Comparison of 90nm and 65nm Logic Synthesis of a SAD Configurable VLSI Architecture

Ismael Seidel, Bruno George de Moraes, André Beims Bräscher, José Luís Güntzel

Embedded Computing Lab. (ECL),

Federal University of Santa Catarina (UFSC), Florianópolis, Brazil

{ismaelseidel, brunogm, andre.brascher, guntzel}@inf.ufsc.br

Abstract — **This paper evaluates the impact of the technology node on the area, performance and power consumption of a configurable VLSI architecture for Sum of Absolutes Differences (SAD). Such architecture may be configured to take benefit from pel decimation, trading off quality for energy. The proposed architecture was synthesized for 90nm and 65nm technologies assuming both nominal and Low-Vdd/High-Vt (LH) cases. Results showed that pel decimation itself is responsible for up to 60% of energy efficiency with a negligible area and power overhead with respect to a non-configurable reference SAD architecture. Our best result, using pel decimation 4:1 in 65nm/LH spends only 2.6 pJ for each 4x4 block, which corresponds to about 7.4 times less energy when compared with our proposed architecture in 90nm/nominal operating in full SAD mode.**

***Keywords*— Video Coding, SAD, Subsampling, VLSI, Low-power, Configurable.**

I. INTRODUCTION

Video compression is a computationally intensive task that usually demands high performance. The high performance requirement, by its turn, results from the real time constraints inherent to applications such as video capture. Therefore, in some applications it might be necessary to sacrifice quality to meet timing constraints. Moreover, the use of video compression became *a must* in Personal Mobile Devices (PMDs), such as point-and-shoot cameras, smartphones and videogame consoles. In those cases, reducing the number of operations in video encoding helps to prolong battery lifetime.

In most codecs, each frame of a video sequence is divided into smaller blocks which are submitted to Inter frame and Intra frame prediction techniques to explore temporal and spatial redundancies, respectively. Both predictions rely on coding the residue data (i.e., the differences) between similar blocks. Motion Estimation (ME) occurs in Inter frame prediction and corresponds to the most computational intensive task of the whole compression process. For each block, ME searches for the most similar candidate to be used as reference for Motion Compensation (MC) to reconstruct the original block. In such coding, a similarity metric must be used to guide the search for the most similar blocks (i.e., block matching).

The Sum of Absolute Differences (SAD) [1] is the most widely used similarity metric because it is very fast and very appropriate for VLSI (Very-Large Scale Integration) coder implementations, since it relies only on basic arithmetic operations (addition, subtraction and module). Nevertheless, the number of SAD operations required for video compression will increase dramatically, as soon as high and ultra-high definition video formats (e.g., H.264/AVC [2] and HEVC [3]) are adopted. This will make performance and power consumption requirements of VLSI SAD engines even more stringent. A pragmatic strategy to circumvent this problem is the adoption of the so-called pel decimation algorithm [4].

In pel decimation, the pixels are sampled regularly over the search area. Such subsampling improves the performance of the encoding, but it lowers the correlation between candidates, resulting in a decrease of prediction quality. On the other hand, the regularity of subsampling patterns allows for low cost and high performance VLSI implementations that also lead to significant power savings. It is important to observe that pel decimation does not dismiss the use of an efficient search algorithm, but should be used along with it to achieve more significant speedups and power savings. Indeed, several pel decimation-based search strategies are found in the literature, such as APS [5], GEA [6] and QME [7].

As main contribution, this paper evaluates the impact of two different technology nodes in the logic synthesis of a previously proposed SAD VLSI architecture [8] that can be configured to perform pel decimation in different patterns. The architecture cost, performance and energy efficiency are evaluated through delay and power estimates obtained by logic synthesis with Synopsys Design Compiler tool [9] for both 90nm and 65nm commercial standard cell libraries. A 4-input non-configurable fully combinational SAD architecture was also synthesized to serve as reference. Finally, we re-synthesized both the configurable and non-configurable architectures by using low supply voltage and high threshold voltage (Low-Vdd/High-Vt) [10] in order to evaluate the impact of such low power techniques on area, performance and power of the considered architectures. In addition to that, the impact of technology node was also evaluated.

This paper is organized as follows. Section II describes the SAD calculation, introduces pel decimation and comments on the resulting quality decrease. The synthesized architectures are described in Section III. Section IV outlines the achieved synthesis results. Conclusions are presented in Section V.

