresponding to the next position of the search area. Therefore, this block compares SADs partition by partition, keeping the minimum SADs and the positions of the search area corresponding to those minimums. After comparing the SADs from the last search area location, the minimum SADs for each partition and the associated motion vectors are obtained.

4 Implementation results

The proposed architecture is designed as a pipeline process shown in Fig. 7. The memory reading process and shift registers propagation require only one clock cycle. The PUs use one cycle, the SATB requires twelve additional clock cycles, and the comparison block needs one additional clock cycle. So, the proposed architecture requires 63 clock cycles to perform the initial load of the shift registers, 15 clock cycles to load the pipeline, and then as many clock cycles as positions the search area has.

Our proposal has been modeled in VHDL, and it has been synthesized, simulated, and implemented on the Xilinx FPGA, Virtex-7 XC7VX550T-3FFG1158. The correctness of our design was tested and verified with the HEVC HM 14 reference model [17].

To evaluate the performance and efficiency of our design, we have parametrized our IME architecture to

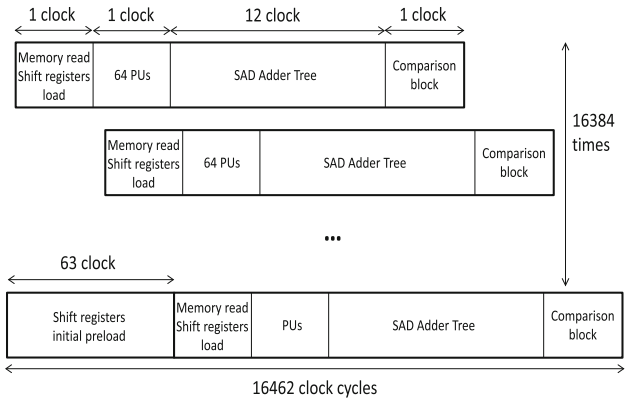


Fig. 7 Pipeline process of the proposed architecture

allow different configurations, such as (a) the maximum CTU size with values of 64 × 64 and 32 × 32, and (b) the size of the search area of the reference frame with values defined as the double size of the CTU, 80 % of the double size of the CTU, and the same size as a CTU.

Firstly, we proceed to test our proposal with the Virtex 7 FPGA technology. In Table 2, we show (a) the resulting operating frequency (clock), (b) the number of clock cycles for each CTU (latency), and (c) the system throughput in terms of the maximum frame rate under different video formats (1080p, 2K, and 4K), for different configurations of CTU and search area sizes. Our design can operate at the frequency of 247 and 318 MHz for a 64 × 64 CTU and a 32 × 32 CTU, respectively. It enables the encoder to carry out the IME process with a 64 × 64 CTU size and a search area of 128 × 128 pixels (as the HM14 reference model [17] establishes), obtaining a throughput of 30 fps at 2K video formats (2K@30fps). Our proposal is able to process video in real time for both 1080p and 2K resolutions in all tested configurations, and also with 4K video formats if the search area size is the same as the CTU size, as can be seen in Table 2.

In Tables 3 and 4, we show the resources used to implement our proposal for maximum CTU sizes of 64 × 64 and 32 × 32, respectively, on a Virtex-7 FPGA. In both tables, we show the resource usage of each block of the proposed architecture, as a resource usage profile. As can be seen, the slice area required by flip-flops and LUTs increases (≈×4) linearly with the increase of the maximum CTU size, as expected. In terms of flip-flops, the SATB is the block that uses the most amount of them (around 40 % of the total) in both configurations. This is due to the 12-stage pipeline design of the SATB. Moreover, calculating the distortion among pixels needs 50 % of the LUTs, due to the amount of subtractions in absolute value required in this process, being 1024 operations performed at each clock cycle. Regarding the required memory for storing the search area reference pixels, 36 and 9 kB memories are used in the case of a 64 × 64 CTU and a 32 × 32 CTU, respectively. On a Virtex-7, a BRAM block has a capacity of 36 kb. So, the slice area demanded by the used BRAMs also increases (≈×4) when going from 32 × 32 to the 64 × 64 maximum CU size.

