A Novel VLSI Architecture of Approximate Arithmetic Units for Video Encoding for DSP Applications

T. Anwesha [1] M. Devendra, M.S, (Ph. D) [2]
[1]PG Scholar, Dept of ECE, GPREC (Autonomous), Kurnool, Andhra Pradesh, India.
[2]Associate Professor, Dept of ECE, GPREC (Autonomous), Kurnool, Andhra Pradesh, India.

Abstract
The research group over the most recent couple of years from the field of approximate computing has gotten important consideration, especially with regards to different signal processing. Image and video compression algorithms, like JPEG, MPEG and so on, which can be essential to realize very power–efficient implementations of these algorithms. However, existing approximate designs commonly fix the level of hardware approximations statically and are not flexible to information. This paper tends to this issue by proposing a reconfigurable approximate design for MPEG encoders that upgrades power consumption with the goal of keeping up a specific Peak Signal-to-Noise Ratio (PSNR) threshold for any video. Toward this end, we design reconfigurable adder/subtract or modules (RABs), which can balance their level of approximation and later integrate these blocks in the motion estimation and discrete cosine transformation modules of the MPEG encoder. We propose two rules for consequently tuning the estimation level of the RABs in these two modules during runtime based on the attributes of every individual video. Dynamically altering the level of hardware approximation depends on the input video regards the given quality bound PSNR corruption across over various videos while accomplishing a power saving a dual mode full adder is more worthy than the full adderμ when contrasted with existing implementations.

Keywords: *Approximate circuits, quality configurable, approximate computing, low power design.*

I. INTRODUCTION

Digital signal processing (DSP) blocks from the foundation of different multimedia applications utilized as a portable gadgets. The greater part of the DSP blocks actualizes image and video compression algorithms. Approximate computing designs use the way that a little relaxation in output correctness can bring about essentially more straightforward and lower implementations. In any case, most approximation hardware models proposed so far experience the effects of the restriction that, for broadly changing input parameters, it turns out to be difficult to give a quality bound on the output, and now and again, the output quality might be extremely degraded. The fundamental purpose behind this output quality variation is that the degree of approximation (DA) in the hardware architecture is settled statically and cannot be altered for various inputs. One conceivable cure is to receive a preservationist approach and utilize a low DA in the equipment with the goal that the output exactness is not greatly influenced. In any case, such a conservative approach will, as expected, to a great degree affect the power savings also.

MPEG Compression Scheme

MPEG is most favored video compression method in present day video applications and devices. MPEG-2/MPEG-4 models, videos can be compressed to little sizes. MPEG utilizes both inter frame and intra frame encoding for video compression. Intra frame encoding includes encoding the whole frame of information, while inter-frame encoding utilizes predictive and interpolative coding strategies as methods for accomplishing compression. The inter-frame version utilizes the high temporal redundancy between adjacent frames and just encodes the distinctions in data between the frames, this leads to greater compression ratios. Furthermore, motion compensated interpolative coding downsizes the information advance using bi-directional forecast. For this situation, the encoding happens based on the contrasts between the present frame and the previous and next frames in the video grouping. MPEG encoding includes three sorts of frames:

1) I-frames (intra frame encoded);

2) P-frames (prescient encoded); and

3) B-frames (bi-directional encoded).

As obvious from their names, an I-frame is encoded totally as it is with no information misfortune .An I-frame regularly goes before every MPEG information stream. P-frames are developed utilizing the contrasts between the present frame and the

instantly going before I or P frame. B-frames are produced with respect to the nearest two I/P frames on either side of the present frame. The I, P, and B frames are moreover compressed when subjected to DCT, which helps to eliminate current inter-frame spatial redundancy however much as could reasonably be expected. A major segment of the inter frame encoding is spent in calculating motion vectors (MVs) from the computed variations. Each non encoded frame is split into small macro blocks (MBs), ordinarily 16x16 pixels. Every MV has a related MB.

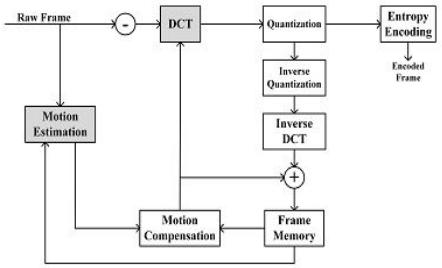


Fig 1.MPEG encoder block diagram.