II. SUM OF ABSOLUTE DIFFERENCES

The SAD calculation is expressed by (1), where *Ori* is the original block and *Ref* is the reference block. From such equation the simplicity of the calculation is evident.

$$SAD = \sum_{i=0}^{M}\sum_{j=0}^{N}\left|Ori_{i,j} - Ref_{i,j}\right| \qquad (1)$$

In [11] Chen et al. present a 720p H.264/AVC [2] coder architecture in which the SAD calculation responds for 33% of gate-count. In [12] Liu and Zaccarin describe a variable block size ME architecture which SAD engine corresponds to roughly 79% of total gate count.

Considering the intensive use of SAD operations and the pressure for more energy-efficient coder implementations, the design of low power SAD (or another similarity metric) architectures is of utmost importance for video compression. The works [13] and [14] present high performance SAD architectures, but do not take into account aspects of power and energy efficiency.

A more accurate analysis on SAD architectures is presented in [15, 16]. The authors synthesized SAD architectures with a variety of parallelism and pipeline configurations. In order to further reduce the energy per SAD calculation the authors also synthesized the architectures for target frequencies. They reported the 4x4 pixel (16 input) architecture with higher degree of parallelism and less pipeline stages as the most energy-efficient one. According to the authors, the pipeline registers are the main contributors to power consumption.

A. Pel Decimation and its quality impact

The full-search algorithm is recognized as the one providing the best block matching results [7]. This method performs a SAD calculation for all candidates in a search window. Despite the simplicity of each single calculation it results in a very high computation effort.

As an attempt to reduce the number of calculations, several works have proposed algorithms that restrict the searching area. Another effective means of reducing the number of calculations relies on using the Pel Decimation algorithm [4] [7] [12] [17] and [18].

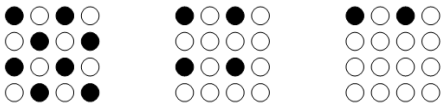

Fig. 1 Example of 2:1, 4:1 and 8:1 commom pel decimation patterns. Adapted from [17]

According to [4], in pel decimation the pixels are taken regularly over the search area, resulting in a subsampling. When pel decimation is applied, a decrease in prediction quality is expected, since the correlation of each pixel in the block is lost. Although the sampling pattern is not defined in the algorithm, there is a tendency on using regular ones as those depicted in Fig. 1. Due to its regularity and effectiveness in reducing the number of calculations and high regularity, pel decimation reveals itself very appropriate for VLSI implementation.

A detailed quality assessment of several pel decimation patterns and ratios is presented in [19]. The authors performed statistical analysis on the final encoded video quality (PSNR [1] and DSSIM [20]) using analysis of variance (ANOVA). For the two pel decimation ratios considered in the current work, the work in [19] reported a maximum of 2.5% PSNR and 8.2% DSSIM loss (in average) for 2:1 patterns respect to full sampling and 2.8% PSNR and 14.1% DSSIM loss for 4:1 patterns with respect to full sampling. Such percentages are within a 95% confidence interval. The work in [19] also demonstrated that as the video resolution increases the quality losses resulted from higher pel decimation ratios becomes more negligible.

III. Our Configurable SAD Architecture

The architectures used in our synthesis comparison were previously described in [8]. The most important details to known are further mentioned. It is important to notice that the chosen block size was 4x4 as other block sizes can be easily derived by compositions of this block size.

For the configurable architecture there are two smaller block resolutions to take into account in addition to full sampling. One with eight pixels, for pel 2:1 and another with only four pixels, for pel 4:1. In the latter case there are only four parallel inputs (for each block) to maximize the energy efficiency. Having no minor cases, the architecture with four parallel inputs became the most suitable for minimizing the energy consumption. A 12-bit output register must be also added along with an extra adder.

When using pel 2:1 the speedup with respect to full SAD is 40%. Using pel 4:1 results in 33% of speedup with respect to pel 2:1. In comparison to the full SAD there is a speedup of 60% when using pel 4:1. These proportions are kept when considering energy efficiency.

It is important to consider access times of a coupled memory to design low power systems. In [21-23] it is shown that current memories, even at low power mode, have access times in the range of 0.4 to 3.4ns. Also these memories show power consumption to vary greatly in accordance with these times. A specialized sample driver should run at least with double frequency than SAD tree to maximize energy efficiency. This driver aggregates data for SAD calculation in accordance to a subsampling pattern and ratio. Considering the 3.4ns access time of the slower and power saving memory, the recommended frequency for our SAD architecture should be smaller than 147 MHz. For maximum throughput the fastest memory should be used, and the maximum frequency of our SAD module should be less than 1.25 GHz.

