Study of Electro-migration (EM) on Full Custom VLSI SRAM Designs and Development of an EM Assessment Methodology for Memory Compilers

Major Project Report

Submitted in partial fulfillment of the requirements

for the degree of

Master Of Technology

 \mathbf{in}

Electronics & Communication Engineering

(VLSI Design)

By Anand Dhonde (13MECV06)



Electronics and Communication Engineering Branch Electrical Engineering Department Institute Of Technology Nirma University Ahmedabad-382481 May 2015

Study of Electro-migration (EM) on Full Custom VLSI SRAM Designs and Development of an EM Assessment Methodology for Memory Compilers

Major Project Report

Submitted in partial fulfillment of the requirements

for the degree of

Master Of Technology

 \mathbf{in}

Electronics & Communication Engineering

(VLSI Design)

By Anand Dhonde (13MECV06)

Under the guidance of

External Guide Mrs. Shafquat Ahmed Engineering Specialist, ST Microelectronics India Ltd. Internal Guide Dr. N. M. Devashrayee Program Co-ordinator, Nirma University.



Electronics and Communication Engineering Branch Electrical Engineering Department Institute Of Technology Nirma University Ahmedabad-382481 May 2015



Certificate

This is to certify that the Major Project entitled "Study of Electro-migration (EM) on Full Custom VLSI SRAM Designs and Development of an EM Assessment Methodology for Memory Compilers" submitted by Dhonde Anand Bharatkumar (13MECV06), towards the partial fulfillment of the requirements for the degree of Master of Technology in VLSI Design, Nirma University, Ahmedabad is the record of work carried out by him under our supervision and guidance. In our opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of our knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. N. M. Devashrayee Internal Project Guide

Dr. P. N. Tekwani Head of EE Dept. **Dr. N. M. Devashrayee** PG Coordinator(VLSI Design)

Dr. Ketan Kotecha Director, IT-NU

Place : Ahmedabad

Date:



STMicroelectronics India Pvt Ltd.

Plot No-1, Knowledge Park-III Greater Noida -201308 Uttar Pradesh, India

This to certify that Mr. Dhonde Anand Bharatkumar (13MECV06), student of M.Tech EC (VLSI Design), Institute of Technology, Nirma University is undergoing internship in our Design Methodology team under SMEM Group from 1st July 2014. He has successfully completed his project entitled "Study of Electro-migration (EM) on Full Custom VLSI SRAM Designs and Development of an EM Assessment Methodology for Memory Compilers".

Signature:

Date & Place:

Declaration

This is to declare that

- a. The thesis comprises my original work towards the degree of Master of Technology in VLSI Design at Nirma University and has not been submitted elsewhere for a degree.
- b. Due acknowledgment has been made in the text to all other material used.

- Anand Dhonde 13MECV06

Acknowledgements

I am overwhelmed with a joy and willing to express my wholehearted thanks and hope that I would have gratified you all by means of my project report.

First of all, I would like to express my profound gratitude to my guide, Dr. N. M. Devashrayee, P.G. Coordinator of VLSI Design, Institute of Technology, Nirma University, Ahmedabad, for his guidance and support during my work.

I would also like to forward my thanking tribute to Mrs. Shafquat Ahmed (Team Head) and Mr. Promodkumar (Manager) for giving me an opportunity to work with them. I would like to thank the whole back end team & SMEM team for their help during my work in ST Microelectronics. Throughout the training, they all have given me much valuable time & advices on my project work which I am very lucky to benefit from. Without them, this project work would never have been completed.

- Anand Dhonde 13MECV06

Abstract

With each progressing technology node, the devices are scaled down to improve the performance and can be densely packed in comparatively smaller area. The active devices are getting better due to scaling but at the same time the passive components like interconnects are getting worse. That is because as the number of devices are increasing in smaller areas of ICs, the number of metal interconnects, their length and their stacks are increasing resulting in more complex and congested routing. Hence the problems like electro migration (EM), IR drop and crosstalk are magnified for the technology node below 0.1 um which were not that much significant for the technology above the 0.1 um. The IR drop is causing the voltage to drop significantly in the interconnects itself such that the devices will not have enough supply voltage to function correctly hence causing the functional failure instantly while the EM may cause the interconnects to short or open eventually after the IC is functional for sufficient long time. So the EM is issue regarding the reliability of the whole IC, because the failure of one of interconnects may cause the whole IC to fail. In today's era, nearly 70% of the chip area is occupied by the memory itself. So it is very important that the designer have the knowledge of the severity of the EM on the memories when designing the IC and it is therefore recommended to consider the electro migration aware physical design at the design stage itself. In this project the objective is to develop the methodology to predict the worst case EM violation of the power nets in any given memory instance of the specific memory compiler. The purpose of reporting the worst case EM violation is that this is the highest EM violation and all the other violations will be less severe than this one on the memory instance. The designer need to solve for the worst violation first before going to the next EM violation. The phenomenon of electro migration and the factors affecting the same is studied first. Then the tools required for the EM analysis is understood and used to generate the power EM violation data as the benchmark data. I have selected the single port low leakage type of SRAM memory compiler in 28nm FD-SOI technology. Finally, power EM assessment methodology is proposed to estimate EM violation in a memory instance based on its physical characteristics.

Abbreviations

IC	Integrated Circuit		
ASIC	Application Specific Integrated Circuit		
\mathbf{EM}	Electro Migration		
HDL	Hardware Description Language		
RTL	Register Transfer Logic		
\mathbf{CDL}	Cictuit Description Language		
DRC	Design Rule Chcek		
\mathbf{GDS}	Graphical Database System		
LVS	Layout Versus Schematic		
\mathbf{PVT}	Process Voltage Temperature		
SoC	System on Chip		
ROM	Read Only Memory		
\mathbf{SRAM}	Static Random Access Memory		
Memcell	Memory Cell		
\mathbf{SNM}	SNM Static Noise Margin		
I/O	Input Output		
N_B	Number of bits		
N_w	Number of words		

Contents

C	ertifi	cate	i
D	eclar	ation	iii
A	ckno	wledgements	iv
A	bstra	let	v
A	bbre	viations	vi
Li	st of	Figures	x
1	Intr	roduction	1
	1.1	Front End Design	1
	1.2	Back End Design	2
		1.2.1 Device folding	3
		1.2.2 Source Drain sharing	4
	1.3	Initial Ramp Up	8
2	Intr	roduction to Memories	11
	2.1	Classification of Memories	12
	2.2	SRAM Memory Architecture	12
	2.3	SRAM Cell Description	15
	2.4	Read Operation of SRAM cell	16
	2.5	Write Operation of SRAM cell	18
	2.6	SNM of SRAM cell	18

