Design of High Speed in Memory Serializer/Deserializer with Integrated Sense Amplifier

SRINIVASAN MUTHUKRISHNAN

MASTER´S THESIS

DEPARTMENT OF ELECTRICAL AND INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING | LTH | LUND UNIVERSITY

Design of High Speed in Memory Serializer/Deserializer with Integrated Sense Amplifier

Srinivasan Muthukrishnan
sr3046mu-s@student.lu.se

Department of Electrical and Information Technology
Lund University

Supervisor: Babak Mohammadi
babak.mohammadi@eit.lth.se

Examiner: Pietro Andreani
pietro.andreani@eit.lth.se

September 11, 2019

Printed in Sweden
Tryckeriet i E-huset, Lund

Abstract

On chip communication between different intellectual properties (IPs) such as memories, processors present inside integrated circuits(IC's) using serial interconnect instead of parallel interconnect is explored in order to reduce the interconnect area. Different blocks such as serializer/deserializer, peripheral circuit (to read data from embedded memory IP's) required to implement serial interconnect is discussed. 32-bit data packet was transmitted in the serial link operating at 8 GHz. The energy required to transfer single bit of data was 675fJ. 28nm CMOS BULK technology was used to design the circuits presented in the thesis. This research was carried out in Xenergic AB

Acknowledgements

I would like offer my special thanks to my supervisor Babak Mohammadi,for being patient with my work and also providing me this opportunity to work with him. I deeply appreciate the time he has spent with me to complete this thesis. I would also like to thank my fellow colleagues at thesis place who helped me to complete this project.

I would also like to thank the professors Joachim, Henrik, Pietro at EIT, LTH for their kind help to complete this master program. I also owe special thanks to Dr. Ali Zaheer and Siyu tan for their guidance during the IC project. I would also like to thank my friends who partnered me in different courses for the completion of this master degree.

I am forever indebted to three of my friends Shahi, Ghanshyam, Anup without whom this master journey would not be a possibility. I m also thankful for my parents for their help in this journey.

I dedicate my thesis to Shri Narayana- the eternal witness beyond the manifest, and to the timeless masters of the Advaitha Vedanta (Non-dualistic Idealism) philosophy.

Popular Science Summary

With the advent of IoT(Internet of Things), it is estimated about 30.73 billion [1] devices to be connected together in the year 2020. For this purpose, application specific integrated circuits are preferred for various sensing applications. For such applications, several IPs such as processors, memories are embedded into single chip. In such designs, data needs to be physically transfered from one unit to another. In order to maintain throughput, parallel interconnect is preferred. But as the width of the data bus increases, the interconnect area increases, causing unnecessary area overhead.

One solution to reduce the interconnect area is to use serial interconnect instead of parallel interconnect. There are some problems associated with design of serial link. First, requirement of additional circuit to perform the task of serilization and deserialization at the source and terminal unit respectively. Second, Latency in data transfer due to the addition of serializer/deserializer circuits. Third, requirement of high speed clock.

This thesis explores the possibility of bidirectional serial data transfer between two IPs such as processor and memories. The thesis also includes design of peripheral circuits such as sense amplifier to read data from different type of embedded memories such as eDRAM and SRAM and to integrate it with the serial link. The initial idea was to save area in peripheral circuit of memory by integrating latch type sense amplifier to the serial link so that additional flip flop after sense amplifier can be avoided.

The initial evaluation showed that serial link can reduce the interconnect area by 80% and can increase depending on the scalability of data bus width. This on chip communication using serial link can be used as a way of area optimization for integrated circuits.

Contents

1 Introduction **1**
1.1 Thesis Motivation . . . 1
1.2 Thesis Aim . . . 2
1.3 Project Specifications . . . 3
1.4 Thesis Organisation . . . 3

2 Relevant Concept **5**
2.1 Bit Cell . . . 5
2.2 Write Function . . . 6
2.3 Read Function . . . 7
2.4 Memory Macro . . . 7
2.5 Serial Link . . . 8

3 Background Research **13**
3.1 Data Sensing Models . . . 14
3.2 Single ended dynamic sense amplifier . . . 16
3.3 Voltage Latch Sense amplifier . . . 18
3.4 Serializer . . . 21
3.5 Deserializer . . . 23

4 Circuit Design **25**
4.1 Current Mode Sense Amplifier . . . 25
4.2 Latch mode sense amplifier . . . 27
4.3 Serializer Design . . . 28
4.4 Deserializer design . . . 29
4.5 Clock divider . . . 30
4.6 Transmission Driver . . . 31
4.7 RLC modeling . . . 32
4.8 Receiver . . . 33

5 Layout **35**
5.1 Symmetry and Matching . . . 36

6 System Integration and Test Bench **39**

6.1 eDRAM 39
6.2 SRAM 40
6.3 Serializer test bench 41
6.4 Deserializer test bench 41

7 Simulation 43

8 Conclusion and Future Work 55
8.1 Conclusion 55
8.2 Future Work 55

Bibliography 57

A SERDES 59

List of Figures

1.1 Shared bus architecture 1
1.2 Serial link 2
1.3 System model 2

2.1 Bit cell of a SRAM 5
2.2 Bit cell of a eDRAM 6
2.3 Memory Macro 7
2.4 Bit line modeling of bitcells in a single column in a macro [10] 8
2.5 Analysis model for fig.2.4 8
2.6 A typical SERDES system 9
2.7 Serializer Idea 9
2.8 Encoding Schemes used in Serial link 10
2.9 Transmitter Output Swing 11
2.10 Clock and data recovery waveform 12

3.1 Histogram of bitline swing 13
3.2 Current sensing scheme [12] 14
3.3 Voltage sensing scheme [12] 14
3.4 Single ended voltage sensing scheme 15
3.5 Dynamic Single ended Sense amplifier [13] 16
3.6 Black-box representation of the current conveyor 17
3.7 Current comparator basic idea 17
3.8 Strong arm latch and its small signal equivalent 18
3.9 Voltage build up in cross coupled pair 18
3.10 Conventional latch sense amplifier 19
3.11 High impedance latch sense amplifier 20
3.12 Leakage in Bitline 21
3.13 MUX based Serializer 21
3.14 Tree based Serializer 22
3.15 Clock Divider 22
3.16 Tree based de-serializer 23

4.1 Current sensing unit 26
4.2 Current comparator 26

4.3 Proposed voltage latch sense amplifier 27
4.4 2:1 serializer 28
4.5 DFF 29
4.6 Deserializer Circuit 29
4.7 Deserializer area saving 30
4.8 Clock divider block diagram 30
4.9 CVSL based latch [17] 31
4.10 Buffer 31
4.11 Waveform of serial data 31
4.12 Transmission driver 32
4.13 RC Model of Serial Link 33
4.14 Receiver design [6] 33
4.15 Serial Link 34

5.1 Proposed floor plan showing part of SRAM bank 35
5.2 Proposed layout plan for Serializer/De-Serializer 36
5.3 Layout of Serializer and Deserializer 37

6.1 eDRAM Test Bench 39
6.2 SRAM Simulation 40
6.3 Test bench for Serializer 41
6.4 Test bench for Deserializer 41
6.5 Test bench for whole system 42
6.6 Serial link circuit 42

7.1 Reading 0 44
7.2 Reading 1 45
7.3 PDP at various process corners 46
7.4 Monte Carlo Plots 46
7.5 Transient voltage level during read process 47
7.6 Comparison of Power delay product 48
7.7 MC simulation Plot 49
7.8 Sizing of SAE transistor for process variation 49
7.9 Sensitivity dependence 50
7.10 Serializer plots 51
7.11 Deserializer plots 52
7.12 Serial link plots 53
7.13 Area Comparison 54
7.14 System comparison 54

A.1 Connection of 32:1 Serializer 59
A.2 Connection of 1:32 Deserializer 60

List of Tables

1.1 Design Specifications for serial link 3

4.1 Parasitic values 33

6.1 Value stored at the nodes 40

7.1 Performance Summary 43
7.2 Monte Carlo Output 46
7.3 Performance summary for modified latch type sense amplifier for 900mv VDD 48
7.4 Performance summary of modified latch type sense amplifier for 600mv VDD 48
7.5 MC Result 49
7.6 Comparison of two serial link implementation 53

Chapter 1

Introduction

1.1 Thesis Motivation

Rapid development in very large scale integrated(VLSI) industry in last two decades has resulted in embedding many different IPs such as memories, processors together into a single chip. One such application is, shared bus architecture where data needs to be physically transfered from one unit(IP) to another. For such cases, parallel interconnects are used to provide high throughput. Parallel interconnect area increases with increase in data bus. Moreover, delay uncertainty introduced by repeaters, layout parasitic effects and process variations have also a big impact on the quality and speed of transmission as well as in signal degradation along the interconnect [2].

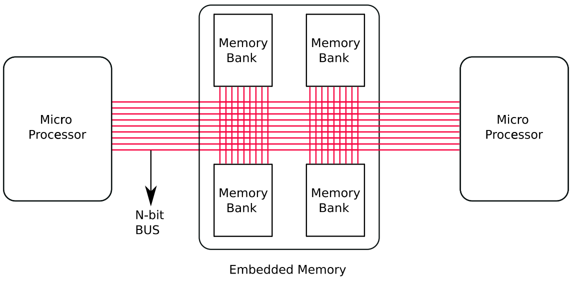


Figure 1.1: Shared bus architecture

To reduce interconnect area and also to reduce the area penalty with respect to scaling of data bus width, serial interconnect is proposed for on chip communication. Serial interconnect requires additional circuits to transmit the data from source to destination.

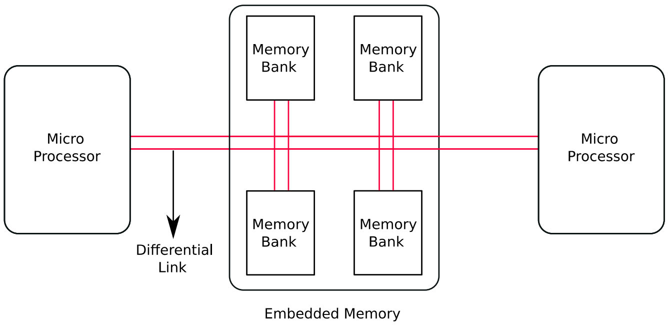


Figure 1.2: Serial link

1.2 Thesis Aim

This thesis aim to estimate the area required to implement bi-directional serial interconnect by modelling data transmission between a processor and an embedded memory as shown in Figure 1.3 and scale it for multiple units (IPs). This thesis scope also includes design of peripheral circuit required to read data from an embedded memory such as eDRAM (embedded dynamic random access memory), SRAM(static random access memory) and integrate it with the serial interconnect to transfer of data. Serializer, deserializer, transmitter and receiver circuit constitute the serial interconnect. The read speed for the sense amplifier is considered around 250 MHz.

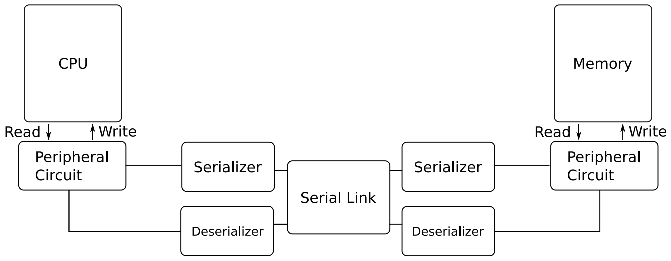


Figure 1.3: System model

1.3 Project Specifications

The main focus on this thesis is to read bits of data from memory bank, serialize the data and send it to the serial link and at the other end deserialize the data and write it into the desired location. The entire design is developed in 28nm CMOS BULK technology. It is important to note that the serial link is bi-directional, due to which each units needs to have both serializer and deserializer blocks. Based on read speed of the sense amplifier, specifications are set for the serial link. Specifications for serial link can change depending upon the application requirement.

Parameter	Specification
Parallel throughput (PT)	250 MHz
Data bus width(DBW)	32 bit
Serial Link Frequency	(PT*DBW) = 8 GHz

Table 1.1: Design Specifications for serial link

1.4 Thesis Organisation

The outline of this thesis is listed below:

- **Chapter 2**: Relevant concept required to have an understanding of working of memory and Serial link
- **Chapter 3**: background research on current design for the blocks designed in this thesis
- **Chapter 4** Circuit design for each different blocks explaining about its functionality
- **Chapter 5** Layout design
- **Chapter 6** System Integration
- **Chpater 7** Simulation results
- **Chapter 8** Conclusion and Future Work

Chapter 2

Relevant Concept

RAM (Random Access Memory) is a volatile storage element in digital systems which can be quickly accessed by the device's processors. Read and write operations are much faster in RAM than in other kind of non-volatile storage devices such as hard disk drive(HDD), solid state drive (SSD) or optical drive.

2.1 Bit Cell

Bit cell is a storage element present inside a RAM which stores single bit of data. There are different types of bitcell depending on the number of transistors present inside it. Generic SRAM design has 6 transistors where as DRAM has 1 and eDRAM has 3 transistors in their design.

2.1.1 SRAM

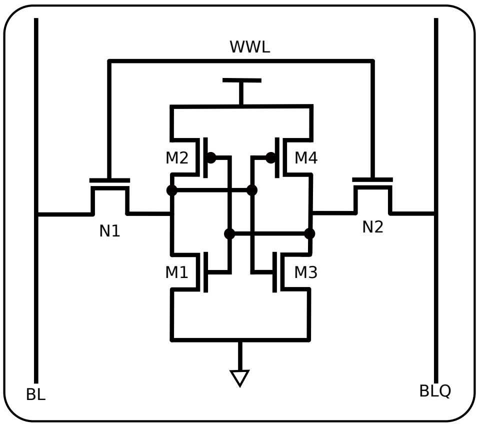


Figure 2.1: Bit cell of a SRAM

Figure 2.1 shows a prevailing 6T SRAM cell architecture commonly used among ICs [6]. In this configuration, SRAM memories rely on a pair of cross-coupled (positive feedback) inverters M1,M2,M3,M4 to latch data, making high speed operations possible at the cost of bigger area per bit. Transistor N1,N2 are called as pass transistors are they act as a pass gate between the latch and bitline BL and BLQ. One of the bitline BL contains actual data whereas the other bitline BLQ contains its complement data.

2.1.2 eDRAM

Embedded dynamic random access memory (eDRAM) is a type of RAM (Random Access Memory) which is completely embedded in the application-specific integrated circuit (ASIC) [7]. The ASIC can include the microprocessor as well. It has a huge advantage over standard DRAM in terms of speed, power requirements, as it is integrated into the IC itself. It has varied applications from gaming to AI based image processing system.

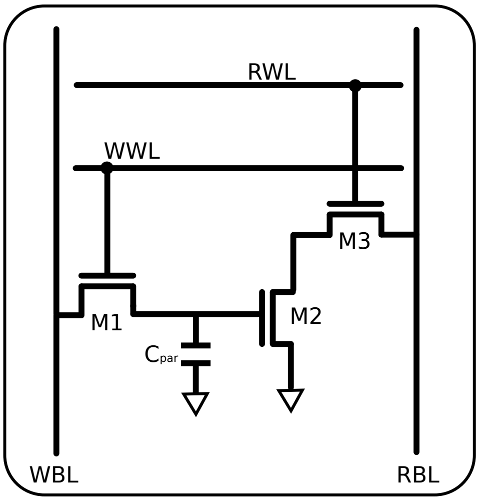


Figure 2.2: Bit cell of a eDRAM

Figure 2.2 shows a single unit of eDRAM. Here, C*par* is used to store the data written into the cell, and it needs to be refreshed continuously, as in case of DRAM to store the written value. Constant refresh is required because of leakage of data from the parasitic capacitance due to leakage current. Leakage current is the current flow inside a transistor when it is not in ON state ie working mode.

