Design Of Prominent Single Precision 32-Bit Floating-Point Adder Using Single Electron Transistor Operating At Room Temperature

by

Ankur Sharma
201711034

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF TECHNOLOGY

In

INFORMATION AND COMMUNICATION TECHNOLOGY

To

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

JULY, 2019

Declaration

I hereby declare that

i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
ii) due acknowledgment has been made in the text to all the reference material used.

_____

Ankur Sharma

Certificate

This is to certify that the thesis work **Design of prominent single precision 32-bit floating point adder using single electron transistor operating at room temperature** has been carried out by Ankur Sharma for the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology under my supervision.

_____

Dr. Rutu Parekh
Thesis Supervisor

Acknowledgments

First and the foremost, I would like to express my sincere gratitude to my thesis supervisor, Dr. Rutu Parekh for her untiring guidance and support. Dr. Rutu Parekh's patience, motivation and immense knowledge helped me all throughout the research. This work would not have been possible without her support and guidance. I would also like to extend thanks to my evaluation committee members Dr. Anjan Ghosh and Dr. Yash Agrawal and Dr. Tapas Kumar Maiti for conveying their views and suggestions. Their suggestions helped me to analyze different aspects of the problem and hence improve upon my work. I would like to thank my friends for encouraging ideas, stimulating discussions and constantly helping me throughout my thesis and for all the technical and non-technical help. I heartily thank my seniors for their help and support.I would like to thank my family for encouraging me throughout my life. Finally, I would like to thank all the staff members of DA-IICT for providing easy access of the college resources.

Contents

Abstract

The floating-point (FP) arithmetic plays the most important role in computer systems. Many of the digital signal processing systems use floating-point algorithms for floating-point computation, arithmetic operation, and real-number manipulation, and each operating system is essentially responsible for special floating-point cases such as underflow and overflow. Here we are using Single-electron transistor (SET) for floating-point addition. SET offers new functionalities that have no CMOS counterpart. This further results in miniaturization with better performance. Arithmetic operations and approximation calculations of real numbers on modern-day computers are done using floating-Point unit using CMOS technology. In the earlier days, each computer manufacturer had used their own implementation for arithmetic operations in their computer. For a long period, arithmetic operations between different computers varied on bases, significant and exponent sizes, formats, etc. And every company kept implementing their own model until the IEEE 754 standard defined a universal format. This research aims to implement a 32-bit binary floating-point adder using SET according to this IEEE 754 standard. Floating-point addition is the most difficult activity as it incurs more delay and power consumption. We compare the performance of SET based floating-point adder and the CMOS (16nm) based floating-point adder using the simulation results in terms of power and delay. SET is utilized here to make new developments which is difficult to accomplish by CMOS. By using SET based floating-point adder, addition becomes faster while using less power. For simulation, CADENCE virtuoso is used. According to our results, SET based FP addition uses 79.70% less power and gives 97.67% faster results than CMOS based FP addition.

List of Symbols

C_{g1}	Gate capacitance
C_{g2}	Back Gate capacitance
q	Electronic charge
ε_o	Permittivity
C	Capacitance
C_q	Quantum capacitance
e	Electronic charge
C_{TS}	Tunnel capacitance
C_{TD}	Tunnel capacitance
R_{TS}	Tunnel resistance
R_{TD}	Tunnel resistance
R_S	Source resistance
R_D	Drain resistance
C_J	Junction capacitance
R_T	Tunnel resistance
V_c	Channel potential
V_{gs}	Gate to source voltage
V_{ds}	Drain to source voltage
S	Dirac voltage
E	sheet charge
F	Dielectric layer width
p	hole concentration

List of Acronyms

ASIC	Application Specific Integrated Circuit
CMOS	Complementary Metal Oxide Semiconductor
DIBL	Drain-Induced Barrier Lowering
SET	Single Electron Transistor
FP	Floating Point
NaN	Not a Number
ED	Exponent Module
ITRS	International Technology Roadmap for Semiconductors
MOS	Metal Oxide Semiconductor
VLSI	Very Large Scale Integrated circuits

List Of Figures

List of Tables

Chapter 1
Introduction

De-escalation of the MOSFET devices and many integrated devices has accomplished the edge of micrometer and nanometer, because of the advancement in technology of microelectronics and diminished the MOSFET's size [1]. The least transistor measurements will reach beneath 10nm by 2020 according to the new reports of international technology roadmap for semiconductor (ITRS) because of nonstop scaling down of the measurements on integrated circuits and semiconductor chips [2] which is the reason of many problems like as short channel effects, costly lithography, and when we are doping there would be some fluctuations, and ultrathin gate leakage will take place. As stated by Moore's Law and the driving of technology in nano-scale routine, nano-electronic solutions will be most wanted, overcoming the physical and economic barriers of recent technologies.

Nonetheless, the forecast of ITRS [2] explains that by the following decade because of the scaling demerits there will be part of the arrangement law of si based electronics devices, e.g. fabrication inconsistencies and very high power density at a nano-meter level. The conglomeration of short channel effects (SCE) commonly named as hot carrier injection, barrier lowering (DIBL) induced by drain and impact ionization also at such low dimensions leakage current has adverse effects on the performance of the device [4]. The rising significance of short channel effects when the device is nanoscale level prompts the debasement of gate control over the channel current, so ' off ' current is increased in a device, and static power utilization also increased in a device.

The channel length is most strongly scaled for high-speed devices, where, this issue is most severe. To deal with these scaling issues and by new electronic devices which are following Moore's law, for example, FINFET, Carbon nanotubes, non-traditional CMOS devices like double gate transistor, Spin transistors and SET devices have developed in the recent years [5]. As SET consumed very low power, and also set has high integration-density and other exceptional functionalities, SET will be turning into a most appropriate option for CMOS

technology in future for electronic devices. In all the present devices, single-electron transistor (SET) has been a fascinating progressively more consideration. SET has many features like small size, it dissipates extremely low power, and it is faster then MOSFET and has the same device structure as MOSFET. While creation and demonstrating methods for SETs are getting developed, for practical use for SET we are creating SET-based circuit design tools that turns out to be exceptionally fundamental and basic [3].

Most of the VLSI and digital circuit applications need to compute floating-point addition operation. Floating point addition is employed as a basic building block of such circuits. Floating point addition is used for various applications in the field of signal processing, communication, and VLSI systems. As the addition is involved in all of the processing instructions, the performance of the system depends on how effectively the hardware is performing the multiplication. So, it is necessary to develop a high-performance floating point adder.

Arithmetic utilizes a formula based representation of real numbers in computing FP (floating point) because FP uses an approximation method so it can support a trade-off between range and accuracy. Recently, in apps for applications like image processing, medical, human robotics area, simulation, and so on, the need for very high-speed, accuracy and high-precision computing has increased. Floating-point computing is best suited for these apps because, despite the wide dynamic range, it keeps operational accuracy high. Not only for high-speed purpose but also small-scale FP devices are in excellent demand for microcomputer applications [6].

Exponent subtraction, alignment shifting, fraction addition, conversion, standardization shifting, rounding, and post-normalization traditionally involved in implementing FP addition. Generally speaking, these steps are sequential and involve two shifts and three additions. Higher efficiency is generally accomplished by decreasing the algorithm's critical path to the maximum amount of serial activities.

The primary goal of the job carried out here is to use the Cadence Virtuoso design environment to introduce and simulate 32-bit single precision floating point adder for SET and CMOS and to compare their efficiency at 800 mV and operate at room temperature.

1.1 Need of Floating point Operations

On computers, there are several methods of representing real numbers. Floating-point notation, especially the normal IEEE format, is the most popular manner to represent an approximated real numbers in computers as it is handled effectively in most big computer processors. Fixed point positions a radix point in the center of the numbers somewhere and is equal to the use of integers representing parts of some unit. A fixed-point has a set representation window that limits the representation of very big or very small figures. Also, if two big numbers are split, fixed-point is susceptible to a loss of accuracy. Floating-point solves a number of issues with representation. Floating-point uses a kind of accuracy "sliding window" suitable to the number scale. This makes it easy to represent figures ranging from 10^{15} to 10^{-15}. The benefit of FP notation over conventional fixed-point notation is that a much larger and wider kind of values can be supported. In scientific notation, floating-point representation reflects the most prevalent alternative. Scientific notation is a base number and an exponent. 123.456, for instance, could be depicted as 1.23456 or 102. FP numbers are generally shown as the sign bit (S), exponent field (E), and significands bits from left to right into a computer datum. In computing, floating-point defines a scheme that represents numbers that are too big or too tiny to be represented as integer numbers. The scaling base is usually 2, 10 or 16. The typical amount that can be precisely represented in the form: significant bits × $base^{exp}$

"Floating point" term refers as a name suggest that the radix-point can "float;" that is, it can be incorporated anywhere relative to the significant numbers. In the inner representation, this position is stated individually, and floating-point notation can thus be considered as a scientific notation computer.