	2.7	Various SRAM Architectures	20
		2.7.1 Basic SRAM Architecture	20
		2.7.2 Split Core Architecture	21
		2.7.3 Bank Architecture	22
	2.8	Memory Compiler	23
3	Ele	ctro Migration Basics	24
	3.1	Factors affecting the EM	26
		3.1.1 Temperature	26
		3.1.2 Interconnect width \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots	27
		3.1.3 Interconnect length	27
		3.1.4 Input voltage & frequency	28
	3.2	Damage caused by EM	28
	3.3	Power & Signal EM	29
4	Too	ols Used For EM	30
	4.1	Virtuoso	30
	4.2	XA-RA	30
		4.2.1 Inputs	30
		4.2.2 XA-RA main setup	32
		4.2.3 Outputs	33
5	Pov	ver EM Assessment	34
	5.1	Need of EM assessment	34
	5.2	Methodology of power EM assessment	35
	5.3	Virtual supply assessment	35
	5.4	Problems resolved in the power EM assessment	36
6	Res	sults & Future Scope	38
7	Cor	nclusion	40

List of Figures

Back end design flow	2
Device folding in two parts	3
Device folded in two parts	4
Common source sharing of folded device	4
Source drain sharing in NAND gate	4
Scan chain schematic	8
Layout of scan chain	9
Simulation result of scan chain	10
Types of memories	12
Memory architecture	14
Voltage sense amplifier	15
6T SRAM cell	16
Setup for SNM measurement	19
SRAM single core memory	20
SRAM split core memory	21
pi network representation	22
SRAM Bank architecture	23
Effect of technology on interconnects	24
Electro migration induced voids & hillocks $\ldots \ldots \ldots \ldots \ldots \ldots$	25
Electro migration dependency on temperature	26
Power and signal EM	29
	Back end design now

LIST OF FIGURES

4.1	Setup of XA-RA for EM analysis	31
4.2	Output of XA-RA	33
$5.1 \\ 5.2$	Memory cuts needed for the assessment	35 36
6.1	Power EM assessment difference	38

Chapter 1

Introduction

Application Specific Integrated Circuit (ASIC) is the heart of most of the electronic devices around us. ASIC unlike a general purpose IC, are designed to serve a certain predefined functionality only. As ASIC is made for one particular purpose only, it has designed by much optimization in terms of area, power, performance, etc. In the process of ASIC design, the standard cell libraries are used as the basic elements. These standard cell libraries in turns are having the basic design elements like gate, MUX, flip flop etc., which are optimized in terms of all means. Using these libraries the ASIC design will be carried out in two major parts: Front end or logical design and back end or physical design.

1.1 Front End Design

In the front end design, the logical design from the specifications will be developed. The Hardware Description Language (HDL) will be used to make the logical model of the design which is called the (Register Transfer Level) RTL code of the design. The device is assumed as black box and the functionality of the device will be coded in the RTL code in terms of the inputs, outputs and the relation between inputs and outputs. This RTL design is independent of the technology which we will going to use. From this RTL code, the synthesis will be carried out. Synthesis is responsible for converting the RTL description into a structural gate level based netlist and hence will be technology specific. From this gate level schematic, lower hierarchy in terms of the transistor level schematic will be developed according to the design technologies we are using. This is the circuit that would be supplied to the back end team for the physical design flow.

1.2 Back End Design



Figure 1.1: Back end design flow

The back end or physical design flow for ASIC is responsible for the conversion of transistor level schematic to the layout that can be fabricated as the IC. As shown in the figure above, the transistor level schematic already present from the front end designers, the Circuit Description Language (CDL) is exported from the Virtuoso Command Interpreter Window (CIW). This CDL is simply having the connectivity information of the active & passive devices, that means is CDL is a text form of representing the schematic. Using this CDL as the netlist, the SPICE netlist is made. In this SPICE netlist input stimuli to the designed circuit, the parameters to probe as output and some other simulation options are kept in addition to the original CDL netlist. This will help us to do the pre layout simulation. The ideal functionality of the design, without considering the parasitic information, will be checked by this simulation. From this simulation we can check whether the design is as per our requirements or not. If pre layout simulation is satisfactory then the next step is to draw the layout of the given schematic. Layout is the top view of the design that would go to actually fabricate on the IC. All the placement of active devices & their proper connectivity using metal layers are carried out in this layout drawing.

CHAPTER 1. INTRODUCTION

There are different techniques to save the area in the IC while drawing the layouts, which are crucial at the time of design of ASIC. Device folding & source drain sharing are some of the techniques.

1.2.1 Device folding

When we have to make the layouts of huge devices for providing large current (like word line drivers), the width of the devices should be very high compared to the other devices because as per the current equation given below, if we increase the width the current capability of the MOSFET increases.

$$I = \mu C_{ox} \frac{W}{L} \left[\left(V_{GS} - V_{th} \right) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

Large transistors of the layout can be split into smaller ones and then shorting the corresponding terminals such as to make the devices of required channel width and channel length as shown in figure 1.2. There are advantages of folding the transistors in multiple transistors. One significant advantage is: The poly resistance has broken into smaller multiple poly, so the voltage drop that happens at the gates of the devices will be less. Device gets proper supply at their respective gates & channel formation will be uniform along the width of the devices. The process variations will be less as same width of the devices will be carried. For the folding of the devices, the transistor with width X can be folded in two parts each with width X/2 & same channel length as of the original transistor as shown in the figure below. If the current originally was I, after folding also the current will remain I only. This is because the current from each transistor M1 and M2 will be I/2 as the width is half of the original one.



Figure 1.2: Device folding in two parts



Figure 1.3: Device folded in two parts

1.2.2 Source Drain sharing

After doing the folding, the source and/or drain sharing will be used to save significant diffusion area from the layout. For the above case of folding the transistors in two smaller width transistor, the common source regions of both transistors will be merged in one diffusion only as shown in figure 1.4 below. This will effectively save one source region area in the layout.



Figure 1.4: Common source sharing of folded device



Figure 1.5: Source drain sharing in NAND gate

Also consider the pull down network of a simple 2 input NAND gate as shown in the figure 1.5. We can see that the source of the M1 is connected to the drain of the M2. Here also we can share the diffusion drain and source of adjacent transistors because the source and drain terminals of the MOSFET are used interchangeably. As we are reducing the diffusion area it is obvious that the parasitic resistance and capacitance associated with the devices will also decrease. Using the device source and drain sharing concept, we can save the leakage that occurs from the devices. The overall aspect ratio for the blocks can be adjusted as we are adjusting the transistors' widths.

The Design Rule Check (DRC) is the step that needs to be checked in parallel while the layout drawing is in progress. The DRC is to take care whether the layout is actually possible to fabricate using particular technology or not. The DRC rules are already given by the foundry. If the design of layout is DRC cleaned, it is most likely to be fabricated as it is as the layout looks. After the DRC is cleaned, the Graphical Design System (GDS) is exported from the CIW of the Virtuoso. The GDS, as the name suggest, is removed nothing but a database format of the design which represents the layout itself in the binary file format. This file is used by the foundries to fabricate the actual silicon chips. The GDS contains the information of the layout like text labels and geometric shapes including the coordinates in hierarchical form of the design. This file can be used to regenerate whole design which may be used in sharing layouts, transferring artwork between different tools, or to directly create the mask for the fabrication of the layouts.

