Static Random Access Memory - SRAM

Dr. Lynn Fuller

Webpage: http://people.rit.edu/lffeee
Microelectronic Engineering
Rochester Institute of Technology
82 Lomb Memorial Drive
Rochester, NY 14623-5604

Email: Lynn.Fuller@rit.edu
Department webpage: http://www.rit.edu/kgcoe/microelectronic/
This PowerPoint module has been published using Adobe Presenter. Please click on the Notes tab in the left panel to read the instructors comments for each slide. Manually advance the slide by clicking on the play arrow or pressing the page down key.
Introduction
Memory Organization
Bitline Capacitance
Column Pullups
Column Select
Write/Read Operation
Write/Read Circuitry
Sense Amplifiers
References
Homework
Read-Only Memories (ROMs) - These are used to store information that will not change during the life of the system. They are permanently programmed during manufacture.

Nonvolatile read-write Memories (EPROM, EEPROM) - These devices retain the information stored in them when the power is turned off. They can be erased but usually much slower than they can be written. The number of erase/write cycles may be limited.

Dynamic Random Access Memories (DRAMs) - Information is stored as charge on a capacitor. The stored charge will eventually leak away so DRAMs must be periodically refreshed. Typically DRAMs are refreshed every 5-50 milli seconds. One transistor one capacitor per cell.

Static Random Access Memories (SRAM) - These devices store information in two cross-coupled inverters. Such a memory does not need to be refreshed. CMOS SRAM is low power. The SRAM cell requires six transistors making it fewer bits per chip than DRAM.
INTRODUCTION

NAND Flash has smallest unit cell, then DRAM and SRAM.
Cell Layout size has been shrinking as technology improves
These pictures show transistor gates and active regions. The metal interconnect is above the transistors and is not shown.
FINFET SRAM REDUCES CELL SIZE
MEMORY ORGANIZATION

The peripheral circuits include multiplexers, decoders, sense amplifiers, column precharge, data buffers, Read-Write circuits, and more. These circuits have to work for millions of storage locations.
6 TRANSISTOR SRAM CELL
In both read and write operations both bitlines are pulled up to near VDD. The circuits used to precharge the bitlines depends on the type of sensing that is used in the read operation. Shown here are three different pullup circuits. A balance transistor between both bitlines ensures that both bitlines are at the same voltage after precharge. Once the precharge is completed the word lines goes high and one of the bitlines remains high while the other goes low. The millivolt difference is amplified by the sense amplifier. Current sense amplifiers require transistors that are always on as show in the two circuits on the right.
Using the following parameters we will estimate the capacitance of the word and bitlines. Assume we have a 16Kbit memory organized as 2048 words each of 8-bits. The wiring capacitance is 0.2fF/um of length. Assuming a length of 1mm we add 200fF for each Bitline and only 3.3fF for each Wordline. (these wordlines are only 8 cells long) (note: selected memory organization will change results significantly)

The two pass transistors in each cell that are connected to the wordline have W/L = 0.5um/0.1um for an area of 0.05um2 and using 50Å gate oxide thickness results in a capacitance of 0.345fF/transistor and 0.69fF/cell. For an 8-Bit word the wordline capacitance is 8 times 0.69fF plus the wiring capacitance which is 3.3fF for a total of 8.8fF

Wordline Capacitance = 8.8fF
Wordline Capacitance = 8.8fF (from previous page)

The bitlines are connected to the drain/source of one transistor per cell. In 100nm technology this results in transistor D/S area of $0.5\,\text{um} \times 0.35\,\text{um} = 0.175\,\text{um}^2$ and perimeter of $\sim 1.7\,\text{um}$. With doping levels $\sim 3\times 10^{17}\,\text{cm}^{-3}$ the width of the space charge layer is $0.08\,\text{um}$ giving junction capacitance of $0.45\,\text{fF/cell}$. For 2048 cells connected to the bitline the total capacitance per bitline is $0.922\,\text{pF}$ plus the wiring capacitance of $200\,\text{fF}$ for a total of $1.1\,\text{pF}$.