1.2 Tool Used

Simulation Software: Cadence Virtuoso Analog Design environment has been used. Cadence Virtuoso Environment is Characterization and Validation Software, it also enables fast and accurate entry of design concepts, including handling design intent in a manner that flows naturally into the schematic. By using this virtuoso setting, you can abstract and imagine the many interdependencies of an analog, RF, or mixed signal model to understand and assess their

impact.

1.3 Thesis Organization

Chapter 1 introduces the suggested job performed during the work of the thesis. Chapter 2 defines SET's thorough knowledge, operating principle, manufacturing and configuration of SET as P-MOS-SET or N-MOS-SET. Chapter 3 discusses the 32-bit Floating-Point adder algorithm and comprehensive architecture. Chapter 4 shows the results of simulation and compares the adder based on the technology node SET and CMOS 16 nm. The conclusion is summarized in Chapter 4 at the end.

1.4 Literature Review

SET provides controlling electronic charging at the rate of one electron through single-electron charging or 'coulomb blockade' impact. These devices work by regulating the transfer of charges to nanometer-sized conductive areas or' islands' across tunnel obstacles. The energy required to charge one electron can be sufficiently big to affect the tunneling process. As soon as the 1950s, the option was recognized that a nanostructure's single-electron charging energy could affect the tunneling of even one electron into the nanostructure C.J. in 1951. K.K. in the mid-eighties Likharev and his colleagues anticipated the impacts of single-electron charging in tunnel junctions of nanometer-scale in excellent detail. Progress in nanofabrication methods had resulted in the capacity to manufacture well-defined, nanometer-scale, island and tunnel junctions at this point. This resulted in 1987 to the first low-temperature demonstration of a constructed single-electron device, Fulton and Dolan's single-electron transistor [7].

1.4.1 Single Electron Transistor (SET)

To learn about new technology like SET I have read many papers to learn how it works and its principle and the properties of SET and also learn about the limitation of SET.

Konstantin K. Likharev *et al* describes the single-electron devices physics and also, as well as their perspective and current applications. It also tells about recent research which is done in

this field has generated which may revolutionize memories (RAM) and data-storage technologies [7].

Rutu Parekh, *et al* have shown that how single-electron circuits can be integrated with the CMOS to achieve 3-D integration of nanoelectronic devices heterogeneously. How SET-CMOS hybrid circuit analysis and design at room temperature and in the end tells about the performance of SET and CMOS 22 nm node that how SET outperformed the CMOS [8].

M. A. Bounouar, *et al* describes the DGSET room temperature use and potential in digital logic applications and circuits. This article describes the nature of the DG-SET nano-devices double gate for this purpose. This document offers a full-custom layout of DG-SET-based basic blocks such as SRAM architecture, ALU and Look-Up Table (LUT) [5].

Y. Ono, *et al* describes that SETs have created interest in future nanoelectronic systems owing to their low power consumption, high density and other new functionalities and have become an appealing option for CMOS. [9].

A. Beaumont, C. Dubuc, J. Beauvais, and D. Drouin *et al* tells that from the 1980s Single Electron Devices (SEDs) have been attracting significant amount of attention since it is evident that they can be used to produce low-power logic systems and high-performance sensors. [10].

K. S. Pavu and J. Jacob, ***et al*** showed that SET can be considered as the new transistor of the century. With features like **low power dissipation and intake capability along with very high-speed** SET offers a potential alternative to CMOS circuits [11].

H. Inokawa, A. Fujiwara, and Y. Takahashi, ***et al*** describes the discreteness of electronic charging on Coulomb Island may be linked to multivalued activities in their article. SETs have therefore become very appropriate for any of the multivalued logic [12].

K.-W, *et al* multiband filtering circuits are only possible with SET with its distinctive behavior such as adverse differential transconductivity and regular present oscillation [13].

1.4.2 Floating Point (FP) Adder

IEEE Standard Board and ANSI *et al* the IEEE floating-point standard, finalized in 1985, it was a scheme for FP arithmetic and it was an effort to provide numbering scheme requiring floating-point calculations on different pcs to generate the same outcome [14].

Quach and Flynn *et al* describe an FP adder using a compound mean and adder with two inputs with a row of half adders and a duplicate carrying chain. [15]

N. Burgess *et al* traditional techniques can be discovered in Omondi, which describes algorithms on the sequence of meaning activities: swap, change, add, normalize and round. They also talk about how to build quicker FP adders [16].

Javier D. Bruguera and Tomas Lang *et al* serial calculations such as additions can be decreased by putting additional adders (or sections thereof) in conjunction with the calculation of speculative outcomes for the subtraction of exponents (Ea –Eb and Eb–Ea) and rounding (Ma+Mb and Ma+Mb+1) and then choosing the right outcome. [17].

E. Hokenek and R. Montoye *et al* a leading zero anticipator can be used to calculate the standardization distance to provide a one-bit standardization change in conjunction with the meaning added. [18].

S. Oberman, H. Al-Twaijry, and M. Flynn *et al* a variable latency architecture has been suggested, resulting in an addition based on one, or more clock cycle information. This means, however, that the host system can benefit from up to three concurrently evolving outcomes, which is a significant planning challenge [19].

CHAPTER 2

Single Electron Transistor (SET)

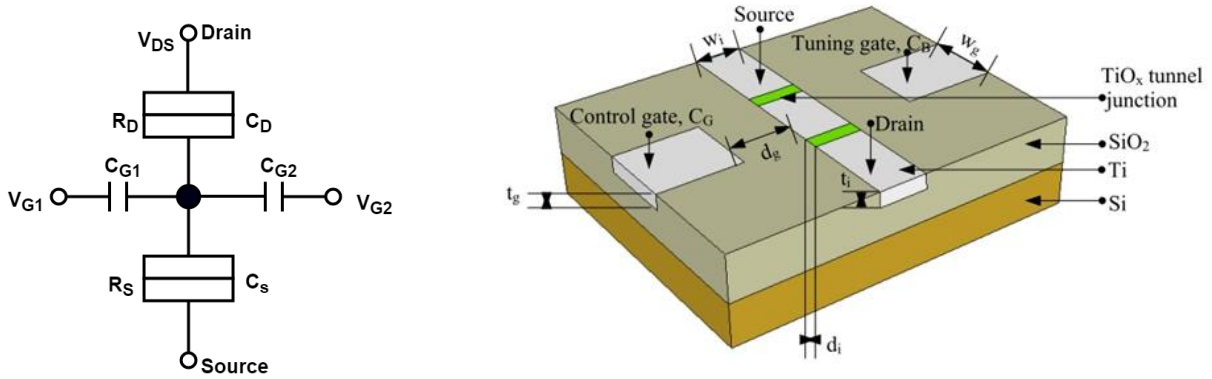
SET is basically a switching tool for the flow of electrons within the machine using regulated electron tunneling. In SET, a nano-dimensional island replaces the channel that is present in MOSFET. Any logical condition can be described by e if we use SET as a circuit construction block.

SET operates on two **Quantum Mechanical** principle :

1. **Coulomb Blockade**
2. **Quantum Tunneling**

2.1 Single electron transistor fabrication

Consider a conducting region which is fabricated in metal or doped semiconductor having dimension less than 100 **nm** which is called the island, this island is near to the other two electrodes and inserting ultra-thin dielectric in between then it works as simple double tunnel junction. If we add the third terminal which is an electrostatically coupled to the island this configuration will work as **Single Electron Transistor** [20].



(a) (b)

Fig. 2.1 (a) Single Electron transistor Symbol and (b) Schematic of SET [8]

As shown in figure 2.1 (a)SET is a 3 terminal device. It has an extra terminal which is called **'back gate'**. Here, $\mathbf{C_{G1}}$ and $\mathbf{C_{G2}}$ are the gate and back gate capacitances. $\mathbf{C_{TS}}$, $\mathbf{C_{TD}}$ are tunnel capacitances are of source and drain respectively and $\mathbf{R_{TS}}$, $\mathbf{R_{T\,D}}$ are tunnel resistances associated with the source and drain respectively.

In the basic structure of a SET, it consists of two tunnel junctions and one conductive island which is sandwiched between source and drain of SET, is depicted in Fig.2.1(a). For a proper task of a SET device requirements are:

1) To impound the electron in the island tunnel junction resistances drain resistance and source resistance (R_D, R_S) should be greater than the Rq , and

2) For avoiding electron tunneling because of thermal emission then the charging energy of the island capacitance should be greater than the available thermal energy.

