CHAPTER 1

INTRODUCTION TO LOW POWER VLSI DESIGN

1.1 INTRODUCTION

The most important aspect of Moore's Law is that it has become a universal predictor for the growth of entire semiconductor industry. From Moore's law it is understood that the performance of an IC doubles every 18 months. This will increase the number of transistors used and hence increases the area and power consumption of the circuit. In general small area and high performances are two conflicting constraints. The IC designer's activities have been involved in trading of these constraints is shown in Figure 1.1.

Recently power dissipation is becoming an important constraint in design process. In that Low power design is becoming a new era in VLSI technology, as it impacts many applications. With the increase in speed, mobility and miniaturization of electronic devices, the power consumption of these devices has become major design factor. Especially for mobile devices the power consumption determines the battery life time. Therefore the designers, consumers, as well as environmental considerations demand a reduction of power dissipation in digital circuits. Digital circuits consist of a number of interconnected logic gates which together perform a function on one or more input signals. Every time an input signal changes it propagates via the gates causing signal transition in every place where the signals propagate. This causes current to charge or discharge the capacitive load of CMOS gates which results in power dissipation.

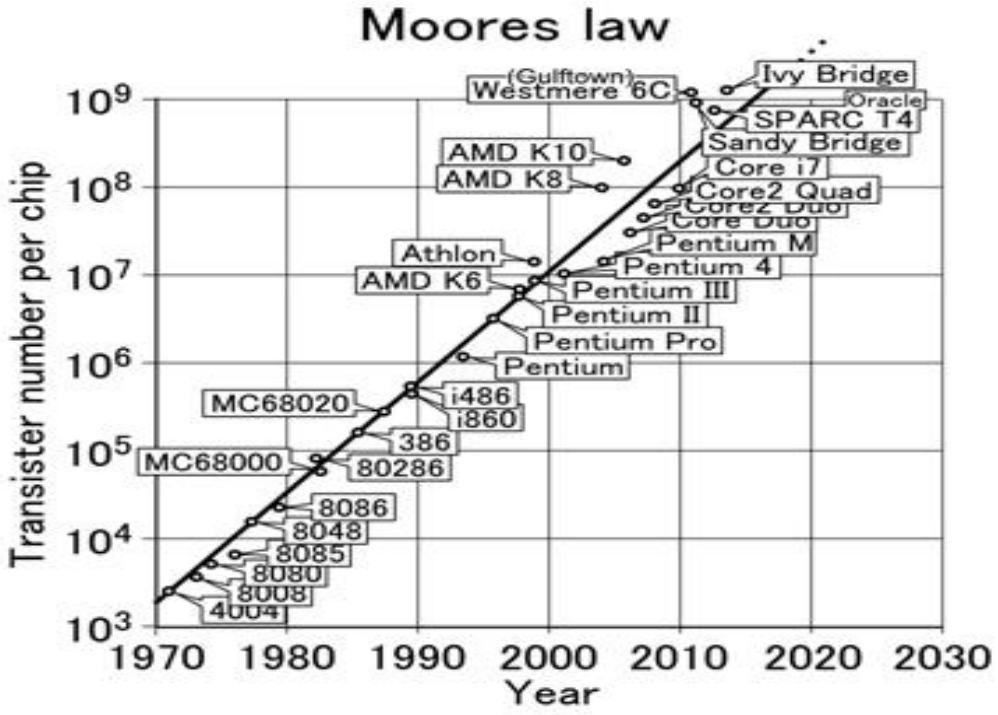


Figure 1.1 Graphical representation of Moore's Law

1.2 OVERVIEW OF VLSI

VLSI is a phrase that stands for "Very Large Scale Integration" is the process of creating an integrated circuit by combining millions of transistors in to a single silicon chip. VLSI field involves packing more and more logic devices into smaller and smaller areas because it is one of the basic building blocks of todays higher end devices and advanced technologies. Circuits that would have taken a full board of space can now be placed into a small space of few millimeters which has opened more opportunity to do things that were not possible before. VLSI circuits are used almost everywhere such as automobiles, different computer systems, car, digital camera, cell-phones and many other electronic products. According to Moore's law, integrated circuits capability increases exponentially every year, especially in computation power, area and yield.

1.2.1 Dealing with VLSI Circuits

Digital VLSI circuits are basically CMOS based. Blocks like latches and gates are implemented which is different from what have been seen so far, but the behavior remains same. Miniaturization involves new ideas to be considered. A lot of thought has to go into actual implementation as well as design. Some of the factors involved are as follows,

- **Circuit delays**

Large complicated circuits running at very high frequencies have one big problem to tackle - the problem of delay in propagation of signals through gates and wire even for few micrometers. Also operation speed takes more time as delays add up.

