An Extensive Review on Fast Multiple-Constant Convolution Circuit

Saim Malik[1], Sharad Mohan Shrivastava[2]

[1]*Mtech. Scholar,* [2]*Reearch Guide*

Department of Electronics and Communication Engineering, SISTEC, Bhopal

Abstract - Many of the algorithms like FFT for high-quality media contents use multiplexers or ROMs or multipliers to compute twiddle multiplications. The above algorithms use Distributed Arithmetic (DA) technique and have the disadvantages of consuming more hardware and power. DSP system design techniques such as folding, pipelining have always improved performance of the systems in terms of hardware, latency, frequency, etc. An FPGA can be configured and reconfigured for different applications which provide precise timing and synchronization, simultaneous execution of parallel task, and rapid decision making. The complete removal of the multiplier circuitry is by far the preferred choice of circuit designers. The distributed arithmetic (DA) method can be successfully applied in order to partition multiplications in simpler shifts and additions. This work presents an extensive survey of literature to improve the performances of DA as well as the performances of general- purpose multipliers in the specific contexts of MCM using Weighted Partitioning.

Keywords- MCM, Fast Multiplierless circuit, Constant Convolution, Weighted Partitioning.

I. INTRODUCTION

Fast Fourier Transforms (FFT) [1] is an algorithm for speedy calculation of Discrete Fourier transform (DFT) of an input data vector used in various signal and image processing applications [3]. The FFT is nothing but a DFT algorithm which reduces the number of computations needed for N points from $O(N^2)$ to O(N log N) where log is the base-2 logarithm using periodicity and property.

In DSP architectures, systematic control circuits are determined by using the folding transformation. In folding technique, time multiplexing of algorithm operations is done, by reducing to a single functional unit. Thus, in DSP systems, folding technique provides a means of trading time for area. Conventional folding technique can be used to reduce the number of hardware functional units by a factor of N at the expense of increasing the computation time or multiplexing time by a factor of N. This technique also helps in data allocation in the required registers. To avoid excess amount of registers that are generated in these architectures while folding, there are techniques to minimize the number of registers needed to implement DSP architectures through folding.

The prime task in Digital Signal Processing (DSP) is modifying a signal mathematically to improve it and make different operations easier for it. It measures, compresses or filters real world analog signals. In current world, Digital Signal Processing (DSP) has become a necessary technology and has taken the place of many traditional analog signal processing systems. There are several advantages of DSP systems such as indifference to fluctuations in temperature, aging effect and component tolerance. Another merit of digital systems is that digital designs can be more densely integrated than analog designs as inferred from past years.

Technological developments have caused an acceleration in the development of the area of DSP. One of them is the devising of an efficient algorithm to calculate the Discrete Fourier Transform. The introduction of programmable digital signal processors (PDSPs) in the late 1970s was another major step in this field. These were able to perform “Multiplication and Accumulation” (MAC) in one clock cycle . It was an important improvement in the “Von Neumann” microprocessor based systems in those years. More improved functions such as memory bank, floating point multipliers, or zero overhead interface to ADC and DAC are included in the modern PDSPs. Digital Signal Processing finds major use in decoding, coding, image compression , audio and speech processing .

In current scenario, there are immense developments in the implementation of Digital Signal Processing (DSP). Discrete Cosine Transform (DCT), Discrete Fourier transform (DFT) and Discrete Wavelet Transform (DWT) are some of the common algorithmic implementations. FPGAs have an edge over DSPs in the way that due to the sequential based architecture, the performance of DSPs are sometimes not up to the mark but FPGAs are efficient as they possess both sequential and concurrent architecture. Discrete Wavelet Transform (DWT) is used for denoising of signals, compression of images and in video processing. Filter banks are used by DWT especially in the case of designing FIR Filters. FIR filters are used widely due to their linear phase and stability. All frequency components in linear phase filters tend to have equal time delay since such filters have constant group delay. A filter having non-linear phase results in phase distortion.