The MVs truly contain data in regards to the relative motion of the MBs in the present frame in the comparison with the reference. These are calculated by obtaining the minimum value of sum of absolute differences (SADs) of an MB as for every one of the MBs of the reference frame. The resultant vectors are additionally encoded along with the frames. Be that as it may, this is not adequate to give a precise clarification of the original frame. Consequently, in addition to the MVs, a residual error is figured, which is then compressed using DCT. It has been demonstrated that the ME and DCT blocks are the most computationally costly segments of a MPEG encoder. The different steps concerned in performing MPEG compression have appeared in Fig. 1

Quality of a Video

The value of the encoding operation can be chosen from the output quality of the decoded video. Target measurements, for example, Peak Signal-to-Noise Ratio (PSNR), SAD, etc, have a decent relationship with the individual methodology of measuring the nature of the videos. Henceforth, we have used the prominent and basic PSNR metric as a method for video quality estimation. PSNR is a full-reference video quality evaluation strategy, which uses a pixel-to-pixel contrast as for the original video. In this paper, PSNR of a video is characterized as the normal PSNR over a constant number of frames (50) of the video.

Approximate adders:

Adders are utilized for calculating the addition (or sum) of two binary numbers. Two normal sorts of adders are the ripple-carry adder (RCA) and the carry look ahead adder (CLA). In an n-bit RCA, n 1-bit full adders (FAs) are cascaded; the carry of every FA is propagated to the following FA, along these delay of RCA grows in the extent to n (or O (n)). An n-bit CLA comprises of n SPGs, which work in parallel to deliver the sum, generate (gi = ai bi) and propagate (pi= ai + bi) signals, and associated with a carry look ahead generator. For CLA, all carries are produced straightforwardly by the carry look ahead generator utilizing just the generate and propagate signals, so the delay of CLA is logarithmic in n (orO (log (n))), subsequently clearly shorter than that of RCA. In any case, CLA requires larger circuit area and higher power dissipation. The carry look ahead generator turns out to be very complex for giant n. The area complexity of CLA is O (nlog(n)) when the fan-in and fan-out of the fundamental gates are fixed. Numerous approximation plans have been proposed by diminishing the critical path and hardware complexity of the accurate adder. An early system depends on a logical operation. In an n-bit logical adder, each sum bit is anticipated by its previous k least significant bits (LSBs) (k < n). A speculative design makes an adder essentially quicker than the conventional design. Segmented adders are proposed. An n-bit fragmented adder is implemented by a few littler adders working in parallel. Subsequently, the carry propagation chain is truncated into shorter fragments. The segmentation is likewise used, yet their carry contributions for each sub-adder are chosen in a different way. This kind of adder is referred to as a carry select adder. Another technique for diminishing the critical path delay and power dissipation of a conventional adder is by approximating the full adder; the approximate adder is generally connected to the LSBs of an accurate adder. In the sequel, the adders are isolated into four classifications.

II. LITERATURE SURVEY

There has been a great deal of work in building energy-efficient video compression schemes. A large number of them are identified with the particular instance of a MPEG encoder. Distinctive techniques of power-reduction incorporate algorithmic adjustments, voltage over-scaling, and imprecise

computation of measurements. The presentation of approximate computing strategies has opened up totally new opportunities in building low-power video compression models. Approximate computing techniques accomplish a lot of power savings by presenting a little measure of error into the logic block. Different approaches for approximation incorporate error presentation through voltage over scaling, intelligent logic manipulation, and circuit rearrangements utilizing don't care-based enhancement methods. The strategies present imprecision by supplanting adders with their approximate counterparts. The approximate adders are acquired by intelligently deleting some of the transistors in a mirror adder. An imperative point to note is that these approximate circuits are hardwired and cannot be altered without recombining the whole circuit. There additionally exists circumstance of approximations presented in a MPEG encoder. The majority of them accomplish the inherent error resilience of the motion estimation (ME) algorithm which brings about little quality corruption. For instance, Moshnyagaet al. utilizes a bit width compression procedure to decrease power consumption of video frame memory. He and Liou and Heet al. utilize bit truncation to introduce approximations in the ME block of a MPEG encoder. An adaptive bit masking technique is proposed, where the authors propose to truncate the pixels of the present and previous frames required for ME relying on the quantization step. Be that as it may, such a coarse-grained input truncation is applicable just to the particular instance of ME and gives undesirable outcomes for different blocks, like, discrete cosine transform (DCT), which requires a finer regulation over the error appear, helps in keeping up better control over application-level quality measurements while at the same time receiving the power utilization rewards of hardware approximation. Our proposed strategy can naturally change the degree of hardware approximation progressively in view of the video attributes. Additionally, such dynamic reconfiguration likewise furnishes users with a control handle for fluctuating the output nature of the videos and the power utilization for the battery-powered multimedia devices.

