International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.5., 2015
(Sept.-Oct.)

RESEARCH ARTICLE

ISSN: 2321-7758

LOW POWER FIR FILTERS USING IMPRECISE ADDERS

SUNDARA K CHAKRAVARTHY M[1],K RAJASEKHAR[2]

[1]PG Scholar, Department of Electronics and Communications
University college of Engineering, JNTUKakinada, Kakinada, Andhra Pradesh
[2]Assistant Professor, Department of Electronics and Communications
University college of Engineering, JNTUKakinada, Kakinada, Andhra Pradesh.

ABSTRACT

Human beings gather information from erroneous outputs. Hence the output of the system need not be accurate but relatively accurate or imprecise. Designing Imprecise adders can greatly reduce the power, area and delay of the device. The approach mentioned performs the reduction in transistor level. Various imprecise adders are applied in the design of FIR filter and relevant power saving levels are calculated. An approximate Ripple Carry Adder of various approximation levels are designed and the error model is calculated. The adder is implemented on 3 different FIR filter models using 16- bit audio signals and the SNR value is calculated.

Keywords—ripple carry adder, low power, mirror adder.

INTRODUCTION

Multimedia DSP algorithms mostly consist of additions and multiplications. Multiplications can be treated as shifts and additions, thus, adders can be considered as basic building blocks for these algorithms. Interestingly, most DSP algorithms used in multimedia systems are characterized by inherent error tolerance. Hence, occasional errors in intermediate outputs might not manifest as a substantial reduction in the final output quality. Use approximate FA cells only in the LSBs, thus ensuring that the final output quality does not degrade too much. Most of the previous techniques have worked on either the logic level, algorithmic level or the gate level. Here the operation is performed on the Transistor level, which is easy to perform and replicate producing better results. As the transistor count reduces the total area reduces and the delay from the input to the output decreases too. Similarly the capacitance value due to the transistors also reduces which results in the power reduction of the entire system. When these FA cells are used in the construction of large DSP applications, a consider number of transistors gets decreased and hence a large power is saved. If the output error is considerable this technique is preferred.

The contributions in the paper are listed as:

1. A mirror adder which is accurate is established for comparison with the remaining imprecise adders
2. Various approximations have been established with reduction in transistor count, their truth tables are constructed and verified with accurate adder.
3. The area and the capacitance values of individual imprecise adders are listed.
4. The mean error and variance are plotted with respect to number of LSB's that are replaced with approximate adders.

5. An FIR filter is designed based on the Multiple Constant Multiplication approach. The adders are then replaced with the imprecise adders. Outputs of accurate adder is compared with all the imprecise adders and the pass band and stop band errors are calculated and listed. The power saving factor is also calculated.
6. An imprecise Ripple carry adder with 32 bits is constructed and the normalized error due to fifth approximation on various set of LSB's is plotted. Hence calculated the maximum number of LSB, which allows the output with allowable error value.
7. The SNR value for various approximated FIR filters is calculated with comparison on the standard filter on a set of audio signals.

IMPRECISE ADDERS AND THEIR ERRORS

Imprecise Adders

A conventional Full adder is designed in the mirror adder format. Five different imprecise adders are deduced from the conventional adder by removing some of its transistors. Here the main objective is to avoid short or open circuits for any set of input combinations and to reduce the size of the structure. The structures and their respective truth tables are as shown.

The conventional Mirror Adder contains 24 MOS transistors, which provide perfect output of sum and carry out for a input combination. A set of transistors have been removed to deduce a new imprecise adder. Similarly four more imprecise adders are designed.

The number of transistors in conventional and five imprecise adders are 24, 18, 14, 11, 11 respectively.