Now the design needs to be Layout Versus Schematic (LVS) clean, which needs the GDS & CDL as inputs & it is obvious as LVS need to do the comparison between the layout & schematic. The need of this check is to verify whether the drawn layout is matching to the schematic or not. The LVS will match all the ports, devices, nets, connectivity, etc. from schematic to the layout.

If the design is LVS clean, then the next step is to do the parasitic extraction of the layout drawn. Parasitic extraction is the step to calculate the parasitic effects in both the designed devices and the metal interconnects of an electronic circuit. Detailed device parameters, parasitic capacitances and parasitic resistances are called parasitic. The reason behind the extraction is to know how much will be the effect of the parasitic once the IC is fabricated in actual circumstances. There are various formats of the extracted netlist, Detailed Standard Parasitic Format (DSPF) is the one among them. There are also different extraction methodologies depending on which factor we want to extract the parasitic.

CHAPTER 1. INTRODUCTION

Extraction Methodologies

Extraction methodology actually specifies the methodology the interconnect extractor will follow during the extraction process. There are various methodologies of the parasitic extraction depending on the users' requirements. Depending on the user selects, the extracted netlist will be having different extracted parasitic parameters, like resistance, capacitance or both of them. Here are a few methodologies of the extraction:

Mode C Extraction:

In this type of extraction, only the parasitic capacitance that exists between the net and the ground will be extracted. So each net will report only one capacitance.

Mode Cc Extraction:

In this extraction methodology, one net will report one normal capacitance with respect to ground & other coupling capacitances those are associated with the nets lying in the proximity of the original net.

Mode R Extraction:

This is the simplest kind of extraction methodology, in which only the parasitic resistances associated with all the nets will be reported in the extracted netlist.

Mode RC Extraction:

In this method, the resistance associated with the net as well as the capacitance that exist between net & ground will be reported in the extracted netlist.

Mode RCc Extraction:

In this type of extraction, the resistance associated with the nets, the capacitance with respect to ground & the coupling capacitances with respect to the other nets in the proximity of the original net will be reported. So we can say that this is the exhaustive and accurate extraction methodology which contains all the parasitic information. But in this case, the extracted netlist will be very huge.

In the post layout extraction flow, there is also provision of the threshold of the resistance & capacitance. The tool will skip the nets which have resistance and/or capacitance less than the threshold specified for the extraction.

EXTRACTION CORNERS:

The corners, or process, voltage, and temperature (PVT) conditions, are defined by the designer, and the circuit behavior is analyzed at these corners. Corners generally represent the worst and best case scenarios of the process variations and in turn attempt to simulate the worst case circuit performance or timing characteristics. These corners are an attempt to represent the maximum variation that is possible between any two die due to normal manufacturing tolerances. In technology 0.1 um and above, the timing path of any design is dominated by the cell delays only. That means RC of the cells are higher than the RC of the interconnects most of the times. But below 0.1 um technology node, the contribution of the interconnect delay in a timing path become significant and the coupling capacitance also alter the values of the timing path. There are 5 parasitic corners currently present:

- *Cbest:* It refers the minimum parasitic capacitance and hence represents smallest delay for paths.
- *Cworst:* It represents the maximum parasitic capacitance and hence presents the largest delay for paths.
- *RCbest:* This refers to the corner which minimize interconnect RC delay, that is the product of parasitic resistance and capacitance is minimum.
- *RCworst:* This is the corner which maximize interconnect RC delay.
- *Typical:* This will be the nominal value of interconnect resistance and capacitance.

Once the parasitic extraction is done, the extracted netlist (either DSPF or SPICE format) is included instead of the original CDL netlist in the SPICE file for the post layout simulation. The parasitic capacitances and resistances extracted according to how your layout is designed might be critical in affecting the actual performance of your design. In order to get an idea of how the design would work from your layout, you should perform a post-layout simulation from the extracted view. This simulation will be more accurate to the actual IC that would be fabricated.

1.3 Initial Ramp Up



Figure 1.6: Scan chain schematic

As the initial introduction to all required tools, I practiced the back end flow on the design of scan chain of one of the memory instances. The transistor level schematic is as shown in figure 1.6. This schematic is used to generate the CDL and to do pre layout simulation using eldo. From the transistor level schematic, the Cadence Virtuoso layout editor is used to draw the layout. The technology used is 28nm FD-SOI. The area optimized layout is made by using device folding and source/drain sharing as shown in figure 1.7. Also the DRC and LVS are cleaned on this layout using the Calibre tools. The parasitic extraction is done on RCc methodology and nominal (typical) extraction corner using the internal tool PLSkit and synopsys's tool starRCxt. This parasitic netlist is then used to do the post layout simulation of the layout I have made that of the scan chain.



Figure 1.7: Layout of scan chain



Figure 1.8: Simulation result of scan chain

Chapter 2

Introduction to Memories

A data storage element is to be expected to have the factors such as low cost, high speed performance, high packing density, low power dissipation, non-volatility, easy testability and reliability. Before the era of transistors, memory storage was based on the magnetic tape kind of technology where bits of digital data were stored on magnetic tapes. This kind of technology was having limitations in terms of cost, performance, reliability and speed. Hence the memories based on semiconductor devices and integrated circuits with their smaller sizes were responsible for a new era in data storage technology. Semiconductor memories are the electronic circuits which store digital information in large amount, hence are important modules in modern integrated circuits. These semiconductor memories also offered their ability for continued improvements in terms of speed and reliability by using smaller and smaller devices with each progressing technology node.[9]

In today's era of System on Chip (SoC), mainly the components are fabricated on a single chip or IC including the memories. Memories are very important part of the SOC design because they occupy around 70% of the total chip's area. The increasing demand for superlative performance from semiconductor circuits has driven the technology and hence the memory development towards more compact and complex design rules and consequently towards higher data storage densities, which is essential demand. Memories manufactured in ST Microelectronics are broadly classified into three main categories namely, Read Only Memory (ROM), Single Port Static Random Access Memory (SPSRAM) and Dual Port Static Random Access Memory (DPSRAM).

2.1 Classification of Memories

Semiconductor Memories are classified according to the the type of data access it has and type of data storage and as shown in figure 2.1. The Semiconductor Memories are majorly divided in two parts, Random Access Memory (RAM) and Read Only Memory (ROM). ROM circuits as the name implies allow only the access to the previouslystored data and don't allow changes in the stored information contents during its normal operation. ROMs are non-volatile memories, i.e., the data storage function are not lost even when the power supply is turned off. ROMs are again sub classified depending on the type of data storage methods, such as mask-programmed ROMs, Programmable ROMs (PROM), Erasable PROMs (EPROM), and Electrically Erasable PROMs (EEPROM). Random Access Memory circuits, unlike the ROMs, must permit the modification of data bits stored in the memory core, as well as their access when user require the contents. This kind of memories mostly are volatile, i.e., the stored data will be lost when the power supply voltage is turned off. Based on the operation type of individual data storage cells, RAMs are classified into two main categories: Static RAMs (SRAM) and Dynamic RAMs (DRAM). The reason behind the name RAM is: you can access any storage element of the RAM irrespective of the location of the same. Unlike the FIFO or LFSR, in which the access of the storage element will be in the series only.



