ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

64 Bit×64 Bit Multiprecision Multiplier for Operands Scheduler with Dynamic Voltage Scaling

B.Ravi Teja
M.Tech Student,
Department of ECE,
CMR Institute of Technology,
Hyderabad, Telangana, India.

Muni Praveena Rela
Associate Professor,
Department of ECE,
CMR Institute of Technology,
Hyderabad, Telangana, India.

Abstract:

In this paper, we present a multiprecision (MP) reconfigurable multiplier that incorporates variable precision, parallel processing (PP), razor-based dynamic voltage scaling (DVS), and dedicated MP operands scheduling to provide optimum performance for a variety of operating conditions. All of the building blocks of the proposed reconfigurable multiplier can either work as independent smaller-precision multipliers or work in parallel to perform higher-precision multiplications. Given the user's requirements (e.g., throughput), a dynamic voltage/ frequency scaling management unit configures the multiplier to operate at the proper precision and frequency. Adapting to the run-time workload of the targeted application, razor flip-flops together with a dithering voltage unit then configure the multiplier to achieve the lowest power consumption. The single-switch dithering voltage unit and razor flip-flops help to reduce the voltage safety margins and overhead typically associated to DVS to the lowest level. The large silicon area and power overhead typically associated to reconfigurability features are removed. Finally, the proposed novel MP multiplier can further benefit from an operands scheduler that rearranges the input data, hence to determine the optimum voltage and frequency operating conditions for minimum power consumption. Experimental results show that the proposed MP design features a 28.2% and 15.8% reduction in circuit area and power consumption compared with conventional fixed-width multiplier. When combining this MP design with error-tolerant razor-based DVS, PP, and the proposed novel operands scheduler, 77.7%–86.3% total power reduction is achieved with a total silicon area overhead as low as 11.1%. This paper successfully demonstrates that MP architecture can allow more aggressive frequency/ supply voltage scaling for improved power efficiency.

Index Terms:

Computer arithmetic, dynamic voltage scaling, low power design, multi-precision multiplier.

I.INTRODUCTION:

Consumers demand for increasingly portable yet high-performance multimedia and communicationproducts imposes stringent constraints on the power consumption of individual internal components [1]–[4]. Of these, multipliersperform one of the most frequently encountered arithmetic operations in digital signal processors (DSPs) [4]. For embeddedapplications, it has become essential to design morepower-aware multipliers [4]–[13]. Given their fairly complex structure and interconnections, multipliers can exhibit a largenumber of unbalanced paths, resulting in substantial glitchgeneration and propagation [8], [11]. This spurious switchingactivity can be mitigated by balancing internal paths through acombination of architectural and transistor-level optimizationtechniques [8], [11]. In addition to equalizing internal pathdelays, dynamic power reduction can also be achieved by monitoringthe effective dynamic range of the input operands soas to disable unused sections of the multiplier [6], [12] and/ortruncate the output product at the cost of reduced precision[13]. This is possible because, in most sensor applications,the actual inputs do not always occupy the entire magnitudeof its word-length. For example, in artificial neural networkapplications, the weight precision used during the learningphase is approximately twice that of the retrieval phase [14].Besides, operations in lower precisions are the most frequentlyrequired. In contrast, most of today's full-custom DSPs andapplication-specific integrated circuits (ASICs) are designedfor a fixed maximum word-length so as to accommodate theworst case scenario.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Therefore, an 8-bit multiplication computedon a 32-bit Booth multiplier would result in unnecessaryswitching activity and power loss.Several works investigated this word-length optimization. [1], [2] proposed an ensemble of multipliers of different precisions, with each optimized to cater for a particular scenario. Each pair of incoming operands is routed to the smallest multiplier that can compute the result to take advantage of the lower energy consumption of the smaller circuit. This ensemble of point systems is reported to consume the least power but this came at the cost of increased chip area given the used ensemble structure. To address this issue, [3], [5] proposed to share and reuse some functional modules within the ensemble. In [3], an 8-bit multiplier is reused for the 16-bit multiplication, adding scalability without large area penalty. Reference [5] extended this method by implementing pipelining to further improve the multiplier's performance. A more flexible approach is proposed in [15], with several multiplier elements grouped together to provide higher precisions and reconfigurability. Reference [7] analyzed the overhead associated to such reconfigurable multipliers. This analysis showed that around 10%–20% of extra chip area is needed for 8–16 bits multipliers. Combining multiprecision (MP) with dynamic voltage scaling (DVS) can provide a dramatic reduction in power consumption by adjusting the supply voltage according to circuit'srun-time workload rather than fixing it to cater for the worstcase scenario [4].