As shown in figure 2.1 (b) this the SET schematic and this is a 4-terminal device. SET has 2 gates source and drain which are CG1 and CG2. Like to CMOS transistor 1st gate looks like the functionality gate and 2nd gate is explicitly utilized as a back gate for controlling reason [21]. The electrical reaction of SET relies on the parameters of its physical manufacture. The design considers SET logic to function within the hybrid SET–CMOS / SET circuit with CMOS similar voltage range and sufficient driving capacity. At the same time, the parameters of the SET device such as tunnel capacitance C_j, gate capacitance C_g and tuning gate or back gate capacitance C_b must be within the range of manufacture for 300 K. With the increase in temperature, the maximum permitted capacitance decreases. For further analysis of the capability values of the SET tunnel junction, its design range of parameters and tradeoffs. A parallel plate technique can be used to calculate the capacitance value. As illustrated in Fig. 2.1 (b), The door capacitance parameters are w_g, t_g and d_g, while the tunnel junction parameters are wi, ti and di. Table 1 shows the possible range for dimension parameters and the calculated C_j, C_g and C_b values. By varying the physical dimensions of the SET, the desired values of C_j, C_g, and C_b [24, 25] can be obtained. C_j, C_g, and C_b are calculated using the equations using the parallel plate model,

$$c_g, c_b = \varepsilon_r(\mathrm{sio}_2)\varepsilon_o \frac{w_g t_g}{d_g}$$

$\varepsilon_r(Sio_2)$ is the relative permittivity = 3.9 for sio_2 and ε_o is the permittivity of free space where $\varepsilon_o = 8.854\times10^{-12}$ F/m.

$$c_j = \varepsilon_r(tio_2)\varepsilon_o \frac{w_i t_i}{d_i}$$

A SET inverter's higher limit, as well as simulated inherent velocity, is approximately the resistance of tunnel junction, 160 fs response time $R_t \times C_{\Sigma.}$ for the SET parameters of R_t = 1MΩ and C_Σ = 0.16 aF where C_Σ is ($C_\Sigma = 2C_j + C_g + C_b$) [25]. High-frequency SET driving capacity is not restricted by its inherent velocity but by its restricted capacity to drive the high capacity load.

2.2 Working of SET

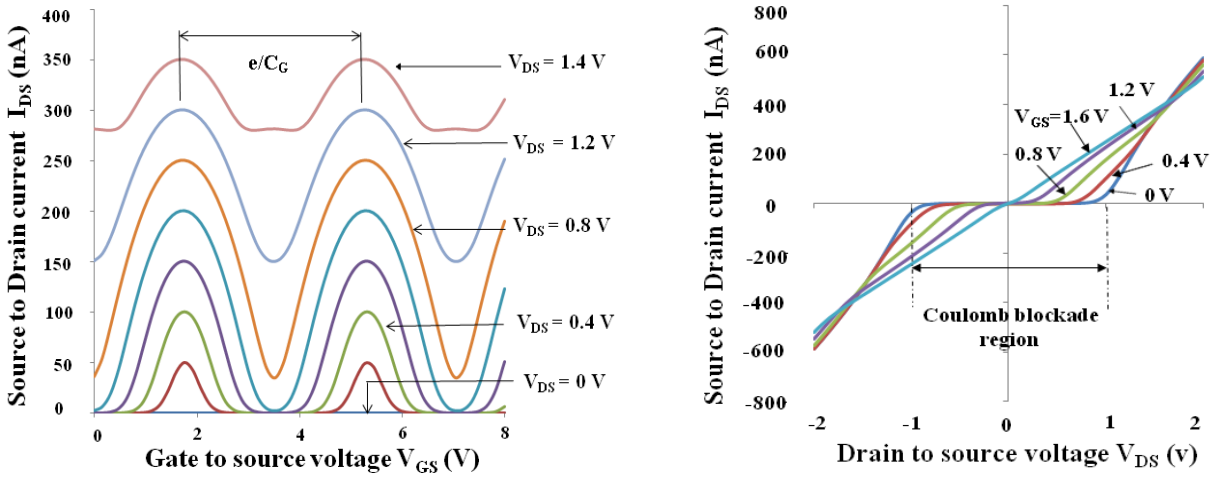
2.2.1 Principle of Operation

Source and island have a number of electrons, and the tunnel junction works as a dielectric that is polarized so that it is capable of polarization. Now, if bias potential V_{DS} is applied by working against this capacity and if this potential or voltage energy is enough to rise above the Coulomb energy E_C so that the electron will tunnel to the island and then again from island to drain terminal. Here, the island's potential will boost by V_{DS} value, which will block other electron's tunneling. If the bias voltage is not enough to overcome E_C, there is a Coulomb blockade and there is no flowing current.

Coulomb Blockade is an impact that handles one electron's tunneling. The single-electron charging energy that is equal to $E_c = e^2/2C_g$ can not be overcome by low bias voltage electron values, so the tunnel junction that acts as a barrier to cross the e^- to reach on the island so because of that no current flows in the device. This is called the **Coulomb Blockade**.

2.2.2 Single Electron Based Transistor characteristics

As needs are the V_{gs} gate voltage is changing and we are fixing the V_{ds} bias voltage, so tunnel junction voltages changes with gate to source voltage V_{gs}. When we increase the voltage on the tunnel junction and the voltage value is greater than the threshold voltage V_{th} then the electron will cross the tunnel junction and will reach the conductive island. At the point when V_{ds} is little at that point tunneling course take on periodicity as the V_{gs}, the current I_{ds} take on coulomb oscillation.



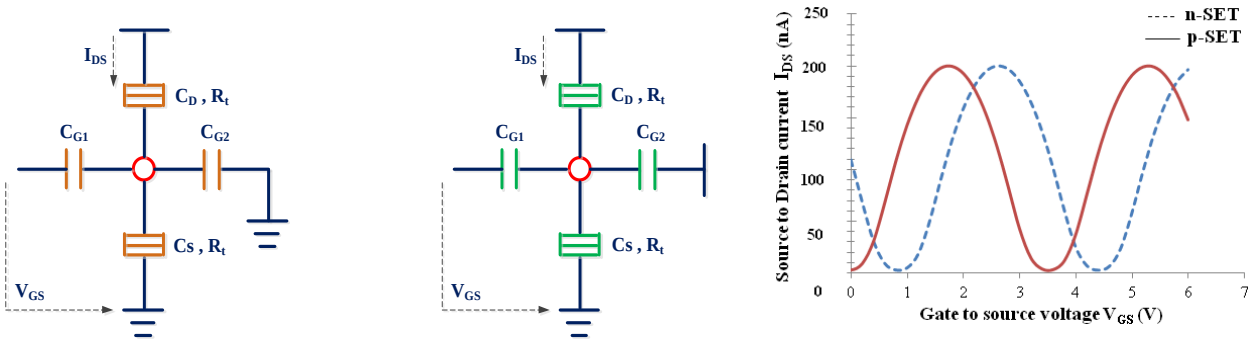
(a) (b)

Fig.2.2 (a) The Id vs Vgs characteristic and (b) The I_d vs V_{ds} characteristic of the SET.

The SET I_{ds}-V_{gs} characteristic shown in the Fig.2.3(a). The Fig.2.3(a) tells that when the number of gates is increasing the oscillation period is changeless, but we can adjust the phase of coulomb oscillation by the gate voltage (v_g).

2.3 Single Electron Based Transistor as P-SET and N-SET

SET displays complimentary behavior if we compare to N-MOSFET and P-MOSFET when the back gate is biased differently which is also known as tuning gate. If we are biasing the tuning or back gate by $\mathbf{V_{DD}}$ then it will result in N-SET and if we are biasing the back or tuning gate by GND or zero voltage then it results in P-SET.



(a) (b) (c)

Fig. 2.3 shows SET configuration as (a) P-SET (b)N-SET. (c)The I_{ds} vs V_{gs} characteristic of the n-SET and p-SET.

2.4 Design Consideration for SET and CMOS 16nm device

For SET device we will see how can we increase the speed and decrease the delay or by increasing the bandwidth by design consideration. these parameter is defined by this work [8]. and The SET characteristics are based on the Mahapatra–Ionescu–Banerjee model (MIB) [26]. We are using 16-nm BSIM predictive model for bulk CMOS [28], and implemented them in Virtuoso Analog Design Environment of CADENCE [27]. So here we have capacitance C_{D1} = C_{S1} = C_{D2} = C_{S2} = C_j, the control gate capacitance, C_{g1} = C_{g2} = C_g; the tuning gate capacitance, C_{b1} = C_{b2} = C_b; so the total capacitance is C_Σ is (C_Σ = $2C_j$ + C_g + C_b) and the tunnel junction resistances, R_{D1} = R_{S1} = R_{D2} = R_{S2} = R_t. The parameters are: C_{g1} = 0.045aF, C_{g2} = 0.050aF, C_d = 0.030aF, C_s = 0.030af, R_t = 1MΩ, T = 300K, and a CMOS 16nm parameter is defined by 16-nm BSIM predictive model [29, 28].

CHAPTER 3

Design of 32-bit Floating-Point Adder (Single Precision)

The following chapter provides a brief description of encoding (numerical) which serves as a standard used to represent arithmetic floating-point, as well as the detailed design of the adder, is also presented.