2.2 Write Function

In order to store a value inside a bitcell, writing operation must be carried out. Initially, the data to be stored is pre-charged to the bitline and then Word line is activated to store the value in to the cell. In case of SRAM, due to its latched architecture, the data remains unaffected as long as new write operation is carried out. In case of eDRAM, in order to avoid loss of written data, constant refreshing is required because of the design shortcomings.

2.3 Read Function

The goal of read operation is to maintain data integrity (stored data to not to be destroyed). To access a data stored inside a bitline, reading process is carried out. Sense amplifier is used to enhance the speed of read operation. Two important cycles are required to read a data. They are, precharge phase and evaluation phase. In precharge phase, the bit-lines are pre-charged to VDD, and during evaluation phase, WL are activated along with SAE(Sense Amplifier Enable) signal in tandem. SRAM has two complementary bitlines, where as eDRAM has single bitline, which needs to be compared with another reference voltage to read the stored value inside the bitcell.

2.4 Memory Macro

It consist of an array of bitcells along with peripheral circuits. Peripheral circuits consist of row, column decoders, precharge circuits which assist in read and write operation. A memory macro consists of 2^n words of 2^m bits each. A word line WL is shared among all the cells in same row and the bitline pairs are shared over each column .In order to keep the parasitics of the bitlines within an acceptable range, n and m are decided upon it. Figure 2.3 depicts a general structure of a memory macro.

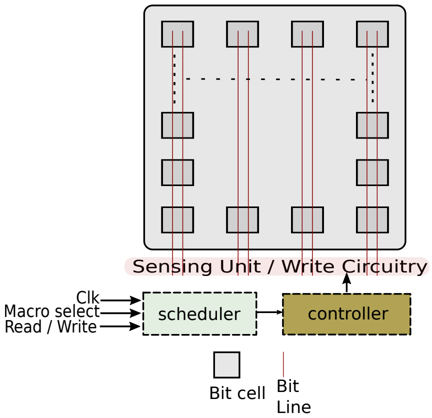


Figure 2.3: Memory Macro

The bit-line modeling of single column in a memory macro can be shown in the Figure 2.4. The bitline load represents the precharge transistors, capacitors represent the bitline capacitance, the long interconnect lines are modeled as distributed RC line since the delay can be represented more accurately [8]. In Figure 2.5, R$_L$ is resistive termination, modeled as input of the sense amplifier. Inductance is neglected as the read frequency is around 200-500 MHz and inductance effect occurs only high frequency [9].

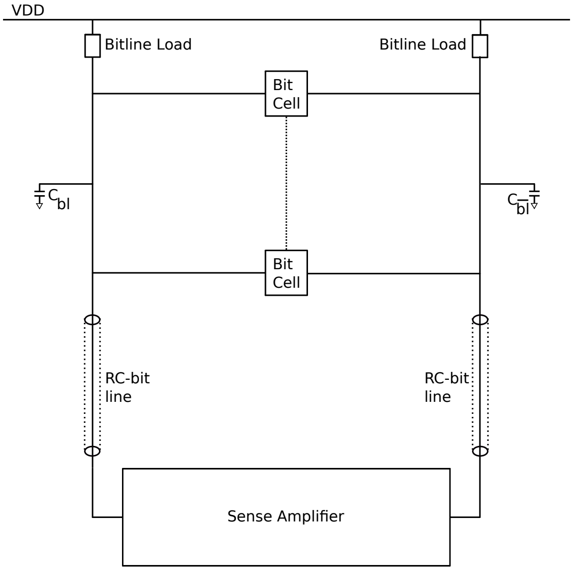


Figure 2.4: Bit line modeling of bitcells in a single column in a macro [10]

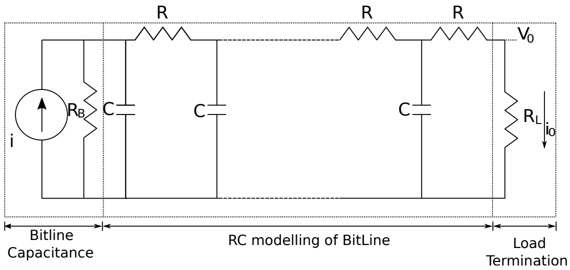


Figure 2.5: Analysis model for fig.2.4

2.5 Serial Link

SerDes(Serializer/Deserializer) system refers to the complete assembly of transmitter, channel and receiver that constitute the high speed serial link. A typical block diagram of the SerDes system is shown in the Figure 2.6. The basic blocks present in a typical SerDes system are Serializer, PLL, Transmitter (TX), Transmission Line with terminations at both ends, Receiver (RX), Clock and Data Recovery(CDR), Deserializer.

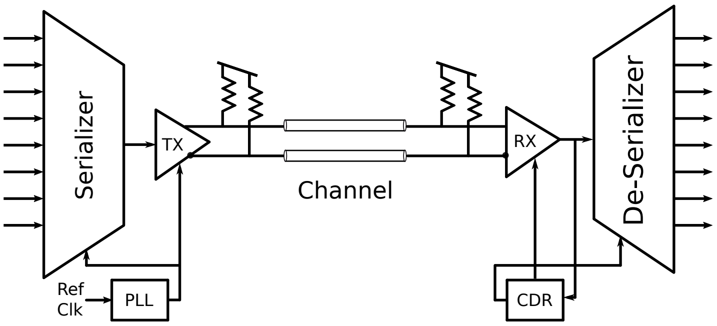


Figure 2.6: A typical SERDES system

2.5.1 Serializer

Number of bitlines present in a memory macro is usually a power of 2 such as 16, 32, 64, etc. Thus, the input to the SERDES system is parallel in nature. At every read cycle, new value is available in the sense amplifier out which in turn is transmitted to the desired location (memory or processor) before another read cycle starts. Thus, a serializer converts a parallel stream of data into a serial stream, suitable for transmission over a high speed serial link [11]. Depending on the number of bits to be serialized, the serializer is termed as N:1 serializer, where N represents the data word length, i.e., the number of input lines. Figure 2.7, represents a common way of implementing serialization process where, an external clock is utilized to complete the process. A common way to implement N bit Serializer, is to use an external high speed clock configured at N times the clock frequency of the data, i.e., for the case of N = 8, and a 1 GHz system clock for the data, then we need 8 GHz external clock to realize the serialization process.

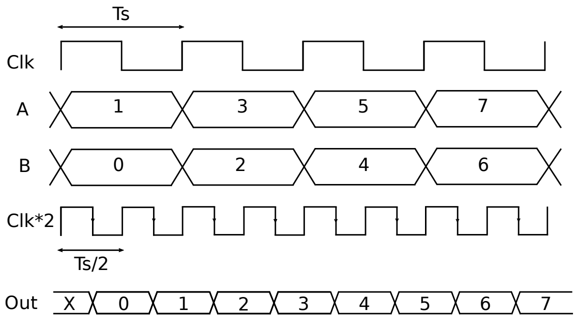


Figure 2.7: Serializer Idea

2.5.2 Phase Locked Loop (PLL)

A PLL is a negative-feedback system whose purpose is to take an input reference clock with reference frequency f_{in}, and produce an output clock with frequency f_{out}, such that $f_{out} = \alpha f_{in}$, where (α >1) is the multiplication factor. PLL is required for clock frequency in Gigahertz, since crystal oscillators can provide a high spectral purity reference clock only upto a frequency of about 200 MHz [11]. Design of PLL is by itself a separate research topic, and is not under the scope of this thesis. We use clock generated from ideal source for simulation purpose.

2.5.3 Encoding and Signaling Scheme

Encoding refers to the process in which the data to be transmitted is mapped onto a different set of data in a recoverable manner. Encoding schemes are implemented, so that data and clock can be easily recovered in the receiver block. There are several encoding schemes which are implemented based on the application requirement. Non-Return to Zero(NRZ), Return to Zero (RZ), Pulse-amplitude Modulation (PAM), are some of the common encoding schemes used in the transmission of data. Figure 2.8, represents different types of encoding schemes. In NRZ, each bit is represented by either high or low state of a pulse, where as in RZ, data goes to high or low for half clock period and returns to common mode voltage for another half clock period. PAM is predominantly used in sine wave based encoding schemes [6]. In this thesis, RZ encoding scheme is used. In addition to encoding, it is also important to describe how the binary digits of 0's and 1's are electrically represented in the channel, and it is known as signaling. Signaling process is realized in the transmitter block. It is important to remember that a complex signaling technique requires an equally complex receiver design. RZ scheme eliminates separate data decoding logic block at the receiver chain which can save area but requires its clock to be twice of the maximum operating frequency of the system.

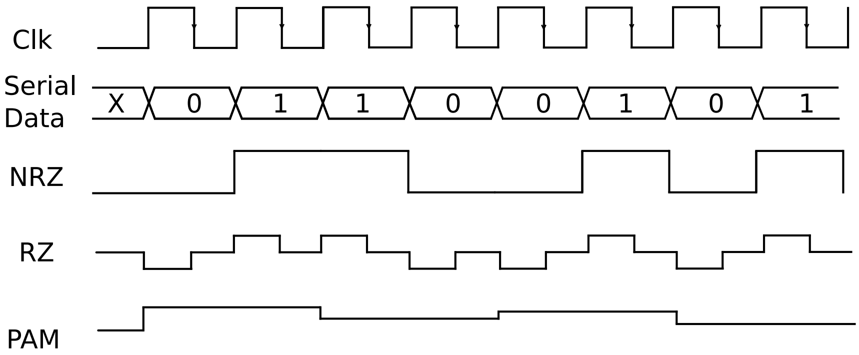
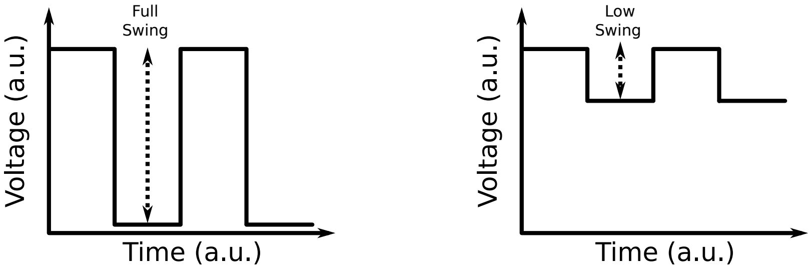


Figure 2.8: Encoding Schemes used in Serial link

2.5.4 Transmitter

In order to improve performance when driving long wires, small voltage swing is preferred rather than having to detect full voltage swing. To take advantage of low-swing signaling, the signal should travel on differential pair of wires [6]. This scheme can be realized by using differential amplifier with a current source at bottom to regulate the voltage swing at output. Figure 2.9 shows the output voltage level on one of the differential wire. Here, in Figure 2.9, we observe the difference between full level voltage swing and low level voltage swing.



(a) Full Swing **(b)** Low Swing

Figure 2.9: Transmitter Output Swing

2.5.5 Transmission Line

Transmission lines are the medium through which data signals are going to travel from transmitter to receiver. Depending on the distance between transmitter and receiver, a distributed RLC (Resitance, Inductance, Capacitance) line [9] is modeled based on the values given in the respective technology files. It is also important to terminate the transmission line at both the ends to maintain dc level and avoid impedance mismatch. Characteristic Impedance is defined as the ratio of the voltage over current present at a given point of the transmission line. With absence of proper termination of transmission line, reflections can occur (signal reflected back to the transmitter from receiver), causing signal distortion. The problem of reflections due to impedance mismatch can be neglected due to the state of the art process technologies [9].

2.5.6 Receiver

Receiver block is used to detect the bit that was sent and recover the clock which was transmitted so that it can be used for de-serialization process. Initially it amplifies the received signal and then recovers the data and clock. A more detailed explanation on receiver design can be found in [6].

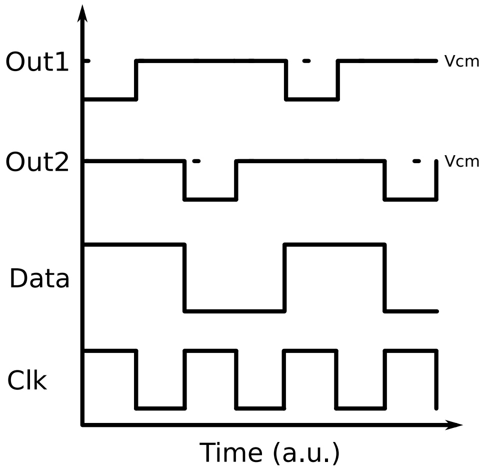


Figure 2.10: Clock and data recovery waveform

Figure 2.10 shows the basic functionality of the receiver circuit. *Out1*, *Out2* represents differential voltage level on the transmission line based on RZ logic, from which *Data* and *clk* can be recovered.

2.5.7 Deserializer

Deserializer performs the complementary function of the serializer. As the serial data at the receiver end needs to be processed in terms of words of specific length, it is pertinent to convert the serial data stream back to its parallel form. This function is performed by 1:N de-serializer, where N represents the word length.

Chapter 3

Background Research

Sense amplifier is an important component in reading the data from a bitcell which determines the yield of the embedded memories. The three main factors controlling the reading process are :

- Read current variation from bitcell which affects the sensing level fed to the Sense amplifier

- Sensing window which affects read operation

- Process variation in sense amplifier

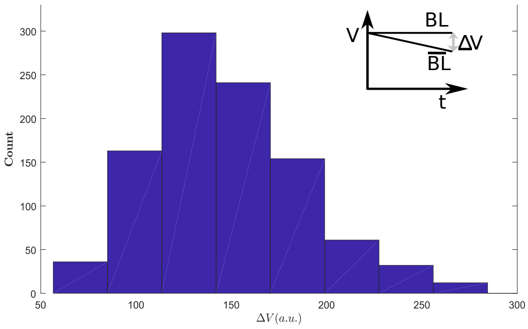


Figure 3.1: Histogram of bitline swing

Figure 3.1 shows the Monte carlo simulation results of Δ V (difference between complementary bitline)in a typical SRAM bitcell which makes it, paramount to design a robust sensing scheme for a wide distribution of voltage difference.

3.1 Data Sensing Models

Unlike other peripheral circuitries in a memory, sense amplifier is an analog design instead of logic design. It is designed to easily differentiate a small voltage difference into logic levels. In this work, we consider two types of sensing schemes: current sensing and voltage sensing. In current mode, we have low input impedance to the sense amplifier where as in case of voltage mode, we have high input impedance.

In current sensing scheme as shown in Figure 3.2, the state of memory cell is read out by measuring the resulting current through the selected memory cell when a read signal is applied. The current on the bit-line is compared to the reference current generated by reference cells, the current difference is amplified by current mode sense amplifiers and they are eventually converted to voltage signals [12].

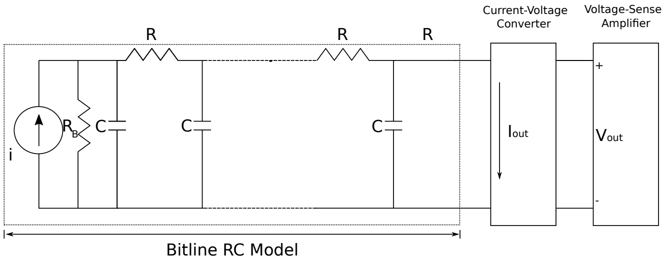


Figure 3.2: Current sensing scheme [12]

In voltage mode sensing, the input resistance of the sense amplifier is infinite, and hence the output signal of the RC modeled system will be an open circuit voltage V_o. Voltage mode sensing scheme is shown in the Figure 3.3.