When adjusting the voltage, the actualperformance of the multiplier running under scaled voltagehas to be characterized to guarantee a fail-safe operation.Conventional DVS techniques consist mainly of lookup table(LUT) and on-chip critical path replica approaches [17]–[19]. The LUT approach tunes the supply voltage according to a predefined voltage-frequency relationship stored in a LUT, which is formed considering worst case conditions (process variations, power supply droops, temperature hot-spots, coupling noise, and many more). Therefore, large margins are necessarily added, which in turn significantly decrease the effectiveness of the DVS technique. The critical path replica approach typically involves an on-chip critical path replica to approximate the actual critical path. Therefore, voltage could be scaled to the extent that the replica fails to meet the timing. However, safety margins are still needed to compensate for the intradie delay mismatch and address fast-changing transient effects [24]. In addition, the critical path may change as a result of the varying supply voltage or process or temperature variations.

I If this occurs, computations will completely fail regardless of the safety margins. The aforementioned limitations of conventional DVS techniques motivated recentresearch efforts into error-tolerant DVS approaches [24]–[27],which can run-time operate the circuit even at a voltage levelat which timing errors occur. A recovery mechanism is thenapplied to detect error occurrences and restore the correct data.Because it completely removes worst case safety margins,error-tolerant DVS techniques can further aggressively reducepower consumption. n this paper, we propose a low powerreconfigurable multiplier architecture that combines MP withan error-tolerant DVS approach based on razor flip-flops [25].The main contributions of this paper can be summarizedfollows.

1)AnovelMPmultiplierarchitecture featuring,respectively, 28.2% and 15.8% reduction in silicon areaand power consumption compared with its conventional32 × 32 bit fixed-width multiplier counterpart. Allreported multipliers trade silicon area/powerconsumption for MP [7]. In this paper, silicon area isoptimized by applying an operation reduction techniquethat replaces a multiplier by adders/subtractors.

2) A silicon implementation of this MP multiplierintegrating an error-tolerant razor-based dynamic DVSapproach. The fabricated chip demonstrates run-timeadaptation to the actual workload by operating at theminimum supply voltage level and minimum clockfrequency while meeting throughput requirements. Priorworks combining MP with DVS have only considereda limited number of offline simulated precision-voltagepairs, with unnecessary large safety margins added tocater for critical paths [9], [10].

3) A novel dedicated operand scheduler that rearrangesoperations on input operands so as to reduce thenumber of transitions of the supply voltage and, inturn, minimize the overall power consumption ofthe multiplier. Unlike reported scheduling works, thefunction of the proposed.

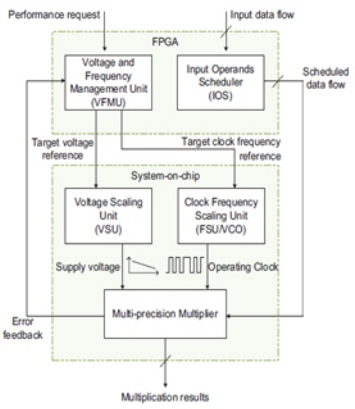

Fig 1: Overall multiplier system architecture.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Scheduler is not task schedulingrather input operands scheduling for the proposed MPmultiplier.The rest of this paper is organized as follows. Section IIpresents the operation and architecture of the proposed MPmultiplier. Section III presents the approach used to reduce theoverhead associated to MP and reconfigurability. Section IVpresents the operating principle and implementation of theDVS management unit.. Section V presentsexperimental results. Section VI presents the operands schedulerunit. Finally, a conclusion is given in Section VII.

II.SYSTEM OVERVIEW AND OPERATION:

The proposed MP multiplier system (Fig. 1) comprises fivedifferent modules that are as follows:
1) The MP multiplier;
2)The input operands scheduler (IOS) whose function isto reorder the input data stream into a buffer, hence toreduce the required power supply voltage transitions;
3) The frequency scaling unit implemented using a voltagecontrolled oscillator (VCO). Its function is to generatethe required operating frequency of the multiplier;
4) The voltage scaling unit (VSU) implemented using a voltagedithering technique to limit silicon area overhead. Itsfunction is to dynamically generate the supply voltageso as to minimize power consumption;
5)The dynamic voltage/frequency management unit(VFMU) that receives the user requirements (e.g.,throughput).