3.1 Representation of Standard Floating Point

Each real number has two parts namely integer part and secondly fraction part; to distinguish between them, a radix point is used. The number of bits assigned to the integer part may differ from the number of a fractional parts. Figure 3.1 shows a generic binary to decimal conversion.

	Integer					**Binary**	**Fraction**			
Binary	---	2^3	2^2	2^1	2^0	.	2^{-1}	2^{-2}	2^{-3}	---
Decimal		8	4	2	1		½	¼	⅛	

Figure 3.1: Binary to decimal conversion.

3.1.1 Representation of Fixed-Point

A fixed-point representation is one where we can represented number after the decimal point by an approximated fixed number of places. Usually, the decimal point is placed next to the slightest important bit, so only the integer part is represented. The main benefit of this sort of representation of fixed point is that they can be implemented with integer arithmetic and small values can be stored. This helps to increase the speed and efficiency of the activities. The main disadvantage is that there is limited or no flexibility in a fixed-point number.

3.1.2 Two’s Complement Representation

The 2's complement notation is used to depict both positive and negative fixed-point numbers. The complement of positive numbers is depicted as easy binaries. A negative number is depicted in such a manner that the response is zero when this number is added to a number

(positive) of the same magnitudes. The most important bit is called the sign bit in the 2's complement. The amount is not-negative, i.e., zero or high if the sign bit equals to zero. If the bit of the sign is 1, the amount is negative or below 0. For calculating a 2's complement add-on or a -ve of a fix binary integer no., the first 1's complement add-on is performed, i.e. all the bits are inverted and then we can add 1 to the result.

3.1.3 Representation of Floating-Point Numbers

In example, a FP number is expressed as ± b.bb... b × ß^{E}, b.bb... b is the mantissa and has q digits are the precision of the given number, and here ß is the base which is "10" for decimal numbers, "2" for binary numbers or "16" for hexadecimal numbers. If ß= 10 and q= 4, then we can write this numbers 0.1 is as 1.000 × 10^{-1}. If ß= 2 and q = 19, then the this no. 0.1 will not be written as above, but it is 1.100110011001100110 × 2^{-4}. Thus the most significant factor in the depiction of the floating-point is the accuracy or amount of pieces used to represent the meanings. Other significant parameters for a certain depiction are E_{max} and E_{min}, the biggest and the lowest encoded exponents, providing the variety of a number.

3.2 FP Representation based on IEEE 754

Arithmetic operations and approximation calculations of real numbers on modern-day computers are done using floating-Point arithmetic. In the earlier days, each computer manufacturer had used their own implementation for arithmetic operations in their computer. For a long period, arithmetic operations between different computers varied on bases, significant and exponent sizes, formats, etc. And every company kept implementing their own model until the IEEE 754 standard defined a universal format. In 1985, the Institute of Electrical and Electronics Engineering (IEEE) released a binary floating-point arithmetic standard, 754 [15]. This standard was necessary to exclude the arithmetic vagaries of the computing industry. This chapter discusses the normal elements used to implement and evaluate different floating-point adder algorithms.

3.2.1 Formats

There are 2 fundamental formats defined in IEEE 754 format, a). 64-bit double-precision,

and b). 32-bit single-accuracy. Table 3-1 demonstrates the differences between the basic aspects of the two.

Table 3.1: Summary of Single & double precision format

Parameter	Single Precision	Double Precision
Format width (bits)	32	64
Precision	23+1	52+1
Exponent width (bits)	8	11
Maximum value	+127	+1023
Minimum value	-126	-1022

Considering only the single-precision format to assess distinct adder algorithms. Single-precision format utilizes sign bit 1-bit, exponent 8-bit and 23-bit to represent the fraction as shown in Figure 3.2.

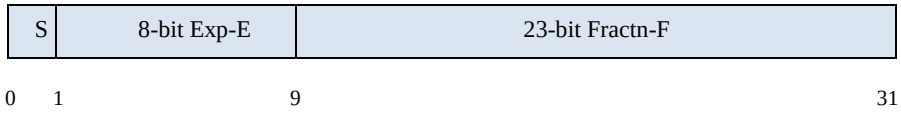


Fig. 3.2: IEEE 754 format for single precision

The single-`precision number is expressed as (-1) (Sign bit) × 2 (Exponent – 127) × (1. Mantissa). For non-negative numbers, the sign bit is either 0 or 1 for negative numbers. To do this, the real exponent is added a bias. For instance, this value is 127, and a stored value of 180 shows a (180-127) or 53 exponents. The mantissa or meaning is made up of a leading bit as well as fraction bits, representing the number's accuracy parts. Special numbers such as zero, denormalized number, ∞, and -1s are reserved for exponent values (hexadecimal) of 00FF and 0000. Table 3-2 summarizes the mapping.

Table 3.2: Denormalized numbers

Exponent	Mantissa	Object Represented
0	0	0
0	Non-zero	denormalized
1-254	anything	FP number
255	0	infinity
255	Non-zero	NaN

3.2.1.1 Normalized numbers

A FP number would be standard if it is actual or given exponent and bias is other than the 00FF and 0000. If the numbers are normalized, then the first bit is considered as 1 which is the left to the decimal point and is not specified in the FP addition representation and therefore also called it is called the hidden-bit. Single-precision encodes therefore only the decreased 23 bits.

3.2.1.2 Denormalized numbers

A FP number will be measured to be denormalized When the exponent field which is shown by E is 0 and the fraction field (F) does not include all 0's then the FP number is a denormalized number. The concealed or implicit bit is forever set to 0. These numbers will tend to fill the gap among 0 and the smallest standardized number.

3.2.1.3 Infinite

In this depiction, infinity is defined by the 0xFF exp field and the entire 0's fraction field.

3.2.1.4 Not a Number (NaN)

In this depiction, NaN is written as by E field of 00FF and the F part which is fraction part

that doesn’t include all 0’s.

3.2.1.5 Zero

In this illustration, zero is written as an exponent part of 0000 and the whole fraction part of zeros. The sign part shows -0 and +0, respectively.

Table3.3: Operations that can generate Invalid Results

Operation	Remarks
Addition/ Subtraction	An operation of the type $\infty \pm \infty$
Multiplication	An operation of the type $0 \times \infty$
Division	Operations of the type 0/0 and ∞/∞
Remainder	Operations of the type x REM 0 and ∞ REM y
Square Root	Square Root of a negative number

3.2.2 Rounding Modes

Rounding requires an infinitely accurate amount and, if needed, changes it to fit in the format of the destination while indicating the inaccurate exception. The default mode is REN and is mostly used in software and hardware for all arithmetic applications. The representable value closest to the infinitely accurate consequence is selected in this mode. If the two closest representable values are equally close to each other, the one which is least and more close to zero will be selected.

3.2.3 Exceptions

The IEEE standard 754 has 5 types of exceptions which are explained below. We can enabled the exceptions by writing a flag or a trap.

3.2.3.1 Invalid operation

It is impossible to perform the specified procedure on the operands. In an adder, when both the numbers are infinity then we get the invalid results.

3.2.3.2 Inexact result

It is defined if the rounded outcome is not accurate or if it overflows without a trap.

3.2.3.3 Division by zero

In this a zero gets divided by zero, the result is ∞.

3.2.3.4 Overflow

The rounding mode used defines the exception to overflow. In REN, overflow happens when the rounded outcome has an exponent equal to 0xFF or is infinite.

3.2.3.5 Underflow

Exception underflow happens when a loss of precision happens. If the implicit outcome bit is 0 and the exponent output is -126 or 0001, the amount is too low to be fully represented in a single accuracy format and the underflow flag is set.

3.3 Standard floating-point algorithm for addition

3.3.1 Algorithm

Figure 3.3 shows the flowchart of the standard floating-point adder algorithm.

1. First, check the exponents are the same or not.
2. If exponents numbers are not the same then we right shift the smaller number.