An interesting analysis of our design can be observed at Table 2 when comparing the results of the 64 × 64 search area size with both CTU sizes. As can be seen, the latency is nearly the same but the throughput of the 64 × 64 CTU size is more than triple than the one obtained with the 32 × 32 CTU size. In terms of resource usage, the 64 × 64 CTU size requires near four times more resources, as shown in Tables 3 and 4. This implies that the use of more resources in the design provides higher throughput in a 4:3 relationship.

4.1 Systems evaluation

In Table 5, we compare our proposal with previous state-of-art architectures implemented on different FPGA platforms for both the 64 × 64 CTU and the 32 × 32 CTU size, and different search area sizes. We have chosen those works whose architectures are comparable to our proposal (i.e., perform the same functionality) and were implemented under FPGA technology. To make the comparison as fair as possible, we have obtained the performance results of our proposal with the same technologies, CTU sizes, and search area sizes as the ones used by the selected candidates. We will consider the system throughput as the main performance result of every proposal under comparison.

Regarding results for the 64 × 64 CTU size, Medhat et al. [3] present a parallel SAD block for the HEVC

Table 2 Throughput for different configurations in Virtex-7

CTU size Search area	64 × 64 128 × 128	64 × 64 104 × 104	64 × 64 64 × 64	32 × 32 64 × 64	32 × 32 52 × 52	32 × 32 32 × 32
Clock (MHz)	247	247	247	318	318	318
Latency	16,462	10,894	4174	4142	2750	1070
Fps at 1080p	32	48	124	39	59	151
Fps at 2K	30	45	116	37	55	141
Fps at 4K	8	12	30	10	15	37

Table 3 Utilization resources for 64 × 64 CTU implementation in Virtex-7

Resources	Flip-flops	LUTs	Memory (kB)
Memory read controller block	36,657 (25.40 %)	36,413 (19.30 %)	36 (100 %)
PUs (distortion computation)	32,768 (22.71 %)	94,208 (49.93 %)	–
SAD adder tree block (SATB)	58,727 (40.70 %)	47,063 (24.95 %)	–
Comparison block	16,150 (11.19 %)	10,980 (5.82 %)	–
Total	144,302	188,664	36

Table 4 Utilization resources for 32 × 32 CTU implementation in Virtex-7

Resources	Flip-flops	LUTs	Memory (kB)
Memory read controller block	10,155 (27.55 %)	9812 (20.22 %)	9 (100 %)
PUs (distortion computation)	8192 (22.22 %)	24,541 (50.57 %)	–
SAD adder tree block (SATB)	14,580 (39.55 %)	11,445 (23.58 %)	–
Comparison block	3937 (10.68 %)	2733 (5.63 %)	–
Total	36,864	48,531	9

Table 5 Comparison of the proposed architecture with state-of-the-art works

Design	Medhat [3]	Proposal 1	D'huys [7]	Proposal 2	Yuan [6]	Proposal 3
CTU size	64 × 64	64 × 64	64 × 64	64 × 64	32 × 32	32 × 32
Search area	104 × 104	104 × 104	64 × 64	64 × 64	48 × 48	48 × 48
Technology	Virtex-7	Virtex-7	Virtex-5	Virtex-5	Virtex-6	Virtex-6
Clock (MHz)	458.7	247	150	159	110	200
AMP	No	Yes	No	Yes	Yes	Yes
Throughput	2K@30fps	2K@45fps	720p@57fps	720p@173fps	1080p@30fps	1080p@43fps
Flip-Flops	39,901	144,302	199,682	178,620	19,744	43,531
LUTs	24,957	188,664	210,158	184,288	55,346	45,752
Memory (kB)	44	36	1229	36	148	9