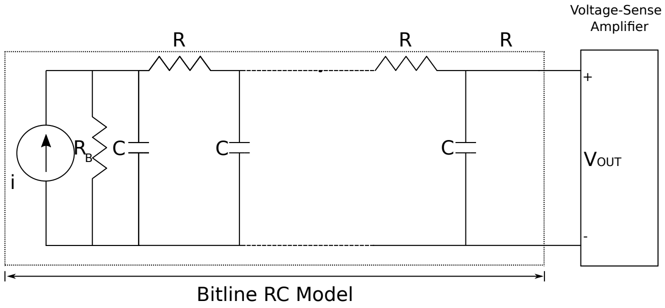


Figure 3.3: Voltage sensing scheme [12]

Delays of current, voltage mode sensing schemes can be given using the following equations from [10]: where R_T, C_T are the total line resistance and capacitance, RB equivalent resistance of the memory cell. t_i, t_v represents the RC delays of current and voltage mode sensing.

$$\delta t_i = \frac{R_T C_T}{2} \cdot (\frac{R_B + \frac{R_T}{3}}{R_B + R_T}) \tag{3.1}$$

$$\delta t_i = \frac{R_T C_T}{2} \cdot (1 + \frac{2R_B}{R_T}) \tag{3.2}$$

3.1.1 Single Ended Sensing

In single ended sensing scheme, data is available in only one bitline, and a reference cell is required to read the correct logic level. This type of single ended sensing is commonly used in eDRAM and two port SRAM design, as both read and write can be done simultaneously. This kind of topology is very useful in high density static RAM design because the sizing of the transistors is done as minimum as possible while maintaining good read/write and memory retention characteristics. In the Figure 3.4, inverter is used as a sense amplifier, but the switching threshold is vulnerable to process variations especially at lower technologies.

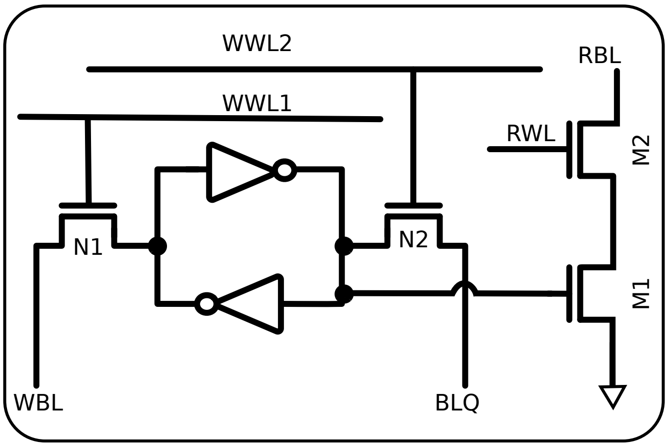


Figure 3.4: Single ended voltage sensing scheme

3.1.2 Dual Ended Sensing

In dual ended sensing scheme, complementary bitlines are available for reading the value stored in a bitcell. This type of memory access is commonly known as single port as only one operation either read/write can be carried out at a single instance.

3.2 Single ended dynamic sense amplifier

Single ended sensing commonly use latch type sense amplifier similar to dual ended, while using one end with a common reference and other being input from a bitline. The basic idea of this design is to generate a reference voltage at half the vdd [9] and precharge it to the bit line, and while reading from the bitline (DRAM or SRAM), voltage drops (to read 0) or increases (to read 1), is compared with the reference voltage to provide the proper decision. The bitline is allowed to drop or increase to certain voltage level so that offset variations in the latch does not cause improper decision. One of the major bottlenecks in this design are area overhead for separate reference system, complexity and reliability at scaled voltages. This concept doesn't work for eDRAM as the voltage doesn't increase while reading 1. A better method was proposed by generating dynamic reference system instead of creating constant reference source which can save area. It comprises a differential sense amplifier circuit and a dynamic reference generation circuit [13]. The differential sense amplifier circuit has two input ends, of which one end is coupled to a data line so as to receive input data with a voltage inp, and inp also feeds back to the dynamic reference voltage generation circuit which generates a dynamic reference voltage vref based on inp, and vref is transmitted to the other end of the differential sense amplifier circuit as shown in the Figure 3.5. The dynamic reference voltage generation circuit includes a PMOS transistor MP0, two NMOS transistor MN0, MN1 that collectively forms a cascade down to GND. The idea behind the design is to generate a complement of the bitline voltage and then feed it to the latch based sense amplifier.

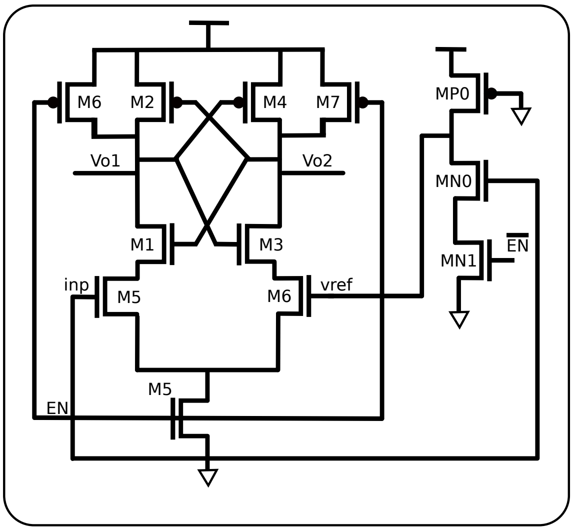


Figure 3.5: Dynamic Single ended Sense amplifier [13]

The above mentioned architecture is suitable for reading in SRAM where there is not much signal degradation. In case of eDRAM, different approach is required, due to which current conveyor based current mode sense amplifier is explored.

3.2.1 Current Conveyor

Current mode sense amplifier is realized using current conveyor circuit. Current conveyor is a 3-port device whose black-box representation can be seen in the Figure 3.6. This has been defined as a device having a virtual short-circuit input port and a unity - gain current transfer characteristic from input to output [14]. Furthermore, the current I_x will be conveyed to the output terminal such that terminal Z has the characteristics of a current source, of value I_x, with high output impedance.

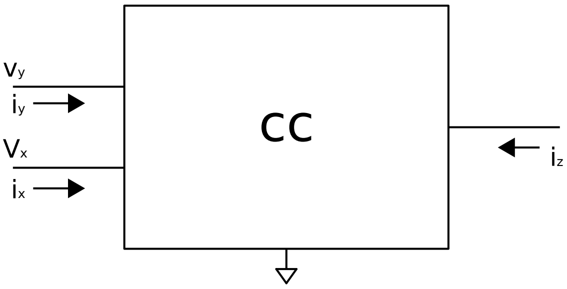
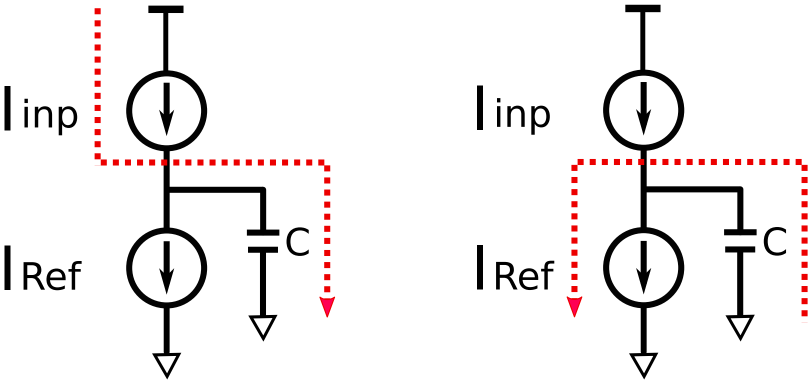


Figure 3.6: Black-box representation of the current conveyor

3.2.2 Current Comparator

It serves as a current threshold detector to discriminate various input current levels. The current mode comparator receives an input signal, in the form of current, and compares it with its own threshold current. The result of the comparison is in the form of an output voltage [15]. Figure 3.7 shows the basic operation of a current comparator during two possible conditions, i.e., when input current is greater than the reference current and when input current is lesser than the reference current.



(a) input > reference **(b)** input < reference

Figure 3.7: Current comparator basic idea

3.3 Voltage Latch Sense amplifier

Cross coupled architecture finds its application in decision making circuits such as comparator as well as sense amplifier. It is also known as strong arm latch [16]. The main reason for its popularity is zero static power consumption, produces rail-to-rail outputs.

Positive feedback connection in the cross coupled pair helps in decision making based on the input given to it. The condition for the latch to make decision properly is that the conductance must be greater than the inverse of the load resistance present in the circuit. Once this condition is achieved, exponential build up of voltage is developed which causes regeneration of voltage level to rail-to-rail. From this basic circuit, latch type sense amplifier evolved. In the Figure 3.8, C represents the parasitic capacitance. The exponential voltage raise is given by the equation 3.3 [17]. From Figure 3.9, exponential voltage is build up based on the initial condition V$_d$(0).

$$V_d(t) = V_d(0) \cdot e^{t(g_m - \frac{1}{R})\dfrac{1}{C}} \tag{3.3}$$

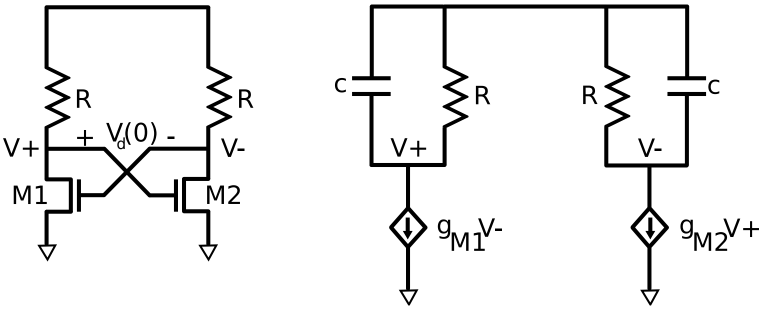


Figure 3.8: Strong arm latch and its small signal equivalent

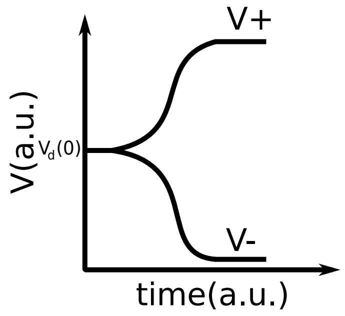


Figure 3.9: Voltage build up in cross coupled pair

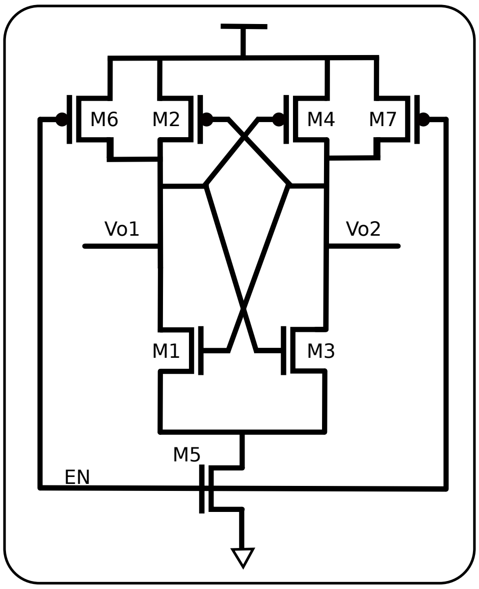


Figure 3.10: Conventional latch sense amplifier

Latch based topology finds its application largely in sense amplifier and comparator as they achieve a fast decision due to a strong positive feedback. Figure 3.10 shows a conventional latch [18]. Two cross-coupled inverters provide positive feedback. The basic idea behind the operation of this architecture is to provide an imbalance (voltage difference from the bitline) which causes the switching of the inverters to logic levels. This design is widely used due to its small area which fits into the pitch requirement of the SRAM macro. Major bottlenecks in this positive feedback is the process variations which can causes the inverter to make decisions based on process variations rather than the imbalance provided to the circuit.

In Figure 3.10, the enable signal EN turns on the amplifier and starts the sensing operation. Here, Vo1 and Vo2 are input and output terminals at the same time. Depending on the polarity of the voltage difference between the nodes Vo1 and Vo2, the sense amplifier will flip in one or the other direction. Therefore, the circuit cannot be connected directly to the bitline since the circuit would attempt to discharge the bitline capacitance during the decision phase and would increase delay and power. A solution to this is to either separate the bitline by a multiplexer or to use pass gates, forming a decoupling resistor. Both devices cause a voltage drop that deteriorates the available input voltage difference. This way the voltage swing at the bitlines can easily reduce by half, resulting in lower speed and margin. Other major problem is the kick-back noise from the previous stage which can cause wrong decision, causing reduction in memory yield.

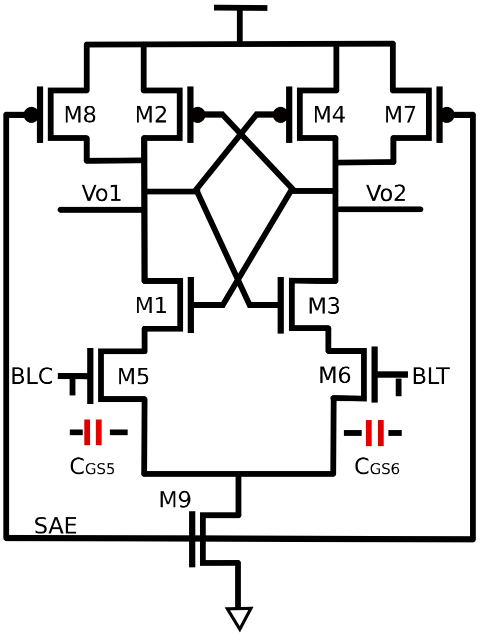


Figure 3.11: High impedance latch sense amplifier

The above mentioned drawbacks does not occur for the latch circuit shown in Figure 3.11, because of high impedance input differential stage [18]. This sense amplifier combines strong positive feedback with a high resistive input. The current flow of the differential input transistors M5 and M6 converts to a large output voltage. Despite negating all the problems in previous design, it still suffers from process variation which can cause wrong decision. This mismatch can induce trip point mismatch among cross-coupled inverters of Sense Amplifier (SA) or current mismatch in the evaluation branch of the SA circuit, resulting in operation failure [19]. The major factors contributing for offset voltage are gain factor, the drain current *(*I_d), the threshold voltage(V_{th}) and the layout of the device. Among this, *(V_{th}) mismatch* and *leakage* has been identified as dominant contributing factor [18].

This latch topology draws high transient currents from the inputs and the supply. These transients become troublesome if a large number of sense amplifiers operate in parallel. The *kickback*(leakage) currents drawn from the input stem from several components. The main contribution occurs when M9 turns on, initially drawing current from C_{GS5}, C_{GS6} reducing the bitline voltage value. Figure 3.12 shows leakage in bit line voltage due to enabling of SAE signal. Due to this, SRAM's suffer from slow read speed or high read failure probability. Thus, developing an SA with greater kickback noise tolerance is a prerequisite to achieve higher-yield. One way of reducing is to drive all the nodes to a known voltage level, to reduce kickback noise. Detailed discussion about the same is mentioned in the next chapter.