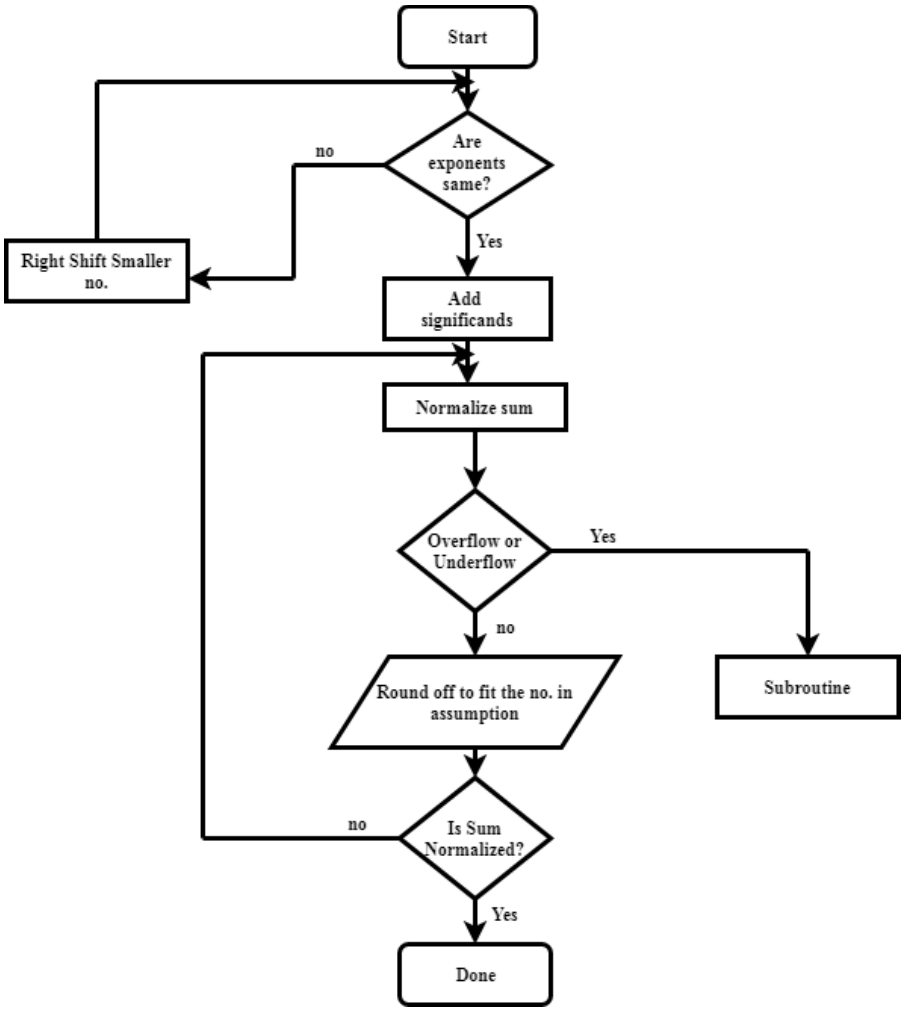


Figure 3.3: Flow chart for adder

3. If exponents are the same we add the significands.
4. After addition, we check that sum is normalized or not.if not then we normalize the sum
5. Now after normalization of sum, we check that there is overflow or underflow in this.
6. If there is overflow or underflow it has to go from the subroutine process.
7. If there is no overflow or underflow then the next step is we'll round off to fit the number in assumptions.
8. After rounding off again we check sum is normalized or not if not then go to step 4 and normalize the sum.
9. If normalized then the process is done.

3.3.2 Micro-Architecture

The conventional floating-point adder has been constructed using the above algorithm. Figure 3.4 shows the comprehensive design micro-architecture. It indicates the primary hardware modules needed for the addition of floating-points. Each module will be provided with a comprehensive description and functionality later in this section. Exponent difference module, correct shift shifter, 2's complement, adder, multiplexer, mux logic, and rounding module are the primary hardware components for a single-precision floating-point adder.

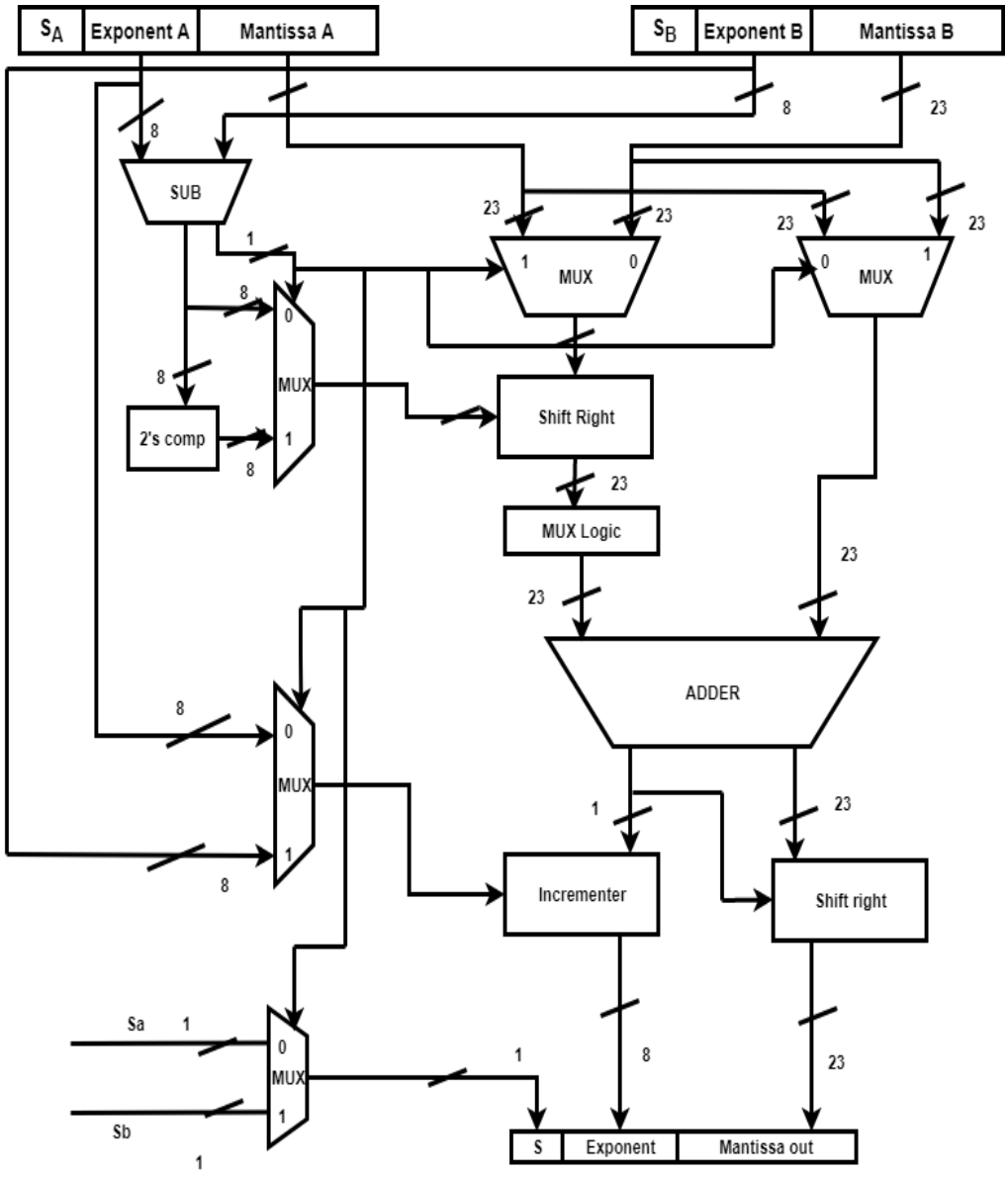


Figure 3.4: the architecture of Adder

3.3.3 Exponent Difference Module

This module has two functions:

1. To calculate the difference between two numbers (8-bit).
2. To check whether e1 is less than e2.

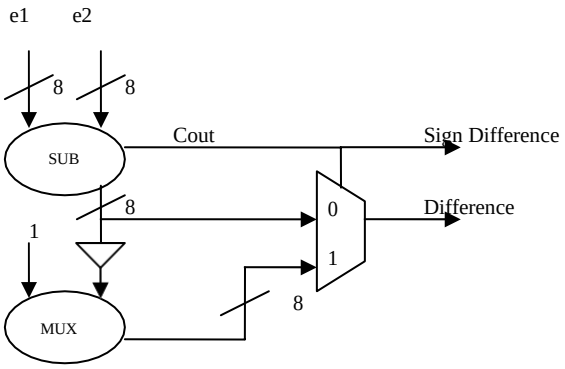


Figure 3.5: Implementation for exp difference

An adder (8-bit) is incorporated to check the difference between the exponents and the carry out determines if e1 is less than e2. If the output is negative, it is complemented and a 1 is added to it give out the difference. For example, we have two exponents it subtracts the numbers and give the difference between two numbers and goes to mux. If the number is negative then it goes to 2's compliment and goes to mux.