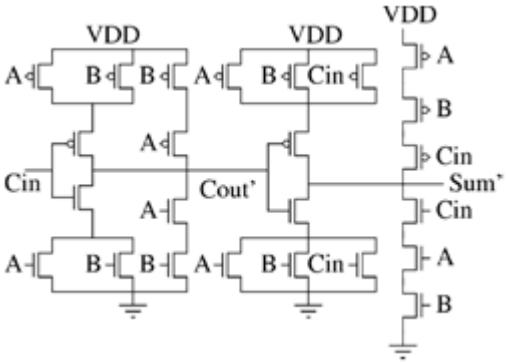


Fig 1: Conventional Mirror Adder

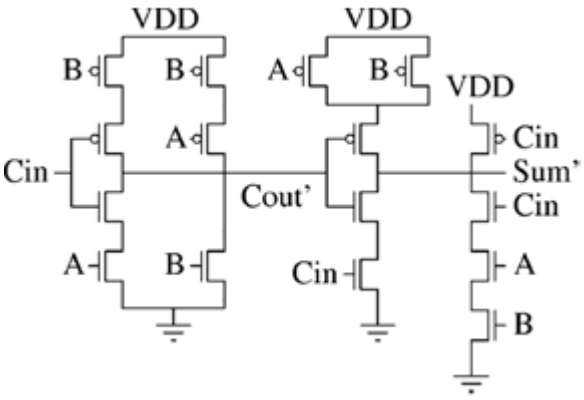


Fig 2: Imprecise Adder 1

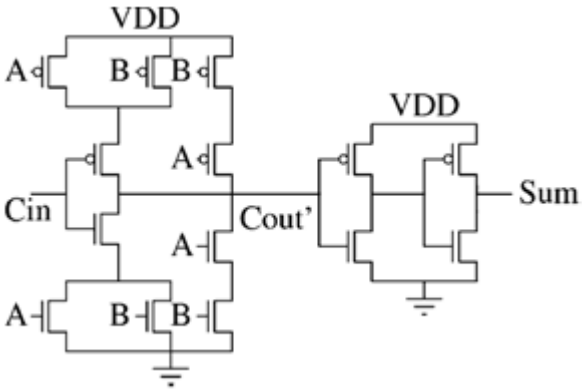


Fig 3: Imprecise Adder 2

The fifth imprecise adder is designed in such a manner that the output directly follows one of the inputs. It means it uses buffers only. Here Sum=A and Carryout = A or Sum = B and Carryout = A.

The area of the different adder structures are based using IBM - 90nm technology. The fifth approximation gets the minimum area value as it is designed using buffers only.

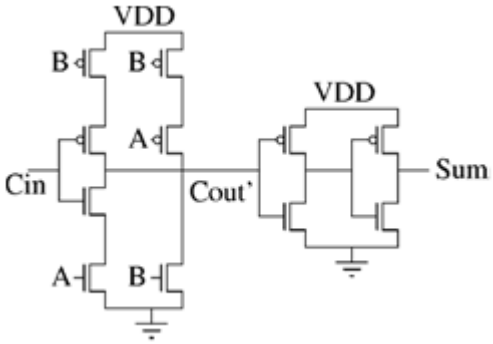


Fig 4: Imprecise Adder 3

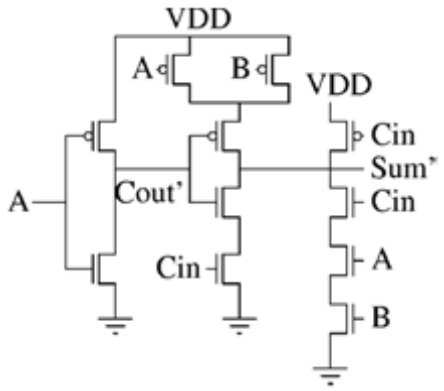


Fig 5: Imprecise Adder 4

Mean Error and variance in the Imprecise adders

For calculating the mean error and the variance, the sum and the carry are first classified as the probabilistic parameters with respect to the input.