Another effect of high frequencies is increased consumption of power. This has two-fold effect - devices consume batteries quickly and an increase in heat dissipation. Taking into consideration that surface areas also have decreased, heat is a major threat to circuit stability.

- **Layout**

Laying out the circuit components is a task common to all branches of electronics. There are many possible ways to do this like, multiple layers of different materials on the same silicon and different arrangements of smaller parts for the same component and so on.

The power dissipation and speed in a circuit present a trade-off and optimizing one affects another. The selection between the two is determined by the layout of the circuit components. This will also affect the fabrication of chips.

1.2.2 The VLSI Design Process

All modern digital designs start with a designer writing a hardware description of the IC (using HDL or Hardware Description Language) in Verilog/VHDL. Both essentially describes the hardware (logic gates, Flip-Flops, counters etc) interconnect of the circuit blocks and its functionality. CAD tools are also available to synthesize circuits based on HDL. VHDL means "VHSIC Hardware Description Language", where VHSIC stands for "Very High Speed Integrated Circuit". By using these languages, circuits are designed at higher-level. One describes behavioral description that means what the circuit is expected to do and the other tells about structure and how the circuit is made. Other languages such as Verilog also work in a similar fashion. Both languages are used to generate a very low-level description that actually spells out how all this is to be fabricated on the silicon chips. Most of the time VLSI designs are classified into three categories.

1. Analog

Analog IC design involves small transistor precision circuits like as amplifiers, data converters, filters, phase locked loops, sensors etc.

2. Asics or application specific integrated circuits

Progress in the fabrication of IC has enabled us to create fast and powerful circuits in very smaller devices. This also means that more functionality could be packed into the same area. These are integrated circuits that are created for specific purposes and each device is created to do a particular job. The common application area is DSP signal filters, in Image compression. Consider the fact that the digital wrist watch normally consists

of a single IC doing all the time keeping jobs as well as extra features like calendar, Low Power GPS Receiver etc.

3. Soc or system on a chip

Soc systems are highly complex mixed signal circuits (digital and analog all on single chip) and very much used in embedded systems, wireless chip etc.

Advantages

1. Less power consumption
2. Less testing requirements at system level
3. Higher reliability
4. Less logic gate are needed here
5. Performance is very high
6. Less consumption work
7. Increase retention time
8. Reduced device area
9. High accuracy

1.3 INTERCONNECT

Interconnect plays an increasing role in determining the entire area of chip, delay, power dissipation and must be considered early, i.e. during the design process itself. Once the devices are fabricated completely, they must be electrically connected to each other to make circuits and communicate with outside world. The wires are used to connect the devices together are called interconnect which play a major role in the performance of modern system.

Figure 1.2 and 1.3 shows the schematic diagram and comparison of interconnects structure. In general interconnect can be classified in to Global interconnect and Local interconnect. Local interconnects are the first or lowest level of interconnect used to connect very short distance. Global interconnect used to connect very long interconnects between the blocks, including power, ground and clocks. Local interconnects have high resistivity than global interconnect. Global interconnects are very long interconnect and also connect different devices and different part of the circuit. So global interconnect always use low resistant materials.