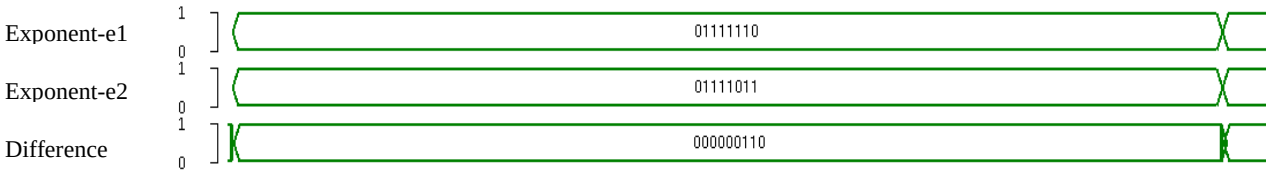
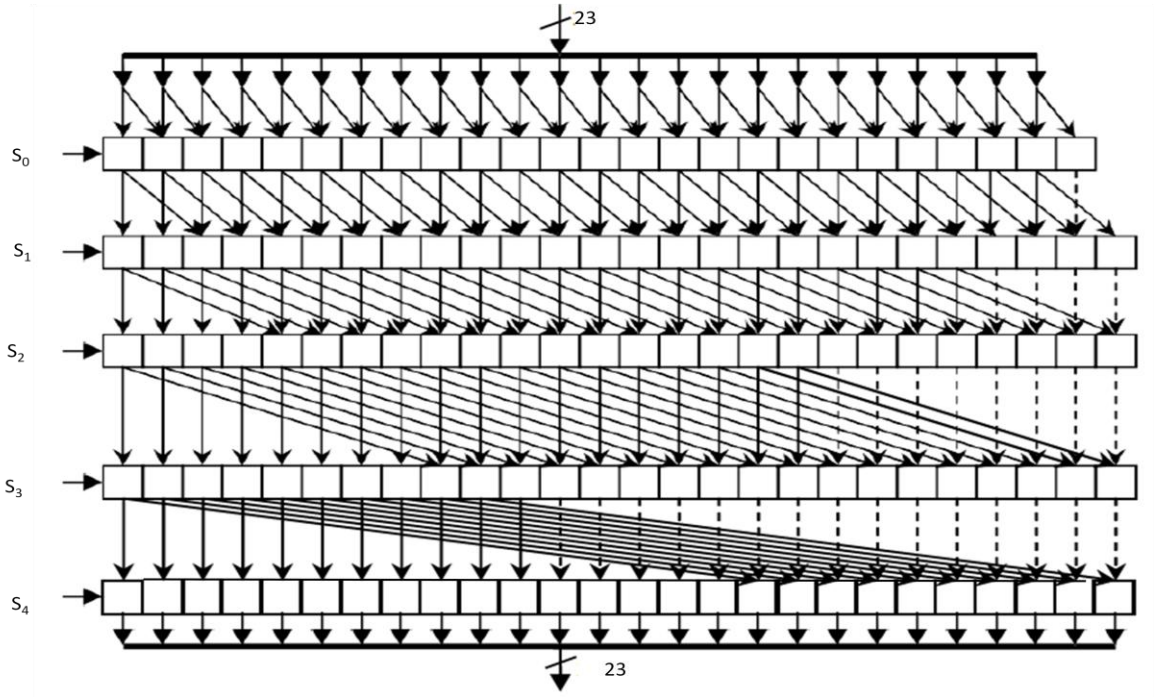


Figure 3.6: Simulation results for exponent difference

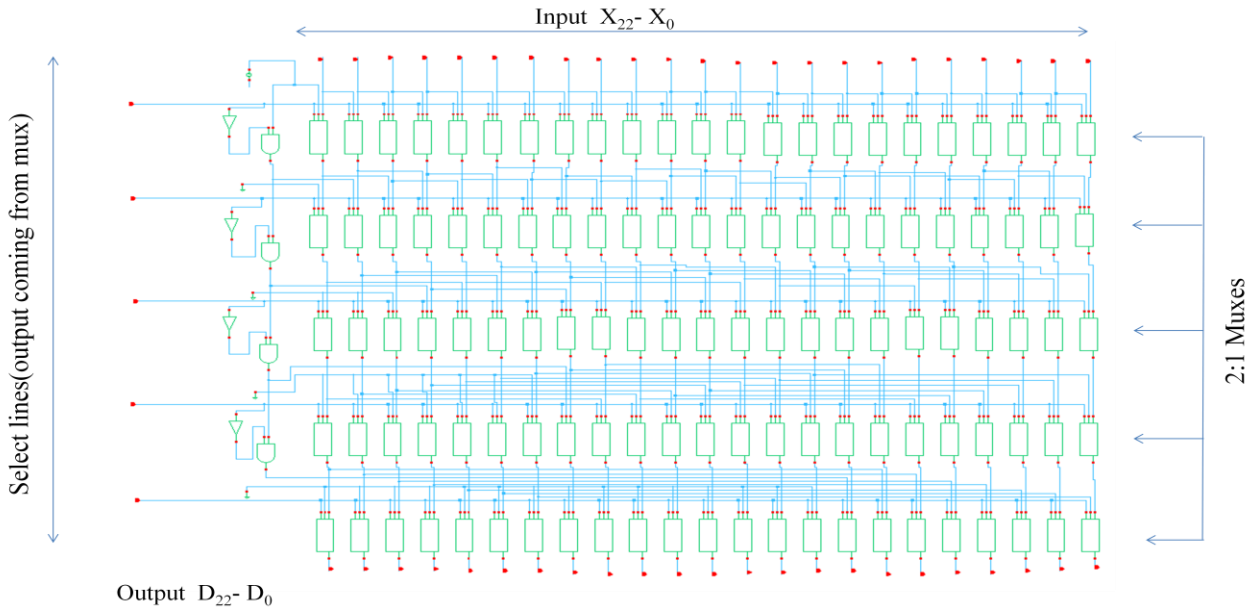
3.3.4 Right Shift Shifter

The right shifter is shifted the smaller number by the significant bits which is the difference of exponent. This will make sure that given two numbers of FP addition will have the exact same exponent and the normal addition of integer can be done.

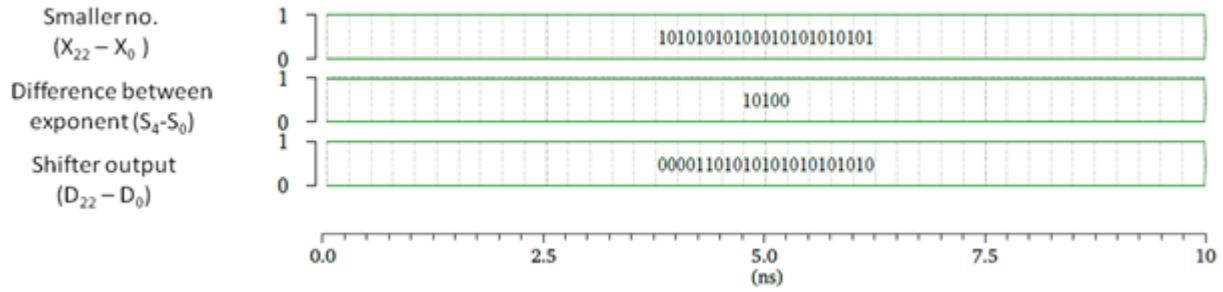
By using the concatenation operation we have designed the bit shift modules of 1, 2, 4, 8, and 16 in this shifter. The concept of this shifter and barrel shifter are the same but the multiplexers are implemented behaviorally in this.



(a)



(b)



(c)

Figure 3.7: (a) The architecture of barrel shifter (b) Implementation barrel shifter in cadence and (c) simulation results for the shifter in cadence.

3.3.5 Rounding

Rounding is done using the result's guard, round and sticky bit. If the guard bit is set, this is done by rounding up and then grounding output LSB if both r & s are zero. If the r & s bit is one or r and either of the last two bits of the result is one, one is added to the output. This step is important to ensure accuracy and omit accuracy loss.

3.4 Example

In this example, we are adding the two numbers with the same sign. If the numbers are not of the same sign then we need to do subtraction of the numbers.

ADDITION
example on decimal value 0.6 and 0.1 which are given in scientific notation:

Number 1 = 0.6 = 0.100110
Number 2 = 0.1 = 0.000110

Now represented the numbers in FP format

First step: make both exponent equal

1.001100×2^{-1}
1.100110×2^{-4}

Second step: then add both the numbers

1.000110×2^{-1}

0.001100×2^{-1}

1.011001×2^{-1} (sum)

Third step: normalize the result

$Sum=-1.011 \times 2^{-1}$

example on FP value given in binary:

Example

0.6 = 0 01111110 00110011001100110011010

0 .1 = 0 01111011 10011001100110011001101

to add these two floating-point representations,

Step 1 If exponents numbers are not the same then we right shift the smaller number.

shift by 01111110

-01111011

00000011 (3) places.

0 01111011 10011001100110011001101 (original)

0 01111100 11001100110011001100110 (1 place shifted and in msb of mantissa hidden bit is shifted)

0 01111101 01100110011001100110011 (2 places shifted)

0 01111110 00110011001100110011001 (3 places shifted)

Step 2: Addition

0 01111110 1.00110011001100110011010 (0.6)

\+ 0 01111110 0.00110011001100110011001 (0.1)

---

0 01111110 1.01100110011001100110011

Step 3: Normalize the output and the. the result is(get the "hidden bit" to be a 1)

0 01111110 01100110011001100110100

In this chapter, we explained FP representation as a standard IEEE 754 format briefly. After that, we also explained the conventional or naive floating-point adder algorithm and also demonstrated FP architecture. Each module of FP adder is logically implemented, discussed and analyzed in this chapter.