The VFMU sends control signals to the VSU and FSUto generate the required power supply voltage and clockfrequency for the MP multiplier.The MP multiplier is responsible for all computations. It is equipped with razor flip-flops that can report timing errors associated to insufficiently high voltage supply levels. The operation principle is as follows. Initially, the multiplier operates at a standard supply voltage of 3.3 V. If the razor flipflops of the multiplier do not report any errors, this means that the supply voltage can be reduced. This is achieved through the VFMU, which sends control signals to the VSU, hence to lower the supply voltage level. When the feedback provided by the razor flip-flops indicates timing errors, the scaling of the power supply is stopped. The proposed multiplier (Fig. 2) not only combines MPand DVS but also parallel processing (PP). Our multiplier-comprises 8 × 8 bit reconfigurable multipliers.

These buildingblocks can either work as nine independent multipliers or a single32 × 32 bit work in parallel to perform one, two or three 16 × 16 bitmultiplications or a single-32 × 32 bit operation. PP can beused to increase the throughput or reduce the supply voltagelevel for low power operation.

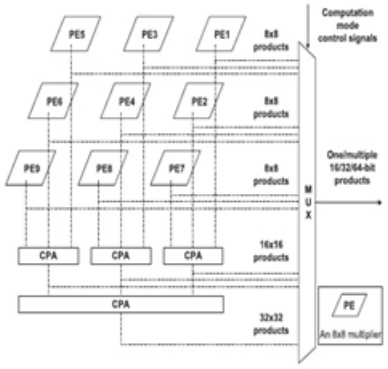

Fig 2: Possible configuration modes of proposed MP multiplier.

III. MP AND RECONFIGURABILITY OVERHEAD:

Fig.3 shows the structure of the input interface unit,which is a sub-module of the MP multiplier (Fig. 1). Therole of this input interface unit (Fig. 4) is to distribute theinput data between the nine independent processing elements(PEs) (Fig. 2) of the 32 × 32 bit MP multiplier, consideringthe selected operation mode. The input interface unit usesan extra MSB sign bit to enable both signed and unsigned-multiplications. A 3-bit controlbus indicateswhethertheinputs are 1/4/9 pair(s) of 8-bit operands, or 1/2/3 pair(s) of16-bit operands, or 1 pair of 32-bit operands, respectively.Depending on the selected operating mode, the input datastream is distributed (Fig. 4) between the PEs to performthe computation. Fig. 5 shows how three 8 × 8 bit PEs areused to realize a 16 × 16 bit multiplier. The 32 × 32 bitmultiplier is constructed using a similar approach but requires3 × 3 PEs. A 3-bit control word defines which PEs workconcurrently and which PEs are disabled. Whenever the fullprecision (32 × 32 bit) is not exercised, the supply voltageand the clock frequency may be scaled down according to theactual workload. To evaluate the overhead associated to reconfigurability andMP, we define X and Y as the 2n-bits wide multiplicand andmultiplier, respectively. XH, YHare their respective n mostsignificant bits whereas XL, YLare their respective n leastsignificant bits. XLYL , XHYL , XLYH, XHYHis the crosswiseproducts. The product of X and Y can be expressed as follows:

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

P = (X H YH)22n+ (X H YL+ X L YH)2n+ X L YL(1)
Where 2n-bit reconfigurable multiplier can be built usingadders and four n bit × n bit multipliers to computeX-HYH, X H YL, X L YH, and X L YL. this would resultin overheads of 18% and 13% for the silicon area and power, respectively. However, if we define [18]
X 1= X H+ X L (2)
Y1= YH+ YL (3)
then (1) could be rewritten as follows
P = (X H YH)22n+(X 1Y1_− X H YH− X L YL)2n+ X L YL(4)
Comparing (1) and (4), we have removed one n × n bit multiplier (for calculating XHYL or XLYH) and one 2n-bit adder (for calculating XHYL + XLYH). The two adders are replaced with two n-bit adders (for calculating XH + XL and YH + YL) and two (2n + 2)-bit subtractors (for calculatingX'Y' − XHYH− XLYL).

