



HAL
open science

Implementation techniques of high-order FFT into low-cost FPGA

Yousri Ouerhani, Maher Jridi, Ayman Alfalou

► **To cite this version:**

Yousri Ouerhani, Maher Jridi, Ayman Alfalou. Implementation techniques of high-order FFT into low-cost FPGA. IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), 2011, Aug 2011, North Korea. pp.1-4. hal-00783028

HAL Id: hal-00783028

<https://hal.science/hal-00783028>

Submitted on 31 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Implementation techniques of high-order FFT into low-cost FPGA

Yousri Ouerhani, Maher Jridi and A. Alfalou, Senior Member, IEEE
 Equipe Vision, Laboratoire L@bISEN, CS 42807, 29228 Brest Cedex 2, France
 e-mail: {yousri.ouerhani, maher.jridi and ayman.al-falou}@isen.fr

Abstract—In this paper, our objective is to detail know-how and techniques that can help the designer of electronic circuits to develop and to optimize their own IP in a reasonable time. For this reason, we propose to optimize existing FFT algorithms for low-cost FPGA implementations. For that, we have used short length structures to obtain higher length transforms. Indeed, we can obtain a VLSI structure by using $\log_4(N)$ 4-point FFTs to construct N-point FFT rather than $(N/8)\log_8(N)$ 8-point FFTs. Furthermore, two techniques are used to yield with VLSI architecture. Firstly, the radix-4 FFT is modified to process one sample per clock cycle. Secondly, the memory is shared and divided into 4 parts to reduce the consumed resources and to improve the overall latency. Comparisons with commercial IP cores show that the low area architecture presents the best compromise in terms of speed/area.

I. INTRODUCTION

The Discrete Fourier Transform (DFT) is one of the most important tools used in Digital Signal Processing applications. It has been widely implemented in digital communication systems such as Radars, Ultra Wide Band (UWB) receivers and many other applications. Computing this operation has a high computational requirement and needs a large number of operations (N^2 complex multiplications and $N.(N-1)$ complex additions). This makes computing and implementation very difficult to realize.

To reduce the number of operations a fast algorithm has been introduced by Cooley-Tukey [1] and called Fast Fourier Transform (FFT). The latter, reduces complexity from $O(N^2)$ to $O(N\log N)$. Other researchers, propose numerous techniques such as radix-4 [2], split radix [3] to avoid radix-2 structure in order to reduce the complexity of FFT algorithm. These architectures are either based on the Decimation-in-Time (DIT) or on the Decimation-in-Frequency (DIF). Several designs based on these architectures were proposed in order to implement these algorithms.

On the other hand, there is a growing interest in Field Programmable Gate Arrays (FPGAs) because of their potential to substantially accelerate computational intensive algorithms such as FFTs. Unfortunately, high order FFT are almost implemented into high cost FPGAs. For example, it is not possible to instantiate 512-point FFT with the Xilinx IP core to implement it in Spartan 3 family.

To meet with this challenge, we present in this paper a VLSI architecture to allow the implementation of high order FFT into low cost FPGAs.

The remainder of this paper is organized as follows. In

section II, definition and two kinds of distributions (spatial and temporal) are introduced. Section III is devoted to the proposed low area architecture. We detail the principle and the structure of 64-point FFT which may be generalized to higher orders. Then, techniques to save area are illustrated. Section IV presents the experimental results and comparisons with IP core and prior works quoted in the literature. Finally, we summarize and conclude this paper in section V.

II. BACKGROUND

A. Definition

For a given sequence x of n samples, the DFT frequency components $X(k)$ may be defined by

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{n.k} \quad (1)$$

where $W_N = e^{-\frac{2j\pi}{N}}$ is the twiddle factor, n and k are respectively the time and frequency indexes, $0 \leq k \leq N-1$, $0 \leq n \leq N-1$ and N is the DFT length.