$P[A(x)=1] = a_x$, $P[B(x)=1] = b_x$

$P[Sum(x)=1] = s_x$, $P[C_{in}(x)=1] = c_x$

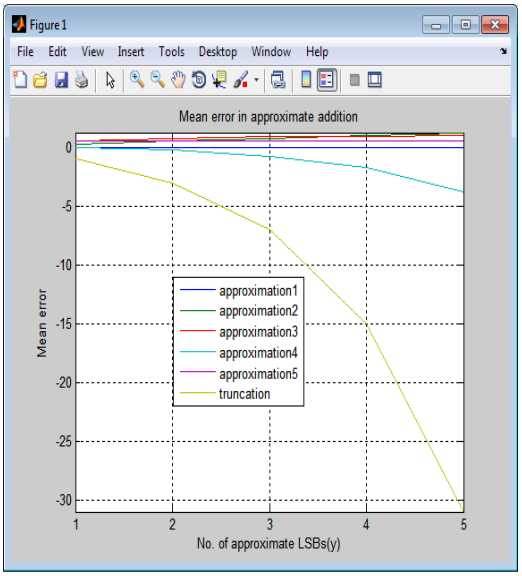
The sum expression and the carryout expression are given based on the equations derived from the full adder circuits. All the expressions for probabilities are listed with respect to the number of LSBs used.

The error is treated as the difference between the output of the imprecise adder to that of the having the actual input bit by bit.

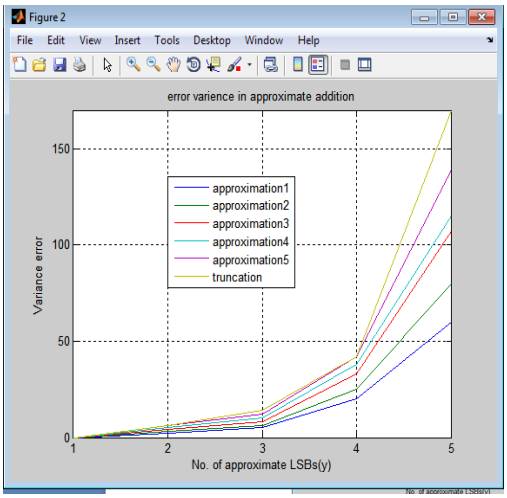
The mean error and variance vs number of imprecise bits used is plotted

	Sum expression	Carry Expression
Imprecise 1	$(1 - 2^{-(2x)})/3$	$(2 - 2^{-(2x-1)})/3$
Imprecise 2	$(1 + 2^{-(x+1)})/2$	$(1 - 2^{-(x)})/2$
Imprecise 3	$(1 + 2^{-(2x+1)})/3$	$(2 - 2^{-(2x-1)})/3$
Imprecise 4	3/8	½
Imprecise 5	½	½

Name of the Adder	
Sum of Accurate Adder	$E(e[x]) = s'_x - ½$
Carry of Accurate Adder	$E(e[y]) = c'_y - (0.5 - 2^{-(y+1)})$
Imprecise Adder 1	$\mu(y) = 0$
Imprecise Adder 2	$\mu(y) = y/4$
Imprecise Adder 3	$\mu(y) = 1 - 2^{-y}$
Imprecise Adder 4	$\mu(y) = 0.25 - 2^{y-3}$
Imprecise Adder 5	$\mu(y) = 0.5$



Area of the imprecise adder is given by the number of transistors and interconnections. Similarly the power is characterized by the number of capacitances. The total number of capacitances at each input node of imprecise adder is listed. It is clear that the number of capacitances decrease relevantly from the first imprecise adder to the last adder.



The total amount of capacitance per each input is given by the total gate capacitance and the drain

Inputs			Approximate Outputs								Choice 1		Choice 2	
A	B	C_{in}	Sum_1	C_{out1}	Sum_2	C_{out2}	Sum_3	C_{out3}	Sum_4	C_{out4}	$Sum = A$	$C_{out} = A$	$Sum = B$	$C_{out} = A$
0	0	0	0 ✓	0 ✓	1 ×	0 ✓	1 ×	0 ✓	0 ✓	0 ✓	0 ✓	0 ✓	0 ✓	0 ✓
0	0	1	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	0 ×	0 ✓	0 ×	0 ✓
0	1	0	0 ×	1 ×	1 ✓	0 ✓	0 ×	1 ×	0 ×	0 ✓	0 ×	0 ✓	1 ✓	0 ✓
0	1	1	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	1 ×	0 ×	0 ✓	0 ×	1 ×	0 ×
1	0	0	0 ×	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	0 ×	1 ×	1 ✓	1 ×	0 ×	1 ×
1	0	1	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	1 ×	1 ✓	0 ✓	1 ✓
1	1	0	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	0 ✓	1 ✓	1 ×	1 ✓	1 ×	1 ✓
1	1	1	1 ✓	1 ✓	0 ×	1 ✓	0 ×	1 ✓	1 ✓	1 ✓	1 ✓	1 ✓	1 ✓	1 ✓