The proposed architecture, as well as a 4-input non-configurable combinational (“non-config”) architecture, used as reference, were described in Verilog and logically synthesized for 90nm [24] and 65nm [25] TSMC standard cell libraries using Synopsys Design Compiler [9] in Topographical mode. Both architectures were also synthesized with Low-Vdd/High-Vt (hereinafter referred to as “LH”) for both 90nm and 65nm technologies.

IV. Syntheses Results

To better organize our syntheses results, we chose to assign an alphabetical index for each synthesis case according to technological node (90nm or 65nm), architecture (non-configurable or configurable) and synthesis for nominal or LH

case. These indices are shown in Table I. Also, as we first performed syntheses for maximum frequency, we kept the same indices for the target, indicating either as the results are for maximum or target frequency.

TABLE I
SYNTHESIS INDEX

	Architecture			
	Non-config		Config	
	90nm	*65nm*	90nm	*65nm*
Nominal	(A)	*(B)*	(C)	*(D)*
Low-Vdd/High-Vt	(E)	*(F)*	(G)	*(H)*

Table II and Table III show the area and power results as given by the synthesis tool for all 16 syntheses. The synthesis results for maximum frequency are shown in Table II whereas the results for the target frequency of 66.67 MHz are shown in Table III.

The maximum frequencies achieved by the nominal synthesis for both "non-config" and "config" architectures were of 800 MHz for 90nm and 1.25 GHz for 65nm. It is important to notice that considering the memories presented in Section III, the maximum recommended frequency is achieved by the 65nm synthesis. In this case the whole system can work at the maximum frequency and yet the configurable architecture will only spend the minimum of energy for the required operation. In the case a higher frequency is to be achieved, using these memories will increase the system's latency because the SAD becomes idle, waiting for the next pixels. This will increase energy consumption due to leakage.

When applying LH, the 90nm maximum frequency gets limited on 270.27 MHz. It presents 66% of frequency degradation. In the case of 65nm/LH synthesis the maximum frequency was 675.67 MHz, which corresponds to a lower frequency degradation of 46%.

The area reported in all cases was smaller for the 65nm synthesis when comparing with their equivalents in 90nm (the column straight in the left). The dynamic power results also tend to be smaller for the 65nm synthesis but, considering the maximum frequency achieved, the LH synthesis reports indicates more dynamic power. It can be noticed that in nominal synthesis, the maximum frequency is about 1.6 times higher for 65nm with respect to 90nm whereas in the LH synthesis this value is about 2.5 times higher. It can explain the larger dynamic power dissipation. When considering the target frequencies, the dynamic power is always smaller for 65nm syntheses.

The leakage power is higher for the 65nm node as expected. Even when applying LH it is more prominent than in 90nm cases, both for maximum and target frequencies. The total power reports show that in almost all cases 65nm consumes less power, but this is not true for the maximum frequencies when using LH. Again, it is important to notice that the LH synthesis had presented a 2.5x higher frequency for the 65nm compared to 90nm.

To get a better view of the energy efficiency of all syntheses, the graphics in Fig. 2 e Fig. 3 present the energy spent for each block. Fig. 2 presents the maximum frequency results whereas Fig. 3 presents the target frequency results.

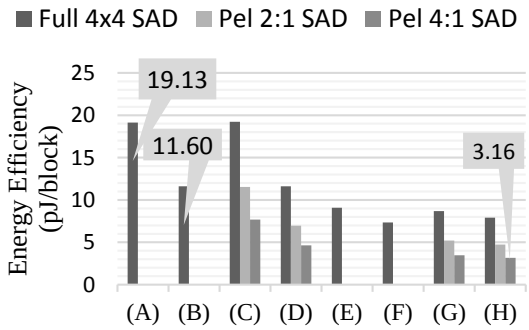

Fig. 2 Energy efficiency for maximum frequency

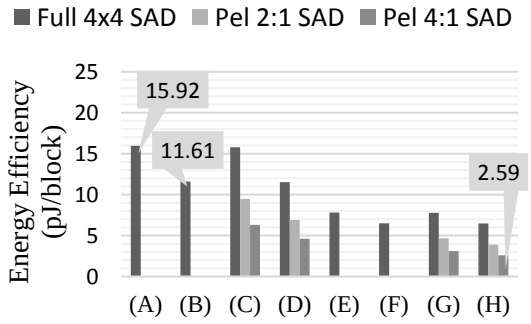

Fig. 3 Energy efficiency for target frequency

It can be seen that in nominal syntheses the 65nm presents a better energy efficiency result than 90nm. This behavior is kept in LH syntheses, but the differences are now smaller. It should be also noticed that are only little differences when comparing the configurable and non-configurable architectures against each other. The lowest value of energy/block is in the case of 65nm/LH synthesis of the configurable architecture (H) when using 4:1 pel decimation. Also for full sampling this synthesis demonstrates to be the most energy-efficient.