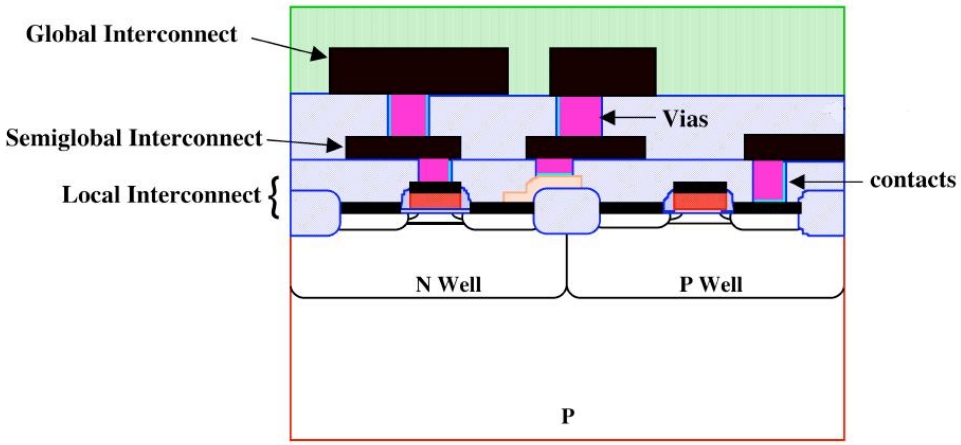


Figure 1.2 Interconnect structure

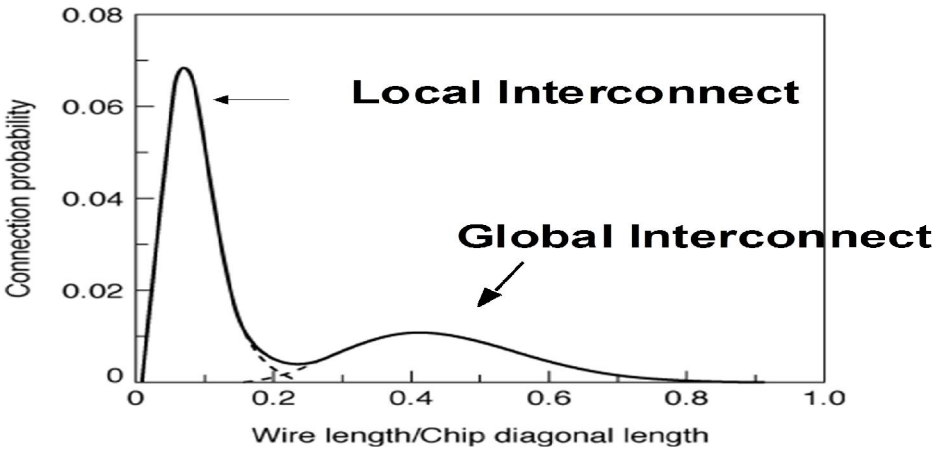


Figure 1.3 Comparison of interconnects, Rabaey 2003

1.3.1 Requirements of the Interconnection Materials

1. Low resistivity of conductors
2. Low capacitance => low dielectric constant
   - Low RC delay
   - Low cross talk
   - Low power dissipation (CV^2f loss)
3. Resistance to electro migration
4. Ease of deposition of thin films of the material
5. Ability to withstand the chemicals and high temperatures required in fabrication process
6. Ability to be thermally oxidized
7. Good adhesion to other layers - low physical stress
8. Stability of electrical contacts to other layers
9. Ability to contact shallow junctions and provide low resistance
10. Good MOS properties
11. Ability to be defined into fine patterns - dry etching

1.3.2 Interconnect Resistance

VLSI chips are just not transistors and they need to be connected properly. Analysis of interconnect is becoming as important as transistors in modern process. Wires are also as important as transistors. They affect Speed, they conceive power, and they can wear out too. In CMOS a resistive metal such as Nichrome may be added to produce high value and quality resistors. Its accuracy further improved by laser trimming the resulting resistors on each

chip to some predetermined test specification. Sheet resistance value in the KΩ / square is normal. The resistors have excellent temperature stability and long term reliability. Until the 180 nm generation, most wires were aluminum, modern process often use copper because of low resistivity. Figure 1.4 shows interconnect resistance geometry here W is the width of the wire l is the length and t is the thickness. Resistivity of different metals were listed in Table 1.1. From this table silver is not used in modern process because of high cost so nowadays copper is used to design the system. Figure 1.5 shows resistivity of copper wire dimensions. The resistance of a uniform slap of conducting material can be expressed as

$$R = \frac{\rho l}{wt} \tag{1.1}$$

Where ρ is the resistivity. This expression can be written as

$$R = R_o \frac{l}{w} \tag{1.2}$$

Where R_o is the sheet resistance and has units of Ω / square

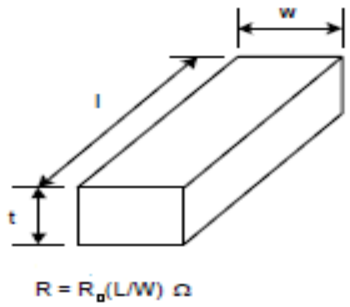


Figure 1.4 Resistance geometry

Interconnect resistance can be reduced by the following methods

1. By using materials with low resistivity
2. Using reduced wire length
3. Using of less vias

4. Use metal instead of polysilicon even for short distance routing

5. Use silicide coating to reduce polysilicon resistance

Table 1.1 Resistivity of different metals

Sl.No	Metal	Bulk resistivity (μ Ω *cm)
1	Silver(Ag)	1.6
2	Copper(Cu)	1.7
3	Gold(Au)	2.2
4	Aluminum(Al)	2.8
5	Tungsten(W)	5.3
6	Molybdenum(Mo)	5.3