In a 32-bit multiplier, we can thussignificantly reduce the design complexity by using two 34-bitsubtractors to replace a 16 × 16 bit multiplier. We actuallyneed two 16 × 16 bit multipliers (for calculating XHYH andXLYL) and one 17 × 17 bit multiplier (for calculating X_Y _).

To evaluate the proposed MP architecture, a conventional32-bit fixed-width multiplier and four sub-block MP multipliersare designed using a Booth Radix-4 Wallace treestructure similar to that used for the building blocks of our MPthree sub-block multiplier. These multipliers are synthesizedusing the synopsis design compiler with AMIS 0.35-µmcomplimentary metal-oxide-semiconductor (CMOS) standard cell technology library. The power simulations are performed at a clock frequency of 50 MHz and at a power supply of 3.3 V.

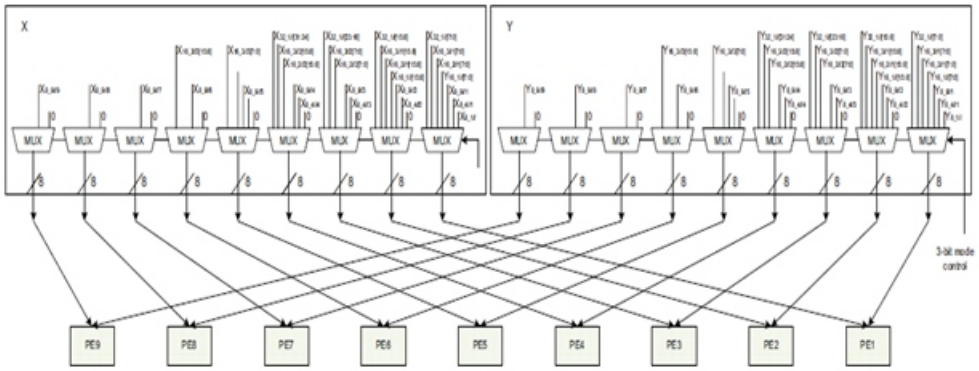

Fig.3: Structure of input interface unit

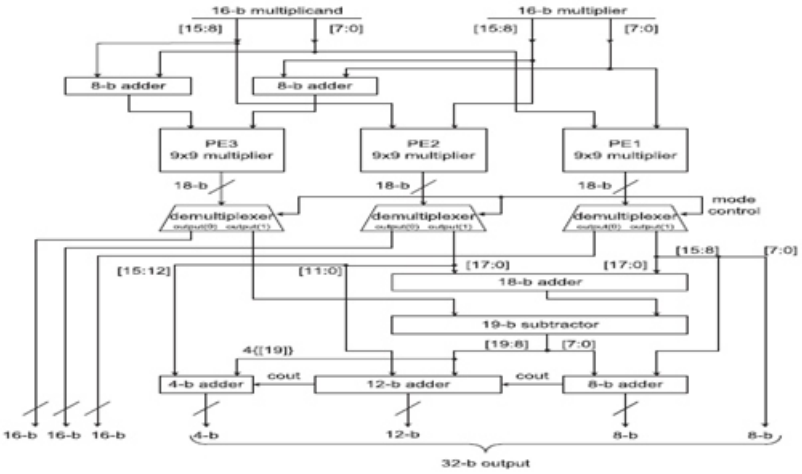

Fig.4:Three PEs combined to form 16 × 16 bit multiplier.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

IV. DYNAMIC VOLTAGE AND FREQUENCY SCALING MANAGEMENT

A. DVS Unit:

In our implementation (Fig. 1), a dynamic power supply and a VCO are employed to achieve real-time dynamic voltage and frequency scaling under various operating conditions. In [28],near-optimal dynamic voltage scaling can be achieved whenusing voltage dithering, which exhibits faster response time than conventional voltage regulators. Voltage dithering uses power switches to connect different supply voltages to the load, depending on the time slots. Therefore, an intermediate average voltage is achieved. This conventional voltage dithering technique has some limitations. If the power switches are toggled with overlapping periods, switches can be turned on simultaneously, giving rise to a large transient current. To mitigate this, nonoverlapping clocks could be used to control power switches. However, this may result in system instability as there are instances where all supply voltages are disconnected from the load.