In order to reduce hardware cost, all the bits of an input data can be processed in a parallel manner using Look Up tables. But in Distributed Arithmetic structure, bitwise operation is performed. The input data is processed bit by bit, first processing the least significant bit and then rest of the bits. There is substantial reduction in hardware as all bits have to pass through same architecture. In other words, if there is a N-bit parallel design, the DA method requires only (1/N)th of the hardware resources. As a result only one clock cycle is required for execution while about N cycles are required in serial execution. However, for the serial design the time-hardware product is smaller than the parallel design because the propagation delays are generally smaller as compared to the parallel structure.

II. WEIGHTED PARTITIONING

Given an un-weighted graph G with V vertices and E edges and given a number k, the Graph Partitioning problem is to divide the V vertices into k parts such that the number of edges connecting vertices in different parts is minimized given the condition that each part contains roughly the same number of vertices. If the graph is weighted, i.e. the vertices and edges have weights associated with them; the problem requires the sum of the weights of the edges connecting vertices in different parts to be minimized given the condition that the sum of the weights of the vertices in each part is roughly the same. The problem can be reduced into that of bisection where the graph is split into two parts and then each part is further bisected using the same procedure recursively. The problem addressed in this work is that of bisecting the given graph according to a given ratio.

Also, the input graph is assumed to be un-weighted. However, this assumption is just at the implementation level and does not in any way change the underlying algorithms.

It has been shown that the Graph Partitioning problem is NP-hard [1] and so heuristic based methods have been employed to get sub-optimal solutions. The goal for each heuristic method is to get the smallest possible cut in reasonable time. Graph Partitioning is an important problem since it finds extensive applications in many areas, including scientific computing, VLSI design and task scheduling. One important application is the reordering of sparse matrices prior to factorization. It has been shown that the reordering of rows and columns of a sparse matrix can reduce the amount of fill that is caused during factorization and thus result in a significant reduction in the floating point operations required during factorization. Although the ideal thing is to find a node separator rather than an edge separator, an edge separator can be converted into a node separator using minimum cover methods.

The multi-level graph partitioning algorithm reduces the size of the graph gradually by collapsing vertices and edges over various levels, partitions the smallest graph and then uncoarsens it to construct a partition for the original graph. Also, at each step of uncoarsening the partition is refined as the degree of freedom increases. In this work the multi-level graph partitioning algorithm has been discussed the Fiduccia Mattheyses algorithm for refining the partition at each level of un-coarsening. The smallest graph is cut and the partition gets projected and refined as it moves up to the original biggest graph as demonstrated in figure 2.1.

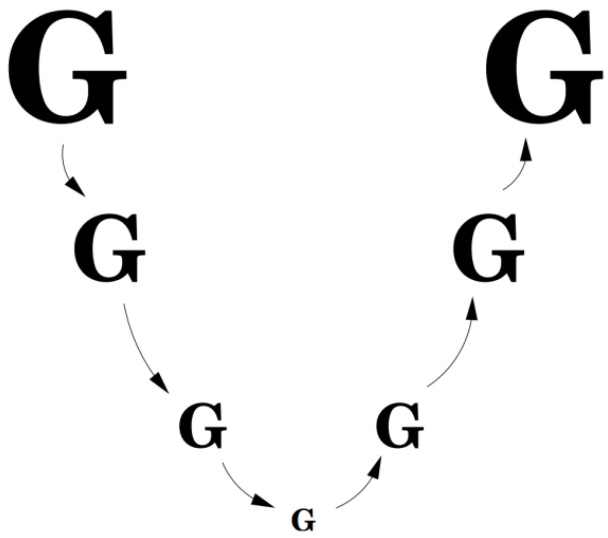

Figure 2.1 Multi- Level Coarsening and Uncoarsening.

In the multi-level approach the coarsening phase is important. If a graph is folded in such a way that the properties of the graph are preserved i.e. the coarse graph is a smaller replica of the fine graph, a good cut of the coarse graph translates into a good cut of the fine graph [2]. It is possible that the properties of the graph are not preserved when handling dense nodes and may result in a very unbalanced coarse graph. Present a heuristic for handling dense nodes that preserves the properties of the graph during coarsening and gives a more balanced coarse graph in fewer coarsening steps.