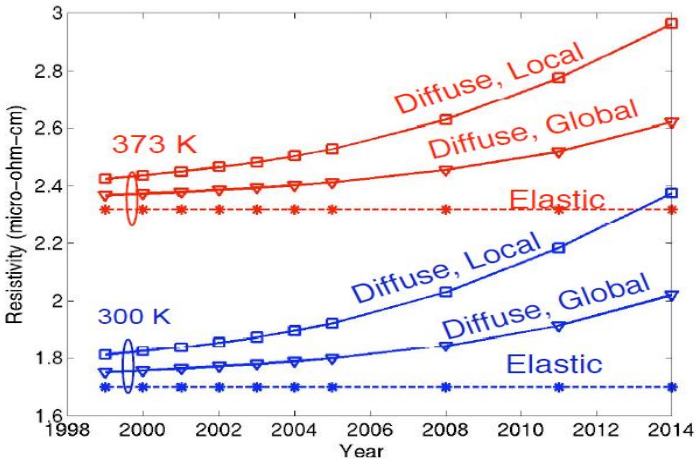


Figure 1.5 Resistivity of Cu wires dimensions recommended by ITRS'99

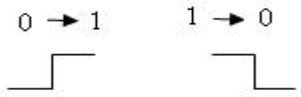
1.3.3 Interconnect Capacitance

Low power design not only needed for portable applications but also to reduce the power of high performance systems CMOS has been recognized as the technology for VLSI because of its unique advantages. It is suitable for low power systems demanding high speed operations one

important goal in achieving low power CMOS device is the reduction of capacitance. The interconnect capacitance estimation is however a difficult task even after technology mapping due to lack of detailed place and route information. Approximate estimation can be obtained by using information derived from a placement solution or by procedural interconnect models. There are three types of capacitance formed between interconnects are shown in Figure 1.6.

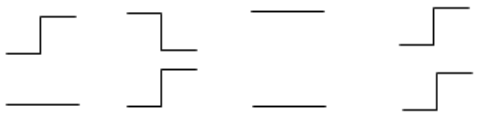
1.3.3.1 Area capacitance

It is formed between substrate and any metal layer. It is also named as self capacitance or metal to substrate capacitance



1.3.3.2 Mutual coupling capacitance

Capacitance between metal lines of the same metal layer is called inter wire or coupling



1.3.3.3 Fringe capacitance

It is formed between non overlapping side wall of one conductor and surface / side wall of a second conductor on the same layer or different layer from the first one

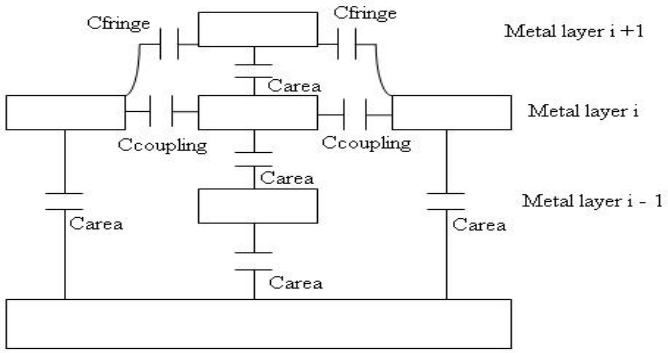


Figure 1.6 MOS capacitance model

1.4 BUS MODEL

The scaling of wire dimension not only affects delay time but also all related interconnect performance such as power dissipation, reliability and bandwidth for local and global interconnects. The ever decreasing interconnect cross section dimension gives rise to increase in resistance (Arti Joshi et al 2014). VLSI chips are not just transistors alone, but they also have to be connected and also should be supplied with power and ground. In today's technology wires form complex geometries that introduce parasitic elements like resistive, capacitive and inductive. The parasitic effect affects circuit performance by increasing signal propagation delay, energy dissipation, power distribution and also introduces extra noise sources which affect circuit reliability. Bus energy model is classified in to lumped RC model and distributed RC model

1.4.1 Lumped RC Model

Here the lumped total resistance of the wire in to single resistor 'R' and combines the global capacitance of the wire in to a single a capacitor 'C' is shown in Figure 1.7 and the different lumped interconnect models displayed in Figure 1.8.