Let us consider $N = M.T$, $k = s + T.t$ and $n = l + M.m$, where M, T are integer and $s, l \in \{0, 1 \dots M-1\}$ and $t, m \in \{0, 1 \dots T-1\}$. Applying these considerations in (1), we obtain (2)

$$X(s + T.t) = \sum_{l=0}^{M-1} \sum_{m=0}^{T-1} \left[x(l + M.m)W_{M.T}^{((l+M.m)(s+T.t))} \right] \quad (2)$$

It can be found that (2) is equivalent to

$$X(s + T.t) = \sum_{l=0}^{M-1} \sum_{m=0}^{T-1} \left[x(l + M.m)W_{M.T}^{((l+M.m)(s+T.t))} \right] \quad (3)$$

And finally, (3) can be rewritten

$$X(s + T.t) = \sum_{l=0}^{M-1} W_M^{l.t} \sum_{m=0}^{T-1} \left[W_{M.T}^{l.s} x(l + M.m)W_T^{m.s} \right] \quad (4)$$

Equation (4) means that it is possible to realize N-point FFT by first decomposing into one M-point and one T-point FFT where $N = M.T$, and then combining them. To illustrate this by example, we take the 64-point as a case study after that we can make generalization to a higher order. To perform 64-point FFT we may choose $M = T = 8$. Then equation (4)

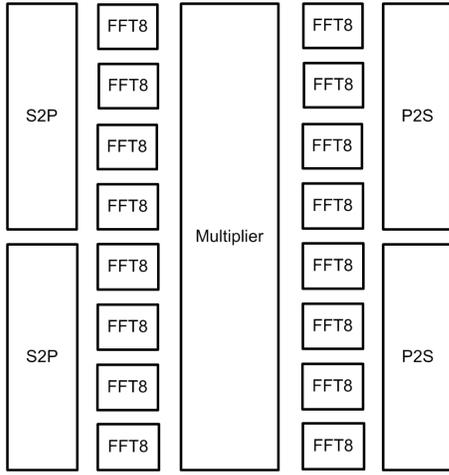


Fig. 1. Signal Flow Graph of the spatial distribution FFT architecture

can be written as in [4] by

$$X(s + 8.t) = \sum_{l=0}^7 W_8^{l.t} \sum_{m=0}^7 [W_{64}^{l.s} x(l + 8.m) W_8^{m.s}] \quad (5)$$

Equation (5) means that is possible to express the 64-point FFT by two-dimensional structure of 8-point FFT.

The processing element of higher order FFT according to equation (5) is the 8-point. Hence, the performance of high length depends in 8-point performance. The choice of 8-point FFT structure becomes crucial. In this work, the 8-point FFT architecture used is the Split Radix DIT because of its lower number of arithmetic operations.

B. Spatial distribution

One possible realization of the 64-point FFT is presented in the Signal Flow Graph (SFG) of Fig. 1. It can be observed that computing 64-point FFT is composed on five levels. The first level is composed of two serial to parallel blocks used to store real and imaginary part of data presented in a serial way. the second floor is composed of 8 blocks of 8-point FFT Split Radix DIT. The third block contains 49 complex multipliers used to compute non trivial complex multiplication. The fourth is similar to the second one. the last level is composed of two parallel to serial blocks gives data in a serial way.

At the 64th clock cycle all input data are ready to be proceeded. After 5 clock cycles, the 8-point FFT outputs are available and multiplication can be started. Block multiplier needs 2 clock cycles to perform the 49 complex multiplications. The 64-point FFT outputs are available 5 clock cycles after the last stage of 8-point FFT transformation.

Hence, the main advantage of this architecture is the high speed and low-latency. However, the implementation of this architecture on FPGA needs high memory, high number of complex multipliers and complex adders. Therefore, this architecture is not suitable for low cost FPGA such as Spartan 3 family.

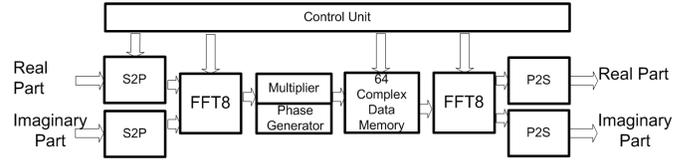


Fig. 2. Signal Flow Graph of the temporal distribution FFT architecture

C. Temporal distribution

Another possible realization of the 64-point FFT is illustrated in Fig. 2. According to this structure, the first stage is realized by one block of 8-point FFT rather than 8 as in Fig. 1. Similarly, the third stage is performed by only one block of 8-point FFT rather than 8. Consequently, the control unit in Fig. 2 plays an important role to synchronize all the treatments. This architecture performs FFT in a pipeline way.