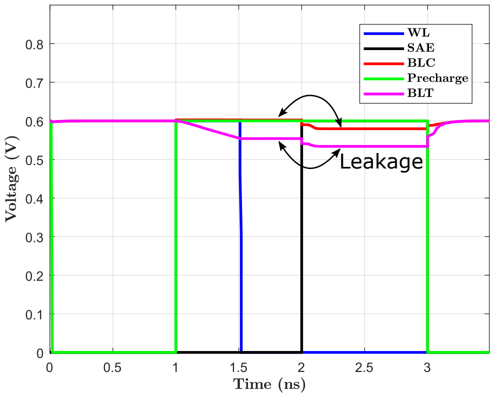


Figure 3.12: Leakage in Bitline

3.4 Serializer

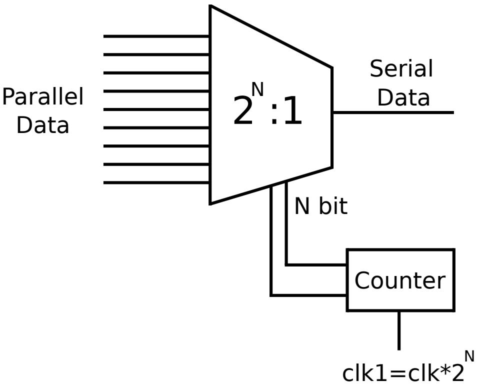


Figure 3.13: MUX based Serializer

Multiplexer(MUX) based serializer requires the highest clock of the system to 2^N times the frequency of the data signal, where N is total number of data bits to be transferred. This sometimes leads to bottleneck for static CMOS design as we require the system fast enough to operate at such high frequency. A simple multiplexer based serializer is shown in the Figure 3.13. Operating MUX at high frequency leads to increase in power consumption since the dynamic power is directly proportional to the frequency of operation.

$$P_{dynamic} = \frac{1}{2} \cdot CV^2 \cdot frequency \tag{3.4}$$

In order to overcome the above issues, a tree-based topology with dual edge triggering is proposed in this thesis. A simple 8 bit tree structure based design is shown in Figure 3.14. The major components used in this design is D-Flip Flop (DFF), multiplexer, clock divider.

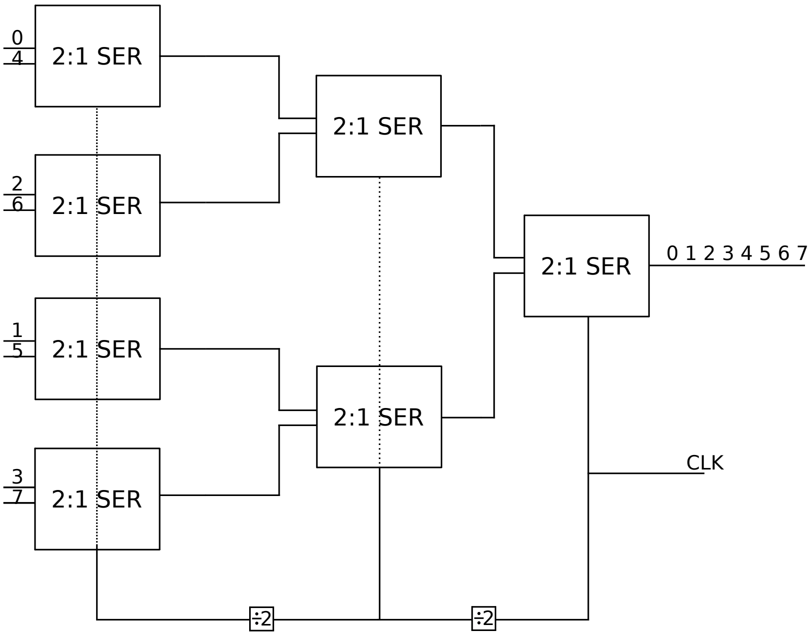


Figure 3.14: Tree based Serializer

Divide by two block in Figure 3.14 represents clock divider circuit. Clock divider (÷2) circuit is designed using DFF as shown in Figure 3.15.

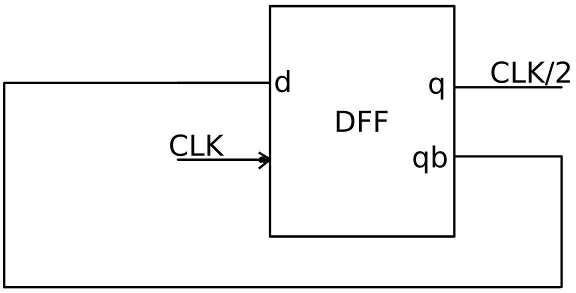


Figure 3.15: Clock Divider

3.5 Deserializer

Deserializer performs the function that is complement of serializer, so similar tree based architecture is followed. A simple 8 bit de-serializer implementation is shown in the Figure 3.16. It also requires clock divider by 2 circuit similar to serializer.

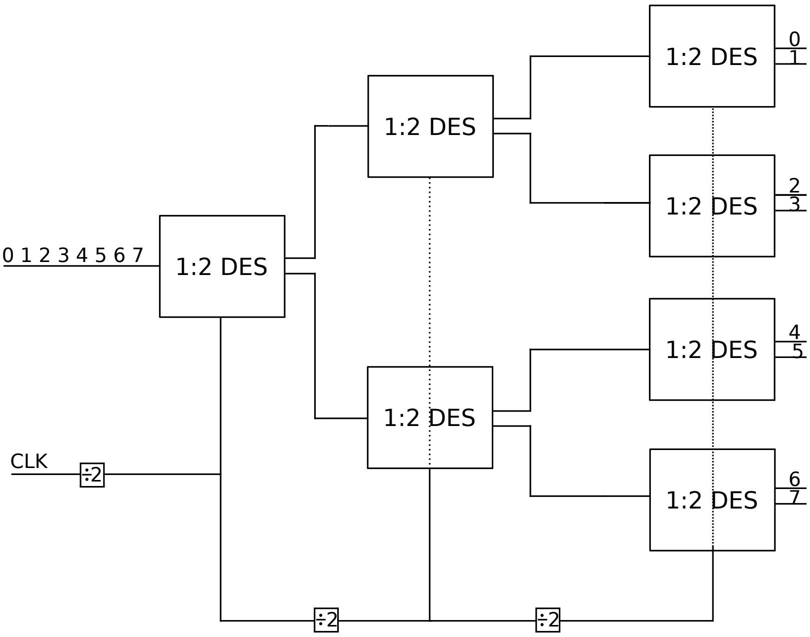


Figure 3.16: Tree based de-serializer

Chapter 4

Circuit Design

Two different topologies are considered for design of sense amplifier. In case of single ended sense amplifier, current mode topology was chosen in order to explore the current mode processing in decision making. For dual ended sense amplifier design, a modified latch type sense amplifier is designed to lessen the effect of offset variations. The main consideration in the design of sense amplifier are

- No static current flow
- Low kickback noise
- Sensitivity with respect to Process, Voltage, Temperature (PVT) variations
- Isolation between input and output nodes

4.1 Current Mode Sense Amplifier

Current mode sense amplifier requires two stages to perform read operation from a bitcell. As mentioned in previous chapter, it requires current sensing and current comparison circuit.

4.1.1 Current Sensing

Current sensing unit is designed using the concept of current conveyor circuit as shown in the Figure 4.1. This is a positive feedback based circuit which provides unity gain transfer characteristics from input to output. The sizing is done such that while reading 0 no current flows in the circuit and while reading 1 circuit conducts. So, while reading 0 (high voltage at CBL), the transistors M1, M2, M3, M4 turns off, resulting in OUT to be at high voltage. In case of reading 1 (low voltage at CBL), transistors M1, M2, M3, M4 turns on resulting in OUT voltage pulled towards ground. OUT signal is given as input to the current comparator stage which differentiates between 1 and 0. This circuit is entirely novel with respect to conventional single ended based sensing schemes which uses reference voltage to compare with the bitline voltage.

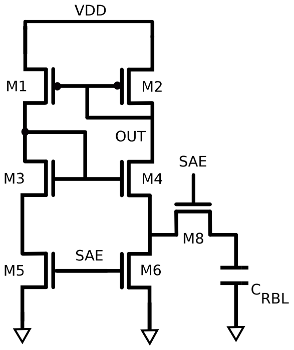


Figure 4.1: Current sensing unit

4.1.2 Current Comparator

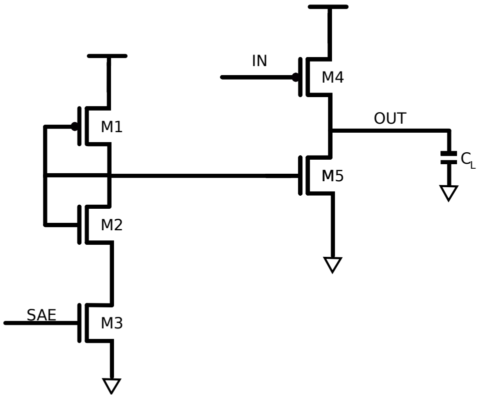


Figure 4.2: Current comparator

Figure 4.2 shows conventional CMOS current mode comparator. It comprises the reference threshold current generating circuit from M1, M2, M3 mirrored at M5. The output of current sensing unit is mirrored at M4. When the input current is greater than the reference current, the difference in current is used to charge the capacitance C_L, whereas when the input current is lesser than the reference current, the difference in current discharges the charge stored in the capacitance C_L. As a result, OUT is high when input current is greater than the reference current and low when input is lower than the reference current. The output from this circuit is connected to buffer to drive larger load. Here, reference current generating circuit has switch M3 which is used to save power during non read process. Instead of generating reference current, bias voltage can be provided directly to the transistor M5 which can cause area and design overhead due to complexity of designing separate voltage bias. This design is more suitable for memories having single bitline with high capacitance (eDRAM [20]) rather than for the memory (SRAM) having complementary bitlines, due to constant static current flow during read cycle.

4.2 Latch mode sense amplifier

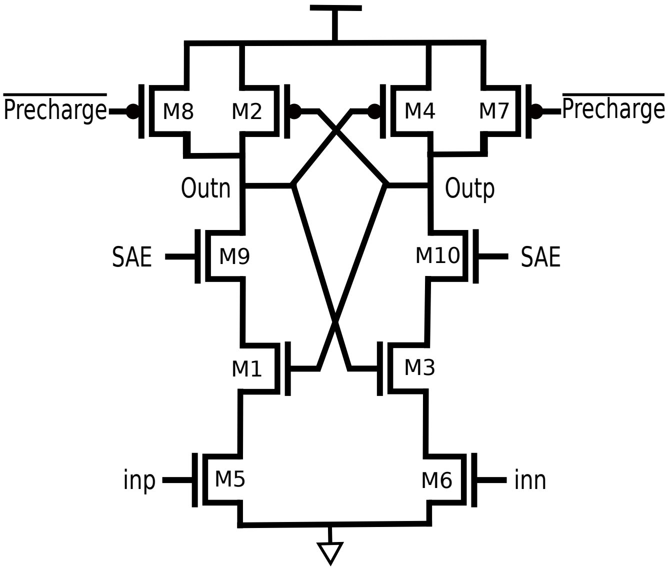
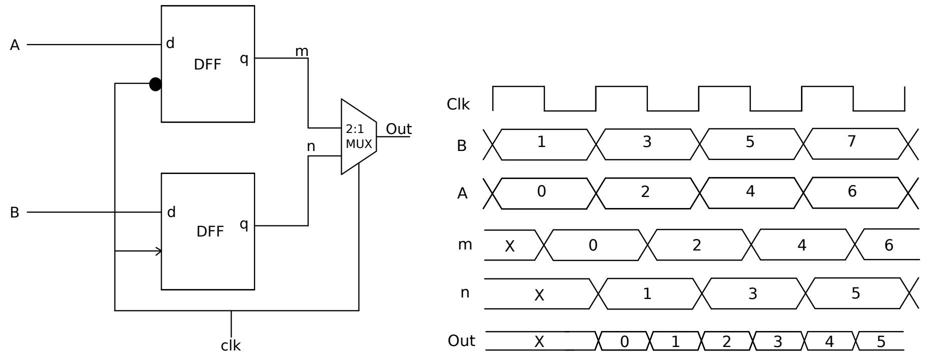


Figure 4.3: Proposed voltage latch sense amplifier

The proposed latch type sense amplifier is shown in the Figure 4.3 which has reduction in kickback noise (leakage). The input and and output is moved away from each other to reduce the leakage in bitline. This circuit has all its nodes driven to known location before *SAE* signal is enabled. This reduces offset caused due to PVT variation. The operation of the circuit is as follows. During precharge state, *Outp Outn* are precharged to *VDD* and the drain of M5, M6, M1, M3 are driven to *GND*. When *SAE* is raised to high, depending on the input at M5 and M6, the output nodes *Outp* and *Outn* changes to complementary logic levels, i.e., if *inp* > *inn*, then *Outn, Outp* changes to *0, VDD* respectively. As the transistors M9, M10 are in path of decision making circuit, the process variations comes into account if its not properly sized. The maximum offset variations the sense amplifier can handle, depends upon the speed, area and power consumption. Sensitivity of this sense amplifier to PVT variations depends on the operating voltage and the voltage at which bitlines are precharged. The trade off between area, speed and power consumption is such that, higher the difference in voltage (between bitlines) lower the area, where as lower difference in voltage(between bitlines) results in more power and area. Sense amplifier design purely depends on read speed requirement.

4.3 Serializer Design

As mentioned in chapter 1, we transfer 32 bit of data from one unit to another. Considering 250 MHz as parallel throughput frequency, we require a clock running at 4GHz to implement 32:1 tree based serialization process. Thirty one 2:1 serializer(2:1 SER) units are required to realizer 32 bit serialization process. Circuits used in the 2:1 SER are DFF and MUX. Figure 4.4 shows the design of 2:1 SER block, which consist of negative edge and positive edge trigger flip flop and 2:1 MUX. The MUX is sampled with half cycle delay after the data is latched (seen m, n and Out waveform in Figure 4.4b) in order to accommodate for the setup and hold time of the DFF.



(a) 2:1 SER

(b) Serializer waveform

Figure 4.4: 2:1 serializer

4.3.1 DFF

Flip Flop(FF) propagates the data at its input to the output depending upon the type of triggering it is designed for. Positive or negative edge trigger FF can be realized by changing the the clock signal present in the Figure 4.5. It shows transistor level design of DFF. MUX is designed based on transmission gate logic as given in [9].

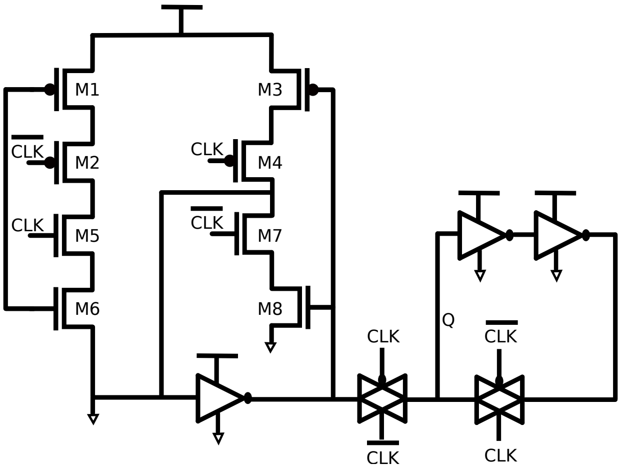
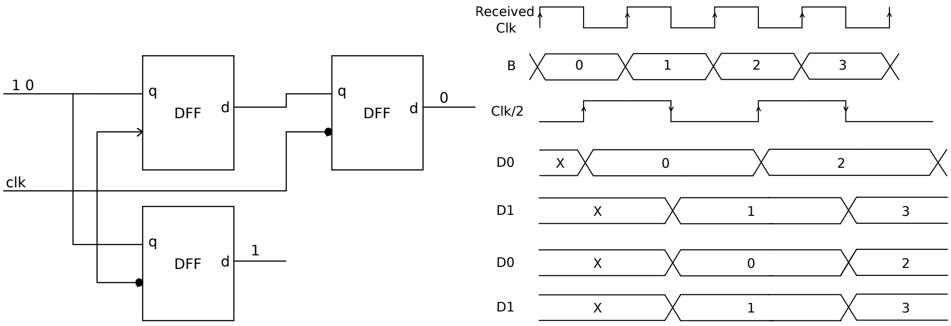


Figure 4.5: DFF

4.4 Deserializer design



(a) 1:2 DESER

(b) De-Serializer waveform

Figure 4.6: Deserializer Circuit

Similar to serializer, 1:2 Deserializer (1:2 DESER) block is used to realize 32:1 deserialization process. Deserializer is realized using DFF as shown in Figure 4.6a. Additional DFF (negative edge trigger for first row in Figure 4.6a), to have data latched at the same time at negative clock cycle as described in the Figure 4.6b (Last two data).