III. RELATED WORK

G. D. Licciardo, C. Cappetta, L. Di Benedetto and M. Vigliar,[1] A new radix-3 partitioning method of natural numbers, derived by the weight partition theory, is employed to build a multiplierless circuit that is well suited for multimedia filtering applications. The partitioning method allows conveniently premultiplying 32-b floating-point filter coefficients with the smallest set of parts composing an unsigned integer input. In this way, similar to the distributed arithmetic, shifters and recoding circuitry, typical of other well-known multiplier circuits,

are completely substituted with simplified floating-point adders. Compared to the existent literature, targeted to both field-programmable gate array and std_cell technology, the reported solution achieves state-of-the-art performances in terms of elaboration velocity, achieving a critical path delay of about 2 ns both on a Xilinx Virtex 7 and with CMOS 90-nm std_cells.

X. Y. Shih, Y. Q. Liu and H. R. Chou, [2] In this research, reported a reconfigurable (RC) fast Fourier transform (FFT) design in a systematic design scheme. The RC design bricks are mainly reported to arbitrarily concatenate to support FFT-point required. Meanwhile, it show three developed design techniques, including six-type RC processing element, systematic first-in first-out reuse arrangement, and section-based twiddle factor generator to elaborate our FFT design. In a design/implementation example, it can support up to 2187 FFT-point manipulation and 48 RC modes. It also supports 32 operating modes defined in 3GPP-LTE standard. In application-specified integrated circuit implementation with TSMC 90-nm CMOS technology, our design work occupies a core area of 1.664 mm2 and consumes 35.2 mW under maximal clock frequency of 188.67 MHz. This research work also has outstanding design performance in terms of speed-area ratio and power-frequency ratio for comparison reference.

Table 1: Summary of Related Work

SR NO.	TITLE	AUTHORS	YEAR	APPROACH
1	Partitioning for Fast Multiplierless Multiple-Constant Convolution Circuit,	G. D. Licciardo, C. Cappetta, L. Di Benedetto and M. Vigliar	2017	A new radix-3 partitioning method of natural numbers, derived by the weight partition theory
2	48-Mode Reconfigurable Design of SDF FFT Hardware Architecture Using Radix-32 and Radix-23 Design Approaches,	X. Y. Shih, Y. Q. Liu and H. R. Chou	2017	A reconfigurable (RC) fast Fourier transform (FFT) design in a systematic design scheme
3	Optimization design approach for multiplierless FIR filter	Ranjithkumar S and Thilagam S	2016	Consider the filter order and sparsity of filter to implement the FIR Filter
4	Design of multiplierless cosine modulated filterbank using hybrid technique in sub-expression space	I. Sharma, A. Kumar, G. K. Singh and H. N. Lee	2016	An optimal design technique for M-channel multiplierless cosine modulated filterbank (CMFB)
5	Truncated ternary multipliers	B. Parhami	2015	based on the symmetric radix-3 digit set {-1, 0, +1},
6	A Novel Hybrid Radix-3/Radix-2 SAR ADC With Fast Convergence and Low Hardware Complexity	M. Rahman, A. Sanyal and N. Sun	2015	(SAR) analog-to-digital converter (ADC) based on the radix-3 and radix-2 search approaches
7	Weighted partitioning of sequential processing chains for dynamically reconfigurable FPGAS,	M. Feilen, A. Iliopoulos, M. Vonbun and W. Stechele	2013	A metric for weighted partitioning of pre-defined processing element sequences
8	A Weighted Partitioning Dynamic Clustering Algorithm for Quantitative Feature Data Based on Adaptive Euclidean Distances	F. d. A. T. d. Carvalho and L. D. S. Pacifico	2008	An iterative four-steps relocation algorithm involving the determination of the clusters representatives (prototypes),

Ranjithkumar S and Thilagam S, [3] In this exploration FIR filter is designed by two algorithms. Consider the filter order and sparsity of filter to implement the FIR Filter. The problem identification handles by the l0-norm minimization which can be implemented by two different algorithms. In first design approach the upper bound value taken into consideration which leads to weighted l0-norm minimization problem, in the second one a problem transformed into another weighted l0-norm minimization. Weighted l0-norm minimization problem can be tackled by the IRLS algorithm to solve the problem. Both Sparsity and filter order design approach provides the best FIR Filter Design. From this reduced coefficient set, it can implement multiplierless FIR filter to achieve the reduced hardware design.