First, input data comes in a serial manner. To perform the computation input data have to be parallelized. This is realized by S2P blocks which are implemented by means of delay registers.

On the other side, the control unit manages the input data addresses. The first 8-point input data has the address in the format $8j$, $j \in \{0, 1, \dots, 7\}$. On the 56th clock cycle these data have been proceeded to the first stage of 8-point FFT. After 5 clock cycles, the 8-point FFT outputs are available and multiplication can be started.

Similarly, on the 57th clock cycle, data indexed $8j + 1$ will be transformed by the first 8-point FFT and after 7 clock cycles, results data will be available at the multiplier output. And so one until the last result of multiplier output which will be available at the 71st clock cycle. These results are stored on the fly on 64-complex data memory.

Likewise, the second 8-point FFT stage will proceed the stored data to compute 64-point FFT.

D. Compromise analysis

Some concluding remarks related to this section have to be drawn. Firstly, decomposing a high length FFT to 8-point FFTs may be done in a spatial or in temporal distribution. In terms of throughput, the two distributions present one complex output per clock cycle since data have to be serialized by P2S component.

On the other hand, the latency which represents the elapsed time to get the first result is the same. In fact, for a given $N = 8^n$ where n is the number of stages, the latency in both architectures may be expressed as $L(N) = N + 7 \log_8 N - 2$. The main difference between the two distributions is the consumed area. Obviously, the second architecture consumes averagely 7 times less area than the first one. The number of 8-point FFT blocks pass from 16 to 2 and the number of non-trivial multiplier pass from 49 to 7. Furthermore, the complex data memory used in Fig. 2 may be avoided by storing the multiplier outputs on S2P registers. Indeed, since input data at address $8j$, $8j + 1$, .. are proceeded one can use these addresses to store the multiplier outputs.

Definitively, the major drawback of the decomposition of high length FFT on 8-point FFTs is related to the hardware consumed resources of the 8-point FFT. Synthesis results of the split radix DIT description of 8-point FFT show that the percentage of occupied slices in Spartan3E XC3S500 is about 30%. Therefore, to design a higher order FFT, the FPGA resources will be overflowed. Another drawback is about the limitation of the number of input with exclusively 8-point FFT elements since $N = 8^n$.

To overcome this problem we replace the 8-point FFT by a 4-point FFT using radix-4 algorithm. This choice is reinforced by the synthesis results of radix 4 in terms of slice occupation which is about 2%.

III. LOW AREA ARCHITECTURE

A. Definition

The N-point FFT equation can be split into three stages according to next equation

$$X(s+Mq+MKp) = \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} x(l, m, k) W_N^{(MKl+Mm+k)(MKp+Mq+s)} \quad (6)$$

For $N = 64$, one possible solution consists on constructing the 64-point FFT according to the temporal distribution by using 8-point, 4-point FFT and 2-point FFTs. The obtained design is not highly structured and inhomogeneous. The second solution consists in constructing the 64-point FFT by three stages of 4-point FFT. For $L = M = K = 4$, 64-point FFT equation can be written as

$$X(s+4q+16p) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{k=0}^3 x(l, m, k) W_{64}^{(16l+4m+k)(16p+4q+s)} \quad (7)$$

B. Optimizations

Using the radix-4 processing element, we can represent the 64-point FFT according to SFG in Fig. 3. The 64-point FFT is composed of a control unit, three blocks 4-point FFT units, two blocks multipliers units with two phase generator units and a complex 64-point memory unit. The control unit, indeed of managing the FFT4, multipliers and memorizing unit, it is used also to generate addresses of the inputs and the outputs of each block.

1) *Radix-4 modification*: Outputs of such algorithm are presented in next equations

$$\begin{aligned} X(0) &= \overbrace{x(0) + x(2)}^A + \overbrace{x(1) + x(3)}^C \\ X(1) &= \overbrace{x(0) - x(2)}^B - \overbrace{j(x(1) - x(3))}^D \\ X(2) &= x(0) + x(2) - x(1) - x(3) \\ X(3) &= x(0) - x(2) + jx(1) - jx(3) \end{aligned} \quad (8)$$

The SFG of the radix-4 structure is illustrated in Fig. 4. It is shown that radix-4 algorithm is composed of 8 complex additions/subtractions.