Figure 2.1: Types of memories

2.2 SRAM Memory Architecture

The memory circuit is called static if the data stored on it can be retained till the sufficient power supply voltage is provided to the memory. That means unlike dynamic memories, static memories will not require a periodic refresh operation. The basic 1 bit data storage cell, that is called memcell, in static RAM arrays is made up of a simple latch circuit which in turn is designed using two simple inverters. This with two stable operating states. Depending on the stored state of the latch circuit, the data being held in the memory cell will be represented either as a logic "0" or as a logic "1". To access (read and write) the information stored in any one bit memory cell via the bit line, we need the switch at the end of cell, which will be controlled by the corresponding word line, i.e., the row selection signal.

The whole memory block may be broadly divided in two parts: main array and supporting circuits. The memory array is made with the repetitive use of memory cell (memcell), which is the basic storing element of the memory block. The supporting circuits to main memory array, also called as periphery, are designed for the operation of the memory array like reading and writing to required memcell(s).

General memory architecture is shown in figure 2.2. The whole memory block may be divided in 3 major parts as mentioned below:

1. Memory Array :

The memory array consists of individual memcells arranged in a fashion to have an array of horizontal rows and vertical columns. Each memcell will be kept such as to share one same connection with the other memcells in the same row. This common connection among the rows of the memory array is called the word lines. One another common connection of memcell with the other memcells in the same column is also existing in the memory array. This is known as the bit line. In $2^M \times 2^N$ capacity of memory array, there are 2^M rows called word lines and 2^N columns called bit lines. Each memcell is capable of storing one bit of digital information.

2. Row & column decoder:

If there is $2^M \times 2^N$ size of memory array, then there will be 2^M word lines to address each row and 2^N bit lines to access each column. Now to access a particular memcell, corresponding word line and bit line has to be enabled. To accomplish this task, M to 2^M row decoder and N to 2^N column decoder are used which will have M bits of row address and N bits of column address as inputs respectively.



Figure 2.2: Memory architecture

3. Sense amplifier:

As there are lots of devices connected to the bit lines, the capacitive load of bit line is very high. For reading of the memcell, if we allow to discharge the bit lines completely which were pre charged well before the reading operation, it takes too much time to discharge and again to pre charge due to high capacitance. This will limit the speed of the memory. We need to use some mechanism by which the pre charged bit lines can be discharged very fast. This mechanism is called sense amplifier. Its function is to detect the stored data from the selected memcell.



Figure 2.3: Voltage sense amplifier

A simple voltage sense amplifier is shown in figure 2.3. This kind of sense amplifier is used to detect the differential amount of voltage between two bit lines. The sense amplifier is turned on by asserting the sense signal high when there is sufficient amount of differential voltage on one of the bit lines by naturally allowing it to discharge. Suppose logic 0 is stored in the memcell. As the word line is asserted, the access transistor causes the bit line to discharge slightly. As there is enough difference in the bit & \overline{bit} line, the sense amplifier is turned on, the bit line now discharging through the sense amplifier & due to the positive feedback, the discharge rate is high in the case of sense amplifier.

2.3 SRAM Cell Description

SRAM stands for Static Random Access Memory. This type of memory can retain the stored data as long as power supply is available to the memory chip, unlike non-volatile memory where no constant power needs to be supplied for data retention. The name random access arises from the fact that in an array of SRAM cells any given cell can be read or written in any order, accessing a new cell doesn't depends on which cell was last accessed. The application of SRAM is when we require high speed because SRAM devices offer extremely fast access times but are more expensive to produce. Generally, SRAM is used for the cache memories where we need higher speed of data access. The basic static RAM cell consists of two cross-coupled inverters and two access transistors, commonly known as the 6 transistor (6T) memcell as shown in figure below, which is capable of storing 1 bit digital data.



Figure 2.4: 6T SRAM cell

The transistors which form cross coupled inverters, M1, M5 and M2, M6, will be acting as one bit storage element of SRAM. The design efforts are behind the area minimization of this 1 bit storage element because in the memory array, there will be repetition of number of such cells.

The gate terminals of two MOSFETs transistors M3 & M4 are connected to the word line at their respective gate terminals and the bit lines at their source or drain terminals. Hence these two transistors will be acting as the access transistors of the memcell because the access of the memcell to the outside world will only possible if only if the word line is active (i.e. M3 and M4 are on). The word line is used to select the cell while the bit lines are used to perform read or write operations on the memcell. Internally, the cell holds the stored value on one side and its complement on the other side (q and \bar{q} as shown in figure). The two complementary bit lines are used to improve speed and noise rejection properties of the memcell.

2.4 Read Operation of SRAM cell

When the word line is not activated, the cross coupled inverters are holding the data in their stable states. Suppose in above figure, logic 0 is stored at the q node and logic 1 is stored at the \bar{q} node in the memcell. Hence M1 is on and M2 is off.

Initially, bit and \overline{bit} lines are charged to a supply voltage by the help of bit lines pre charge pMOS transistors. After that the word line which was kept in deactivated state in the stand by mode of the memcell, is raised to supply voltage which will ensure that the access transistors M3 and M4 are turned on. Hence current will now start to flow through the MOSFETs M3 and M1 to ground. This current will be now responsible for slowly discharging the bit line which was already charged in the pre charge stage earlier. During the discharge of the bit line capacitance, the \overline{bit} line voltage remains unchanged since there is no path to ground through the MOSFET M2. The slight voltage difference between bit and \overline{bit} lines will be fed to sense amplifier such as to generate a valid low voltage output. This properly generated output from sense amplifier will be then stored in output data buffer.

The problem in read operation is that, as current flows from MOSFETs M3 and M1, it will raise the output voltage at node q as at point q, some finite amount of voltage will be developed due to the on state resistance of M1. If this voltage is enough high, this might turn on M2 and bring down the voltage at node \bar{q} . However the voltage at node \bar{q} may drop a little only but if it fell below threshold of the device (V_{TH}) , the memcell state will be flipped. To avoid flipping the state of the memcell when reading operation is carried out, the voltage at node q should be controlled by sizing M1 and M3 appropriately. This is accomplished by making the conductance of M1 about 3 to 4 times that of M3. This will in turn ensure that the amount of read current that charges node q will be lower than the amount of current that discharges the same node q such as the drain voltage of M1 does not rise above (V_{TH}) .

One of the other parameters while designing the read cycle is to provide sufficient cell current so as to discharge the bit line properly within 20% to 30% of the cycle time. The contradictory part over here is: The cell current is very low and the bit line capacitance is very huge. The voltage will drop very slowly at b. The rate of change of the bit line can be approximated as follows:

$$I_{cell} = C_{bit} \frac{dV}{dt}$$

Obviously I_{cell} controls the rate at which the bit and \overline{bit} lines discharges. If a rapid full swing discharge is desired, I_{cell} is made large. However, the transistors M1 and M3 would have to be larger. Since there are millions of such cells, the area and power of the memory would be correspondingly larger. Instead, a different approach is taken, attaching a sense amplifier to the bit lines to detect the small difference, V between bit and \overline{bit} and produce full-swing logic high or low value at the output. The trigger point relative to the rising edge of the word line, for the

enabling of the sense amplifier is chosen by based on the response characteristics of the amplifier.