For energy efficiency regarding the target frequency, the obtained values shows to be the best cases, but in some cases the values are almost the same than for maximum frequencies. This happens more often in the 65nm syntheses. The best values again are from the (H) synthesis. The best one is obtained when using pel decimation 4:1, which is about only 2.59 pJ/block.

V. CONCLUSIONS

This paper presented a comparison between two technologic nodes (90nm and 65nm) for the logic synthesis of configurable architecture for SAD calculation with pel decimation capability and also for its non-configurable equivalent. Previous synthesis results [8] for 90nm had highlighted that the impact of configurability is negligible and in this work we show that this behavior also holds for 65nm syntheses.

TABLE II
SYNTHESIS RESULTS FOR MAXIMUM FREQUENCY

		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
	Technology (nm)	90	65	90	65	90	65	90	65
	Frequency (MHz)	800	1250	800	1250	270.27	675.67	270.27	675.67
	Area (μm²)	5011.88	3114.00	5104.31	3098.00	5845.19	3336.80	5622.93	3728.80
Power (μW)	**Dynamic**	1517.80	1426.40	1524.90	1427.50	244.47	490.05	233.69	527.09
	Leakage	12.75	24.03	13.05	24.00	1.09	5.73	1.03	6.95
	Total	1530.55	1450.43	1537.95	1451.50	245.57	495.78	234.72	534.05

TABLE III
SYNTHESIS RESULTS FOR TARGET FREQUENCY

		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
	Technology (nm)	90	65	90	65	90	65	90	65
	Frequency (MHz)	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67
	Area (μm²)	3025.61	2037.20	3027.73	2020.40	3710.05	2041.20	3717.10	2050.00
Power (μW)	**Dynamic**	100.20	64.39	99.06	64.07	51.54	40.29	51.22	40.18
	Leakage	6.22	13.01	6.20	12.73	0.58	3.02	0.58	3.03
	Total	106.42	77.40	105.27	76.81	52.12	43.31	51.80	43.21

Our best result, using pel decimation 4:1 in 65nm/LH spends only about 2.6 pJ for each 4x4 block, which corresponds to about 7.4 times less energy when compared with the configurable architecture in 90nm/nominal operating in full SAD mode.

ACKNOWLEDGMENT

This work was partially supported by the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) and by the Brazilian Council for Scientific and Technological Development (CNPq) through the "Brazil-IP Project".

REFERENCES

[1] I. E. G. Richardson, H. 264 and MPEG-4 video compression: video coding for next-generation multimedia. John Wiley & Sons Inc, 2003.

[2] ITU-T, "H.264 corrigendum 1," jan 2009.

[3] K. Ugur et al., "Low complexity video coding and the emerging HEVC standard," in Picture Coding Symposium (PCS). IEEE, Dec. 2010, pp. 474–477.

[4] P. M. Kuhn, "Fast mpeg-4 motion estimation: Processor based and flexible vlsi implementations," The Journal of VLSI Signal Processing, vol. 23, pp. 67–92, 1999.

[5] H.-K. Jung, C.-P. Hong, J.-S. Choi, and Y.-H. Ha, "A vlsi architecture for the alternative subsampling-based block matching algorithm," IEEE Transactions on Consumer Electronics, vol. 41, no. 2, pp. 239–247, may 1995.

[6] Y.-W. Huang, S.-Y. Chien, B.-Y. Hsieh, and L.-G. Chen, An efficient and low power architecture design for motion estimation using global elimination algorithm. IEEE, may 2002, vol. 3, pp. III–3120–III–3123.

[7] K. Lee, H. Chin, H. Hsu, and C. Jen, "QME: an efficient subsampling-based block matching algorithm for motion estimation," in Proceedings. ISCAS '04, vol. 2. IEEE, May 2004, pp. II– 305–8 Vol.2.