In order to reduce the number of complex multipliers, after each 4-point FFT and to keep the pipeline way in computation of the design we modify the 4-point FFT architecture.

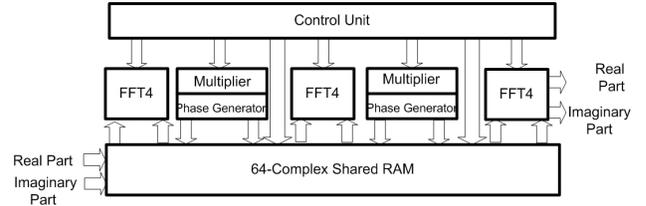


Fig. 3. Signal Flow Graph of the proposed low area 64-point FFT architecture

Usually, the radix-4 is computed as multi-inputs multi-outputs system. This structure requires 4 multipliers in one clock cycle. It is true that this structure presents a high speed design, but almost a P2S block is used to serialize data. For these reasons, we rectify the architecture in order to have one multiplier per clock cycle. So, the resulting design have one complex input and give one complex output per clock cycle as represented in Fig. 4. Intermediate signals A, B, C and D used in the diagram are indicated to understand the parallel computing.

2) *Sharing memory*: For each output of the 4-point FFT block the phase generator generates the correspondent twiddle factor and the multiplier unit performs the complex multiplication and stores the result on 64 complex data memory. This last will be reused and shared between all the blocks as it is shown on Fig. 3.

Usually, computing 64-point FFT based on 4-point FFT needs 3 complex memories. In our architecture we use only one complex 64-point. Moreover, this memory is divided into four small 16-point complex memories in order to improve the latency. Indeed, the problem behind this consists in using one shared memory with only one writer port. This is impossible since a part of data already saved in the memory are not used. Furthermore, if we use a dual port memory, this will be synthesized as BRAM blocks which are oversize and available in limited number in low cost FPGAs.

IV. EXPERIMENTAL RESULTS

A. Synthesis results

In table I some comparison results with recent works in terms of latency are illustrated. Functional verification is carried out using Xilinx ISE and FPGA implementation on Spartan 3E XC3S500 FPGA from Xilinx. In [4], authors have proposed a similar architecture as in Fig. 2 and obtained a latency of 79 against 76 in the proposed design of section II-C. This difference comes from the block multiplier which is implemented by delay registers in [4].

For the low area design of Fig. 3, we obtain better result than Xilinx IP core [5]. This is mainly due to the memory division into 4 small memories.

Regarding the consumed resources, operating frequency and power consumption some comparison has been made between our proposed architecture and Xilinx IP core and are presented in Table. II .

For the 64-point FFT, the consumed cell area by the proposed design is 29% smaller than consumed resources in Xilinx IP core. And, for the 256-point FFT our proposed BRAMless

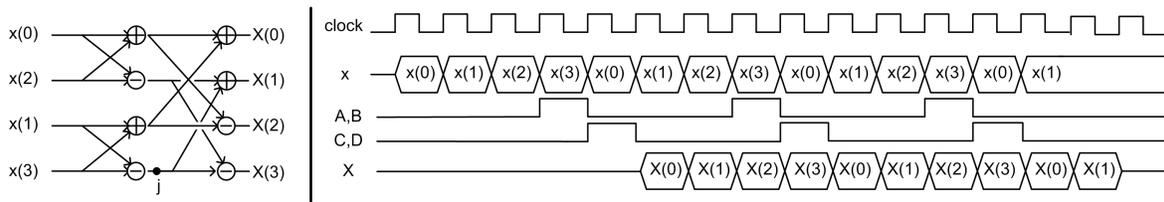


Fig. 4. SFG of the modified radix-4 algorithm and the corresponding timing diagram

TABLE I
64-POINT FFT LATENCY COMPARAISON

Low Latency		Low Area	
[4]	design of Fig. 2	Xilinx IP	design of Fig. 3
79	76	192	152

TABLE II
FPGA IMPLEMENTATION COMPARISON

	Proposed		Xilinx IP	
	64	256	64	256
Length	64	256	64	256
Slices	758	1110	1063	1702
Slices Flip Flops	1080	1442	1764	2654
BRAM	0	0	0	1
Mult18x18	8	12	8	12
Estimate static Power (mW)	76	76	76	76
Maximum Frquency (Mhz)	170	116	219	215

design consumes 35% less than the Xilinx FFT which has one BRAM block. In fact, the shared memory was divided into four small memories which are synthesized as distributed memories. These memories are implemented on LUTs without using BRAM.