2.5 Write Operation of SRAM cell

The operation of writing 0 or 1 is carried out by forcing one of the bit lines, either bit or \overline{bit} low while the other bit line remains at V_{DD} . For SRAM memcell taken above, to write logic 1, \overline{bit} is forced low, and to write logic 0, bit is forced low.

Consider the case where we need to write a 0 on node \bar{q} (i.e. on the *bit* line). The memcell must be designed such a way as the resistance of M4 is lower than M6. This should be ensured so that the drain of M2 is pulled down to logic 0 value. Consider the node q bar as the capacitive node which is initially charged to supply voltage before the write operation begins. When \bar{bit} bar line is pulled low to write a 0 at node \bar{q} bar, the current starts flowing from \bar{q} to bit bar line via M4. If the conductance of M4 is higher than that of M6 then only the node \bar{q} will start falling in terms of voltage. This in turn starts a feedback between two inverters connected in a latch fashion. Gradually, M1 turns off due to the discharge of node q bar. This will cause the voltage at node q to rises to V_{DD} . At the same time, M2 turns on and assists M4 in pulling output \bar{q} to the low value. When the cell fully flips to the new state, the word line will be returned to its logic low level in the standby mode.

2.6 SNM of SRAM cell

A key figure of merit for an SRAM cell is its static noise margin (SNM), because the stability of SRAM circuit depends on the SNM of the memcell. The SRAM cell immunity to static noise is measured in terms of SNM that quantifies the maximum amount of voltage noise that can be tolerated at the cross-inverters output nodes without flipping the cell.

The set up to measure the SNM of the given memcell during read operation is shown in figure 2.5. As in the case of the read operation, bit and \overline{bit} lines are set to supply voltage and then the word line is activated by pulling the voltage of the same to supply voltage. Two external power supplies, E and VX are kept at the internal node of the memcell. These voltage sources will now be acting as the internal DC noise sources. E is voltage dependent voltage source whose value is depending on the value of the voltage source Vx. The connection of these internal noise voltage sources are such that they support each other to flip the data stored in the memcell as speedy as possible to get the worst case SNM of the cell. The voltage VX is increased slowly and the voltage at which the memcell flips its state is called the SNM of that cell.



Figure 2.5: Setup for SNM measurement

In addition to the read current, leakage current and SNM; there are lots of other important parameters associated with the memcell. Data retention voltage is the minimum power supply voltage to retain the content stored on memcell as it is in the standby mode. Of course the data retention voltage should be greater than the threshold voltage. Write margin is defined as the minimum bit line voltage required to flip the state of an SRAM cell.

The smaller the write margin, the harder it is to write into the cell. Read Margin is represented as the minimum differential voltage needed to be developed between Bit and \overline{bit} to read the content of the memcell correctly. For Sense amplifier to pull bit and \overline{bit} line to correct logic level (V_{SS} for 0 and V_{DD} for 1), the minimum differential signal is needed to be developed which will be sensed by sense amplifier.

2.7 Various SRAM Architectures

2.7.1 Basic SRAM Architecture

A memory block is an arrangement of SRAM memory cells (memcell) in rows and columns (which forms the memory array or core), with decoder block, the input / output (I/O) blocks and the control circuit for the memory cell matrix. The figure below represents a very basic architecture of memory. To access the memcell, we need to select the memcell and this task is accomplished by the row decoder (rowdec) block. The horizontal lines in memory array represent the word lines which will be activated from the rowdec. The input / output (I/O) block is used for the read and write circuits. The vertical lines in memory array represent the bit lines. The memcell circuits are arranged to share connections in horizontal rows (word lines) and vertical columns (bit lines). The function of the control block is to generate the signals for the operation of the memory.

Rowdec	Memory Array
Control	Inuput/Output

Figure 2.6: SRAM single core memory

Memory cut is a single memory instance with a given configuration of the number of words (N_W) , number of bits (N_B) & mux size (m). The memory capacity is given by the product of N_W & N_B . The function of column mux is to maintain the aspect ratio of the memory. If we don't use the column mux & if the memory size is $2^M \times 2^N$ then there are 2^M rows & 2^N columns used. For the same $2^M \times 2^N$ capacity of memory, if we use column mux = 4 then the physical rows = $2^M \div 4$ & the physical columns = $2^N \times 4$. Now the memory block's height is divided by 4 & the width is multiplied by 4. User can select the required column mux size according to the size & capacity of memory cut required.

One of the parameters while designing the memory instances is the memcell leakage current. If there are lots of physical rows in a memory instances, all those except one row (which is actually activated for the read/write operation) will not be activated & hence there will be addition of huge leakage current in that case. Suppose we have 2048 words, 16 bits & MUX 4 case of memory instance. In this case, we are having 512 physical rows among which one will be activated at a time & 511 other rows will be contributing to the leakage current. If the cumulative leakage current of 511 number of rows are more than the read current of the memcell then the access transistor might turn on & the memcell will be accessed even if we haven't intended to access it. So there will be the upper limit to the number of physical rows from the maximum amount of leakage current that is generated. In the worst case, the maximum cumulative leakage current from all the memcell which are not activated should always be smaller than the read current of the memcell.

$$I_{leak} = \frac{I_{read}}{number of physical rows}$$

The problem in this basic architecture is increasing load on the word lines as the memory size is increasing. Also the word line load will be increased if we are using the column mux. Split core architecture gives solution to this problem.

2.7.2 Split Core Architecture

The problem with the basic architecture of memory is that when the memory size is increasing, the load of the word lines in terms of the resistance and capacitance also increases. The increased load will then reduce the speed of the operation of the memory as the worst case memcell which is located at the last part of the word line will violate the timing constraints. If by any means we can reduce the resistive and the capacitive load on the word lines, we can control the delay and hence more capacity of memory can be practically possible in single memory cut.

Memory Array	Rowdec	Memory Array	
Inuput/Output	Control	Inuput/Output	

Figure 2.7: SRAM split core memory

A method is proposed to just split the core in two parts and use the same word line driver in the center to drive the common word line which is running to both the core as shown in figure 2.7. We are driving the word lines from the center that is rowdec. So, we are providing two paths for the driver and each of the paths will see half of the resistive and capacitive loads as compared to the case in which there is only a single core. We can represent the load on a word line as shown in figure 2.8 (a) for single core memory, while figure 2.8 (b) represents the case of the split core memory. In case of a single core, the rising time will be governed by the time constant T=RC and T=RC/4 for split core at nodes B1 and B2[6].



Figure 2.8: pi network representation

It is obvious from the above figures and the value of the time constant T for both the case that though the points B1 or B2 and B start rising almost at the same time there is a time gain in both the nodes reaching the required voltage level of V_{DD} .

2.7.3 Bank Architecture

In this architecture, number of rows are divided into groups, where each group is called the bank. Each bank is locally controlled by the local control block. Multiplexer at the top level will be used to select one particular bank from the available banks. This architecture requires the bank address in addition to the column an row addresses. The purpose of this architecture is to break the bit lines into several smaller parts and increase the performance of the design. Of course this architecture is providing this benefits with the penalty of some area overhead in terms of multiplexing.