integer-pel full search architecture without supporting AMP modes with a search area of 104 × 104 pixels. They used the Virtex-7 technology, and their design can operate at the frequency of 458.7 MHz. The operating frequency of our proposal with the same technology and configurations is almost two times lower. However, our architecture is capable of processing 45 fps at 2K video formats instead of 30 fps as obtained by the proposed design in [3]. Therefore, our proposed architecture is 1.5× as fast as the one proposed in [3] using the same search area size and considering all the AMP partition modes, contrary to [3], where AMP partitions are not calculated. This is due to the fact that our design takes advantage of the minimal latency to perform the same operations as we have an efficient pipeline design. Therefore, our system achieves higher throughput, reaching real-time processing for 2K video resolutions at 45 fps, and being on the way to accomplishing the same goal for 4K video formats, where 12 fps were obtained.

On the other hand, D'huys [7] proposes a reconfigurable design for HEVC motion estimation which can operate at the frequency of 150 MHz. His architecture is compared with our proposal, setting a common search area size to 64 × 64 pixels and the Virtex-5 technology. The operation frequency of our proposal is 159 MHz, achieving system throughput of 20 fps at 4K and 75 fps at 2K video formats. Our design significantly improves the performance of the architecture presented in [7], which is able to process a lower resolution video (720p) at 57 fps. If the video resolution is set to 720p, our architecture is capable of processing 173 fps. So, our architecture presents good balance between the maximum frequency and pipeline processing design, taking advantage of the low latency by leveraging all available resources.

Regarding results for the 32 × 32 CTU size, in Table 5, we show the comparison results between our proposal (implemented on a Virtex-6 FPGA) and the integer motion estimation design found in [6], both with a search area size of 48 × 48 pixels. The most significant feature, worthy of attention, is that our proposal can provide a higher operation frequency, achieving throughput of 43 fps at 1080p and 40 fps at 2K resolution, whereas the architecture presented in [6] is able to achieve 30 fps at 1080p video formats, using a similar amount of FPGA resources.

Considering the presented results, our architecture shows an efficient implementation of available resources in

FPGA, overcoming the performance of previous state-of-the-art architectures.

4.2 HEVC R/D performance and time profiling

To better understand the capabilities of IME hardware devices, we have performed a set of tests to analyze the benefits of including an IME FPGA-based accelerator, like the one proposed here, in the HEVC reference software in terms of speedup and observe how both parameters, the CTU size, and the search area size impact on the R/D performance of the HEVC encoder. To perform these tests, we have used the HEVC HM 14 reference model [17] working with the main profile and low-delay configuration mode. The HEVC reference software was compiled with Visual Studio 2010 with the default compiler options and run over a PC platform with an Intel Core i7-3770 CPU 3.40 GHz with 16 GB RAM. Three video sequences from the HEVC common conditions video set were selected: (s1) Racehorses at 832 × 480 resolution (30 fps), (s2) Basketball Drive at 1920 × 1080 (50 fps), and (s3) People On Street at 2560 × 1600 (30 fps).

The experiments were performed using different search area sizes (128 × 128, 104 × 104, 64 × 64, 52 × 52, and 32 × 32) and CTU sizes (64 × 64 and 32 × 32).

Tables 6 and 7 show all data gathered for CTU sizes of 64 × 64 and 32 × 32, respectively. The first row shows the total time (in seconds) required to encode each video sequence (10 frames). The second row shows the percentage of the total time needed by the IME software module using a full search algorithm (% IME time SW). These percentages vary from 62 to 96 % depending on the video sequence, the search area size, and the CTU size. As was expected, the time required by the IME software module decreases as both the search area size and the CTU size do. Rows three and four show the number of CTUs per second that can be computed by software (CTU/s SW) and hardware (CTU/s HW) versions of the IME module. As can be seen, these values also depend on the search area size and maximum CTU size, and in the case of the IME software module, also depend on the video sequence.