III. PROPOSED SYSTEM

Reconfigurable Adder/Subtract or Blocks

Dynamic variation of the DA should be possible when each of the adder/subtract or blocks is provided with at least one of its approximate copies and it can switch between them according to necessity. This reconfigurable design can incorporate any approximate version of the adders/sub tractors. As a reference, Gupta et al. proposed six various types of approximate circuits for adders. Be that as it may, it likewise should be given that the extra area overheads required for developing the reconfigurable approximate circuits are negligible with adequately huge power power savings. As cases, we have picked the two most normal techniques exhibited, to be specific, truncation and approximation 5, for approximating the adder/subtract or blocks. The latter one can likewise be conceptualized as an improved version of truncation as it just transfers the two 1-bit inputs, one as Sum and alternate as Carrying Out (Choice 2). In case A, B, and Cin are the 1-bit inputs to the full adder (FA), at that point, the outputs are Sum=Band Cout=A. The resultant truth-table demonstrates that the outputs are correct for more than half of all input combinations, in this manner it provide the better approximation mode than truncation. The proposed scheme replaces every FA cell of the adders/subtractors with a dual- mode FA (DMFA) cell (Fig. 5) in which every FA cell can work either in fully accurate or in some approximation mode relying upon the condition of the control signal APP.

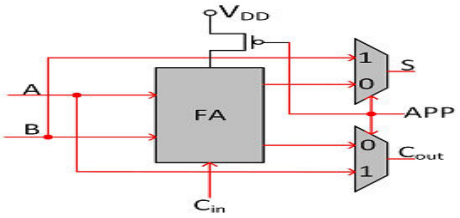


Fig 2.1-bit DMFA

A logic high estimation of the APP signal signifies that the DMFA is working in the approximate mode. We term these adders/sub tractors as RABs. Note that the FA cell is power gated while working in the approximate mode. Synthesis and evaluation of power consumption of a 16-bit RCA were performed in Synopsys Design and Power Compiler and the corresponding results are described in Table I.

Original FA(μW)	DMFA Accurate Mode(μW)	DMFA Approximate Mode (μW)
1.53	1.74	0.01

TABLE-I: Power consumption of Different Modes

Our experiments have demonstrated a negligible contrast in the power utilization of DMFA when worked in both of the two approximation modes. Subsequently, with no loss of generality, approximation 5 was decided for its higher likelihood of giving the right output result than truncation, which invariably outputs 0 independent of the input. Fig. 5 demonstrates the logic block diagram of the DMFA cell, which replaces the constituent FA cells of an 8-bit RCA, as appeared in Fig.3.2. What's more, it additionally comprises of the approximation controller for creating the approximate select signals for the multiplexer.

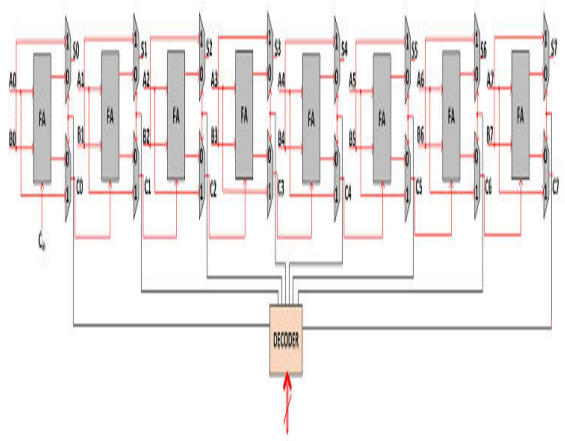


Fig.3. 8-bit reconfigurable RCA block

A multimode FA cell would give even a superior other option to the DMFA for the purpose of controlling the approximation size. Be that as it may, it likewise increases the complexity of the decoder block utilized for express the right select signs to the multiplexers and additionally the logic overhead for the multiplexers themselves. This undermines the essential target as the vast majority of the power savings that we get from approximating the bits are lost. Rather, the two-mode decoder and the 2:1 multiplexers have insignificant overhead and furthermore give adequate summon over the approximation degree.