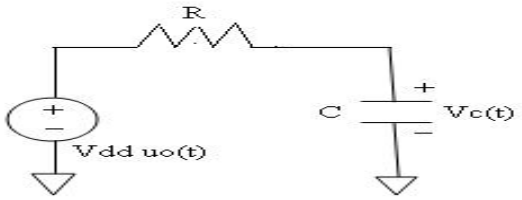


Figure. 1.7 Lumped RC Model

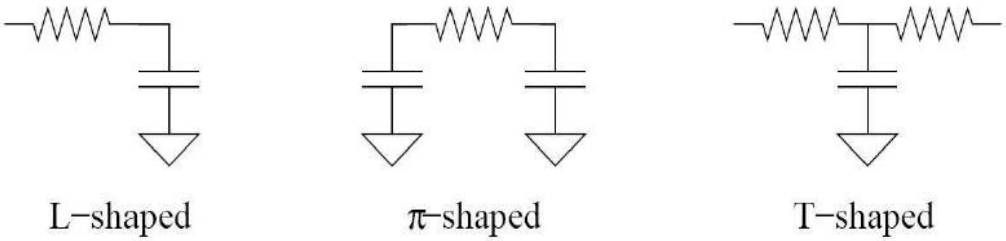


Figure 1.8 Lumped interconnect models

From the Figure 1.7.

$$Iint = C\frac{dVc(t)}{dt} \tag{1.3}$$

$$\frac{Vdd\,u(t) - Vc(t)}{R} = C\frac{dVc(t)}{dt} \tag{1.4}$$

By applying Laplace Transform,

$$Vdd\frac{u(t)}{R} - \frac{Vc(t)}{R} = C\frac{dVc(t)}{dt} \tag{1.5}$$

$$\frac{Vdd}{SR} - \frac{Vc(s)}{R} = Cs\,Vc(s) \tag{1.6}$$

$$Vc(s) = \frac{Vdd}{RC} \cdot \frac{1}{s\left(s+\frac{1}{RC}\right)} \tag{1.7}$$

$$Vc(s) = Vdd\left[\frac{1}{s} - \frac{1}{\left(s+\frac{1}{RC}\right)}\right] \tag{1.8}$$

$$Vc(t) = Vdd\left[1 - e^{-\frac{t}{RC}}\right]uo(t) \tag{1.9}$$

Time delay (50% and 90 %)

$$Vc(t) = Vdd\left[1 - e^{-\frac{t}{RC}}\right] \quad (1.10)$$

$t_{0.5}$ = 0.693RC

$t_{0.9}$ = 2.3 RC

1.4.2 Distributed RC Model

Figure 1.9 shows distributed RC bus model here a wire can be represented in to several RC sections. The delay of the circuit is approximated as;

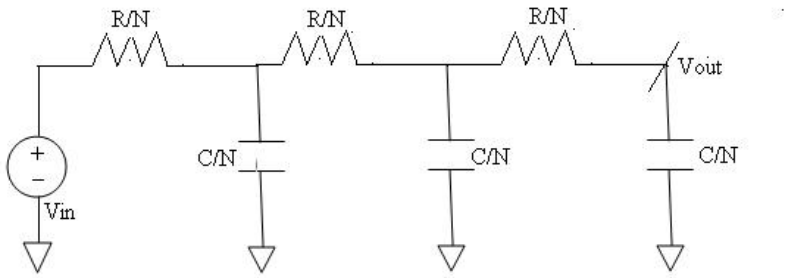


Figure 1.9 Distributed RC model

$$T_d = \sum_{j=1}^{n} \frac{C}{N} \sum_{k=1}^{n} \frac{R}{N} \quad (1.11)$$

$$T_d = \frac{C}{N} * \frac{R}{N}\left(\frac{N(N+1)}{2}\right) \quad (1.12)$$

$$T_d = RC\left(\frac{N+1}{2N}\right) \quad (1.13)$$

T_d = RC / 2 for N = ∞ (1.14)

N – is the number of sections

If N becomes large the sections reduces to

$$T_d = \frac{RC}{2}\left(\frac{r1}{w}\right)\left(Calw + 2Cp(l+w)\right) \quad (1.15)$$

$$T_d = \frac{1}{2} rCal^2 \quad (1.16)$$

r – Resistance per unit length

c – Capacitance per unit length

l – Wire length

Here RC effects dominate for very long wires because the delay is proportional to the square of the length. Doubling the length of the wire will quadruples the delay is listed in the Table 1.2.