So by looking at the information provided in Tables 6 and 7, we could assess that the IME module is a bottleneck in the HEVC reference software. Therefore, if the IME software module is replaced by our FPGA-based device, the overall encoding time will be significantly reduced. For example, for a high-resolution video sequence like PeopleOnStreet (s3) and setting the CTU size to 64 × 64 and the search area size to 128 × 128 (default values in the HEVC reference software), the total encoding time (10 frames) will be reduced from 38 h to 2 s, since the motion estimation module takes around 95 % of the overall encoding time.

To reduce the hardware complexity, allowing faster versions with reduced power consumption, the CTU size and the search area must be reduced as much as possible. However, this may cause performance degradation in the encoding process, decreasing the overall video quality and/or reducing the compression rate. To evaluate this aspect, we will analyze the impact of these parameters on the R/D (rate/distortion) performance. In Fig. 8, we show the video quality of the test video sequence RaceHorses (s1) for each CTU and search area sizes at different compression levels (QP values). As can be seen, there are slight differences between the CTU size, being greater the difference as the compression rate increases. Differences between search areas are negligible. Although R/D differences may depend on the video content, similar results were obtained for the other two video sequences tested.

Finally, we also have performed a profile of the IME HEVC with another motion search algorithm, which is available in the HEVC reference software (diamond-like search). This algorithm is used by default in the reference software and it is about 90 times as fast as the full search algorithm, with the disadvantage that it does not guarantee finding optimal MVs, and as consequence video quality could be affected. As can be seen in Table 8, the inclusion of our IME hardware module will speed up the IME computation of diamond-like search algorithm 230 and 700 times for 32 × 32 and 64 × 64 CTU sizes, respectively.

After performing the whole analysis, a trade-off should be taken to determine which configuration better adapts to the application requirements (low power consumption,

Table 6 Time profile of the IME HEVC for a 64 × 64 CTU

Search area	128 × 128			104 × 104			64 × 64		
Video sequences	s1	s2	s3	s1	s2	s3	s1	s2	s3
Encoding time SW (s)	13,670	61,602	135,970	9392	42,050	92,863	4117	17,881	39,933
% IME time SW	95	96	95	90	94	93	83	86	85
CTU/s SW	0.24	0.26	0.23	0.38	0.39	0.35	0.91	1.00	0.89
CTU/s HW	14,993	14,993	14,993	22,625	22,625	22,625	59,172	59,172	59,172
HW gain	62,260	57,767	64,800	59,621	58,312	65,384	64,856	59,115	66,477

Table 7 Time profile of the IME HEVC for a 32 × 32 CTU

Search area	64 × 64			52 × 52			32 × 32		
Video sequences	s1	s2	s3	s1	s2	s3	s1	s2	s3
Encoding time SW (s)	3652	15,892	35,295	2634	11,303	25,293	1378	5748	12,911
% IME time SW	85	87	86	79	81	81	60	63	62
CTU/s SW	4	5	4	6	7	6	14	17	15
CTU/s HW	76,923	76,923	76,923	115,607	115,607	115,607	297,619	297,619	297,619
HW gain	20,383	17,293	19,457	20,654	17,384	19,675	21,096	17,664	19,912

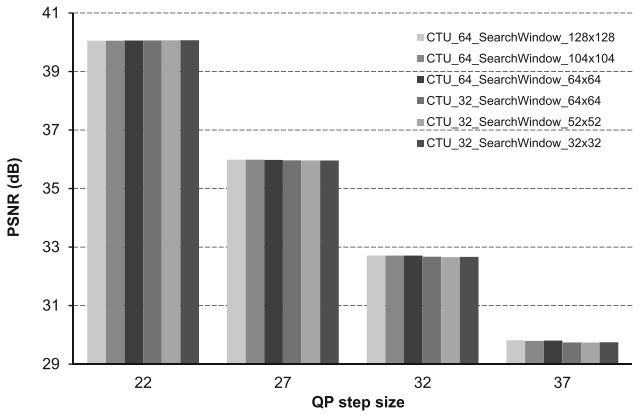


Fig. 8 R/D performance for different CTU and search area sizes for the RaceHorses sequence

Table 8 Average CTU IME time with 2 × CTU search area size

CTU size	Full search SW	Diamond search SW	Full search HW
64 × 64	4.11 s	4.65E−02 s	6.67E−05 s
32 × 32	2.48E−01 s	3.05E−403 s	1.30E−05 s

encoding time, compressed video quality). The use of hardware accelerators designed in FPGA platforms like the one proposed here are mandatory when real-time UHD video encoding is the objective.