The requirement for multiple supplies can also result in system overhead. To address these issues, we implemented a single-supply voltage ditheringscheme [Fig. 6(a)], which operates as follows. When the supplyvoltage (Vn) of the multiplier drops below the predefinedreference voltage (Vref), the comparator output (Va) toggles.Therefore, the VFMU turns on the power switch via Vctrl,for a predefined duration Tc = 5 μs. The chosen value for theoff-chip storage capacitor Cs is 4.7 μF. This value is chosen toachieve a voltage ripple magnitude of 50 mV [Fig. 6(b)] witha charging current set to 50 mA, hence to limit the resistivepower loss of the dithering unit to less than 1% of the totalpower consumption. The value of Cs is a tradeoff betweenripple magnitude, tracking speed, and area/power overheads.Fig. 6(b) shows experimental results for the voltage control loop.

B. Dynamic Frequency Scaling Unit:

In the proposed 32 ×32 bit MP multiplier, dynamic frequency tuning is used to meet throughput requirements. It is based on a VCO implemented as a seven-stage current starved ring oscillator. The VCO output frequency can be tuned from 5 to 50 MHz using four control bits (5 MHz/step). This frequency range is selected to meet the requirements of general purpose DSP applications.

The reported multiplier can operate as a 32-bit multiplier or as nine independent 8-bit multipliers.For the chosen 5–50 MHz operating range, our multiplier boasts up to 9 ×50 = 450 MIPS. The simulated power consumption for the VCO ranges from85 (5 MHz) to 149 μW (50 MHz), which is negligible compared with the power consumed by the multiplier. Fig. 7 shows experimental measurements showing the transient response for the worst case frequency switching (from 50 to 5 MHz). Clock frequency can settle within one clock cycle as required.

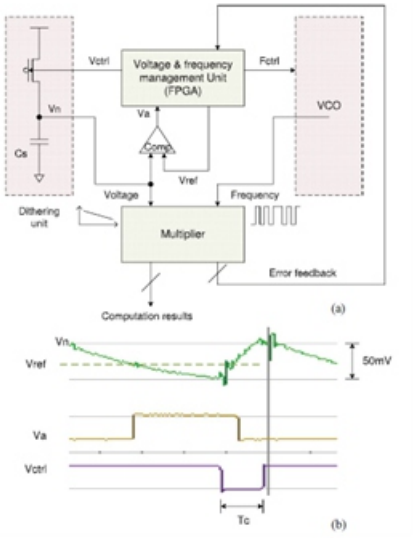

Fig. 6. (a) Proposed single-header voltage dithering unit and voltage andfrequency tuning loops. (b) Experimental timing results from voltage dithering unit.

V. INPUT OPERANDS SCHEDULER:

Here, we present three different algorithmsto reduce this overall power consumption. Each of thesealgorithms constitutes a different approach to process themixed-precision data held in the operands buffer (Fig.5). Theperformance of each algorithm is evaluated using a mixed precision-data set of 120 000 randomly operands, with athird corresponding to each precision (8, 16, 32and 64-bit).In the following, the specified throughput Tp for the proposed 64× 64 bit multiplier is 64 F (Mbits/s), where F is themultiplier's operating frequency.

Algorithm A:

In the first algorithm, the multiplier throughput Tp = 64 F iskept constant by fixing the operating frequencies (f32, f16,or f8) of each precision-data group (32, 16, or 8-bit) to f32 = F f16 = F/2 f8 = F/4 (10)

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

WhereF is the multiplier's operating frequency. This isbecause the throughput in 8 × 8 bit multiplication mode is fourtimes that of the 32 × 32 bit multiplication mode and doublesthat of the 16 × 16 bit multiplication mode, as a result of themultiplier PP. The minimum supply voltage (Vmin32, Vmin16or Vmin8) associated to each operating frequency (f32, f16or f8) is determined through a Vmin–fLUT. Algorithm Ashows its limitations when 32-bit operands are processedinitially all N32 operands of the data-block are processed, the supply voltage (Vn) needs to decreaserapidly from point A (Vmin32) to point B (Vmin16) at which allN16 16-bit operands of the data block should be processed. If N16 is too small, most 16-bit operands will be actual-lyprocessed in Sections A and B, that is at a voltage pos-siblymuch higher than the minimal Vmin16 level. Simi-larly 8-bitoperands of the data block could be processed in Sections Cand D, B-C, or even A-B for the worst case. This contributes to increasing Pcompu_overhead.