capacitance. let Cdn and Cdp be the drain diffusion capacitances respectively. If the pMOS transistor has three times the width of the nMOS transistor, then Cgp ≈ 3Cgn and Cdp ≈ 3Cdn. Let us also assume that Cdn ≈ Cgn. In a multilevel adder tree, the Sum bits of intermediate outputs become the input bits A and B for the subsequent adder level. The output capacitance at each Sum node is Cdn + Cdp. The schematic of the conventional MA in Fig. 1 can be used to calculate the input capacitances at nodes A,B and Cin. Thus, the total capacitance at node A can be written as (Cdn + Cdp) + 4(Cgn + Cgp) ≈ 20Cgn. Similarly, the total capacitance at node B is (Cdn + Cdp) + 4(Cgn + Cgp) ≈ 20Cgn, while the capacitance at node Cin is (Cdn + Cdp) + 3(Cgn + Cgp) ≈ 16Cgn. The values of the total capacitance is listed based on the assumptions.

	Node A	Node B	Node C_{in}
Conventional	20	20	16
Imprecise 1	12	15	13
Imprecise 2	12	12	8
Imprecise 3	8	11	8
Imprecise 4	12	8	9
Imprecise 5	4	8	0

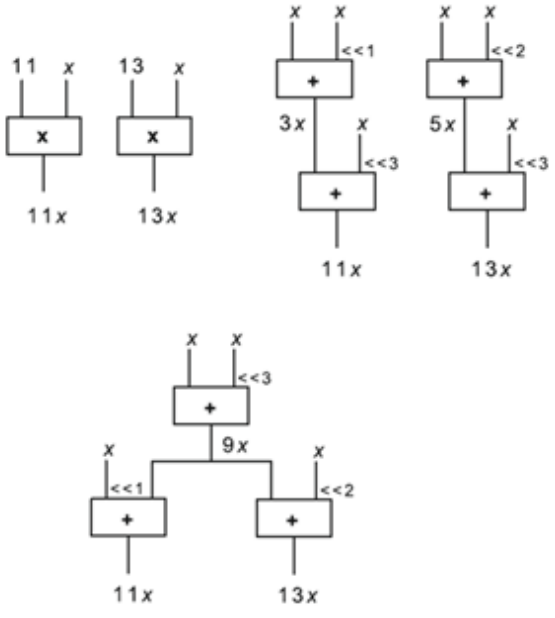
MCM and CSE Methods

In designing of a hardware structure, the shifts and adds are generally used to design the multiplications which are generally treated as constants throughout the structure. This is done based on two reasons:

(i) the multiplication design is expensive.

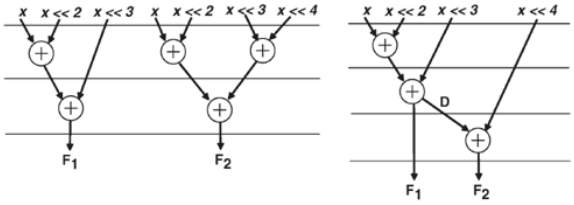
(ii) the same constants are used throughout the structure and hence they can be pre determined and designed using MCM operation. Here Multiple Constant Multiplication (MCM) is used to determine the minimum number of additions/subtractions required for the constant multiplications to be implemented.

The MCM helps to reduce maximum number of operations and thus the area and power consequently. This is achieved by commonly sharing the partial products. An example is presented as shown in the figure.



Common sub expression elimination method is used to reduce the number of repeated terms which are the partial products. It searches for identical expressions and uses a single variable to replace the entire repetition and thus reduce the number of extra terms used. A single variable is used in place of the entire repetition. This concept is mainly used in the compiler theory.

