Implementation of Floating Point Multiplier Using VHDL

Bhagyashree Hardiya
Student, Dept. of ECE
Institute of Engineering and Science, IPS Academy
Indore (MP), India
Hardia.golchi@gmail.com

Virendra Singh Rathore
Dept of Electronics and Communication Engg.
Institute of Engineering and Science, IPS Academy
Indore (MP), India
Virendra883@gmail.com

***Abstract*— In this paper, multiplication of the floating point numbers described in IEEE 754 single precision floating point multiplier is done using VHDL .Implementation in VHDL(VHSIC Hardware Description Language) is used because it allow direct implementation on the hardware while in other language we have to convert them into HDL then only can be implemented on the hardware. In floating point multiplication, adding of the two numbers is done with the help of various types of adders but for multiplication some extra shifting is needed.This floating point multiplication handles various conditions like overflow, underflow, normalization, rounding. In this paper we use IEEE rounding method for perform the rounding of the resulted number.This paper reviews the implementation of an IEEE 754 single precision floating point multiplier developed by many researchers.**

***Keywords*— Floating point unit, XILINX ISE 8.1i, model-sim, Floating point arithmetic, Booth multiplier, IEEE rounding method, serial by parallel adder.**

I. INTRODUCTION

In IEEE 754 [1] Floating point is the binary representation of the real numbers. There are many ways to represent the number system however non-integer representation has gained widespread use i.e. floating point. The typical floating point number can be represented exactly is of the form:

$$Significant\ digits \times base^{exponent} \quad (1)$$

A. *Fixed Point Representation*

A fixed point number representation is a real data type for a number that has a fixed number of digits after the decimal point ('.'). A value of a fixed-point data type is essentially an integer that is scaled by a specific factor determined by the type.The scaling factor is usually a power of 10 (for human convenience) or a power of 2 (for computational efficiency). However, other scaling factors may be used occasionally.

B. Floating Point Representation

The term floating point refers to the fact that a number's decimal point can "float"; that is, it can be placed anywhere relative to the significant digits of the number. This position is indicated as the exponent component in the internal representation, and floating point can thus be thought of as a computer realization of scientific notation. As an example, an exponent of (−3) and significand of 1.5 might represent the number 1.5 × 2^3= 0.1875.

Floating Point Format

1 bit	8 bit	23 bits

Sign bit **exponent** **mantissa**

Floating point number wordlength[2] may be of two types :

1. Single precision floating point number consists of 32 bits
2. Double precision floating point number consists of 64 bits.

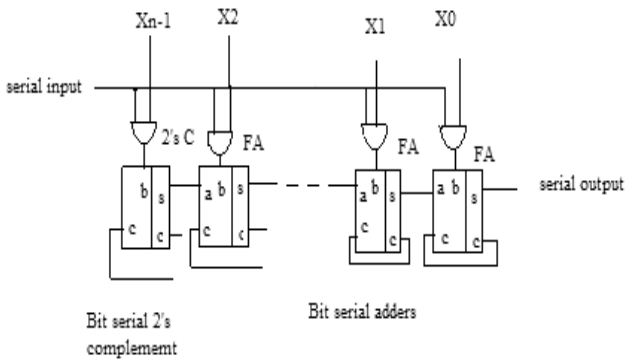
The format for single precision floating point number is shown in figure above.

In this paper we make use of only single precision floating point multiplier because of less complexity .The exponent is a signed number represented using the bias method with a bias of 127. The term biased exponent refers to the unsigned number contained in bits 1 through 8 and unbiased exponent means the actual power to which 2 is to be raised. The fraction represents a number less than 1, but the significand of the floating-point number is 1 plus the fraction part. In other words, if e is the biased exponent and f is the value of the fraction field, the number being represented as:

$$1.f * 2e–127 \quad (2)$$

C. Serial by parallel Booth Multiplier

The common multiplication method is “add and shift” algorithm. To reduce the number of partial products to be added, Modified Booth algorithm [2] [3] is one of the most popular algorithms. The simple serial by parallel booth multiplier is particularly well suited for bit serial processors implemented in FPGAs without carry chains because all of its routing is to nearest neighbours with the exception of the input. The serial input must be sign extended to a length equal to the sum of the lengths of the serial input and parallel input to avoid overflow, which means this multiplier takes more clocks to complete than the scaling accumulator version. This is the structure used in the venerable TTL serial by parallel multiplier.



Serial by parallel booth multiplier

The general architecture of the serial/parallel multiplier is shown in the figure above. One operand is fed to the circuit in parallel while the other is serial. N partial products are formed each cycle. On successive cycles, each cycle does the addition of one column of the multiplication table of M*N PPs. The final results are stored in the output register after N+M cycles. While the area required is N-1 for M=N.

D. IEEE Rounding method

Four rounding modes are dictated by the IEEE 754 standard: the round to RN mode, the round to positive infinity (RP) mode, the round-to-minus infinity (RM) mode, and the RZ mode. The RP mode can be implemented as the RI mode for positive numbers and the RZ mode for negative numbers. Similarly, the RM mode can be implemented as the RZ mode for positive numbers and as the RI mode for negative numbers.

To perform IEEE rounding [4] using a conventional algorithm, we have to perform the following:

Step 1) C and S computation: Compute C and S to a precision of 2N bits by generating the partial products and reducing them with a partial product reduction network.

Step 2) R term computation: Compute R by adding up C and S using a CPA.

Step 3) Pre-round normalization: Normalize R if R.m=1by right shifting R by 1-bit and adjusting the exponent appropriately.

Step 4) Rounding bits computation: Compute the g and s bits.

Step 5) Rounding: Denote the higher order N bits of R as Rh. Based on the rounding mode and rounding bits, add a rounding one unit-in-the-last-place (ulp) to Rh when necessary.