As we use two flip flops one after another, area is saved by combining two D Latch into one as shown in the Figure 4.7.

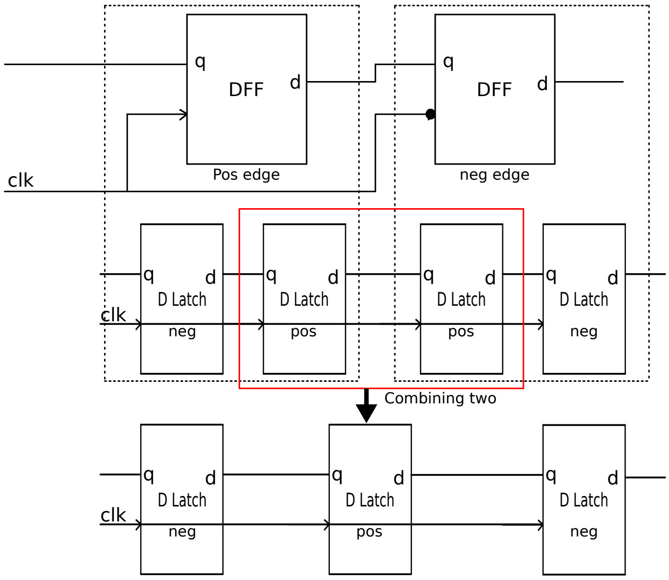


Figure 4.7: Deserializer area saving

4.5 Clock divider

In this design, we have the highest clock frequency at 8 GHz as discussed in chapter 2 (Maximum clock *2 - RZ coding scheme). The serializer/deserializer block uses frequency between from 4 GHz-250 MHz, for which 5 clock dividers are used. Block diagram of clock divider design in shown in the Figure 4.8.

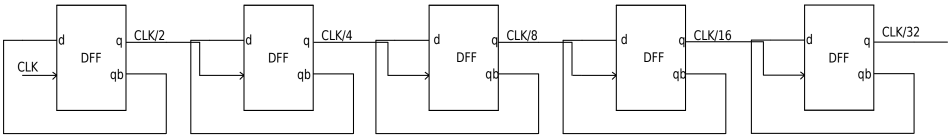
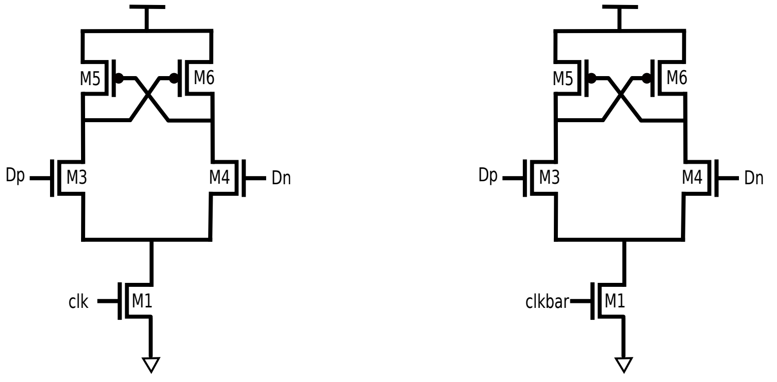


Figure 4.8: Clock divider block diagram

Dual edge trigger increases the load for the clock twice which requires special clock dividers circuit instead of conventional CMOS logic. Cascode voltage switch logic (CVSL) based latch is designed from which flip flop is realized. This design is ideal for operating at frequency less than 10 GHz [17]. Reset signal is not depicted in the Figure 4.9.



(a) D latch, positive clock **(b)** D latch, negative clock

Figure 4.9: CVSL based latch [17]

After clock generation, buffer is used between the clock and the clock input to serializer/deserializer to accommodate for the parasitic capacitance present in these blocks and also to reduce the delay of clock signal. Buffer design is designed as described in [9]. Here, 2,4,8 describes the increase in sizing of the inverter chain from an unit size of an inverter.

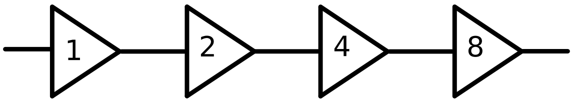


Figure 4.10: Buffer

4.6 Transmission Driver

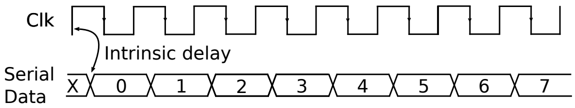


Figure 4.11: Waveform of serial data

Once data is serialized it needs to be transmitted based upon RZ coding scheme. We use current mode driver, where the output of serializer is converted into differential signal as *Ap An* and connected to the input of the driver. In this configuration, transistors M3 and M4 act as complementary switches which steer the total tail current I$_{tail}$. The differential output swing between the nodes *outp outn* is between *VDD* and *VDD* - $I_{tail}R_D$ where R$_D$ is output resistance of the diode connected transistor. RZ scheme is realized when current source M1 draws constant current when the $\overline{clk}$ is high and no current when $\overline{clk}$ is low. We sample at

$\overline{clk}$ rather than at *clk* in order to accommodate for the intrinsic delay present in the clock divider circuits as seen in Figure 4.11. This differential driver consumes less power compared with inverter drivers at multi-Gb/s data rate [21]. Figure 4.12 depicts the circuit design of a differential driver.

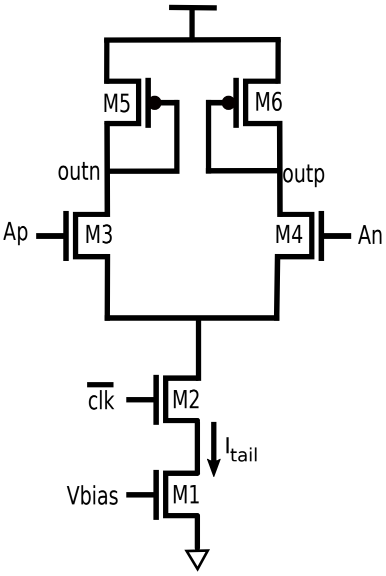


Figure 4.12: Transmission driver

4.7 RLC modeling

As mentioned in section 2.5.5, a serial interconnect with a distance of 100 μm, which classifies as short and local range transmission line [8] is modeled as shown in Figure 4.13. This is the distance that exist between two different units (IPs). We use a distributed RLC model since the operating frequency is considered high and delay can be represented more accurately [8]. RLC values were calculated from the technology data sheet for using metal 8 wire, with $w = 4 * w_{min}$. L value was neglected since the inductive part of impedance becomes equal to resistive component at a frequency around 168 GHz, which is obtained by solving the following expression [9]:

$$\omega l = 2\pi f l = r \tag{4.1}$$

Passive Components	Value
R	0.25 Ω/μm
C	0.137 fF/μm
L	.213 pH/μm

Table 4.1: Parasitic values

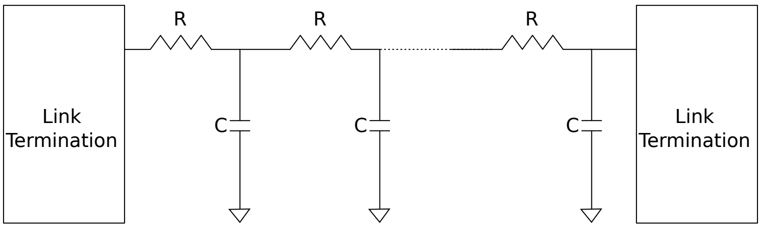


Figure 4.13: RC Model of Serial Link

4.8 Receiver

Receiver in the serial link is used to recover the *Data* and the *clk* sent from the transmitter. The receiver block designed in [6](shown in Figure 4.14), is reused in order to verify the functionality of data transfer between two units. Detailed discussion about the self-biased sense amplifier and three stage differential amplifier can be found in [6]. It is important to note that the self biased amplifier sets the DC level for the serial link at the receiver side.

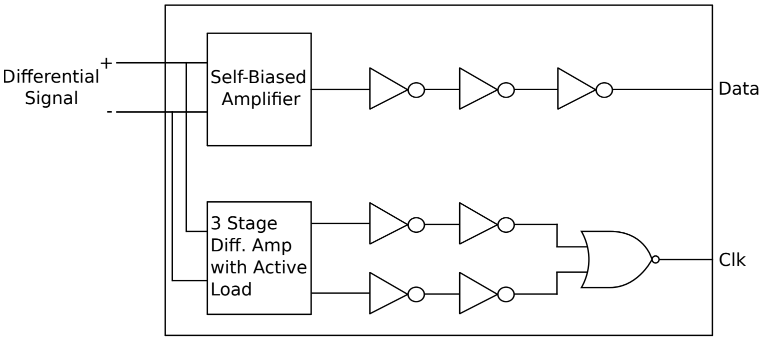


Figure 4.14: Receiver design [6]

In order to maintain DC level at both the ends of serial link, source follower (also known as voltage shifter) circuit is used at the transmission side. Serial link concept for unidirectional data transfer is shown in the Figure 4.15 where the input for transmitter is given from serializer output where as the output of receiver circuit is provided to the deserializer. For bidirectional transfer, transmission gates are added inside the serial link which is further discussed in chapter 6.

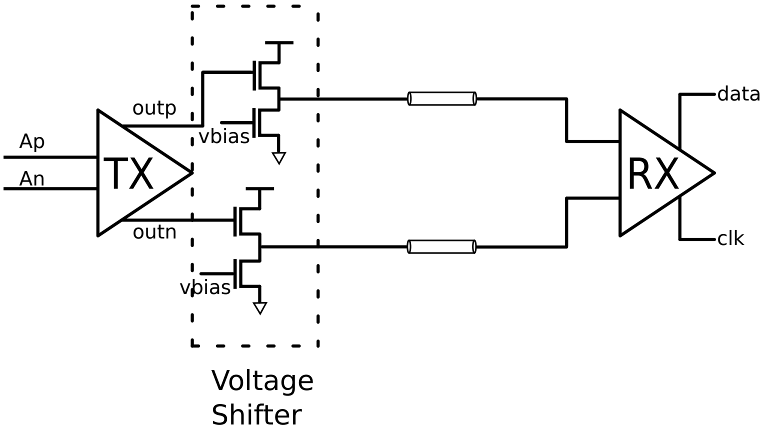


Figure 4.15: Serial Link

Chapter 5

Layout

The next stage after the circuit design is to transfer the same into layout level. This is required, to extract the parasitics in order to analyze the overall performance of the design.

In this thesis, layout of sense amplifier, serializer/deserializer is drawn. Serial link is simulated at schematic level. The main consideration in layout design were matching, symmetry and area. Specifically in sense amplifier design, its width was considered to be a multiple of the pitch of the bitcell. Here, maximum width of the sense amplifier layout was allowed to be 4 times the pitch of the bitcell as shown in Figure 5.1.

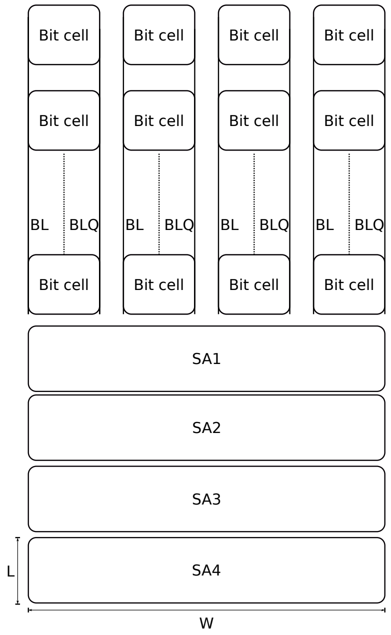


Figure 5.1: Proposed floor plan showing part of SRAM bank

The main parameters to consider in the design latch mode sense amplifier is sensitivity, in order to overcome mismatch. To have higher sensitivity, L in the layout have to be increased, which can cause unnecessary area and power headroom. The sensitivity can be traded off based on the speed requirement of the design.

As we try to build bi-directional link, serializer and de-serializer is required to be in both the sending and receiving end. The layout plan of these blocks is as shown in the Figure 5.2. The area occupied serializer and de-serializer are same and is around 361.73 μm^2. No separate power line routes were made for this layout.

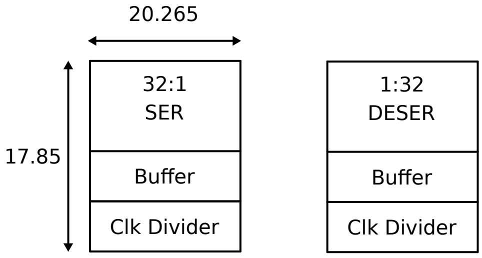


Figure 5.2: Proposed layout plan for Serializer/De-Serializer

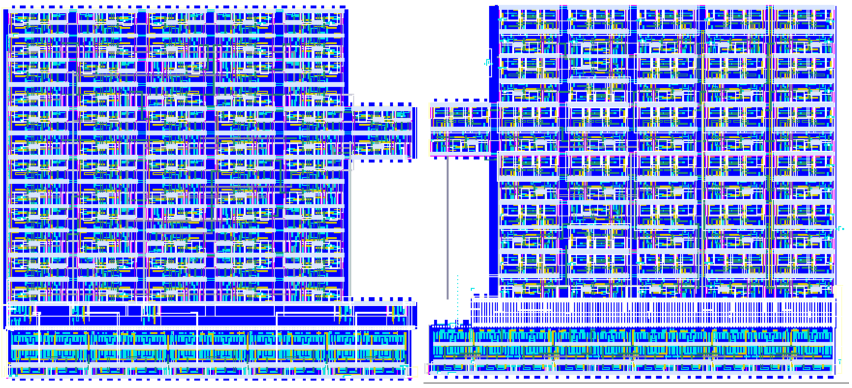
5.1 Symmetry and Matching

Analogue systems demand many more layout precautions so as to minimize the effects such as mismatches, noise, etc [22]. It is important, to place the transistors in a symmetric pattern and maintain an uniform environment around them so that all of them are affected in the same way.

The following guidelines were followed to maximize the matching in between critical transistors:

- Interdigitization technique (sharing common source/drain) was used to reduce the S/D (source/drain) junction area and the gate resistance. As a rule of thumb, fingers width was chosen as 10 times as that of its length.
- Common centroid layout was used to improve the matching of the differential transistors.
- Antenna effect was ameliorated by making sure that only small area of metal was tied to the gate of transistors.
- For the design of digital blocks, Euler path was used in order to reduce the area.

Layout of serializer, de-serializer is shown in the Figure 5.3. Compact layout design was drawn in order to use minimal area.