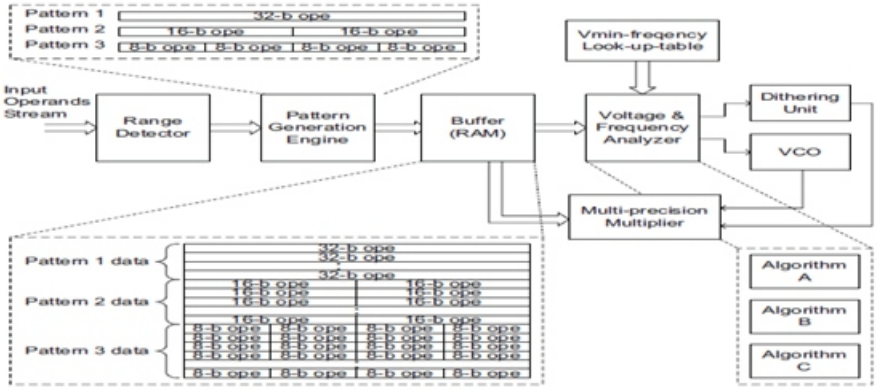

Fig.5: Block diagram of IOS

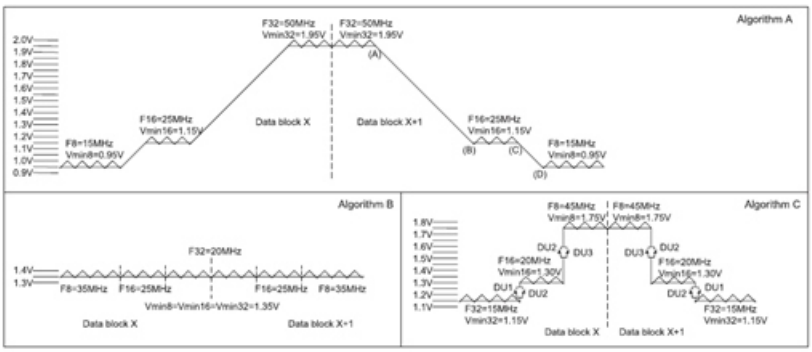

Fig.6: Operation principles of operand scheduling algorithms A, B, and C. Data Block X and Data Block X+1 refer to two-consecutive operand datablocks subsequently stored into the RAM, respectively.

Algorithm B:

This algorithm removes all transitions of the power sup-plyvoltage by making Vmin32, Vmin16, and Vmin8 equal and adjustingf32, f16, and f8 such that the overall throughput is keptunchanged. We thus need to have the following:

$$\frac{64N_{32} + 128N_{16} + 256N_8}{\frac{N_{32}}{f_{32}} + \frac{N_{16}}{f_{16}} + \frac{N_8}{f_8}} = 64F \qquad (11)$$

From a LUT, we can obtain the $V_{min}-f$ relationship as follows:

$$V_{min32} = \psi_{32}(f_{32}) \quad (12)$$
$$V_{min16} = \psi_{16}(f_{16}) \quad (13)$$
$$V_{min8} = \psi_8(f_8). \quad (14)$$

As algorithm B keeps the supply voltage constant

$$\psi_{32}(f_{32}) = \psi_{16}(f_{16}) = \psi_8(f_8) = V \quad (15)$$

the operating frequencies f32, f16, and f8 can be deter-minedby using (11) and (15). For example, when F is set to50 MHz, the values for V , f32, f16, and f8 are found tobe 1.35 V, 20 MHz, 25 MHz, and 35 MHz, respective-ly.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Algorithm C:

Although Algorithm B removes power supply voltage transitionsby setting a single-voltage level V, there may bebetter power saving combinations of power supply voltagesand operating frequencies: (Vmin32, f32), (Vmin16, f16), and(Vmin8, f8). The aim of algorithm C is to find such an optimumfor reduced power consumption. To limit complexity, we willonly seek to minimize the dynamic power dissipated as a resultof the computation

$P = CV2f$

$= Cm32V2\ min32f32 + Cm16V2min16f16 + m8V\ 2min8f8$

$= \chi(f32, f16)$

Given that the Vmin– f relationships are known (12)–(14),one could find the minimum of the above equation for thespecified throughput (11). For example, when F is set to50 MHz, the values for (Vmin32, f32), (Vmin16, f16), (Vmin8, f8)are found to be (1.15 V, 15 MHz), (1.30 V, 20 MHz), and(1.75 V, 45 MHz), respectively. When consideringDVS, razor, RAM, and dedicated scheduling circuitry,algorithm B exhibits the least power consumption, with anoverall power reduction of 86.3%, compared with the standard 32 × 32 bit fixed-width multiplier. However, it requires two additional dithering units to generate all three discrete power supply levels Vmin32, Vmin16, and Vmin8and thus remove transitions among these different supply levels. This increases the total silicon area overhead to 27.1%. Therefore, algorithm B provides the most attractive tradeoff with 81.5% reduction and a silicon area overheard of just 11.9%.