Table 1.2 α and β for lumped and distributed network

Voltage	Lumped RC (α)	Distributed RC (β)
0 - 50%	0.69	0.38
0 - 60%	1	0.5
10% - 90%	2.2	0.9
0 - 90%	2.3	1

1.5 POWER ESTIMATION

Power dissipation of CMOS circuits strongly depends on the total load capacitance. Based on it is classified in to static power dissipation and dynamic power dissipation Static power consumed during the standby mode of a CMOS design .CMOS Transistors typically have some amount of sub threshold leakage current even when Transistors are not turned on. The drain to source leakage current is the main component of static power consumption.

$$P_S = 1/T \int_0^T iDD(t) VDD \, dt \quad (1.17)$$

Dynamic power is consumed by the switching activity of the CMOS design during charging and discharging the interconnect load capacitance

$$P_D = E\ C_L.V_{DD}^2.f \quad (1.18)$$

Load capacitance is the sum of three components is shown in Figure 1.10.

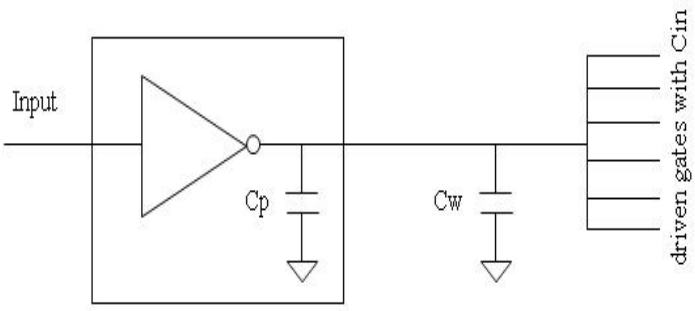


Figure 1.10 Load capacitance of CMOS gate

1. Total input capacitance C_{in}
2. Parasitic output capacitance C_P
3. Wiring capacitance C_w

$$C_L = Cin + C_P + C_W \quad (1.19)$$

1.5.1 Estimation of C_{in}

The total input capacitance C_{in} of the driven gates can be evaluated by summing the input capacitance of all the receiving gates

$$C_{in} = \sum_{i=0}^{N} C_{gatei} \quad (1.20)$$

The gate capacitance of the receiving gate can be approximated by

$$C_{gate} = C_{ox} \sum_{i=0}^{N} (WL)_i \quad (1.21)$$

1.5.2 Parasitic Capacitance

The parasitic capacitance is nothing but the summation of all overlap capacitance and junction capacitance.

$$C_p = C_{ovp}^{'} + C_{ovn}^{'} + C_{djp} + C_{djn} \quad (1.22)$$

1.5.3 Wiring Capacitance

The simple model of wiring capacitance is based on the parallel plate model is given by

$$C_w = lwC_{wa} \quad (1.23)$$

Research work analyzes efficient methods to transmit data through interconnect via buses .Thus the data's are initially compressed then coded using different data coding algorithm , finally transition estimator estimate ,count, compare the least transition coding method for better transmission. The following three methods are discussed in this thesis.

1. Data compression
2. Data Coding
3. Hamming distance estimator

1.6 DATA COMPRESSION

Data compression involves the development of a compact representation of information in all its form by exploiting structure or redundancy in the data. Figure 1.11 shows basic compression principle here compression algorithm is used to compress the information in the required bit level and the decompression algorithm is used to retrieve the original information. More popularly a compression scheme consists of two main

groups namely lossless compression and Lossy compression. Lossless compression preserves all the information in the data being compressed and the reconstruction is identical to the original data. In lossy compression some of the information contained in the original data may be lost. Even though lossy compression may lead to more effective compression, in some important situation details cannot be lost. Examples such as text or messages are preserved as it is without any lose for some legal reasons in military and medical applications. So a lossless compression technique like Simple byte compression algorithm is implemented in this research work.

The reason data compression is needed is that most of the information generated and used in digital form, is in the form of numbers, represented by bytes of data. If the application is multimedia, the number of bytes required to represent the data can be very large. This huge number of data is coded and transmitted as it is and increases the encoder and decoder complexity. This effect will increase Transition activity, area and delay and power dissipation. To overcome these effects a simple byte compression algorithm is implemented here. Simple byte compression algorithm is originally implemented for text, message and bulky data transmission. Here the algorithm is designed to compress up to 8 bytes of information into 5 bytes information and can get high compression ratio compared to other existing methods.