CHAPTER 4
Design and Simulation Results

In this chapter, we will show the design and simulation results of SET based FP adder and CMOS 16 nm based FP adder in terms of delay and power in cadence virtuoso

4.1 Introduction to Cadence Virtuoso Tool

Virtuoso is a custom design tool supplied by the Cadence environment. Here, for any circuit, the fundamental building block is the NAND gate as it has low power consumption compared to any other logic gate. The application of the NAND gate for CMOS technology is performed using the BSIM16 library in virtuoso. This library uses a node of 16 nm technology. The channel length of the cells used for models. Delay and power different entities are calculating by Virtuoso tool which is described as below :

1. **Power -**

In a specific circuit, there has always been a problem to decrease power to increase its battery life. There are two kinds of energy basically: leakage and dynamic. leakage power is dissipated when the inputs are stable while Dynamic power is the power produced during the transitions in the model.

2. **Delay -**

The delay mentioned at this juncture is mainly the delay in propagation that we can calculate this from the outcomes of the simulation by identifying the difference between 50 percent of the input and 50 percent of the output times.

Dynamic power can not be directly calculated from simulations results, because no input can have a change transitions only. So, we can calculate total power by giving constant inputs as leakage power and giving pulses as inputs. By subtracting Leakage Power from Total Power, dynamic power can be calculated.

For a circuit, the transient response is acquired by choosing the' tran' choice from the simulator

analysis window. Time for analysis is set and pulses are provided for inputs. You can find power from the calculator by ‘average of the (supply voltage x current through it)’ this formula or from the Result Browser window.

Static power is measured by analyzing' dc' that is present in the analysis window. Inputs should not be set as pulses for static power measurement but should be set as dc.

4.2 Single precision floating point adder

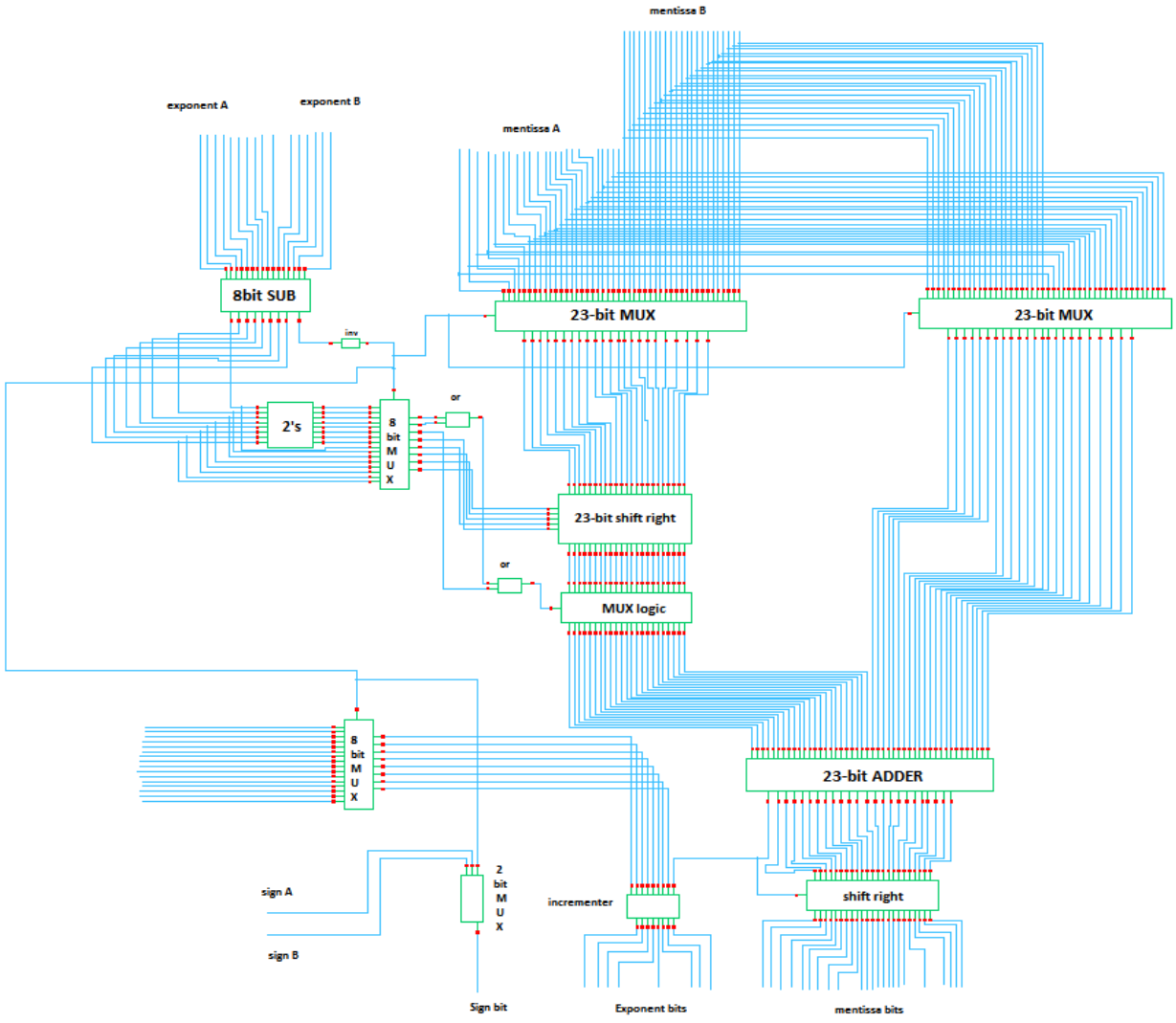


Figure 4.1: 32-bit floating-point adder (in cadence)

4.3 Simulation Results

4.3.1 Parameters

Below the parameter of the CMOS 16nm Single precision Floating Point Adder are pmos length L_P = 14.3nm, pmos width W_P = 17.6nm, nmos length L_N = 13.2nm, nmos width W_N= 17.6nm. Below is the characteristics of the SET. These SETs parameter values are similar which is taken in [10]. As SETs are metallic and made by TiOx/Ti ,so we are using SET modeland using this SET at room temperature (300 K), and created in for the simulations. Design parameters for SET is which are derived in lab for fabrication capability: C_B = back gate capacitance = 50zF, junction capacitance (C_D and C_S) = 30zF, C_G = gate capacitance = 45zF, , tunnel junction resistance R_T = 1M, high voltage = 800 mV, low voltage = 0 V, gain = 1, T = 300 K [8].

Table 4.1: SET parameters for simulation [8]

Parameter	Value
C_{g1}	0.045aF
C_{g2}	0.050aF
C_d	0.030aF
C_s	0.030af
R_t	1MΩ
T	300K

4.3.2 Result

All simulations are performed in the analog design environment of Cadence Virtuoso. The outcomes of the simulation we have been merged to form a bus. Here, the circuit's fundamental construction block is the NAND gate that is implemented separately using both CMOS and SET techniques. As shown below, the results are.

Results

CMOS :

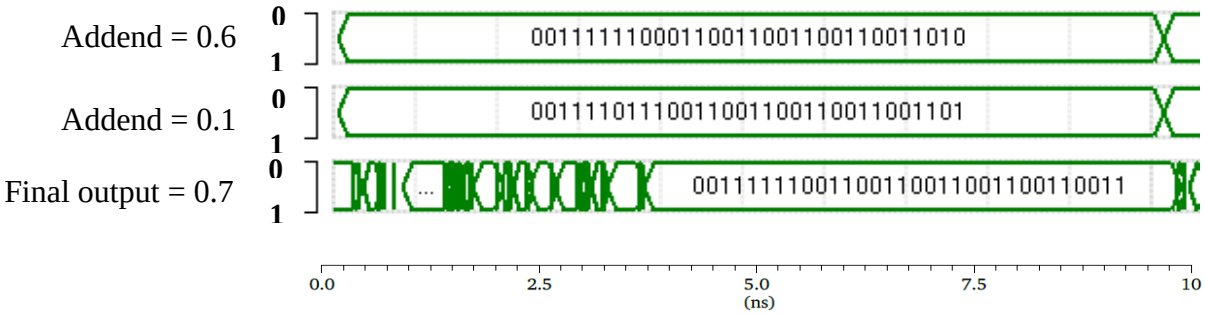


Figure 4.2: Simulation Result for CMOS design

SET :