VI SIMULATION RESULTS:

The simulation of the proposed design is carried out by using Xilinx software. The simulated waveforms are shown in below figure.

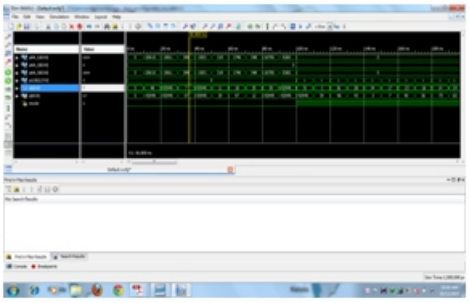

Fig.7: Simulation results of the proposed 64 x 64 bit design in signed decimal

VII CONCLUSION:

We proposed a novel MP multiplier architecture featuring,respectively, 28.2% and 15.8% reduction in silicon area andpower consumption compared with its 32 × 32 bit conventionalfixed-width multiplier counterpart. When integrating thisMP multiplier architecture with an error-tolerant razor-basedDVS approach and the proposed novel operands scheduler,77.7%–86.3% total power reduction was achieved with a totalsilicon area overhead as low as 11.1%. The fabricated chipdemonstrated run-time adaptation to the actual workload byoperating at the minimum supply voltage level and minimumclock frequency while meeting throughput requirements. The proposed novel dedicated operand scheduler rearrangesoperations on input operands, hence to reduce the number oftransitions of the supply voltage and, in turn, minimized theoverall power consumption of the multiplier. The proposed MPrazor-based DVS multiplier provided a solution toward achievingfull computational flexibility and low power consumptionfor various general purpose low-power applications.

REFERENCES:

[1] R. Min, M. Bhardwaj, S.-H. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang,and A. Chandrakasan, "Energy-centric enabling technologies for wirelesssensor networks," IEEE Wirel. Commun., vol. 9, no. 4, pp. 28–39,Aug. 2002.

[2] M. Bhardwaj, R. Min, and A. Chandrakasan, "Quantifying and enhancingpower awareness of VLSI systems," IEEE Trans. Very Large ScaleIntegr. (VLSI) Syst., vol. 9, no. 6, pp. 757–772, Dec. 2001.

[3] A. Wang and A. Chandrakasan, "Energy-aware architectures for a realvaluedFFT implementation," in Proc. IEEE Int. Symp. Low PowerElectron. Design, Aug. 2003, pp. 360–365.

[4] T. Kuroda, "Low power CMOS digital design for multimedia processors,"in Proc. Int. Conf. VLSI CAD, Oct. 1999, pp. 359–367.

[5] H. Lee, "A power-aware scalable pipelined booth multiplier," in Proc.IEEE Int. SOC Conf., Sep. 2004, pp. 123–126.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[6] S.-R. Kuang and J.-P. Wang, "Design of power-efficient configurablebooth multiplier," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 57,no. 3, pp. 568–580, Mar. 2010.

[7] O. A. Pfander, R. Hacker, and H.-J. Pfleiderer, "A multiplexer-basedconcept for reconfigurable multiplier arrays," in Proc. Int. Conf. FieldProgram. Logic Appl., vol. 3203. Sep. 2004, pp. 938–942.

[8] F. Carbognani, F. Buergin, N. Felber, H. Kaeslin, and W. Fichtner,"Transmission gates combined with level-restoring CMOS gates reduceglitches in low-power low-frequency multipliers," IEEE Trans. VeryLarge Scale Integr. (VLSI) Syst., vol. 16, no. 7, pp. 830–836, Jul. 2008.

Authors Profile :

Bottu Ravi Teja is currently pursuing his M.Tech specialization in VLSI system design in CMR Institute of Technologywhich is affiliated to JNTUH in Hyderabad.

Muni Praveena Rela is currently working as Associate Professor in Department Of ECE inCMR Institute of Technologywhich is affiliated to JNTUH in Hyderabad.