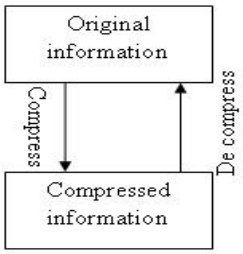


Figure 1.11 Basic compression principle

1.7 DATA CODING

Here input data are coded by using multi coding technique. In this technique data's are coded in eight different ways such as Invert, Swap, invert even position, Invert odd position, Rotate left with invert, Rotate Right with invert, Circular Left Shift and Circular Right Shift. Then 3 bit control signals are added to each coding technique to recover the original data at the decoder

1.8 HAMMING DISTANCE ESTIMATOR

During data transmission via buses, the input data changes from logic '0' to logic '1' and vice versa is called transition activity. Number of transition activity is calculated by using Hamming distance estimator, where Hamming distance'd' is equivalent to Transition activity on buses. So the Hamming distance estimator estimates Hamming distance'd' between the coded input data with reference data. Similarly Hamming distance is calculated for all possible input data .Among all least Hamming distance'd' is transmitted to the output.

1.9 MOTIVATION

In the past, the major concerns of VLSI designers were area, performance, cost and reliability. Power consideration was mostly of secondary importance. In recent years, power is the most emerging factor compared to all designing parameters. Growth of personal computing devices like portable desktops, audio and video based multimedia applications and wireless communication devices demands high speed computation and complex functionality with low power consumption. Power consumption must be dissipated through packaging and cooling strategies are also required when chip power consumption increases. Also there is financial advantage in reducing the power consumed for high performance systems. In addition to

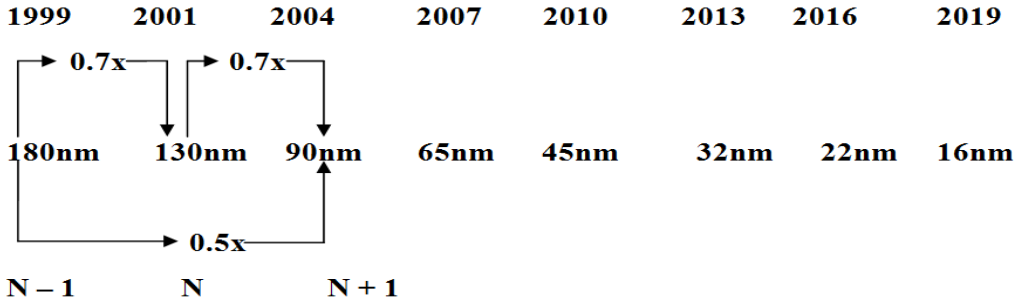
cost, there is an issue of reliability. High power systems often run hot and high temperature tends to aggravate several silicon failure mechanisms. Every 10^{o} c increase in operating temperature roughly doubles a components failure rate. Another crucial driving factor is that excessive power consumption is becoming the limiting factor in integrating more transistors on a single chip or on a multiple chips. If power consumption is not drastically reduced, the heat generated limits the feasible packaging and performance of VLSI circuits and systems. From the environmental point of view, smaller power dissipation will produce smaller heat energy, thus lower the electricity consumption. These will lower the impact on global environment.

The motivation for reducing power consumption differs according to applications. In low power battery operated portable applications, such as cellular phones and personal digital assistance, the goal is to keep the battery lifetime and weight reasonable and the packaging cost low. Power levels below 1-2 W, for instance, enables usage of plastic packages. For high performance, portable computers, such as laptops and notebook computers, the goal is to reduce the power dissipation of the electronic portion of the system to a point which is about half of the total power dissipation. Finally, for high performance, non battery operated systems, such as pcs, set-top computers and signal processors; the overall goal of power minimization is to reduce system cost while ensuring long term device reliability. These different requirements impact how power optimization is addressed and how much the designer is willing to sacrifice in cost or performance to obtain lower power dissipation. According to moore's law the technology parameters were analyzed in the following tables. Based on the technology used, parameters like gate length, oxide thickness, dielectric constant, threshold voltage, etc were listed in the Table 1.3 until the year 2016.

Table 1.3 Forecast technology parameters

Year	Technology Node(nm)	Physical Gate(nm)	Tox (nm)	Dielectric K	V_{dd} (V)	V_{th} (V)	Na (/cm^3)	Nd (/cm^3)	xj (nm)
2001	130	90	3.0	3.7	1.2	0.34	1.0_e16	1.0_e19	67.5
2004	90	53	2.4	3.0	1.1	0.32	1.4_e16	1.4_e19	46.7
2007	65	32	1.7	2.5	0.9	0.29	2.0_e16	2.0_e19	33.8
2010	45	22	1.5	2.0	0.8	0.29	2.9_e16	2.9_e19	23.4
2013	32	16	1.4	1.9	0.7	0.25	4.0_e16	4.0_e19	16.6
2016	22	11	1.3	1.7	0.6	0.22	5.9_e16	5.9_e16	11.4