(a) Serializer Layout

(b) Deserializer Layout

Figure 5.3: Layout of Serializer and Deserializer

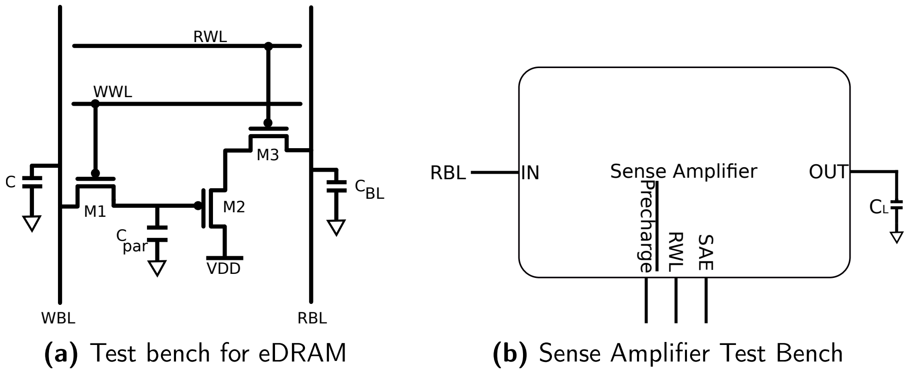
Chapter 6

System Integration and Test Bench

The initial stage of simulation consisted on analyzing the design of sense amplifier, serializer and de-serializer. The test bench to verify the functionality of the fore mentioned circuits are discussed in this chapter. All the control signals required are generated from ideal sources.

6.1 eDRAM

To test the functionality of single ended current mode sense amplifier, p-type based eDRAM (as shown in Figure 6.1a) was used. The operation of eDRAM is as follows: data is initially stored inside the bitcell using write operation and after specific time (refresh time) data is read out. During the read process, bitline C_{BL} is precharged, after which (Read Word Line) RWL and SAE signal is enabled in tandem.

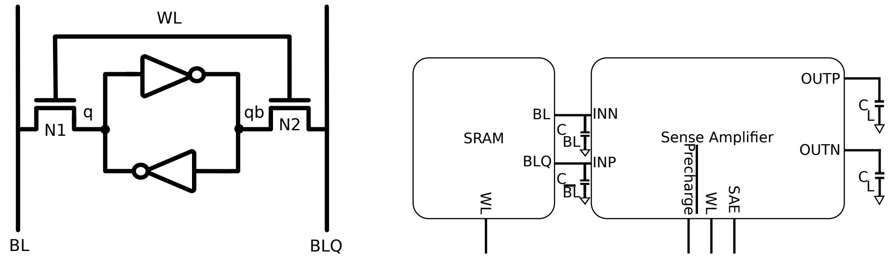


(a) Test bench for eDRAM **(b)** Sense Amplifier Test Bench

Figure 6.1: eDRAM Test Bench

6.2 SRAM

Unlike the eDRAM, SRAM read operation can be carried out without performing write operation i.e., the nodes of the SRAM can be set to high and low based on choice using the option of set initial condition in the ADE L environment in Cadence. For all the simulations for SRAM, the values present in the table 6.1 represent the value stored at its back to back nodes in Figure 6.2a.



(a) Initial Condition for SRAM **(b)** Test Bench for SRAM

Figure 6.2: SRAM Simulation

Node	Value
q	0
qb	1

Table 6.1: Value stored at the nodes

The test bench for the sense amplifier is shown in the Figure 6.2b. Read operation is carried out by pre-charging the bit lines to VDD, and then word line of the bit cell is turned on for the time until which the sense amplifier can make a decision correctly without the effect of process variations. Once it is done, sense amplifier is enabled by turning on the SAE signal. The VDD of the system was varied between 600m and 900mv to check its function at lower voltage level.

This topology is difficult to implement for single ended eDRAM with high bit line capacitance, as one end requires a reference voltage which is quite difficult to generate (area and design complexity). This sense amplifier can used in integration for the serial link as it has built in latch and removes the requirement for additional flipflop.

6.3 Serializer test bench

Seriazlier constitutes three blocks namely clock divider, buffer and tree structural placing of 31 - 2:1 SER block. We integrate all the three blocks into a single system. The internal connection within the serializer is given in the appendix. The inputs to the serializer system is from output of sense amplifier from memory bank. Initial phase in the design of serializer used constant voltage source as parallel input and later replaced by output of memory unit. The test bench for the serializer block is given in the Figure 6.3.

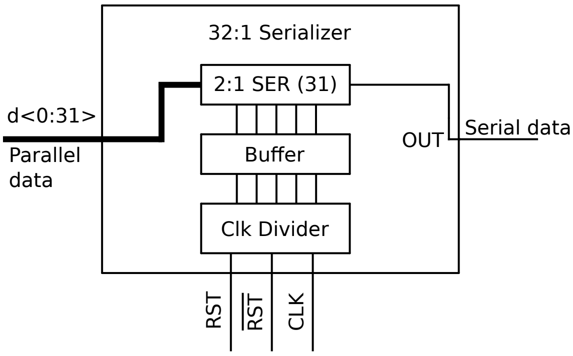


Figure 6.3: Test bench for Serializer

6.4 Deserializer test bench

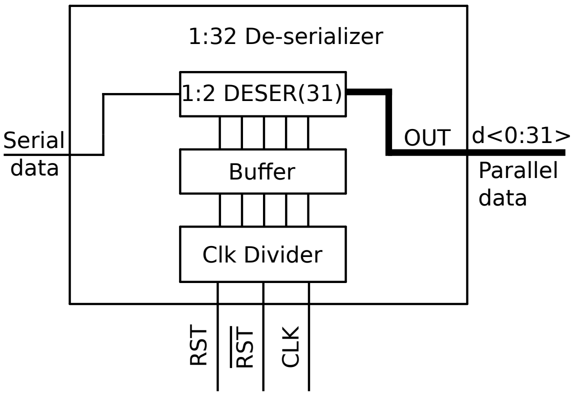


Figure 6.4: Test bench for Deserializer

Deserializer has similar structure to that of serializer block, except having 1:2 DESER instead of 2:1 SER. Input of the Deserializer is from output of the receiver chain present in the serial link. In order to validate the design of deserializer, 32 bit data stream was generated from pulse generating source along with clock and later replaced by receiver chain of Serial Link. The test bench for the de-serializer block is shown in the Figure 6.4.

Once each individual blocks were designed and validated, total integration was done and its shown in the Figure 6.5. The concept used in this test bench is to read data from memory unit using the designed sense amplifier and then transmit the serialized data in the serial link. At the other end, the clk and serial data is recovered from which data is parallelized using de-serializer in order to write the data into the desired unit. Due to the requirement of bi-directional data transfer, we use transmission gate and switches in serial link as shown in Figure 6.6.

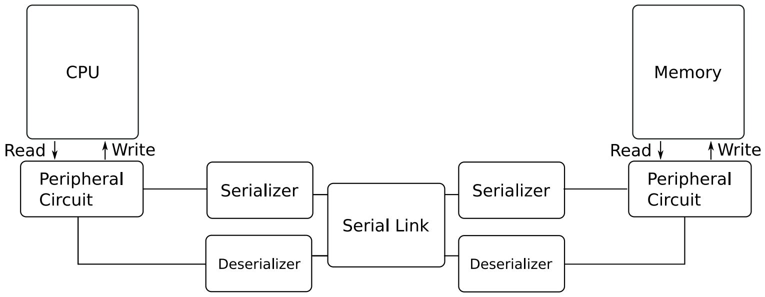


Figure 6.5: Test bench for whole system

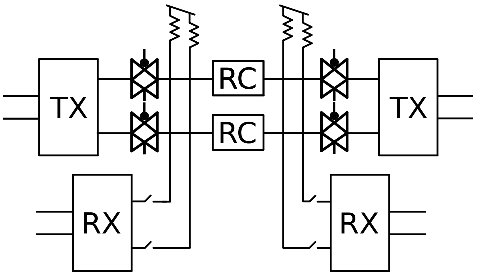


Figure 6.6: Serial link circuit

Chapter 7

Simulation

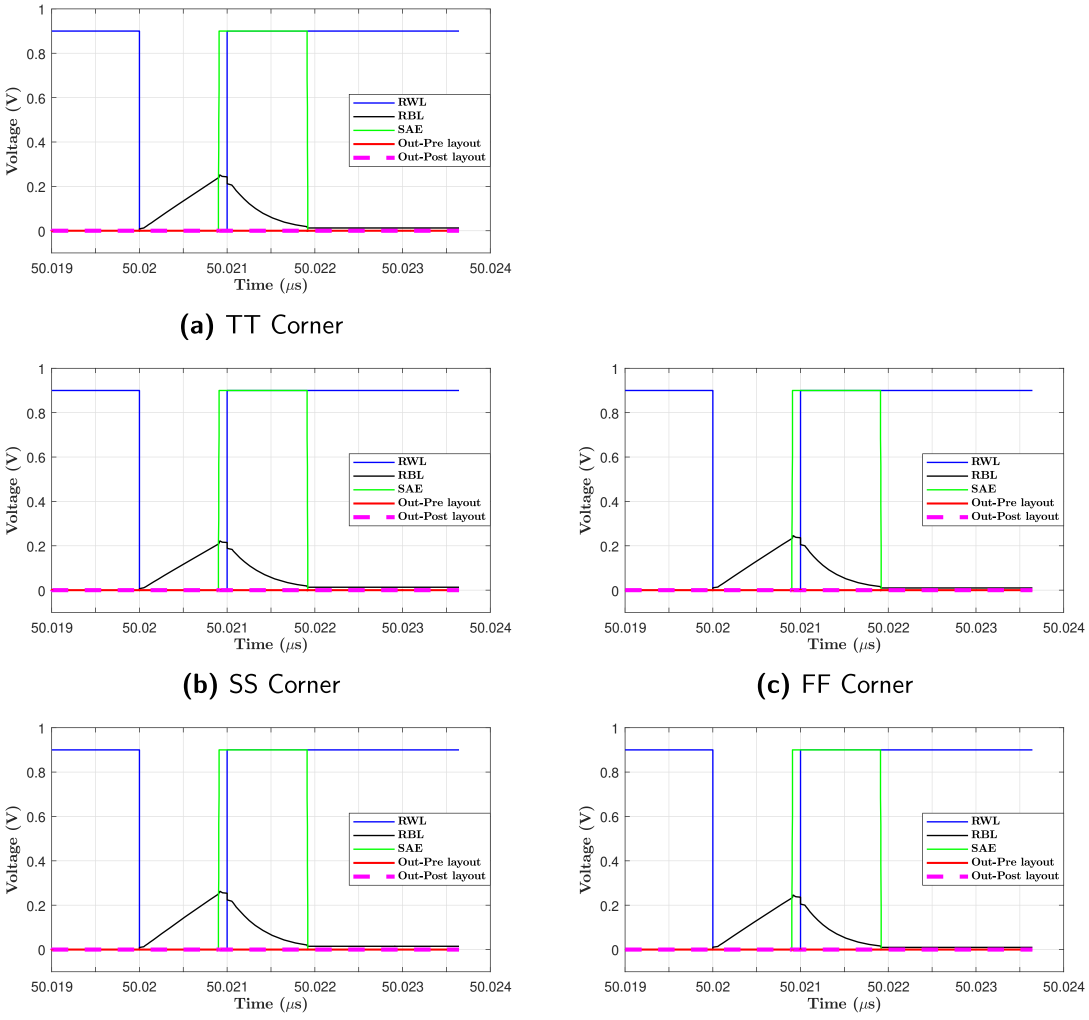
All the following figures and results were obtained with global corner otherwise mentioned.

eDRAM

The eDRAM read simulation for various corners for reading 1 and 0 is shown in the Figure 7.1, 7.2. The layout parasitics has effect on the propagation delay of the output signal. The range for reading 1 was between 40-120m and for reading 0 was between 200-250mv at different corners. Resolution time is defined as the amount of time taken by the output node to raise /fall 50 % of VDD from the moment SAE signal raises to 50 %. Due to this architecture, while reading 0, resolution time cannot be calculated. PDP (Power Delay Product) is a metric that predicts the energy cost of a reading operation. PDP was estimated by multiplying power consumed with resolution time. The table 7.1 represents the performance of the sense amplifier at various corners.

	1			0		
Corners	Resolution time (ps) Schematic	Resolution time (ps) Layout	Power (μw)	Resolution time (ps) Schematic	Resolution time (ps) Layout	Power (μw)
TT	231.5	339.9	26.05	-	-	7.2
SS	219.9	321.9	23.84	-	-	6.6
FF	282.7	442.1	26.55	-	-	10.6
SF	256.0	388.5	25.10	-	-	7.1
FS	216.9	312.1	25.40	-	-	7.6

Table 7.1: Performance Summary



(a) TT Corner

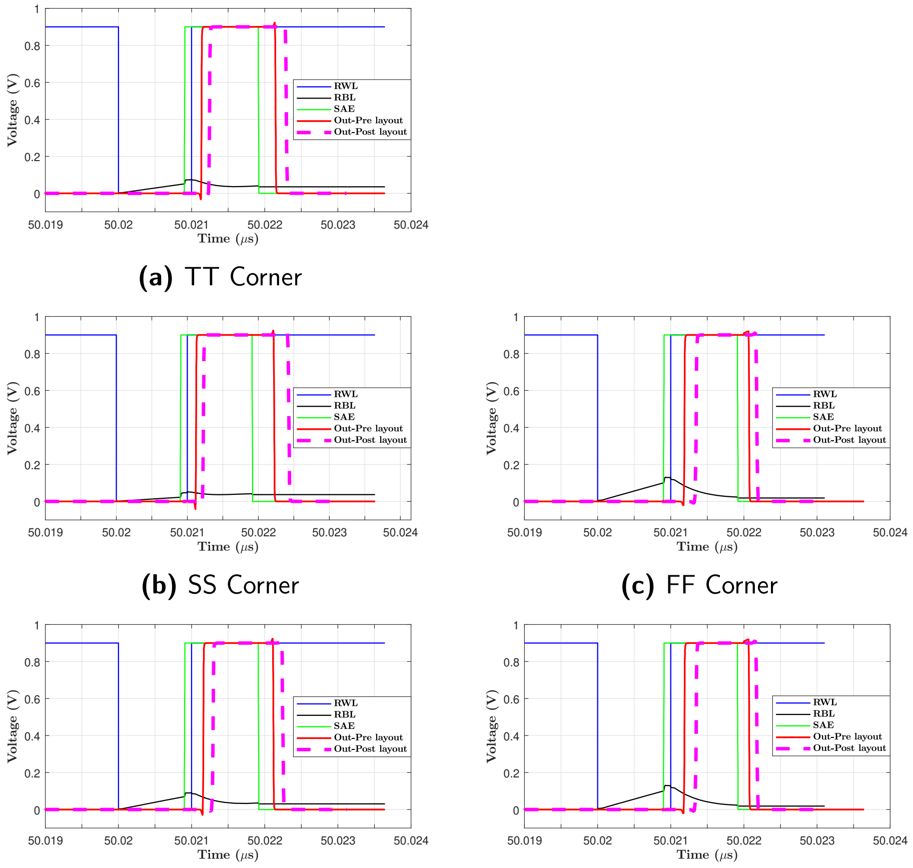
(b) SS Corner

(c) FF Corner

(d) SF Corner

(e) FS Corner

Figure 7.1: Reading 0



(a) TT Corner

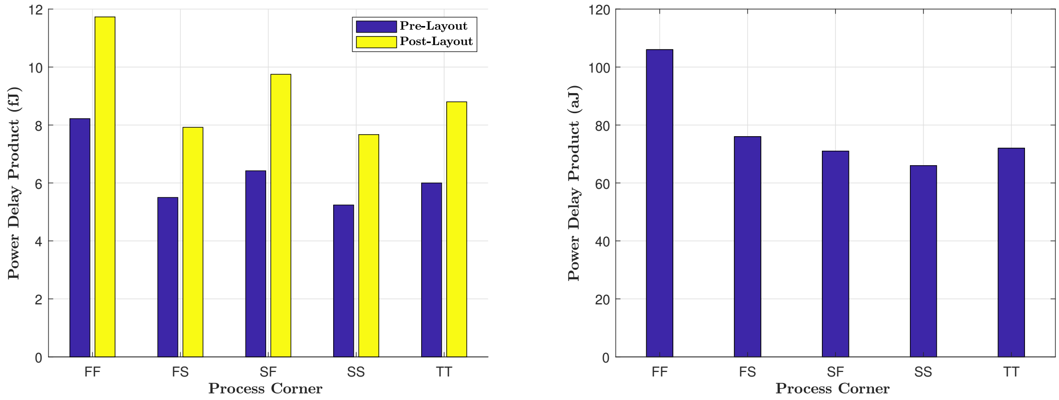
(b) SS Corner

(c) FF Corner

(d) SF Corner

(e) FS Corner

Figure 7.2: Reading 1

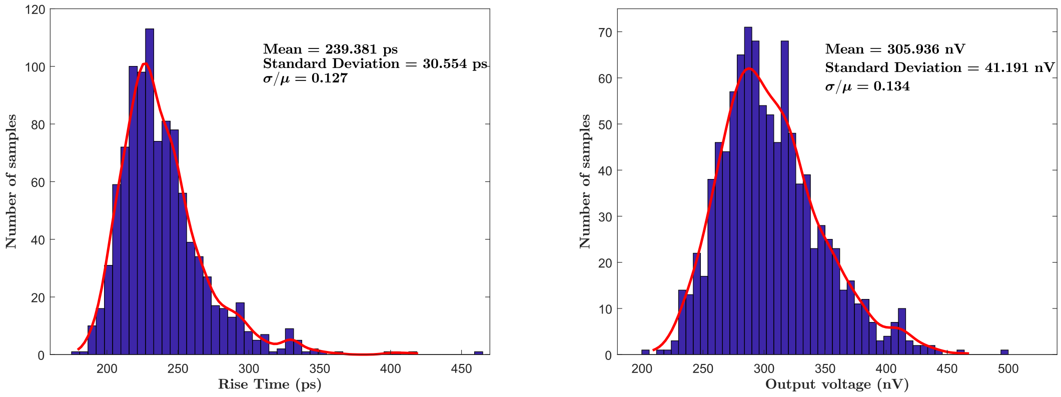


(a) PDP for 1

(b) PDP for 0

Figure 7.3: PDP at various process corners

Monte carlo simulation was simulated for inter-intra die random variations in the sense amplifier to verify its functionality with respect to process variations and mismatch for 1000 runs. In case of read 1, pass rate is calculated based on resolution time where as for read 0, pass rate is counted when the output signal is zero.