This technique to reduce the run length of the bit lines and divided core structure helps in gain in both of speed and power. Here the control and the Input / Output sections are divided. So for a selected word line, the cells of only one bank are activated. Also in case of bit line, the numbers of cells activated are reduced. Thus, a significant improvement is observed in case of word line cap and the bit line cap. There is also reduction in the power consumption to a very significant

Memory Array	Rowdec	Memory Array
Sense Amplifier	Local Control	Sense Amplifier
Memory Array	Rowdec	Memory Array
Inuput/Output	Global Control	Inuput/Output

value. But in this type of architecture, the area used is more and hence a less dense memory is obtained.

Figure 2.9: SRAM Bank architecture

2.8 Memory Compiler

Different designers may ask for various combinations of memory instance according to the configurations of word, bit, mux, memory capacity, etc. Also for various designs, different memory cuts with different aspect ratio & memory size are required. It is practically impossible to generate each & every cut when designers need the memories in their designs. So the idea is to have basic building blocks called leaf cells and then have one kind of setup to create the memory instances according to the user requirements. This kind of setup is called compiler, which will assemble the leaf cells in a memory specific to the designer requirements. The compiler will create the required configuration of memory cut by repetitively using the leaf cells according to the need.

There are different compilers available for different kind of SRAM like single port, dual port, high density, low leakage, etc. Each compiler specifies a range (min - max) of primary parameters like words, bits, mux options and secondary parameters like bank, redundancy, embedded switch, etc. the compiler will generate various kind of outputs like CDL, GDS, etc. for back end, Verilog, tetramax, etc. for HDL modules and interface, Apache, datasheet, etc. for the front end.

Chapter 3

Electro Migration Basics

Modern ICs are extraordinarily complex, containing millions of transistors. As the technology progresses, the performance of the VLSI chips is increasing. The devices and hence the area of ICs are shrinking. But interconnect performance is becoming worse. As the spacing and width of interconnects are decreasing, the problems like electro migration (EM), IR drop and cross talk are becoming more prominent which were not that much of concern for the higher technology nodes. Figure 3.1 (a) represents that the metal interconnect widths are decreasing exponentially. As a result, the overall cross-sectional area of interconnect is shrinking. Figure 3.1 (b) shows that with increasing integration of devices on single chip, total interconnect length is exploding as well. This means that there are more wires on a single chip which will be prone to EM effects. Currents are not scaled proportionally to the shrinking wire widths and, hence figure 3.1 (c) represents that the current density increased drastically as the technology is progressed[3].



Figure 3.1: Effect of technology on interconnects

In order to improve and insure the reliability of integrated circuits, the behavior and failure mechanisms of the constituent materials must be understood. The mechanical stresses present in thin metal films can be substantially higher than that in the bulk metals. Consequently, interconnect reliability has become a field of study. One of the major concerns for interconnect reliability is electro migration. Electro migration is the process of transport of interconnect metal material caused by the gradual movement of the atoms in the conductor due to the momentum transfer between conducting electrons and diffusing metal atoms. Electro migration is a gradual process in which the metal interconnects experiences the atoms transfer from one place to other. Electro migration will generate voids in interconnect from where the metal atoms are depleted and hillocks in interconnect where the metal atoms are accumulated as shown in figure 3.2. The metal interconnect may eventually become open due to generation of voids or interconnect may get shorted to the other metal layer due to creation of hillocks. So electro migration will cause the functional or timing failure to the design hence limiting the reliability of the integrated circuits (ICs).



Figure 3.2: Electro migration induced voids & hillocks

Semiconductors will be having very less electrons and in an intrinsic semiconductor, electro migration does not exist because there just aren't enough charge carriers. However, electro migration can occur in materials when they are so heavily doped such as to work as metals. Due to the operation of the ICs, interconnect may get damaged due to electro migration. The Mean Time To Failure of metal interconnects due to the electro migration is given as[5]:

$$MTTF = AJ^{-n}e^{\left(\frac{E_a}{KT}\right)}$$

Where, MTTF = Mean Time to Failure (Hours) J = Current Density (Amps/cm²) T = Temperature(K)n = Current density exponent Ea = ActivationEnergy, eV $K = Boltzmann'sconstant, 8.62e^{-5}eV/K, and$ A = aconstant

Once the mean life is determined from the test, the currant carrying capacity for a certain operating life is estimated using Black Equation's parameters. The electro migration depends on factors like, supply voltage, load capacitance the metal sees, the temperature at which IC is operating.

3.1 Factors affecting the EM

3.1.1 Temperature

The current permitted in a thin film conductor is a function of temperature. The higher the temperature, the less current is allowed in the conductor. For higher temperature, the random thermal vibrations of metal atoms will be increased, resulting in more collisions of electrons with the metal atoms. Hence the lower current is allowed in interconnects. This can be easily understood by simple current- resistance relationship that is V=IR. The resistance of interconnects are having the positive temperature coefficient. As the temperature increases, the resistance of interconnect can handle to prevent the electro migration.[4]

But the devices in the memory are supplying almost constant current, which exceed the limits of interconnects. The metal interconnects hence undergoes more electro migration at higher temperature compared to lower temperature.



Figure 3.3: Electro migration dependency on temperature

The failures of the metal interconnect lines because of the formation of voids might be occurring due to one other dependence on temperature observed. The figure 2.3 above shows a loop that eventually ends in failures of the metal interconnect lines. Once a void was formed in a metal interconnect line in the IC, interconnect will become narrower at that point compared to earlier case when the void was not formed. Now due to the reduction in the interconnect width, the local current density in that portion of interconnect will increases and as a result, interconnect temperature will be increasing due to Joule heating phenomena. As the temperature of the interconnect increases due to the effect of Joule heating, the growth of the void accelerates, and gradually an open circuit of interconnect may occurs.

3.1.2 Interconnect width

The electro migration is mainly dependent on the electric current density (J) through the metal interconnect. So, the geometry of the metal is important because if the same amount of current is flowing through two metal interconnects with different width, there is more electro migration in case of interconnect with low width & electro migration is less in interconnect with higher width. The current density is inversely proportional to the width of interconnect due to the relationship

$$J = \frac{I}{W.t}$$

As there will be a specified minimum width of the metal interconnects for each technology nodes, there will be a maximum limit of the current density to avoid the electro migration in each technology nodes.

3.1.3 Interconnect length

Blech discovered some important facts in his electro migration experiments. In these experiments, a current was passed through aluminum (Al) stripes of various lengths. The current causes the Al to migrate (drift) in the electron flow direction, generating an area of metal depletion near the negative end of the segment and accumulation near the positive end. Some of his observations were as follows[7]:

- Long metal segments showed a high drift velocity than shorter ones
- At a given current density, segments below a critical length did not drift at all.
- In a given segment length, no drift was observed below a critical current density.

This critical length of interconnect metal below which interconnect will not have electro migration is known as Blech length. The critical current density below which the metal material drift will not happen is called the current threshold of that interconnect.