Table 1.4 Technology nodes from 1999 – 2019



The ever decreasing interconnect cross section dimensions increases resistance. Table 1.4 describes how the technology scales down from 1999-2019. Two year cycle between nodes until 2001, then 3 year cycle begins. The graphical representation of technology scale down shown in Figure 1.12. The scaling of wire dimensions not only deteriorates delay time but also all related interconnect performance such as power dissipation, reliability and bandwidth for local and global interconnects. The on chip power dissipation problem depends with increasing number of long RC time constant of wires, switching activity factor and increase in operating frequency. The reliability issue also becomes very important since future system will require high current density within the reduced wire cross section. Here main aim is to minimize power dissipation by reducing switching activity. One of the methods to minimize the switching activity at the algorithm level is to use an appropriate coding for the signals rather than using of straight binary code.

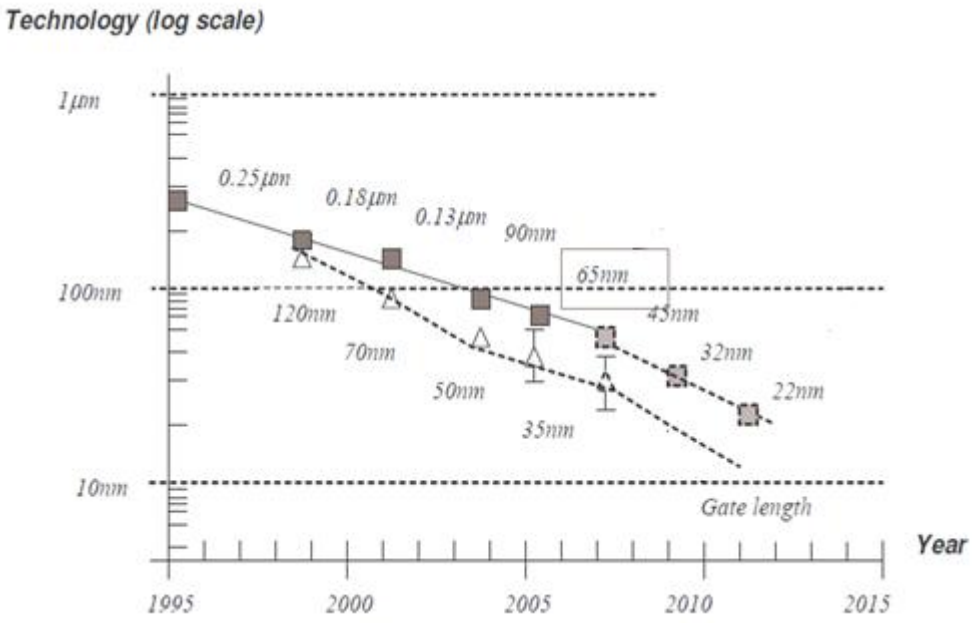


Figure 1.12 Technology scale down

1.10 OBJECTIVE OF THESIS

The main objective of this work is to bring out the low power, high performance and minimized area VLSI design for the present emerging applications. The overall performance cannot be improved by a single technology and is an effort of collected design technologies from semiconductor to final version of system design. This work highlights various methods which are suitable for reduction of power consumption and area of VLSI designs.

The disparity caused due to the resistive and capacitive effects posses a major challenge to VLSI system designers. Much of the power in a bus is dissipated in the process of charging and discharging of high nodal and inter wire capacitances. The majority of the work in this field focuses on minimizing the bus energy transition activity. Here the energy transition activity is minimized by simple byte compression and Multi coding techniques. These techniques are effectively implemented using Xilinx,

ModelSim, Tanner EDA tools. XPower is used to analyze the power calculation.

1.11 ORGANIZATION OF THESIS

This thesis is composed of seven chapters. The overall organizations of the chapters are as follows

Chapter 2 explains various research works about data compression methods, different data coding principles with power dissipation and transition activity.

Chapter 3 explains data compression method to improve compression ratio of input data bytes. A simple byte compression algorithm is implemented, analyzed and compared with different lossless data compression methods.

Chapter 4 describes the concept of data coding principle; also it presents multi coding technique to code the data. Performance is compared with other existing methods

Chapter 5 describes the power estimation techniques, reduction principles and also power estimation tools explained briefly

Chapter 6 presents the transition estimator principles. Also it explains how transitions are calculated by using Hamming distance estimator implemented by full adder logic and pass transistor logic. The performance is compared with other existing methods.

Chapter 7 concludes the research work done and summarizes the finding. It also indicates the potential opportunities that can be extended as future work.