5 Conclusion

In this work, we have presented a fast and efficient IME hardware unit for the HEVC video encoder which (a) supports AMP modes, (b) both CTU and search area sizes are configurable, and (c) is implemented on a Virtex-7 FPGA. The suitability of using FPGAs for implementing the HEVC IME module has been demonstrated in this paper, proposing a design that improves the previous results of other IME hardware systems.

Our FPGA-based design is able to process 2K video formats at 116 frames per second and 4K video sequences at 30 fps, which represents a huge complexity reduction of the HEVC video encoding process, achieving real-time encoding for high-definition video contents and beyond.

We have also analyzed the impact of the maximum CTU and the search area sizes over the encoder complexity and the resulting video quality, showing that the encoder complexity decreases as both the maximum CTU size and the search area do. Furthermore, the maximum CTU size has a minimum impact over the R/D, being more noticeable at high compression rates. In the test video sequences analyzed, the impact over the quality of the search are size is negligible, but it will depend on the video content.

In future work, we are working to include our IME hardware module in the HEVC reference software and perform a complete test over an evaluation platform such as ZYNQ of Xilinx. In addition, we intend to expand the hardware module to perform the fractional-pel motion estimation, or even the SAD unit for intra-mode coding.

Acknowledgments This research was supported by the Spanish Ministry of Economy and Competitiveness under Grant TIN2015-66972-C5-4-R.

References

1. Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol. **22**, 1649–1668 (2012)
2. Sze, V., Budagavi, M., Sullivan, G.J.: High Efficiency Video Coding (HEVC) Algorithms and Architectures. Springer, Switzerland (2014)
3. Medhat, A., Shalaby, A., Sayed, M.S., Elsabrouty, M.: A Highly Parallel SAD Architecture for Motion Estimation in HEVC Encoder. In: IEEE Asia Pacific Conf. Circuits Syst. (APCCAS), pp. 280–283. Ishigaki (2014)
4. Byun, J., Jung, Y., Kim, J.: Design of integer motion estimator of HEVC for asymmetric motion-partitioning mode and 4K-UHD. Electron. Lett. **49**(18), 1142–1143 (2013)
5. Vidyalekshmi V.G., Yagain D., Ganesh Rao K.: Motion estimation block for HEVC encoder on FPGA. In: IEEE Int. Conf.