One important point to be noted here is: designer may smartly take advantage of the Blech's effect. If some long running metal is having higher EM violation, he/she may divide the metal in smaller metals to run to upper or lower metal layers in multilayer metallization. So each segment may be kept well below the Blech length to avoid the electro migration.

3.1.4 Input voltage & frequency

DC current:

$$I = CVf$$

AC current:

$$I = CV \sqrt{\frac{2f}{t_s}}$$

As above relations suggests, the current carrying capacity of interconnect is directly proportional to the voltage, capacitance & frequency. The worst case electro migration can be estimated by the maximum voltage & the maximum frequency of operation of the memory cut.

3.2 Damage caused by EM

If we keep a constant electro migration in the metal interconnect lines of the ICs, it would not cause damage. In steady state operating conditions the only damage caused because of the electro migration should be observed at the beginning and end of the metal interconnect lines and not anywhere else. This is because in the metal interconnect line; whatever be the number of atoms those arrive in a given local volume, the same number of atoms will leave the same volume. Damage to the metal interconnect lines is caused by difference in atomic flux. When the amounts of matter entering and leaving a given volume are unequal, the relevant accumulation or loss of metal material will be resulting in damage. Whenever the number of metal atoms entering into a region is greater than the number of metal atoms those leave the same region of metal interconnect line, the metal material will be gathered over there in the form of a hillocks. And if the number of metal atoms those are leaving the region are greater than the number of metal atoms those are entering the same region, the depletion of metal material will be formed which normally is called voids.

3.3 Power & Signal EM

In the memories there will be lots of devices whose output will be driving the input of the following stages. In all these cases, all the pMOS devices will conduct the current from power supply to the load capacitance for driving the output to V_{DD} as shown in figure 3.4 (a). The nMOS devices will conduct the current from load capacitance to ground for driving the output to ground as shown in figure 3.4 (b). Here the metal interconnects connected to the pMOS and nMOS are having the unidirectional current from power supply and to ground respectively. So these nets are called the power nets. The interconnect connected with the load capacitance will have the bidirectional current due to charging and discharging of the load capacitance as shown in figure 3.4 (c). This is called the signal nets. The electro migration will be more in the power nets which are carrying unidirectional current. The obvious reason for more EM violation in power net is the current is flowing in only one direction and the metal atoms will be drifted to one end of the net only. While the signal nets which carry bidirectional current will be less affected to electro migration as some times the metal atoms will shift in one direction and some times in opposite direction.



Figure 3.4: Power and signal EM

Chapter 4

Tools Used For EM

There are various tools used to analyze the EM on memory compilers. The inputs, outputs and all the setup of the same needed to be studied. This chapter gives a brief overview of the tools used for the power EM assessment methodology.

4.1 Virtuoso

Virtuoso is the tool from Cadence which comes with utilities for designing fullcustom integrated circuits. It includes schematic entry, behavioral modeling (Verilog), circuit simulation, custom layout and extraction. It is used mainly for analog, mixedsignal, RF, standard-cell designs, memory and FPGA designs. We have used the Virtuoso to view the layout & to see the results of EM outputs.

4.2 XA-RA

The electro migration is analyzed using the tool XA-RA which is from the Synopsys. It takes 5 inputs from the user as shown in below figure 4.1

4.2.1 Inputs

1. GDS & CDL of the design

The input GDS & CDL of the input design are kept in respective directory of EM run. We can have number of designs kept in these directories on which we want to run the EM. According to the job limits set from the path file, the jobs will get executed.



Figure 4.1: Setup of XA-RA for EM analysis

2. Cell list

In the cell list, we need to specify on which cut(s) we need to run the EM. The GDS name, top cell of GDS, CDL name, top cell name of CDL, the type of EM (signal or power), the extraction corner, number of mux, number of bits, operating frequency (in nanosecond), etc. In the type of EM, we may provide any one from power or signal depending on the type of EM we want to analyze. We are analyzing the power EM, hence kept EM type to pwrgnd.

3. Common input file

This is one common file which will be applied on all the cells on which we need to run the EM. The common inputs like temperature, compiler type (single port, dual port, ROM, etc.), voltage.

4. Models

The models corresponding to the devices used in the input design should be given in the input so that the simulator of XA-RA will get the information like device characteristics.

5. Path file

The path file has whole paths for all the input files from where the XA-RA will pick up the required files. The paths of input files like GDS, CDL, cell list, stimuli, process and common input file are specified in this file. There are also options given for the XA & extraction job limits, by which we can set the maximum job limits the jobs launched simultaneously.

4.2.2 XA-RA main setup

XA-RA is the setup from Synopsys to analyze the EM/IR on the design. It operates in 3 stages:

The XA-RA tool first calls the wrapper PLSkit for the extraction of the input layout given to it. PLSkit will call Calibre to launch the Layout Versus Schematic (LVS) check. After design is LVS clean, the star-rext is called for the extraction of the input gds. All the parasitic information about the layout in terms of resistance & capacitance are reported in the extracted file that is called Detailed Standard Parasitic Format (DSPF). All the power nets (gndm, vddma & vddmp) present in layout is divided in small segments as the EM is for only power nets.

To run the EM, the generation of stimuli for application on the memory cuts which in turns are specified in the cell list. The stimuli will be generated according to the inputs given by the user.

Now the DSPF file & the stimuli are supplied to the XA-RA for the simulation purpose. All the rules regarding how much current can be allowed in the metal segment depending on the width & length of that segment have been already coded from the information given in the design rule manual. From which it can calculate the threshold limit of current which will not cause electro migration in that particular segment. The actual current flowing through the segment is calculated by the tool using the simulation depending on the PVT given. The electro migration is simply then given by :

violation of the net=Actual current flowing though the net ÷Threshold current limit for that net

If the EM violation value is greater than 1, there will be more current flowing through the segment of metal than the threshold of current. So it will have electro migration as gradually the chip is kept functioning.

4.2.3 Outputs

The XA-RA will generate various output files which will help us to analyze the EM or IR. One of the output files contains the data for each nets, the fields like actual current through net, threshold current, EM violation, net width, net length and EM length of that net is reported. The EM violations are reported in the descending order, so that we can find which is the worst segment for EM violation.

In the second output file of XA-RA, we can map particular segment(s) in the corresponding layout. Calibre is used to load this file to map less number of segments according to the EM violations. The EM violations are reported in the interval of 10% in the calibre window.

If we need to study more number of segments, we need to use third output file from XA-RA. This file need to be loaded in the Cadence Virtuoso to see the results of EM in the form of colour map kind of thing. It gives the pictorial view of the EM violation on the given layout. In the figure shown below, the portions where the segments are highlighted in green color have the least EM violation. As the color changes to yellow, brown and eventually to red, the EM violations gradually increases. The segments blinking with red color have highest EM violations greater than 100%. We can analyze the amount of current flow & location from where also from this file by instantiating the original gds of the design. Because EM is the result of the current, so lots of information can be found from this output.