(a) MC for 1

(b) MC for 0

Figure 7.4: Monte Carlo Plots

Operation	μ (ps)	σ (ps)	Pass	Operation	μ (nv)	σ (nv)	Pass
Read 1	239.381	30.5554	100 %	Read 0	305.936n	41.191	100 %

Table 7.2: Monte Carlo Output

SRAM

The transient simulation of read process in SRAM is shown in the Figure 7.5. Here, we can clearly see there is not leakage in the bitline voltage (BLC,BLT), as the input is isolated well against the output node.

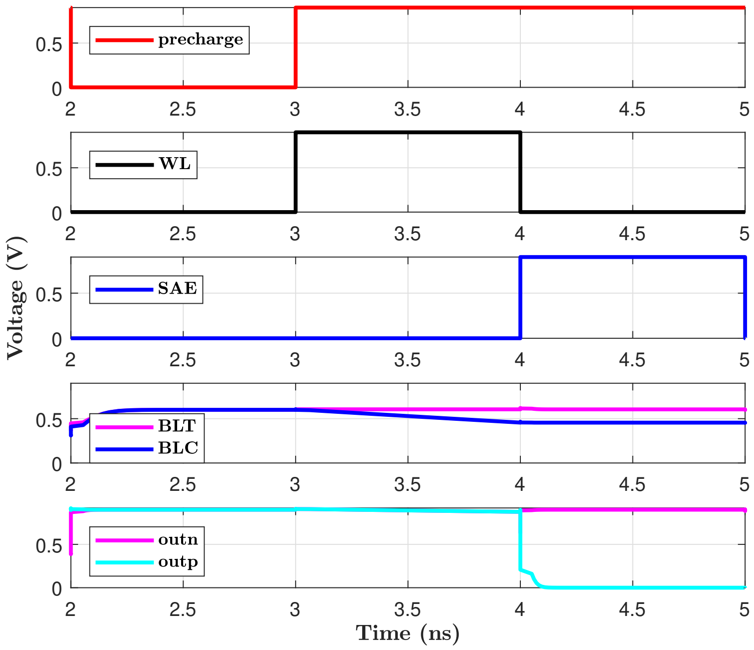


Figure 7.5: Transient voltage level during read process

The performance of the modified latched sense amplifier at various corners are tabulated at two different VDD's. It is observed that the system operates fast at FF corner with more consumption of power where as the slowest performance is at SS corner with low power consumption.

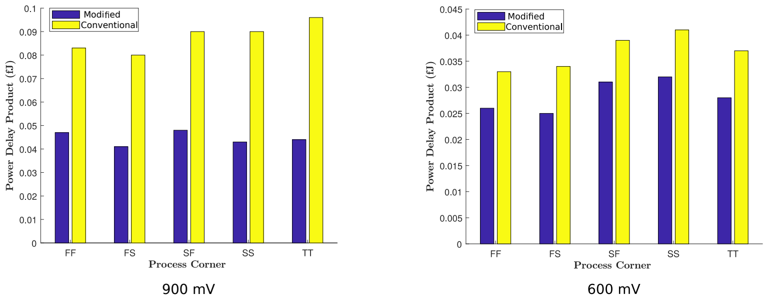
Corners	Resoultion Time (ps)		Power (μw)	
	Schematic	Layout	Schematic	Layout
TT	16.95	29.13	2.565	4.416
SS	18.06	31.31	2.385	4.167
FF	16.32	27.71	2.933	4.845
SF	18.43	31.53	2.600	4.392
FS	15.44	26.7	2.628	4.478

Table 7.3: Performance summary for modified latch type sense amplifier for 900mv VDD

Corners	Resoultion Time (ps)		Power (μw)	
	Schematic	Layout	Schematic	Layout
TT	25.52	45.93	1.103	1.910
SS	31.21	56.7	1.027	1.816
FF	21.51	38.08	2.047	3.457
SF	29.06	52.07	1.097	1.895
FS	22.70	40.77	1.124	1.950

Table 7.4: Performance summary of modified latch type sense amplifier for 600mv VDD

This design was compared with high impedance latch sense amplifier (in chapter 2) by having the same sizing for cross coupled inverters. It is observed that the PDP of the modified circuit is almost half less than that of the conventional design for 900mv VDD and one third less for 600mv VDD. This is because in modified design, we have the pull down network driven to ground before enabling the SAE signal, creating a low impedance path for faster decision.



(a) 900mv VDD

(b) 600mv VDD

Figure 7.6: Comparison of Power delay product

The inter intra die variations simulation for lowest voltage difference ΔV between bitlines for two different voltages are simulated for 1000 runs and result is shown in the Figure 7.7.

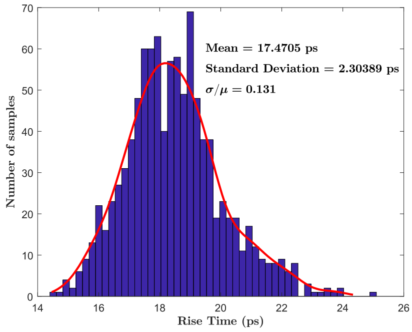


Figure 7.7: MC simulation Plot

Operation	μ (ps)	σ (ps)	Pass
Read	17.4705	2.30389	100 %

Table 7.5: MC Result

It is important to find the impact of the sensitivity of the modified latch type sense amplifier with respect to process variations for two different VDD's. As the SAE transistor(M9,M10) Figure 4.3 remains in the data path of decision making, their sizing dominates the sensitivity. We observe that sensitivity decreases as width increases as shown in the Figure 7.8. Sensitivity was frozen as it met the read speed consideration. In case of requirement of higher sensitivity, sizing must be increased which can add additional area overhead.

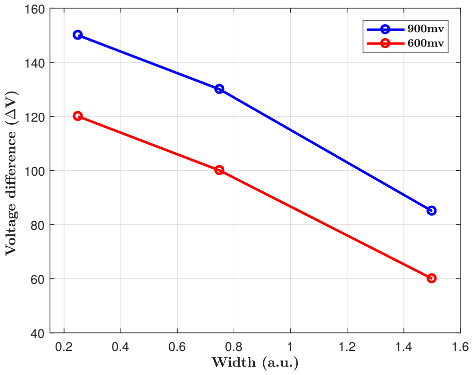
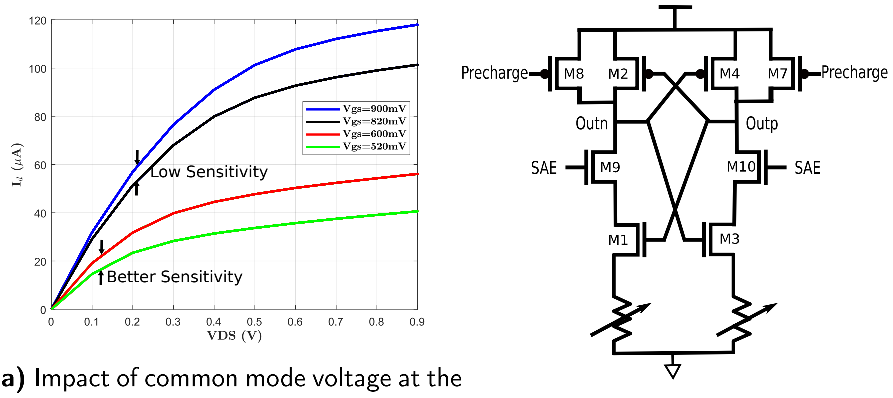


Figure 7.8: Sizing of SAE transistor for process variation

For the same sizing of sense amplifier, we observe different sensitivity for two different VDD. This is because, the sensitivity depends upon the difference in *gm* of the differential pair (operating at triode). For the same voltage offset, higher difference in *gm* is found for 600 mV than to that of 900 mV.



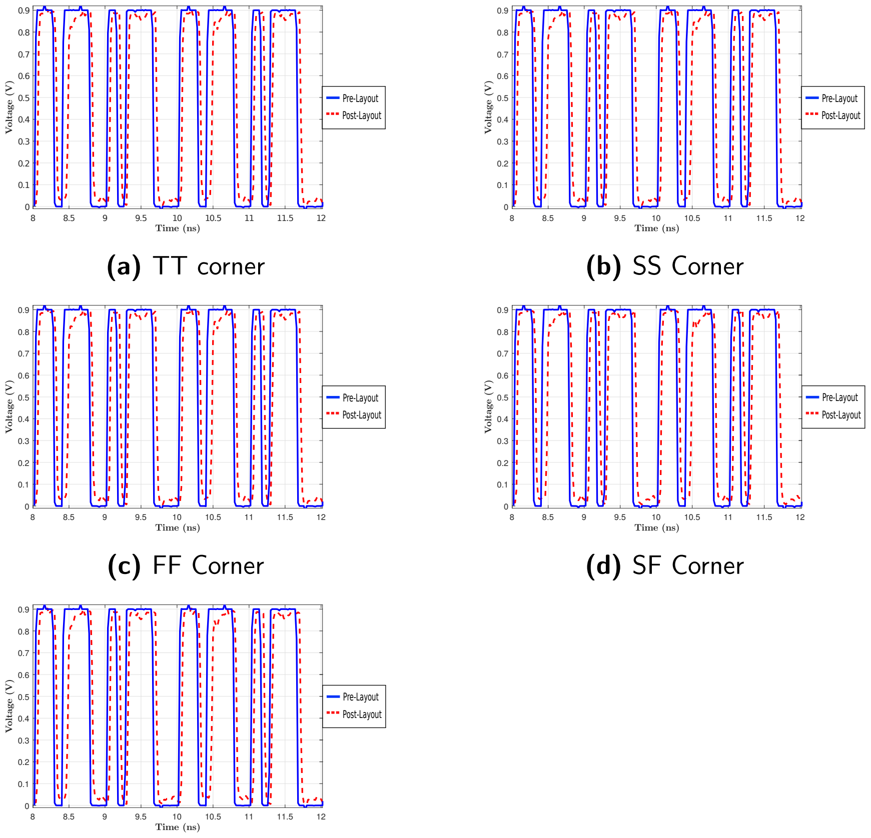
(a) Impact of common mode voltage at the differential input pair

(b) Resistive modeling

Figure 7.9: Sensitivity dependence

Serializer

32-bit serialized data (11011100101110001101110010111000) is shown in the Figure 7.10 for different process corners. As, we observe, layout parasitics has delay in its propagation time. There is an intrinsic delay at the output to that of high speed clock due to clock divider operation around 35-45ps in pre layout and 75-85ps in post layout. This can be negated by proper sampling at the next stage of transmitter. Two clock cycle (slowest clock) delay is observed to obtain the correct data at the output of the serializer. As we sample each unit node in the tree structure at half clock cycle delay and to implement 32 bit serialization process, we require 4 such nodes, which equals $(4 * .5) = 2$ cycle delay.



(a) TT corner

(b) SS Corner

(c) FF Corner

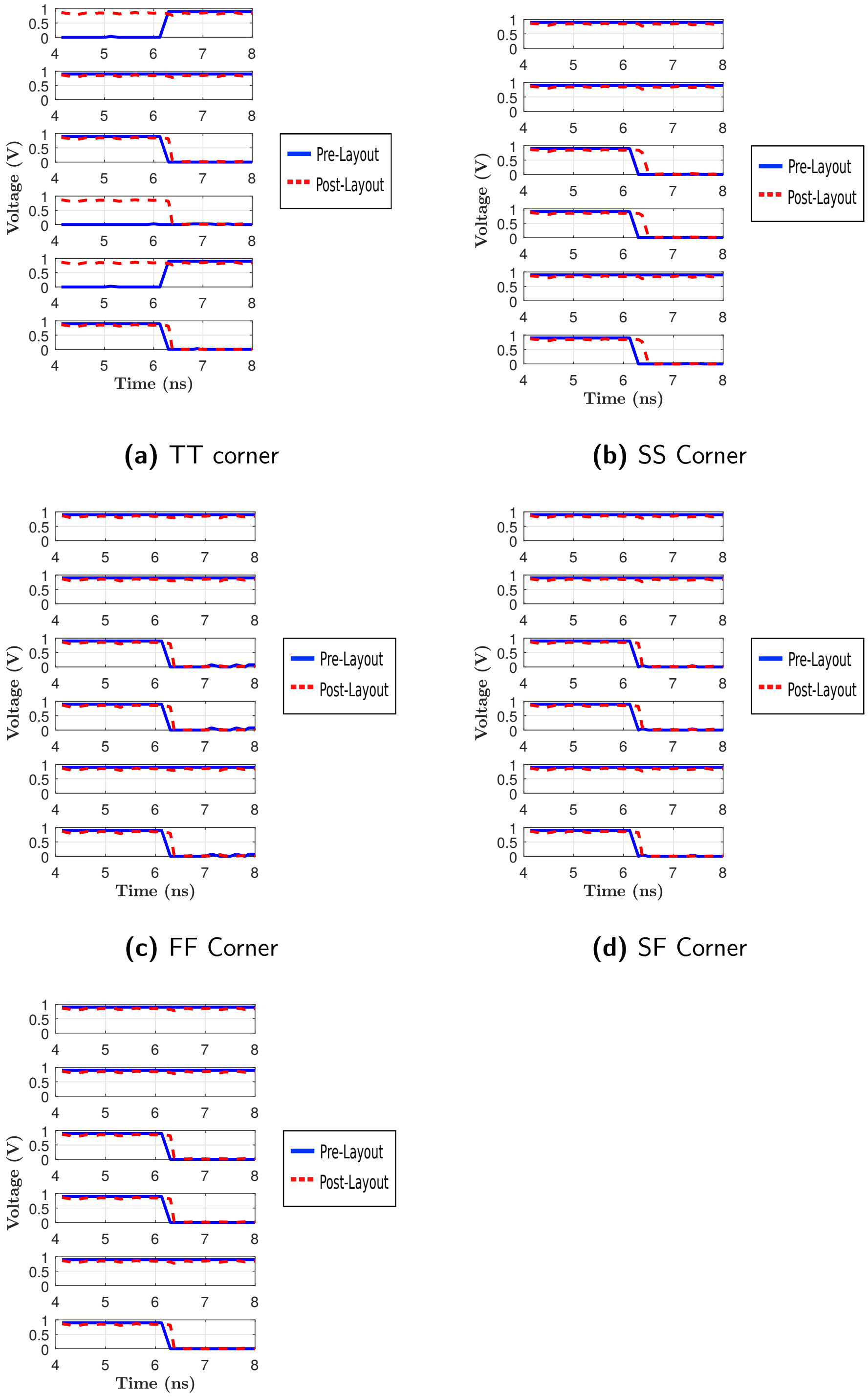
(d) SF Corner

(e) FS Corner

Figure 7.10: Serializer plots

Deserializer

A fraction of 32-bit deserialized data (110010) is shown in Figure 7.11 for various corners. The layout parasitics has effect on propagation delay of the signal. Delay in signal doesn't have any undesirable effect as we wait for writing time before next operation continues. Latency of 1.5 clock cycle (minimum clock) is observed due to the deserializer circuit design and is constant for 1:N bit deserialization process where N is greater than 2.