Step 6) Post-round normalization: If Rh.c=1 as a result of rounding, another 1-bit normalization right shift of Rh is needed, again with the exponent adjusted accordingly. The term Rh is the FCR SV.

Notation

The C and S and terms have 2N bits. The lower order N-1 bits are among the bits that the hardware needs to examine for rounding. The higher order N+1 bits are used to produce the FCR N bit significand; the extra bit is required to account for the 1-bit normalization right shift. For notational simplicity, we define the binary point to be at N-1th bit of C and S. As shown in Fig. 1, the higher order-bit N+1 integer portions of S and C are denoted as Sf and Cf, respectively, and the lower order-bit fractions as and, respectively.

II. Literature Review

Various researchers had contributed in the field of floating point multiplier using VHDL. Some of them are as follows:

a. Paper Title: An Efficient Implementation of Floating Point Multiplier[5]
Publication**:** *International Journal of Engineering Research & Technology (IJERT)*
*A*uthor Name:Mohamed Al-Ashrafy, Ashraf Salem, Wagdy Anis.

This paper presents a implementation of a floating point multiplier that supports the IEEE 754-2008 binary interchange format; the multiplier doesn"t implement rounding and just presents the significand multiplication result as is (48 bits); this gives better precision if the whole 48 bits are utilized in another unit; i.e. a floating point adder to form a MAC unit. The design has three pipelining stages and after implementation on a Xilinx Virtex5 FPGA it achieves 301 MFLOPs.

b. Paper Title: Pipeline Floating Point ALU Design using VHDL[6]
Publication: *ICSE2002 Proc. 2002, Penang, Malaysia*
Author Name: Mamu Bin Ibne Reaz, Md. Shabiul Islam, Mohd. **S.** Sulaiman.

In this project, pipelined floating point multiplication is divided into three stages . Stage **1 checks** whether the operand is 0 and report the result accordingly. Stage 2 determines the product sign, add exponents and multiply fractions. Stage3 normalize and concatenate the product.

c. Paper Title: Optimizing Floating Point Units in Hybrid FPGAs[7]
Publication: IEEE transactions on very large scale integration (VLSI) systems, vol. 20, no. 7, July 2012.
Author Name: ChiWai Yu, Alastair M. Smith, Wayne Luk, Fellow, IEEE, Philip H. W. Leong, Senior Member, IEEE, and Steven J. E. Wilton, Senior Member, IEEE

This paper proposes a novel methodology to determine optimized coarse-grained FPUs in hybrid FPGAs, based on common subgraph extraction. The effect of merging different coarse-grained types into larger FPUs is also studied.

We observe that:

1) The speed of the system is the highest for implementations involving only floating point adders and floating point multiplier.
2) Higher density subgraphs produce greater reductionon area.

3) They provide the best area-delay product.
4) Merging of FPUs can improve the speed of hybrid FPGAs, but results in consuming more area.

III. RESULT

In this paper, all modules are verified in bottom-up approach by simulation through separate test-benches and then test complete design as whole. The results of operation are manually computed and then compared with the simulation results. The generated outputs are in binary form. All binary values are converted back to the decimal format.

IV.CONCLUSION

In this paper, we have seen that the multipliers play an important role in today's digital signal processing and various other applications [8]. With advances in technology, many researchers have tried and are trying to design multipliers which offer either of the following design targets – high speed, low power consumption, regularity of layout and hence less area or even combination of them in one multiplier thus making them suitable for various high speed, low power and compact VLSI implementation. By using serial by parallel Booth multiplier we see that in parallel multipliers number of partial products to be added is the main parameter that determines the performance of the multiplier. To reduce the number of partial products to be added, Modified Booth algorithm is one of the most popular algorithms. Rounding of the resulted number provide a precision multiplication of the numbers by using IEEE rounding method.

V.REFERENCES

[1] IEEE 754-2008, IEEE Standard for Floating-Point Arithmetic, 2008.
[2] L. Song, K.K. Parhi, "Efficient Finite Field Serial/Parallel Multiplication", Proc. of International Conf. On Application Specific Systems, Architectures and Processors, pp. 72-82, Chicago,USA, 1996.
[3] P. E. Madrid, B. Millar, and E. E. Swartzlander, "Modified Booth algorithm for high radix fixed- point multiplication," IEEE Trans. VLSI Syst., vol. 1 , no. 2, pp. 164-167, June 1993
[4] Nhon T. Quach, Member, IEEE, Naofumi Takagi, Senior Member, IEEE, and Michael J. Flynn, Fellow, IEEE" Systematic IEEE Rounding Method for High-Speed Floating-Point Multipliers" IEEE transactions on very large scale integration (VLSI) systems, vol. 12, no. 5, may 2004.
[5] Mohamed Al-Ashrafy, Ashraf Salem, Wagdy Anis, "An Efficient Implementation of Floating Point Multiplier", *International Journal of Engineering Research & Technology (IJERT)*
[6] Mamu Bin Ibne Reaz, MEEE, Md. Shabiul Islam, MEEE, Mohd. S. Sulaiman, MEEE Faculty of Engineering, Multimedia University, 63 100 Cybejaya, Selangor, Malaysia "Pipeline Floating Point ALU Design using VHDL "ICSE2002 Proc. 2002, Penang, Malaysia.
[7] C. W. Yu, A. M. Smith, W. Luk, P. H. W. Leong, and S. J. E. Wilton, "Optimizing coarse-grained units in floating point hybrid FPGA," inProc. ICFPT, 2008, pp. 57–64.
[8] John G. Proakis and Dimitris G. Manolakis (1996), "Digital Signal Processing: Principles, Algorithms and Applications", Third Edition.