DMFA Overhead:

The power gating transistor and the multiplexers of the DMFA are designed to meet the least possible overhead. Our analyses demonstrate that switching power of the CMOS transistors contributes toward the greater part of the aggregate power utilization of the FA and DMFA blocks. Table I introduces the power consumption of FA and DMFA for various modes acquired by exhaustive simulation in Synopsys NanoSim. It demonstrates that the power increments by 0.21 μW when we work DMFA in the accurate mode as contrasted with the original FA block. This distinction in power can be applied predominantly to the raise in load capacitance of the FA obstruct because of the addition of the input capacitance of the interfaced multiplexers. A small portion of the aggregate power is contributed by the additional switching of the multiplexers. Table I likewise demonstrates that the power consumed during DMFA approximation mode is practically irrelevant when contrasted and the accurate mode, which is because of the power gating of the FA blocks by the P-MOS transistor, as appeared in Fig. 5.Reduction in the input switching activity of the multiplexers is likewise an auxiliary reason for this little measure of power. The extra overhead because of switching of the power gating transistor can be neglected, since its switching activity is little because of the idea of our switching algorithm. This is primarily due to the spatial and temporal locality of the pixel values across consecutive frames. The idea of RAB can likewise be reached out to other adder architecture also. Adder architectures, for example, CBA and CSA, which likewise contain FA as the essential building block, can be made precision configurable by directly supplanting of the FAs with DMFAs. Different assortments, as CLA and tree adders, utilize distinctive sorts of carry propagate and generate blocks as their fundamental building units and consequently require some extra alterations to work as RABs. For instance, we implemented a 16-bit CLA comprising of four unique sorts of fundamental blocks (Fig. 8) depending upon the presence of sum(S), Cout, carry propagation (P), and carry generation(G) at various levels. We address the fundamental blocks display at the first (or lowermost) level of a CLA, which have inputs coming in specifically, as for carry look ahead blocks, CLB1 and CLB2. The distinction among them being that CLB1 produces an extra Cout signal contrasted with CLB2. They are comparing dual mode forms, DMCLB1 and DMCLB2, have both S and P approximated by input operand B and both Cout and G approximated by input operand A, as appeared in Fig.4. The fundamental block display at the higher levels of CLA hierarchy is denoted as propagate and generate blocks, PGB1 and PGB2. For this situation, PGB1 produces an additional Cout output as compared with PGB2.

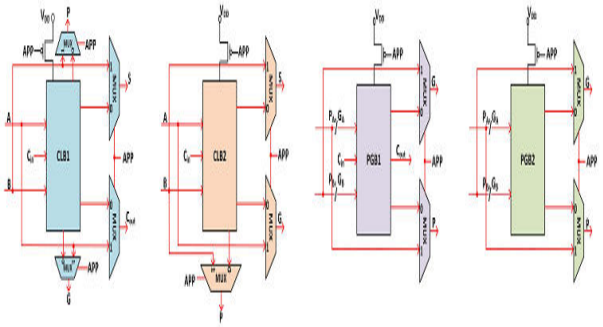


Fig.4. 1-bit dual-mode carry propagates generate blocks.

As appeared in Fig.4, the configurable dual mode forms, DMPGB1 and DMPGB2, utilize inputs PA and GB as approximations for outputs P and G, respectively, while working in the approximate mode

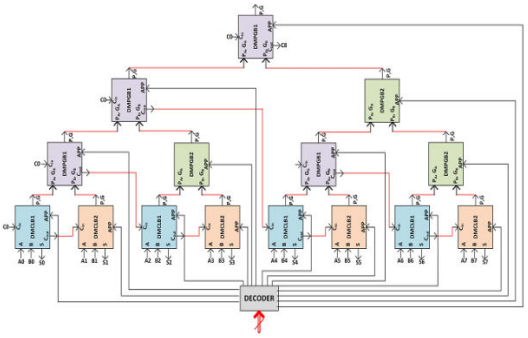


Fig.5. 8-bit reconfigurable CLA block.

These approximations were chosen on trial guaranteeing that the proportion of the likelihood of correct output to the extra circuit overhead for each of the block is large. Table II compresses the outputs of each of the dual mode blocks while working in either approximate or accurate mode