Here the concept is used to reduce the number of Adder circuits in the FIR filter design. An FIR filter consists of constant coefficients that are multiplied with an input repeatedly. So if the constants are pre-computed, then the entire structure for multiplication can be removed. Here some partial terms are formed based on the combination of certain inputs which are in common. Let F1 = X + 4X + 8X = X + X<<2 + X<< 3 and F2 = X + 4X + 8X + 16X = X + X<<2 + X<<3 + X<<4, where "<<" is the left shift operator. Here both F1 and F2 contain X+4X+8X in common. If this term is replaced by a constant, say D, then F1=D and F2=D+16X. Thus the structure can save three extra adders. It is as shown.



FIR FILTER DESIGN

A Finite Impulse Response (FIR) filter is a filter whose impulse response (or response to any finite length input) is of finite duration, because it settles to zero in finite time

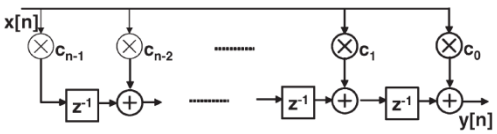
A causal FIR filter of order N is characterized by a transfer function H(z) given by

$$H(z) = \sum_{n=0}^{N} h[n] z^{-n}$$

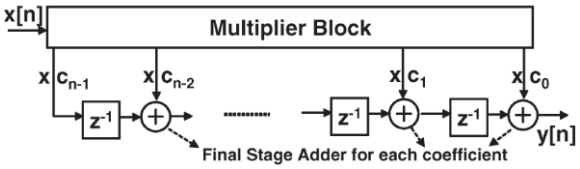
In the time-domain the input-output relation of the above FIR filter is given by

$$y[n] = \sum_{k=0}^{N} h[k] x[n-k]$$

General FIR Filter in transposed form can be as shown:



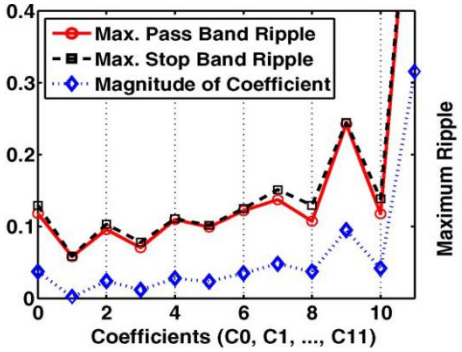
Here the entire set of coefficients are multiplied individually by the input x[n], resulting in a large amount of computation. This can be solved by replacing the multiplications with a suitable number of shifts and adds. The additions in turn can be shared among to reduce their count. The structure can be re drawn as follows where the entire calculation of coefficients is inserted in the Multiplier Block.



The coefficients of an FIR filter are calculated using the FDA tool in the Matlab. Here the coefficients are classified into groups according to their importance. The most important coefficients provide the largest pass band ripple when their value is replaced with zero and again constructed a filter. It can be shown that highest the absolute value of the coefficient, highest is the pass band ripple and highest is the importance of it. The relation between coefficient value and the pass band ripple is plotted as shown.

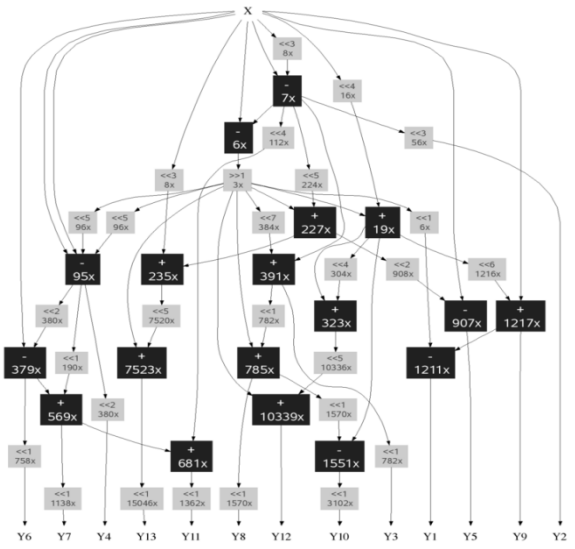
The coefficients are classified into groups where the first five are grouped as l_1, next four as l_2 and remaining as l_3.

Once the coefficients are listed according to their increase of importance, they are converted to their fixed point representation. They are then converted into multiplier block based on shift and add operations.