I. Sharma, A. Kumar, G. K. Singh and H. N. Lee, [4] In this work, an optimal design technique for M-channel

multiplierless cosine modulated filterbank (CMFB) is reported using common sub-expression technique (CSE) and hybrid method with given roll-off factor (RF) and stopband attenuation (As). The key feature of the reported method is utilization of single optimization algorithm to generate optimal quantized and canonic signed digit (CSD) converted coefficients that satisfy the magnitude response of 0.7071 at frequency $\omega = \pi /2M$. CSE is employed to reduce the hardware requirement (adders) of a designed filter. Hybrid technique is based on the concept of particle swarm optimization (PSO) and artificial bee colony (ABC) algorithm. A comparative analysis of different CSE algorithms has been made, and performance of the reported method is evaluated in term of adders.

B. Parhami, [5] Balanced ternary number representation and arithmetic, based on the symmetric radix-3 digit set {-1, 0, +1}, has been studied at various times in the history of computing. Among established advantages of balanced ternary arithmetic are representational symmetry, favourable error characteristics and rounding by truncation. In this study, show an additional advantage: that of lower-error truncated multiplication with the same relative cost reduction as in truncated binary multipliers.

M. Rahman, A. Sanyal and N. Sun, [6] this brief presents a fast-converging hybrid successive approximation register (SAR) analog-to-digital converter (ADC) based on the radix-3 and radix-2 search approaches. The radix-3 approach achieves 1.6 bits/cycle, and the radix-2 approach mitigates the effect of comparator offset and improves the accuracy of the ADC. Incorporating clock gating of comparators and efficient switching of capacitors, the reported hybrid ADC demonstrates promising balance between hardware complexity and speed and can achieve equivalent signal-to-noise-and-distortion-ratio (SNDR) with less capacitors compared with radix-3 SAR ADC. Behavioral simulation-based results verify operation and merit of the reported architecture. An 11-bit 45-MS/s prototype with 5% capacitor mismatch in 180-nm CMOS was simulated in SPICE and achieves 67 dB of SNDR after calibration.

M. Feilen, A. Iliopoulos, M. Vonbun and W. Stechele,[7] Temporal runtime-reconfiguration of FPGAs allows for a resource-efficient sequential execution of signal processing modules. Approaches for partitioning processing chains into modules have been derived in various previous works. Present a metric for weighted partitioning of pre-defined processing element sequences. The reported method yields a set of reconfigurable partitions, which are balanced in terms of resources, while jointly have a minimal data throughput. Using this metric, formulate a partitioning algorithm with linear complexity and will compare our approach to the state of the art.

F. d. A. T. d. Carvalho and L. D. S. Pacifico, [8] this work introduces a weighted partitioning dynamic clustering algorithm for quantitative feature data based on adaptive euclidean distances. The reported method is an iterative four-steps relocation algorithm involving the determination of the clusters representatives (prototypes), the weight of each individual, the distance associated to each cluster and the construction of the clusters, at each iteration. Moreover, the algorithm furnishes automatically the best weight of each individual in such a way that as close it is an individual from the prototype of the cluster it belongs as high it is its weight. Experiments with real and synthetic datasets show the usefulness of the reported method.

IV. PROBLEM STATEMENT

Matrix multiplication is widely used as core operation in various signal processing application like software defined radio. The FFT processor is widely used in DSP and communication applications. It is critical block in the OFDM (Orthogonal Frequency Division Multiplexing) based communication systems, such as WLAN and MC-CDMA receiver. Recently, both high data processing and high power efficiency consumes more power. Due to the nature of non-stop processing at the sample rate, the FFT appears to be the leading architecture for high performance applications. Since these two functions are widely used in various mobile devices they required to have features like low power, lesser area without increase of latency. The design and portable systems requires critical consideration for the averaged power consumption which is directly proportional to the battery weight and volume required for a given amount of time. Most of these portable applications demands high speed computation, complex functionality and often real time processing capabilities with the low power consumption. Portable devices like cellular phones, pagers, wireless modems and laptops along with the limitation of the technology have elevated power.