Basic Block (adder type)	Outputs for App=0(accurate mode)	Outputs for App=1(approximate mode)
DMFA (RCA,CBA ,CSA)	$S=A\oplus B\oplus C_{IN}$ $C_{OUT}=AB+BC_{IN}+AC_{IN}$	S=B $C_{OUT}=A$
DMCLB1 (CLA)	$P=A\oplus B$ G=AB $S=P\oplus C_{IN}$ $C_{OUT}=G+PC_{IN}$	P=B G=A S=B $C_{OUT}=A$
DMCLB2(CLA)	$P=A\oplus B$ G=AB $S=P\oplus C_{IN}$	P=B G=A S=B
DMPGB1 (CLA)	$P=P_AP_B$ $G=G_B+G_AP_B$ $C_{OUT}=G+PC_{IN}$	$P=P_A$ $G=G_B$ $C_{OUT}=G+PC_{in}$
DMPGB2(CLA)	$P=P_AP_B$ $G=G_B+G_AP_B$	$P=P_A$ $G=G_B$

TABLE-II

DUAL-MODE BLOCK OUTPUTS FOR ACCURATE AND APPROXIMATE MODES

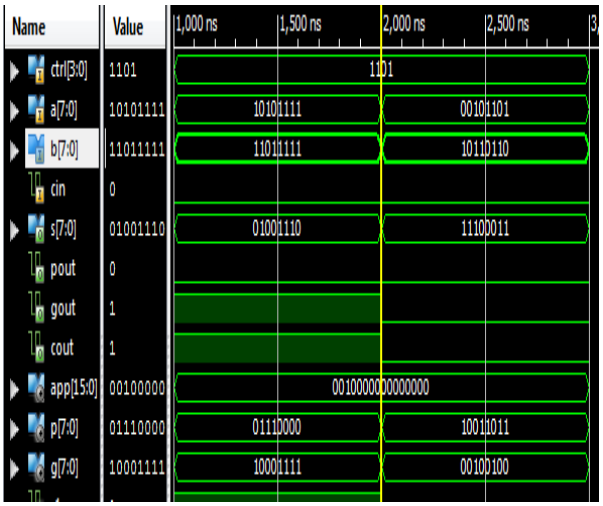
For a reconfigurable CLA, DMCLB1 and DMCLB2 blocks are approximated as per the DA. Notwithstanding, the DMPGB1 and DMPGB2 blocks are approximated just when every single DMCLB1, DMCLB2, DMPGB1, and DMPGB2 block, which has a place with the transitive fan-in cones of the concerned piece, is approximated. Something else, the block is worked in the accurate mode.

For instance, any DMPGB block at the second level of CLA can be made to work in approximate mode, if and just if, both of its constituent DMCLB1 and DMCLB2 blocks are working in the approximate mode. The comparative way is taken after for the blocks residing at higher levels of the tree, where each DMPGB block can be approximated just when both of its constituent DMPGB1 and DMPGB2 blocks are approximated. This architecture can be effectively speculated to another comparable sort CLAs, for example, Kogge–Stone, Brent–Kung, Manchester-convey chain, etc.

IV. RESULTS

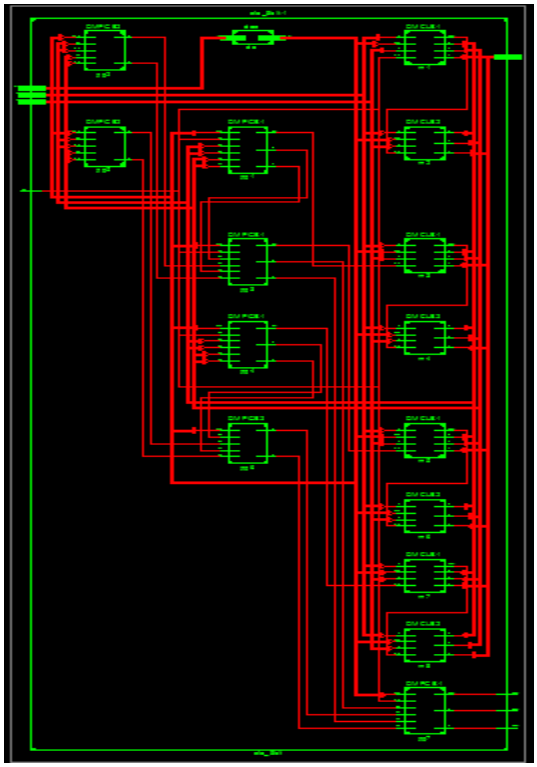
The written Verilog HDL Modules have successfully simulated and verified using Isim Simulator and synthesized using Xilinxise13.2.