Figure 4.2: Output of XA-RA

Chapter 5

Power EM Assessment

In memories, there are long running metal interconnects for the purpose of word lines, bit lines, power sources, etc. This metal interconnects see high capacitive load as they are connected through various devices. As the load is high, the probability of occurrence of electro migration is more in case of memories compared to the other ASIC designs. The power nets are spread as the mesh network in the memories. Due to the unidirectional current flow, the power EM is more dangerous than the signal EM.

Exhaustive electro migration analysis needs huge resources like time, cost & memory space due to large database generated, thereby making it difficult to check EM Compliance for every memory instance that can be generated from a memory compiler.

5.1 Need of EM assessment

The memory cuts are generated from the memory compiler & we don't know which cut will be generated & used by the designer. The designer will not check the EM violations on each & every cut he/she want to use as this will cause wastage of resources like time, cost & memory space due to large data generated. There need to be some methodology by which we can provide the EM violations for the memory cuts which designer want to use. The method to predict the EM violations on any memory cuts using the benchmark cuts' data already available is called the assessment. We have the actual EM violations data of some benchmark cuts & we are predicting the EM violation data of any cuts from these benchmark cuts.

There need to be some benchmark data on which we actually ran the simulations of power EM violation. The cut space of the memory compiler is shown in figure below. The cuts from the memory space will be selected randomly and the trend of the power EM need to be found from these selected memory cuts only. The circles in the following figure represent the memory cuts used for the power EM assessment.



Figure 5.1: Memory cuts needed for the assessment

5.2 Methodology of power EM assessment

Electro migration violation is directly proportional to current density and capacitance of interconnects. So electro migration violation of any current carrying conductor will depend on capacitive load it sees and on the geometry of the memory cut. The parasitic capacitances associated with the nets are directly proportional with the geometry of the memory cut because as the number of words or bits varies, the capacitance of the respective associated nets will also varies. Hence EM violation should vary with physical dimension of memory cut that is with respect to number of words and bits. The important step in this methodology is to establish the relationship of the power EM violations with respect to the words and bits for particular memory compiler. This relationship is established in the main script that works on the output data of the wrapper script, which is working on the actual output of the XA-RA for the benchmark cuts on which we ran the actual simulation of power EM analysis.

5.3 Virtual supply assessment

In the memory cuts, the virtual or derived supplies are used with the aim to avoid the high instantaneous current when the huge drivers turn on and to reduce the leakage currents when the drivers are not activated. One of the cases of the virtual supply is shown below:



Figure 5.2: Virtual supply

These kind of virtual nets are actually power nets because they are the source of power to the other drivers connected to them. Some different kind of strategies are used for these kind of virtual nets. Using this method, the estimation is achieved within accuracy of 10%.

5.4 Problems resolved in the power EM assessment

- 1. First of all, if the actual simulation data on the benchmark cuts' needed to find, I need to understand the whole setup of the XA-RA. What are the inputs and what comes as the outputs of the setup itself is a big issue to be understood first. The outputs of the XA-RA is understood first: how to load them view properly the required entities.
- 2. Then comes the problem with the original setup of the XA-RA. The setup as it was provided can't be used due to the database it is generating is very less and improper with respect to the assessment methodology purposes. The original setup needed to be modified and tuned to have the proper generation of the report files.
- 3. The layouts of the desired cells can't be used as they are. There need to be some modifications done in terms of the labels on them. The modifications on the GDS was carried out before launching the analysis of the power EM.
- 4. For the methodology purpose we have to launch the power EM analysis on benchmark cuts, which are of the order of 20 to 30 cuts. It was not possible to analyze the data of power EM simulation. The script is made in this regard which will use and handle all the data generated by the XA-RA. This script itself includes the methodology of the assessment of the power EM too.

- 5. One of the issue also came up with the XA-RA tool accuracy. One of the fields of the data required in the report file for the assessment methodology is no accurately reported in the output files of the XA-RA. This is convinced to ST internal people and the solution is asked from the EDA vendor. The EDA vendor then provided some sort of plugins and options by which the desired accuracy in XA-RA is achieved and acceptable.
- 6. It is very difficult to use only 20 to 30 memory cuts and to predict the other cuts power EM violation data which are nearly around thousands of memory cuts. The methodology was not working in the unidirectional case earlier when the number of words are varying. The problem ultimately was taken care and the after that the methodology was working properly.

Chapter 6

Results & Future Scope

Having carried all the power EM assessment for selected design range of words and bits and after all the problems were taken care, the assessment results for the selected single port high density type SRAM memory compiler at 1.77 GHz, 1.2 V and 125° C temperature is shown in the figure below:



Figure 6.1: Power EM assessment difference

We can see from the above figure that the power EM assessment methodology accuracy for the selected memory compiler is within 10% of the limit for all the individual blocks, that is the predicted values have the deviation of only 10% from the actual EM violation value. However this limit is in the worst case that is for the

highest electro migration violated net only. But it is natural that the designer needs to take care of the highest violated segment first and then the other segments.

Future Scope:

This methodology can be useful to apply on each and every memory compiler such that the designer will know the worst EM violation on the memory instance. One such kind of framework can be developed from this methodology which is designer friendly. The designer just need to select the compiler, the operating voltage, frequency, temperature, number of bits, number of words, etc. and from the benchmark EM violation data, the estimation of the worst case EM violation on the designer's memory instance will be carried out.

This will save the valuable time to actually simulate the power EM on every memory instance designer wants to use. This in turns prevent the need of higher disk space required to store the results of EM violation data and also avoid to purchase the licenses for the respective tools. Also the deviation of the worst case EM violation reported is only 10% from the actual values, so it is quite practical scenario.

Chapter 7

Conclusion

For the electro migration (EM) violation assessment of the power nets in the memory instance, I took compiler for single port SRAM memcell. I found the actual EM violation data for the benchmark memory cuts. From establishing the relation of the EM violation on the number of words & number of bits, the power EM assessment is carried out. The simulation is carried out at 1.2 V supply voltage, 125° C temperature and 1.77 GHz frequency on the memory instances of selected memory compiler. The deviation of the predicted values of the EM violation to the actual EM violation values are observed below the 10% range in the highest violated net for both the power as well as virtual supply nets. Using this methodology, the resources like time, disk space & EDA licenses can be saved.

Bibliography

- [1] Design Rule Manual of 28 FDSOI technology, ST Microelectronics.
- [2] Single port high density type SRAM memory compiler reference manual
- [3] Bradley Geden, "Understand and Avoid Electro migration (EM) IR-drop in Custom IP Blocks"
- [4] Electro migration For Designers, white paper from Cadence.
- [5] Ahmer Syed, "Factors Affecting Electromigration and Current Carrying Capacity of IC Interconnects"
- [6] Shobha Singh, Shamsi Azmi, Nutan Agrawal, Penaka Phani Ansuman Rout, "Architecture and Design of a High Performance SRAM for SOC Design"
- [7] Christian Witt, "Electromigration in Bamboo Aluminum Interconnects"
- [8] http://www.eetimes.com/document.asp?doc_id=1275855
- [9] http://scholar.lib.vt.edu/theses/available/etd-72198-162528/unrestricted/body.PDF