Coefficient	Absolute Value	Fixed Point
C1	0.0017219515241194	0000000000111000
C3	0.0116238118940803	0000000101111100
C5	0.0231495073798233	0000001011110110
C2	0.0238666379355453	0000001100001110
C4	0.0277034324942943	0000001110001011
C6	0.0347323795277931	0000010001110010
C0	0.0369760641356146	0000010010111011
C8	0.0371467378315251	0000010011000001
C10	0.0415921203506734	0000010101010010
C7	0.0479147131729558	0000011000100010
C9	0.0946865327578015	0000110000011110
C11	0.3155493020793390	0010100001100011
C12	0.4591823044074993	0011101011000110

Now the coefficients in the Multiplier block contains of additions which are 16 bit additions. These additions are carried out by accurate adders or in other words conventional adders. They use large number of transistors and large capacitance values. If these are designed using Imprecise adders, then the total transistor count reduces resulting in power and area savings. This is done by replacing the last bits of adder circuits by the imprecise adders. and thus the total transistor count in the entire circuit is reduced.



This process is repeated for various set of combinations, for different groups. Each circuit is an imprecise FIR filter, and the outputs which are the coefficients are not replica of the original circuit. Hence new set of imprecise coefficients are calculated for a set of inputs. These new coefficients determine the imprecise filter and thus imprecise pass band and stop band ripples are to be calculated. The new ripples differ from the original value of the stop band and pass band ripple. The difference of the ripple in pass band is given as Maximum Pass Band Ripple (MPBR) and that of the stop band is given by the Maximum Stop Band Ripple (MSBR). Different sets of combinations for Imprecise adders are used for different groups and MPBR and MSBR values are tabulated. Here the power savings is also calculated. The table proves that the approximation 5 provides maximum power savings and that of maximum MPBR and MSBR values.

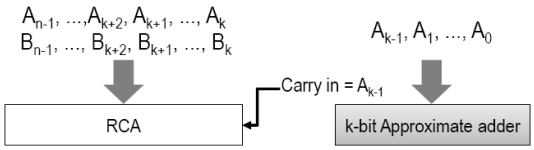
Technique	Optimum [l_1, l_2, l_3]	ΔMPBR(%)	ΔMSBR(%)	Power Savings(%)
Imprecise 1	[10, 10, 10]	2.08	1.48	51.97
Imprecise 2	[10, 10, 6]	5.93	12.74	44.76
	[10, 10, 8]	7.28	12.53	47.88
	[10, 10, 10]	13.43	12.2	50.85
Imprecise 3	[10, 10, 3]	9.66	18.27	42.95
	[10, 10, 4]	9.57	18.26	41.65
	[10, 10, 10]	17.93	18.18	55.14
Imprecise 4	[10, 9, 7]	1.35	1.77	41.00
	[10, 10, 9]	2.73	3.11	47.73
	[10, 10, 10]	1.99	1.82	47.07
Imprecise 5	[10, 10, 10]	0.27	1.92	61.77

Imprecise Ripple Carry Adder

The ripple carry adder is a structure which uses continuous full adder circuits for each bit. It is a simple structure and is of small area. It is slow based on the carry propagation time. The circuit uses full adder circuits and it can be speed up by adding the imprecise adders at LSB locations. Here the imprecise adder 5 is used because of its simple structure, and it has a 50% and 75% chance of correct output when used at the sum and carryout locations.

Here RCA topology is adopted for its advantage of minimum area. Here the approximation is controlled at the Least significant bits and the main aim is the reduction of area and energy. This concept is applied to the MCM digital FIR filters. Here the structure is used to design Multiplier less FIR filters.

A parameter k is consider to decide the best possible chance of approximation. A 32 bit adder structure is considered with k bits being used for imprecise adders. The structure is given as follows:

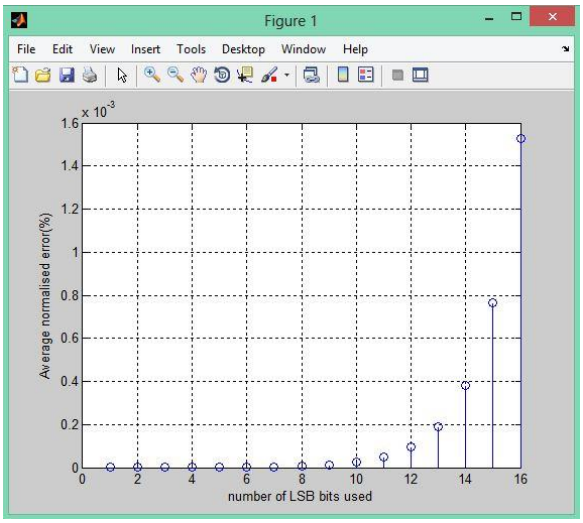


Here various values for k is tested and the respective errors are determined. The appreciable error value is treated as 0.01 to 1 dB. Here the error is calculated based on the value of the normalized error value. It is given by the formula.

The graph between the number of bits and the normalized error is plotted and the best value of k is established to be less than or equal to 11.

$$NE(\%) = 100 \frac{|y - y_{app}|}{y}$$

The concept of Imprecise Ripple Carry Adder is extended to FIR filter. Here the approximation is to be even taken care because of the large circuit. Hence the approximation is not done in calculation of the filter coefficients but also at the adder structure in the delay line of the output. After repeated calculations it is found that the approximation value in the coefficients and the delay line is best fit to 1 and 8 respectively. Three different FIR filters are considered and they are treated with accurate adders and approximate adders to calculate the SNR value. The approximation for perfect limit provided good process for SNR value. The fir filter and its SNR values are as listed.



S.No	Pass band	Stop band	No of taps	SNR in dB
1	0.1	0.15	25	85
2	0.15	0.2	25	79
3	0.2	0.25	25	76

CONCLUSION

Decrease in the number of series connected transistors helped in reducing the effective switched capacitance and achieving voltage scaling. We also derived simplified mathematical models for error and power consumption of an approximate RCA using the approximate FA cells. Using these models, we discussed how to apply these approximations to achieve maximum power savings subject to a given quality constraint. The designs in FIR case studies use approximate adders only, since adders are basic building blocks and massively used in parallel digital FIR filters. For FIR filters designed using the proposed methodology of approximations in the different adders can preserve the signal quality with SNR always higher than 60dB.

REFERENCES

[1]. Low-Power Digital Signal Processing UsingApproximate Adders ,Vaibhav Gupta, Debabrata Mohapatra, Anand Raghunathan, Fellow, IEEE, and Kaushik Roy, Fellow, IEEE

[2]. J. Choi, N. Banerjee, and K. Roy, "Variation-aware low-power synthesis methodology for fixed-point FIR filters," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 28, no. 1, pp. 87–97, Jan. 2009

[3]. N. Zhu, W. L. Goh, W. Zhang, K. S. Yeo, and K. S. Kong, "Design of Low-Power High-Speed Truncation-Error-Tolerant Adder and Its Application in Digital Signal Processing," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 18, no. 8, pp. 1225–1229, Aug. 2010.

[4]. V. Gupta, D. Mohapatra, S. P. Park, A. Raghunathan, and K. Roy, "IMPACT: Imprecise adders for low-power approximate computing," inProc. IEEE/ACM Int. Symp. Low-Power Electron. Design, Aug. 2011, pp. 409–414

[5]. K. K. Parhi, VLSI Digital Signal Processing Systems: Design and Implementation. New York: Wiley, 1999.

[6]. N. Zhu, W. L. Goh, and K. S. Yeo, "An enhanced low-power highspeed adder for error-tolerant application," in Proc. IEEE Int. Symp. Integr. Circuits, Dec. 2009, pp. 69–72.

[7]. L. Aksoy, E. Gunes, and P. Flores, "Search Algorithms for the Multiple Constant Multiplications Problem: Exact and Approximate," Elsevier Journal on Microprocessors and Microsystems, vol. 34, no. 5, pp. 151– 162, 2010.

[8]. C. Yao et al., "A novel common-subexpression-elimination method for synthesizing fixed-point FIR filters," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 51, no. 11, pp. 2211–2215, Nov. 2004.

[9]. R. M. Hewlitt and E. S. Swartzlander, "Canonical signed digit representation for FIR digital filters," in Proc. IEEE WorkShop SiPS, 2000, pp. 416–426.