B. Implementation results

In order to validate the proposed design of Fig. 3, we use a sine wave as test input vector. The frequency of the input signal is set to the quarter of the sampling frequency. To manage data in FPGA, we use Chipscope tool as in [6]. It can be observed in Fig. 5 the variation between matlab simulation and the real performance on FPGA. The mean residual power is equal to -0.1321 dB. Obviously, this is due to the quantization noise since we use a fixed point operators. It should be pointed out that the size of FFT outputs is fixed to 18 bits.

V. CONCLUSION

Techniques to implement high order FFT into low cost FPGAs were presented and validated. After a comprehensive and a comparative study of existing high order FFTs, an optimized architecture of 64-point FFT was proposed. The transition between 64-point and 256-point was exploited. Higher order FFTs could be obtained with the same manner. Our future work for the FPGA implementation will be devoted to the optimization of the block multiplier and the use of the method proposed in [7] to replace embedded multipliers.

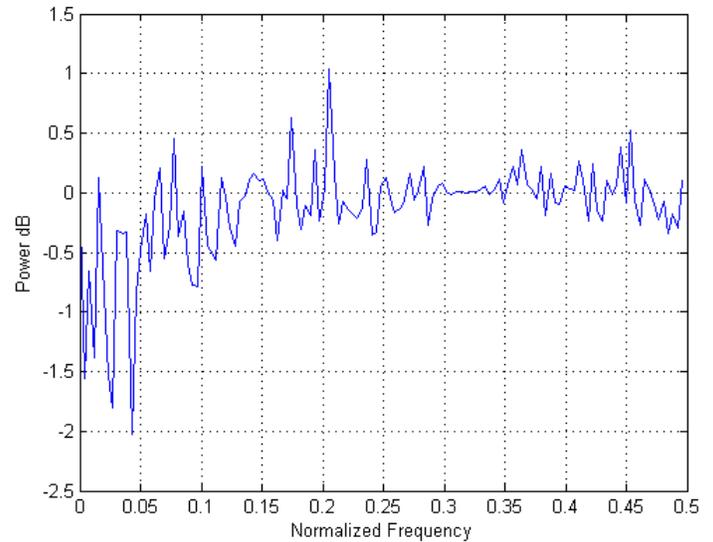


Fig. 5. Residual Power between matlab simulation and real performance on FPGA

REFERENCES

- [1] J. W. Cooley and J. Tukey, *An algorithm for the machine calculation of Complex Fourier series*, *Math. Comput.*, vol. 19, pp. 297-301, April 1965.
- [2] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [3] H. Sorensen, M. Heindeman, and C. Burrus, *On computing the split-radix FFT*, *IEEE Trans. Acoustics, Speech, Signal Process*, vol.34, pp. 152-156, 1986.
- [4] K. Maharatna, E. Grass, and Ulrich Jagldhold, *A 64-Point Fourier Transform Chip for High-Speed Wireless LAN Application Using OFDM*, *IEEE J. Solid-State Circuits*, vol. 39, pp. 484-493, March 2004.
- [5] Xilinx Product Specification, *High performance 64-point Complex FFT/IFFT V.7.0* June 2009 [online]. Available on: <http://www.xilinx.com/ipcenter>.
- [6] M. Jridi and A. Alfalou, *A Low-Power, High-Speed DCT architecture for image compression: principle and implementation*, in *Proc. VLSI Syst. in Chip Conf (VLSI-SoC)*, pp. 304-309, Sept 2010.
- [7] M. Jridi and A. Alfalou, *Direct Digital Frequency Synthesizer with CORDIC Algorithm and Taylor Series Approximation for Digital Receivers*, *Euro Journal of Scientific Research*, vol. 30, No. 4, pp. 542-553, 2009.