V. CONCLUSION

This work discusses literature survey on fast multiplierless multiple-constant convolution circuit based on weighted partitioning. Optimization of parameters at high level multiplication and Fast Fourier Transform designs optimized at algorithm and architecture level by using energy efficient modeling technique. Because of the wide applications and the difficulty of general graph partition problems, extensive research has been done on general graph partition problems and its variations, including both exact and heuristic algorithms. But surprisingly, a little literature found on partitions considering the physical distance between nodes. The underlying logic is that the nodes or locations that are geographically proximal should be in the same partition.

REFERENCES

[1] G. D. Licciardo, C. Cappetta, L. Di Benedetto and M. Vigliar, "Weighted Partitioning for Fast Multiplierless Multiple-Constant Convolution Circuit," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 64, no. 1, pp. 66-70, Jan. 2017.

[2] X. Y. Shih, Y. Q. Liu and H. R. Chou, "48-Mode Reconfigurable Design of SDF FFT Hardware Architecture Using Radix-32 and Radix-23 Design Approaches," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 64, no. 6, pp. 1456-1467, June 2017.

[3] Ranjithkumar S and Thilagam S, "Optimization design approach for multiplierless FIR filter," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, 2016, pp. 1-6.

[4] I. Sharma, A. Kumar, G. K. Singh and H. N. Lee, "Design of multiplierless cosine modulated filterbank using hybrid technique in sub-expression space," 2016 IEEE International Conference on Digital Signal Processing (DSP), Beijing, 2016, pp. 360-364.

[5] B. Parhami, "Truncated ternary multipliers," in IET Computers & Digital Techniques, vol. 9, no. 2, pp. 101-105, 3 2015.

[6] M. Rahman, A. Sanyal and N. Sun, "A Novel Hybrid Radix-3/Radix-2 SAR ADC With Fast Convergence and Low Hardware Complexity," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 5, pp. 426-430, May 2015.

[7] M. Feilen, A. Iliopoulos, M. Vonbun and W. Stechele, "Weighted partitioning of sequential processing chains for dynamically reconfigurable FPGAS," 2013 23rd International Conference on Field programmable Logic and Applications, Porto, 2013, pp. 1-8.

[8] F. d. A. T. d. Carvalho and L. D. S. Pacifico, "A Weighted Partitioning Dynamic Clustering Algorithm for Quantitative Feature Data Based on Adaptive Euclidean Distances," 2008 Eighth International Conference on Hybrid Intelligent Systems, Barcelona, 2008, pp. 398-403.

[9] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers (Sixth Edition). New York, NY, USA: Oxford Univ. Press, 2008.

[10] S. K. Park, “The r-complete partitions,” Discrete Math., vol. 183, no. 1–3, pp. 293–297, Mar. 1998.

[11] Ø. J. Rødseth, “Enumeration of M-partitions,” Discrete Math., vol. 306, no. 7, pp. 694–698, Apr. 2006.

[12] S. Arish and R. K. Sharma, “An efficient floating point multi- plier design for high speed applications using Karatsuba algorithm and Urdvha—Tyriagbhyam algorithm,” in Proc. ICSC, Noidia, India, Mar. 2015, pp. 303–308.

[13] W. M. Chao and L. G. Chen, “Pyramid architecture for 3840 2160 quad full high definition 30 frames/s video acquisition,” IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 11, pp. 1499–1508, Nov. 2010.

[14] M. Potkonjak and M. B. Srivastava, “Multiple constant multiplications: Efficient and versatile framework and algorithms for exploring common subexpression elimination,” IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 15, no. 2, pp. 151–165, Feb. 1996.