[8] I. Seidel, B. G. de Moraes and J. L. A. Güntzel, "A Low-Power Configurable VLSI Architecture for Sum of Absolute Differences Calculation" in 2013 IEEE Fourth Latin American Symposium on Circuits and Systems (LASCAS). IEEE, feb-mar 2013, pp. 1–4.

[9] Synopsys, "Synopsys design compiler, version f-2011.09-sp5-2." 2011.

[10] M. Keating, et al., Low Power Methodology Manual - for System-on-Chip Design. Springer, 2007.

[11] T. Chen, S. Chien, Y. Huang, C. Tsai, C. Chen, T. Chen, and L. Chen, "Analysis and architecture design of an HDTV720p 30 frames/s H.264/AVC encoder," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 6, pp. 673– 688, jun 2006.

[12] Z. Liu, Y. Song, M. Shao, S. Li, L. Li, S. Goto, and T. Ikenaga, "32-Parallel SAD tree hardwired engine for variable block size motion estimation in HDTV1080P Real-Time encoding application," in 2007 IEEE Workshop on Signal Processing Systems. IEEE, oct 2007, pp. 675–680.

[13] J. Vanne, E. Aho, T. D. Hamalainen, and K. Kuusilinna, "A high-performance sum of absolute difference implementation for motion estimation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 7, pp. 876–883, jul 2006.

[14] L. Yufei, F. Xiubo, and W. Qin, "A High-Performance low cost SAD architecture for video coding," IEEE Transactions on Consumer Electronics, vol. 53, no. 2, pp. 535–541, may 2007.

[15] F. L. Walter, C. M. Diniz, and S. Bampi, "Synthesis and comparison of low-power high-throughput architectures for SAD calculation," in 2011 IEEE Second Latin American Symposium on Circuits and Systems (LASCAS). IEEE, feb 2011, pp. 1–4.

[16] F. Walter and S. Bampi, "Synthesis and comparison of low-power architectures for sad calculation," 26th South Symposium on Microelectronics, pp. 45 – 48, Abril 2011.

[17] M. Porto, L. Agostini, L. Rosa, A. Susin, and S. Bampi, "High throughput hardware architecture for motion estimation with 4:1 pel subsampling targeting digital television applications," in Proceedings PSIVT'07. Berlin, Heidelberg: Springer-Verlag, 2007,pp. 36–47.

[18] B. Liu and A. Zaccarin, "New fast algorithms for the estimation of block motion vectors," IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, no. 2, pp. 148–157, Apr. 1993.

[19] I. Seidel, B. G. de Moraes, E. Wuerges and J. L. A. Güntzel, "Quality Assessment of Subsampling Patterns for Pel Decimation Targeting High Definition Video". Accepted on 2013 IEEE International Conference on Multimedia and Expo (ICME 2013), 2013.

[20] A. Loza, L. Mihaylova, N. Canagarajah, and D. Bull, "Structural similarity-based object tracking in video sequences," in Information Fusion, 2006, pp. 1 –6.

[21] M. Qazi, K. Stawiasz, L. Chang, and A. Chandrakasan, "A 512kb 8T SRAM macro operating down to 0.57 V with an AC-Coupled sense amplifier and embedded data-retention-voltage sensor in 45 nm SOI CMOS," IEEE Journal of Solid-State Circuits, vol. 46, no. 1, pp. 85–96, Jan. 2011.

[22] K. Zhang, U. Bhattacharya, L. Ma, Y. Ng, B. Zheng, M. Bohr, and S. Thompson, "A fully synchronized, pipelined, and re-configurable 50mb sram on 90 nm cmos technology for logic applications," in VLSI Circuits, 2003. Digest of Technical Papers. 2003 Symposium on, june 2003, pp. 253 – 254.

[23] G. Fukano, K. Kushida, A. Tohata, Y. Takeyama, K. Imai, A. Suzuki, T. Yabe, and N. Otsuka, "A 65nm 1Mb SRAM macro with dynamic voltage scaling in dual power supply scheme for low power SoCs," in Non-Volatile Semiconductor Memory Workshop, 2008 and 2008 Interna-tional Conference on Memory Technology and Design. NVSMW/ICMTD 2008. Joint, May 2008, pp. 97 –98.

[24] TSMC STANDARD CELL Library RNTCBN90GHP, Taiwan Semi-conductor Manufacturing Company Limited - TSMC.

[25] TSMC STANDARD CELL Library TCBN65, Taiwan Semi-conductor Manufacturing Company Limited - TSMC.