(a) TT corner

(b) SS Corner

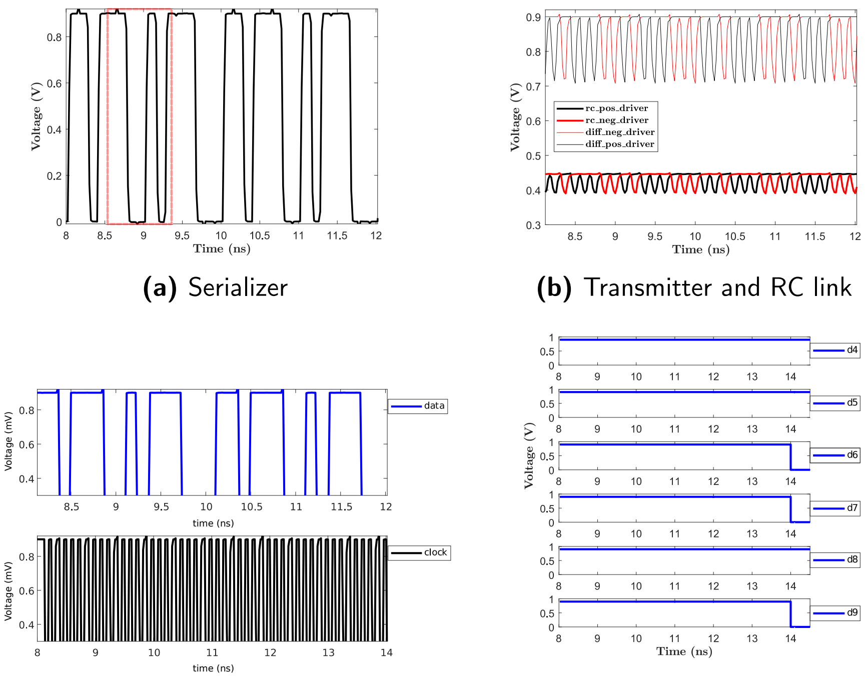
(c) FF Corner

(d) SF Corner

(e) FS Corner

Figure 7.11: Deserializer plots

The overall transient simulation result of data transmission between two units involving serializer, transmission and reception and deserialization process can be observed in the Figure 7.12. Only a part of the deserializer result is shown(data marked in serializer plots). Overall in this topology, it requires 3.5 clock cycle delay of the minimum clock (4 ns) to transfer data serially from one unit to another. To have a comparative study of this work, an already existing high speed serial link for low power memories [6] was compared in terms of power and latency.



(a) Serializer

(b) Transmitter and RC link

(c) Clock and data recovery

(d) Deserializer

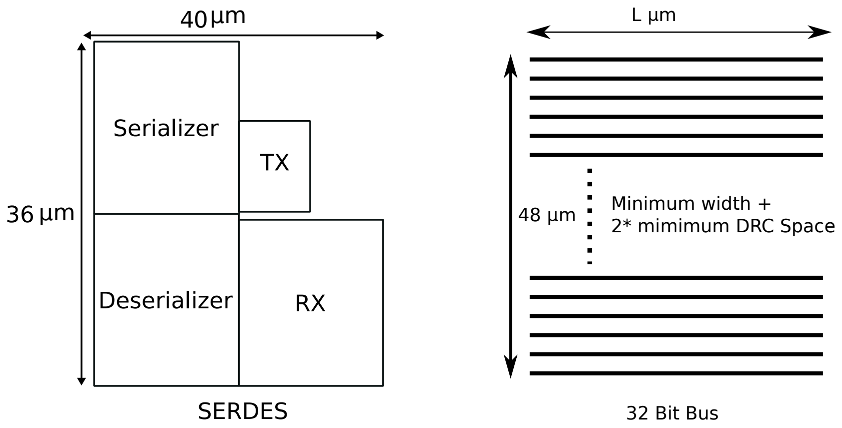
Figure 7.12: Serial link plots

Parameter	This Work	Previous Work [6]
Energy / bit	675fJ	4.93pJ
Latency	3.5 cycle	1 cylce
CMOS Technology	28nm Bulk	28nm Bulk

Table 7.6: Comparison of two serial link implementation

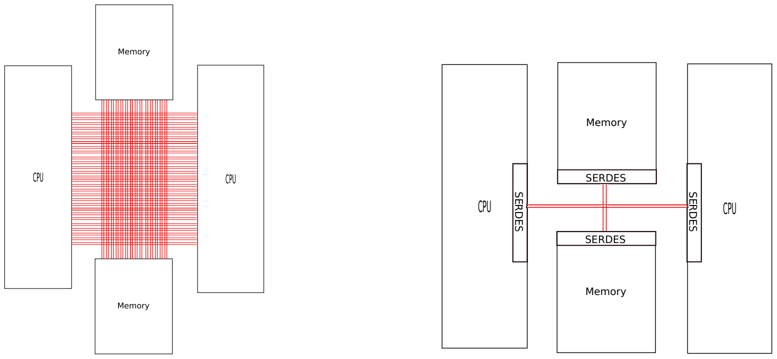
From the table, we can infer that, depending on the energy requirement latency varies. A tree based implementation requires lower energy and higher latency where as normal MUX based [6] (high speed operation) requires more energy and less latency.

The area estimate required to implement serial link is compared with respect to a 32 bit parallel bus having minimum width and distance between two data lines as 2 times the minimum DRC in order to accommodate for cross talk issues as shown in the Figure 7.13b. I make use of already existing receiver circuit layout from [6] to have area estimate for transmitter and receiver circuit. It is observed that with a length over head of 40μm, serial link can reduce the interconnect area upto 80% than in case of parallel links. When the length of interconnect is greater than 200μm, serial link becomes more optimal (to account for length overhead on both ends).



(a) Area estimate for SerDes **(b)** Parallel interconnect

Figure 7.13: Area Comparison



(a) Parallel interconnect **(b)** Serial interconnect

Figure 7.14: System comparison

Chapter 8

Conclusion and Future Work

8.1 Conclusion

In this project, on chip communication between different intellectual properties (IPs) inside an integrated circuit using serial interconnect was explored. Different blocks such as peripheral circuit (to read data from embedded memories), serializer/deserializer, required to implement serial interconnect were discussed. Area estimate between serial and parallel interconnect were examined. The design consumed 675fJ of energy to transfer single bit of data with a latency of 3.5 clock cycle. The serial interconnect system demonstrated to be a good candidate for high performance on chip communication as well as reducing interconnect area. It has the advantage of scalability, being that increasing the number of bits per packet won't increase the routing area.

8.2 Future Work

It is important to explore the feasibility of this implementation in a serial link consisting of multiple units of IPs such as memories and processors. High frequency clock signal was generated from ideal sources, which needs to be generated via PLL. Serializer/Deserializer can be implemented using higher level of implementation (4:1 or 8:1 instead of 2:1) to reduce the latency between data transfer. Serial link was implemented using analog CDR's. A completely different approach of using all-digital CDR needs to explored [11]. In this approach, the incoming analog signal is converted to digital signal using ADC and then the signal processing takes place. Another advantage of all-digital CDRs is that they are fully synthesizable and can be automated using hardware description languages. Serial SRAM required for microcontrollers as off chip memories with lower number of output pins can also be explored.

Bibliography

[1] Anonymous, "Internet of things-number of connected devices worldwide." https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[2] R. Dobkin, M. Moyal, A. Kolodny, and R. Ginosar, "Asynchronous current mode serial communication," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, pp. 1107–1117, 2010.

[3] J. Singh and B. Raj, "SRAM cells for Embedded Systems," Mar 2012.

[4] M. Price, J. Glass, and A. P. Chandrakasan, "27.2 a 6mw 5k-word real-time speech recognizer using wfst models," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 454–455, Feb 2014.

[5] B. Mohammadi, O. Andersson, J. Nguyen, L. Ciampolini, A. Cathelin, and J. Rodrigues, "A 128 kb 7t sram using a single-cycle boosting mechanism in 28 nm fd–soi," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 4, pp. 1257–1268, 2018.

[6] A. C. M. Rasgado, "High-speed serial link for low-power memories," Master's thesis, Lund Univeristy.

[7] Anonymous, "Embedded Dynamic Random Access Memory (eDRAM)." https://www.techopedia.com/definition/11442/embedded-dynamic-random-access-memory-edram.

[8] A. Deutsch, G. V. Kopcsay, P. J. Restle, H. H. Smith, G. Katopis, W. D. Becker, P. W. Coteus, C. W. Surovic, B. J. Rubin, R. P. Dunne, T. Gallo, K. A. Jenkins, L. M. Terman, R. H. Dennard, G. A. Sai-Halasz, B. L. Krauter, and D. R. Knebel, "When are transmission-line effects important for on-chip interconnections?," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, pp. 1836–1846, Oct 1997.

[9] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[10] E. Seevinck, P. J. van Beers, and H. Ontrop, “Current-mode techniques for high-speed vlsi circuits with application to current sense amplifier for cmos sram's,” *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 525–536, April 1991.

[11] R. Sabareeshkumar, “Circuit architectures for high speed cmos clock and data recovery circuits,” Master's thesis, University of Illinois Urbana-champaign.

[12] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, “Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, pp. 994–1007, July 2012.

[13] J. W. M. Meng Fan Chang, Shu Meng Yang, “Single -ended sense amplifier using dynamic reference voltage and operation method thereof.” US Patent 12,191,057.

[14] C. Toumazou, F. Lidgey, and D. Haigh, *Analogue IC Design: The Current-mode Approach.* IEE circuits and systems series, Peregrinus, 1992.

[15] Z. . Kong, K. . Yeo, and C. . Chang, “Design of an area-efficient cmos multiple-valued current comparator circuit,” *IEE Proceedings - Circuits, Devices and Systems*, vol. 152, pp. 151–158, April 2005.

[16] B. Razavi, “The strongarm latch [a circuit for all seasons],” *IEEE Solid-State Circuits Magazine*, vol. 7, pp. 12–17, Spring 2015.

[17] B. Razavi, “The cross-coupled pair - part i [a circuit for all seasons],” *IEEE Solid-State Circuits Magazine*, vol. 6, pp. 7–10, Summer 2014.

[18] B. Wicht, T. Nirschl, and D. Schmitt-Landsiedel, “Yield and speed optimization of a latch-type voltage sense amplifier,” *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1148–1158, July 2004.

[19] B. S. Reniwal, P. Singh, V. Vijayvargiya, and S. K. Vishvakarma, “A new sense amplifier design with improved input referred offset characteristics for energy-efficient sram,” in *2017 30th International Conference on VLSI Design and 2017 16th International Conference on Embedded Systems (VLSID)*, pp. 335–340, Jan 2017.

[20] A. Prieto, “Statistical approach for the design of refresh-free edram with retention timing constraint,” 2019. Student Paper.

[21] A. Tanabe, M. Umetani, I. Fujiwara, T. Ogura, K. Kataoka, M. Okihara, H. Sakuraba, T. Endoh, and F. Masuoka, “0.18-/spl mu/m cmos 10-gb/s multiplexer/demultiplexer ics using current mode logic with tolerance to threshold voltage fluctuation,” 2001.

[22] B. Razavi, *Design of Analog CMOS Integrated Circuits.* McGraw-Hill, first ed., 2001.

Appendix A

SERDES

It is very critical to understand the connection in the tree structure design of serializer/de-serializer to have desired output at their respective stages. Several iterative simulations were made in order to correctly convert the data from parallel to serial and vice versa. Figure A.1,Figure A.2 represents connection inside 32:1 serializer and 1:32 deserializer respectively.

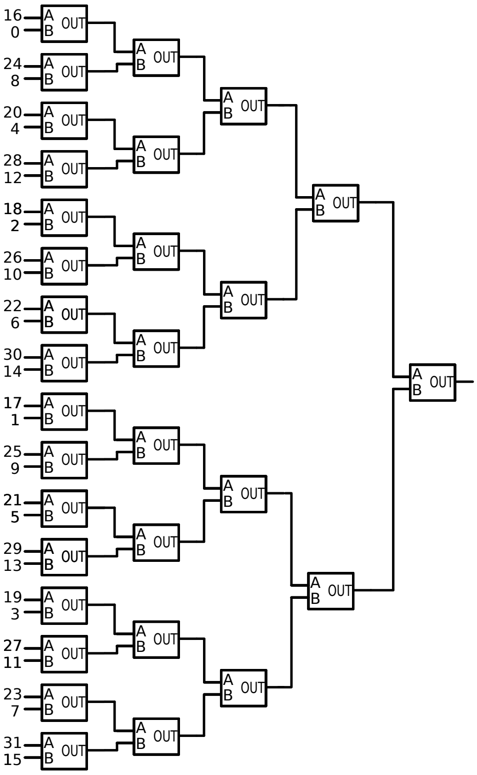


Figure A.1: Connection of 32:1 Serializer

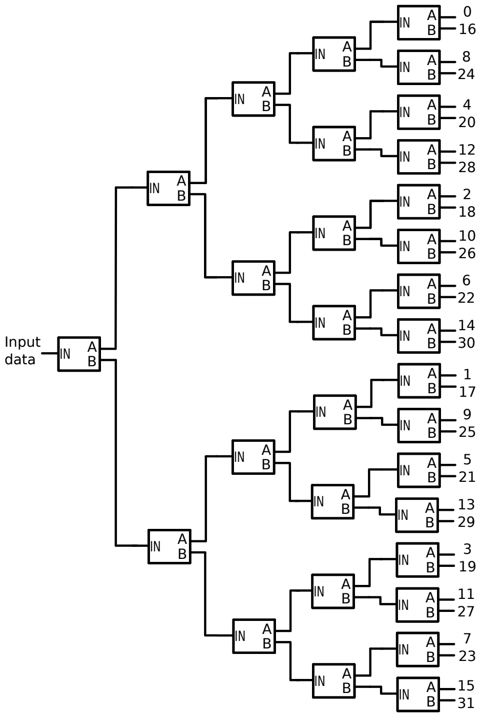


Figure A.2: Connection of 1:32 Deserializer

LUND
UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2019-726
http://www.eit.lth.se

Printed by Tryckeriet i E-huset, Lund 2019