SIMULATION RESULTS

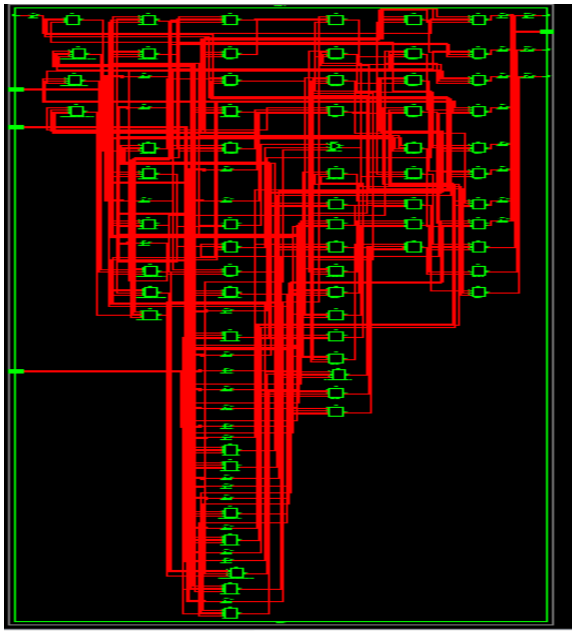


Synthesis results

RTL schematic



Technology Schematic:



Design summary

Device Utilization Summary(estimated values)			
Logic Utilization	**Used**	**Available**	**Utilization**
Number of slices	37	4656	0%
Number of 4 input LUTS	65	9312	0%
Number of bonded IOBS	32	232	13%

Timing Report

Timing Summary:

Speed Grade: -5

Minimum perid : No path found

Minimum input arrival time before clock:No path found

Maximum output required time after clock:No Path found

Maximum combinational path delay:13.835ns

V. CONCLUSION

This paper proposed a reconfigurable approximate architecture for the MPEG encoders that optimize power utilization while keeping up output quality across various input videos. The proposed architecture is depends on the idea of dynamically reconfiguring the level of approximation in the hardware based on the input characteristics. It requires the user to determine just the general least quality for videos has opposed to deciding the level of hardware approximation. Our experiments comes about demonstrate that the proposed architecture brings about power savings equal to a pattern approach that utilizations settled approximate hardware while regarding quality limitations crosswise over various videos. Future work incorporates the fuse of other approximation strategies and stretching out the approximations to other arithmetic and functional blocks.

REFERENCES

[1] M. Elgamel, A. M. Shams, and M. A. Bayoumi, “A comparative analysis of low power motion estimation VLSI architectures," in Proc. IEEE Workshop Signal Process. Syst. (SiPS), Oct. 2000, pp. 149–158.

[2] F. Dufaux and F. Moscheni, “Motion estimation techniques for digital tv: A review and a new contribution,”Proc. IEEE, vol. 83, no. 6, pp. 858–876, Jun. 1995.

[3] I. S. Chong and A. Ortega, “Dynamic voltage scaling algorithms for power constrained motion estimation,” in-proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), vol. 2. Apr. 2007, pp. II-101–II-104.

[4] I. S. Chong and A. Ortega, “Power efficient motion estimation using multiple imprecise metric computations,” in-proc. IEEE Int. Conf.Multimedia Expo, Jul. 2007, pp. 2046–2049.
[5] D. Mohapatra, G. Karakonstantis, and K. Roy, “Significance driven computation: A voltage-scalable, variation-aware, quality-tuning motion estimator,” in Proc. 14th ACM/IEEE Int. Symp. Low Power Electron. Design (ISLPED), 2009, pp. 195–2009.
[6] J. George, B. Marr, B. E. S. Akgul, and K. V. Palem, “Probabilistic arithmetic and energy efficient embedded signal processing,” in-proc.Int. Conf. Compil., Archit., Synth. Embedded Syst. (CASES), 2006, pp. 158–168.
[7] D. Shin and S. K. Gupta, “A re-design technique for data path modules in error tolerant applications,” in Proc. 17th Asian Test Symp. (ATS), 2008, pp. 431–437.

AUTHOR BIOGRAPHIES

Ms. T.Anwesha is Pursuing M. Tech in VLSI and Embedded Systems at G.Pulla Reddy Engineering College, Kurnool, AP, India. She published one Paper in National Conference. Her area of interest is in the field of VLSI.

Mr. M. Devendra received his B. Tech degree in E.C.E Department from SVU, Tirupathi, A.P in 1989,M.S degree in Electronics and control from BITS PILANI,India, in 1996 and currently pursuing his Ph D in the area of Embedded Systems from Rayalaseema University , Kurnool, AP, India. He has 25 years of teaching experience. Currently he is working as Associate Professor in ECE Department, G Pulla Reddy Engineering College, Kurnool, AP, India.