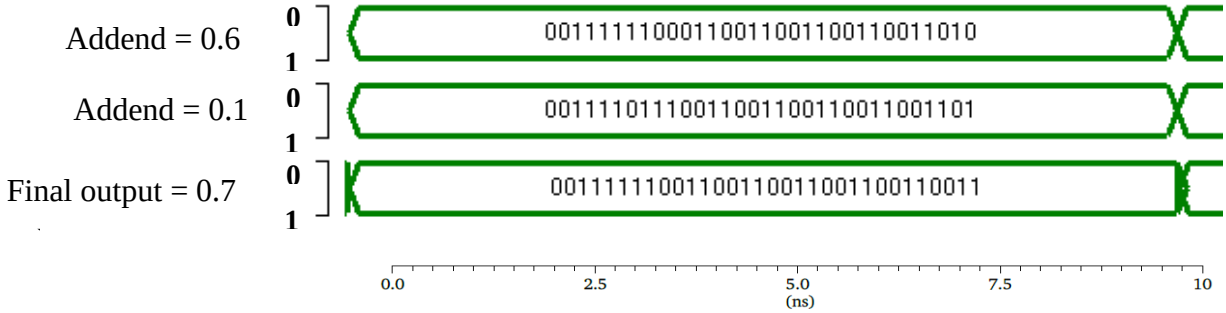


Figure 4.3: Simulation Result for SET design

4.3.3 Performance Analysis

We proposed an efficient single-precision floating-point adder using a single electron transistor (SET). We have also shown the resilience of SET that we can use SET as P-type and N-type by configuring the second gate of SET. As per my results, this SET based FP is energy

Table 4.3: Performance Analysis

	LOGIC	Power(uW)	Improvement	Delay(ps)	Improvement
FP Adder	SET	3.413	79.70%	21	97.67%
	CMOS(16nm)	16.81		904.406	

efficient than its 16nm CMOS FP. Delay and Power have been diminished to a significant amount in the case of SET based FP design. With the given set of parameters and variables, Floating-Point Adder was analyzed Improvisation of performance using SET in percentage is shown in Table 4.3 above. It is noteworthy that Floating-Point Adder based on SET is power effective and has less delay than CMOS(16 nm). SET based Single precision 32-bit FP adder design is 79.70% power efficient and 97.67% faster as compared to 16 nm CMOS based design.

SET consumes less power and less delay if you compared with CMOS (16nm) hence SET offers better performance because SET is a low current device in nA range compared to MOSFET that is in micro amperes. Since power is proportional to square of current it is naturally low compared to CMOS. when a SET is driving SET the delay is less due to tunneling of the current and the dimensions of quantum dot in few nm. So it is faster compared to CMOS where is current is because of drift. But when a SET drives CMOS the delay is more as it has to drive a heavy load with little current. In addition the SET referred in this work is metallic SET. So the parasitics at the interconnection are reduced. Due to this it offers low power with high speed when compared to CMOS.

CHAPTER 5
Conclusion

We proposed an efficient single-precision floating-point adder using a SET in this research. We have also shown the resilience of SET that we can use SET as P-type and N-type by configuring the 2^{nd} gate by making it GND and V_{dd} of SET respectively. As per my results, this SET based FP is energy efficient than its 16nm CMOS FP. Delay and Power have been diminished to a significant amount in the case of SET based FP design. SET is a low current device in nA compared to MOSFET that is in micro amperes. Since power is proportional to square of current it is naturally low compared to CMOS (16nm). When a SET is driving SET the delay is less due to tunneling of the current and the dimensions of quantum dot in the range of few nm. So it is faster compared to CMOS (16nm) where is current is because of drift. But when a SET drives CMOS (16nm) the delay is more as it has to drive a heavy load with little current. SET based Single precision 32-bit FP adder design is 79.70% power efficient and 97.67% faster as compared to 16 nm CMOS based design. By these results, we can say that emerging technology SET can be alternate for conventional CMOS in the future and also could overcome the power consumption and delay issues.

References

[1] G.Zardalidis I.Karafyllidis "SECS: A new single-electron circuit simulator" IEEE Transaction on Circuits and Systems I,vol.55,no 9,2008.

[2] L. Wilson, “International Technology Roadmap for Semiconductors (ITRS),” Semiconductor Industry Association, 2013.

[3] Fengming Zhang, Rui Tang, Yong-Bin Kim “SET-based nano-circuit simulation and design method using HSPICE”. Microelectronics Journal xx (2005) 1–8.

[4] Islam, A. (2015), “Technology scaling and its side effects”, 2015 19th International Symposium on VLSI Design and Test.

[5] M. A. Bounouar, D. Drouin, and F. Calmon, “Towards nano-computing blocks using room temperature double-gate single-electron transistors," in New Circuits and Systems Conference (NEWCAS), 2014 IEEE 12th International. IEEE, 2014, pp. 325–328.

[6] IEEE journal of solid-state circuits, vol. sc- 19, no, 5, October 1984 a CMOS floating-point multiplier Masaru uya, katsuyuki kaneko, and juro yasui.

[7] K. K. Likharev, “Single-electron devices and their applications,” Proceedings of the IEEE, vol. 87, no. 4, pp. 606–632, 1999.

[8] R. Parekh, A. Beaumont, J. Beauvais, and D. Drouin, “Simulation and design methodology for hybrid set-cmos integrated logic at 22-nm room-temperature operation,” IEEE transactions on electron devices, vol. 59 no. 4, pp. 918–923, 2012.

[9] Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, “Si complementary single-electron inverter,” in Electron Devices Meeting, 1999. IEDM’99. Technical Digest. International. IEEE, 1999, pp. 367–370.

[10] Beaumont, C. Dubuc, J. Beauvais, and D. Drouin, "Room-temperature single-electron transistor featuring gate-enhanced on-state current,” IEEE Electron Device

Letters, vol. 30, no. 7, pp. 766–768, 2009.

[11] K. S. Pavu and J. Jacob, “Analysis of discrete conduction spectrum of quantum dots in single-electron transistors," in Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on. IEEE, 2015, pp. 1–4.

[12] H. Inokawa, A. Fujiwara, and Y. Takahashi, “A multiple-valued logic with merged single-electron and mos transistors,” in Electron Devices Meeting, 2001. IEDM’01. Technical Digest. International. IEEE, 2001, pp. 7–2.

[13] K.-W. Song, Y. K. Lee, J. S. Sim, H. Jeoung, J. D. Lee, B.-G. Park, Y. S. Jin, and Y.-W. Kim, “Set/CMOS hybrid process and multiband filtering circuits," IEEE Transactions on Electron Devices, vol. 52, no. 8, pp. 1845–1850, 2005.

[14] IEEE Standard Board and ANSI, “IEEE Standard for Binary Floating-Point Arithmetic,” 1985, IEEE Std 754-1985.

[15] N. Quach and M. Flynn. Design and implementation of the snap floating-point adder. Technical Report CSL-TR- 91-501, Stanford University, Dec. 1991.

[16] N. Burgess “ Prenormalization Rounding in IEEE Floating Point Operations Using a Flagged Prefix Adder” The IEEE Transactions On Very Large Scale Integration(VLSI) Systems, VOL.13, NO.2, February 2005. Pages 266-277.

[17] Javier D. Bruguera and Tomas Lang "Leading one anticipation scheme for latency improvement in single data path floating-point adders", Department of Electrical and Computer Engineering, Spain. Pages 125-133, 2005.

[18] E. Hokenek and R. Montoye. Leading-zero anticipator (lza) in the ibm risc system/6000 floating-point execution unit. IBM J. Res. Develop., 34(1):71–77, Jan. 1990.

[19] S. Oberman, H. Al-Twaijry, andM. Flynn. The snap project: Design of floating-point arithmetic units. In Proc. IEEE 13th Int'l Symp. on Computer Arithmetic, pages 156–

165, 1997.

[20] Z. A. K. Durrani, Single-electron devices and circuits in silicon. World Scientific, 2010.

[21] Raut, Vaishali, and P. K. Dakhole. "Design and implementation of single-electron transistor NBIT multiplier", 2014 International Conference on Circuits Power and Computing Technologies [ICCPCT-2014], 2014.

[22] Z.A.K. Durrani, Single-electron devices and circuits in silicon. World Scientific,2010.

[23] J. Liang, R. Tessier and O. Mencer, "Floating Point Unit Generation and Evaluation for FPGAs," IEEE Symp. on Field-Programmable Custom Computing Machines, pp. 185-194, April 2003.

[24] C. Dubuc, J. Beauvais, D. Drouin, A. Nanodamascene, Process for advanced single-electron transistor fabrication, IEEE Trans. Nanotechnol. 7 (1) (2008) 68–73.

[25] Beaumont, C. Dubuc, J. Beauvais, D. Drouin, Room temperature single electron transistor featuring gate-enhanced ON-state current, IEEE Electron Devices Lett 30 (7) (2009) 766–768.

[26] S. Mahapatra and K. Banerjee, "Analytical modeling of single electron transistor for hybrid CMOS-SET analog IC design," IEEE Trans. Electron Devices, vol. 51, no. 11, pp. 1772–1782, Nov. 2004.

[27] Cadence Design Systems. [Online]. Available: http://www.cadence.com

[28] W. Zhao and Y. Cao, "Predictive technology model for nano-CMOS design exploration," ACM J. Emerging Technol. Comput. Syst., vol. 3, no. 1, p. 1, Apr. 2007, Article 1.

[29] Predictive Technology Model (PTM) http://ptm.asu.edu/.