Recent Advances and Innovations in Engineering (ICRAIE), pp. 1–5. Jaipur, (2014)
6. Yuan, X., Jinsong, L., Liwei, G., Zhi, Z., Teng, R.K.F.: A high performance VLSI architecture for integer motion estimation in HEVC. In: IEEE 10th Int. Conf. ASIC (ASICON), pp. 1–4. Shenzhen (2013)
7. D'huys, T.: Reconfigurable data flow engine for HEVC motion estimation. In: IEEE Int. Conf. Image Processing (ICIP), pp. 1223–1227. Paris (2014)
8. Davis, P., Sangeetha, M.: Implementation of motion estimation algorithm for H.265/HEVC. Int. J. Adv. Res. Elect. Electron. Instrum. Eng. **3**(3), 122–126 (2014)
9. Nalluri, P., Alves, L.N., Navarro, A.: High speed SAD architectures for variable block size motion estimation in HEVC video coding. In: IEEE Int. Conf. Image Processing (ICIP), pp. 1233–1237. Paris (2014)
10. Chen, C.Y., Chien, S.Y., Huang, Y.W., Chen, T.C., Wang, T.C., Chen, L.G.: Analysis and architecture design of variable block-size motion estimation for H.264/AVC. IEEE Trans. Circuits Syst I: Reg. Papers **53**(3), 578–893 (2006)
11. Elhamzi, W., Dubois, J., Miteran, J.: An efficient low-cost FPGA implementation of a configurable motion estimation for H.264 video coding. Springer J. Real-Time Process. **9**(1), 19–30 (2014)
12. Moorthy, T., Ye, A.: A scalable architecture for variable block size motion estimation on field-programmable gate arrays. In: IEEE Canadian Conf. Electrical and Computer Engineering (CCECE), pp. 1303–1308. Niagara Falls (2008)
13. Kthiri, M., Kadionik, P., Levi, H., Loukil, H., Atitallah,,B., Masmoudi, N.: An FPGA implementation of motion estimation algorithm for H.264/AVC. In: IEEE 5th Int. Symp. I/V Communications and Mobile Network (ISVC), pp. 1–4. Rabat (2010)
14. Pastuszak, G., Jakubowski, M.: Adaptive computationally scalable motion estimation for the hardware H.264/AVC encoder. IEEE Trans. Circuits Syst. Video Technol. **23**(5), 802–812 (2013)
15. Pastuszak, G., Trochimiuk, M.: Algorithm and architecture design of the motion estimation for the H.265/HEVC 4K-UHD encoder. J. Real Time Image Process (2015)
16. Lin, Y.L.S., Kao, C.Y., Kuo, H.C., Hen, J.W.: VLSI Design for Video Coding-H.264/AVC Encoding from Standard Specification to Chip. Springer, New York (2010)
17. HEVC software repository HM–14.0 reference model. https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags/HM-14.0. Accessed 2 May 2014 (2014)

Estefania Alcocer was born in Bigastro, Spain, in 1986. She received her M.S. degree in telecommunication engineering in 2010 from the Miguel Hernandez University, Elche, Spain, and she joined the GATCOM research group as Ph.D. student in 2012. Currently, she is an assistant professor in the Department of Physics and Computer Architecture at Miguel Hernandez University, Elche since 2012. Her current research activities are related to image processing, the design of FPGA-based systems and video coding.

Roberto Gutierrez was born in Orihuela, Spain, in 1977. He received his M.Sc. degree in telecommunication engineering in 2003, and the Ph.D. degree in electronic engineering in 2011, both from the Universidad Politecnica de Valencia, Spain. He is an associate professor in the Department of Communication engineering at Universidad Miguel Hernandez, Elche since 2003. His current research interests include the design of FPGA-based systems, computer arithmetic, VLSI signal processing and digital communications.

Otoniel Lopez-Granado received his M.S. in Computer Science from the University of Alicante (Spain) in 1996. Between 1997 and 2003 he worked as programmer analyst in an important industrial informatics firm. In 2003, he joined to the Computer Engineering Department at Miguel Hernandez University (UMH), Spain, as an assistant professor. Then, he received the Ph.D. degree in Computer Science in 2010. In 2012, he was promoted to associate professor. Currently, he leads the GATCOM research group (atc.umh.es) at Miguel Hernandez University. His research and teaching activities are related to multimedia networking (audio/video coding and network delivery).

Manuel P. Malumbres received his B.Sc. in Computer Science from the University of Oviedo (Spain) in 1986. In 1989, he joined to the Computer Engineering Department (DISCA) at Technical University of Valencia (UPV), Spain, as an assistant professor. Then, he received M.S. and Ph.D. degrees in Computer Science from UPV, in 1991 and 1996, respectively. He is a TC member of IEEE Multimedia Communications Group and associate editor of the Signal, Image and Video Processing journal. He was serving as TPC member of several relevant international Conferences related with his main research interests. He is author of more than 160 conference and journal publications and several networking books for undergraduate CS courses. Currently, his research and teaching activities are related to multimedia networking (image/video coding and network delivery) and wireless network technologies (MANETs, VANETs and WSNs).