Bitline Capacitance = $1.1\,\text{pF}$
READ OPERATION

Cbit represents the Bit Line Capacitance
READ OPERATION

The precharge raises the bitline to near VDD volts. Once the wordline, WL is activated the bitlines are connected to the crosscoupled inverter storage cell through a pass transistor. If Q is high QB is low and vise versa. The bitline connected to the low side of the cell inverters will begin to discharge through the pass transistor and the NMOS transistor in the inverter to ground. The voltage on that bitline will decrease slowly while the other bitline will remain high. Once the voltage difference reaches a couple hundred millivolts the amplifier is saturated and the value is stored in a data latch. The wordline is set to zero and the bitlines are recharged high. The cell resets itself to its original state. The transistor sizes for read stability should be such that it does not disturb the original state stored in the cell. The current flows from the bitline through the pass transistor and the NMOS transistor in the inverter. If the NMOS has 3 or 4 times the conductance of the PMOS transistor in that inverter the voltage will remain close enough to low to not disturb the stored value.
READ CIRCUITRY

clk
pc
addr
data
WL
B, BB
QB
Q
COL
sense

Data out

VDD
PC
VDD
M7
M9
M8
WL
VDD
M5
M3
M4
QB
M6
B
Q
M1
M2
M10
Write Driver
M11
SENSE ENABLE
SENSE AMPLIFIER
OUT

Cbit

© November 18, 2016 Dr. Lynn Fuller
READ CIRCUITRY

The precharge raises the bitline to near VDD volts. Once the wordline, WL, is activated the SRAM cell attempts to drive the bitlines to high or low as appropriate. The large capacitance of the bitlines would require a long time to charge or discharge. However the sense amplifier needs only a few 100’s of millivolt difference between B and BB to amplify to obtain the appropriate output voltage. The time required is relatively short. The WL is returned to its standby state and the cross coupled inverters return the Q and QB nodes to the appropriate voltages representing the stored data.
DIFFERENTIAL VOLTAGE SENSE AMPLIFIER

- VDD
- Bitline-1
- Bias
- BL
- Output
- BB
COLUMN SELECTION

Precharge Electronics

Row Decoder

Column Decoder

WL

Column MUX

Sense Amplifiers

Write Enable

Sense Enable

R/W Control

R/W
For current flowing to the right (ie V1>V2) the PMOS transistor will be on if V1 is greater than the threshold voltage, the NMOS transistor will be on if V2 is <4 volts. If we are charging up a capacitor load at node 2, to 5 volts, initially current will flow through NMOS and PMOS but once V2 gets above 4 volts the NMOS will be off. If we are trying to charge up V2 to V1 = +1 volt the PMOS will never be on. A complementary situation occurs for current flow to the left. Single transistor switches can be used if we are sure the Vgs will be more than the threshold voltage for the specific circuit application. (or use larger voltages on the gates)
The operation of writing 0 or 1 is done by forcing one of the bitlines low while leaving the other high. The pass transistors must be wider (~2x) than the PMOS transistors in the inverters to be able to overpower the cell and pull the inverter output low enough to initiate a regenerative effect between the two inverters. Once the inverters have reached their new written state the wordline can be returned to its standby state (low).

For example if Q is 0 and you want to write a 1 then BB is connected to GND while B is left floating (high after precharge). The wordline turns on M5 and M6 and the output Q goes low and eventually Q goes high.
WRITE CIRCUITRY

clk
pc
addr
data
WL
COL
B, BB
QB
Q
WRITE CIRCUITRY

First the columns are precharged to VDD using M7, M8 and M9. Next, the address and data signals are set up and held stable and then the clock is applied. The address signals are converted into column select and wordline activation signals. The data and write signals are applied and then the wordline is enabled. Only one of the two bitlines will be connected to GND, the other remains high from the precharge operation. M13, M14 and M15 are sized to pull down the bitline in a specified time. Once the cell is written the wordline and column select lines return to their standby value.
SRAM CELL WITH PC, READ AND WRITE CIRCUITRY

C\textsubscript{bit} = 1 pF

Rochester Institute of Technology
Microelectronic Engineering

© November 18, 2016 Dr. Lynn Fuller
This is a schematic of the sense amplifier and waveforms for the SRAM Write operation.

L/W for Pass=2/8, NMOS=2/16, PMOS=2/4
Waveforms for the SRAM Write 1 operation.
Waveforms for the SRAM Write 0 operation.
the SRAM Read operation.
L/W for Pass=2/8, NMOS=2/16, PMOS=2/4
Waveforms for the SRAM READ operation.
SRAM LAYOUT SHOWING TWO LAYERS METAL

Active
Poly
CC
Metal 1
Via
Metal 2

Rochester Institute of Technology
Microelectronic Engineering
SRAM LAYOUT SHOWING TWO LAYERS METAL

Note:  PMOS $W < \text{NMOS } W$
L’s same for PMOS and NMOS

Cell layout can be copied and pasted to make array
FINFET SRAM REDUCES CELL SIZE

![SRAM Cell Diagram]

1 SRAM-Cell
0.128 \( \mu \text{m}^2 \)

HKMG-FinFET

HKMG

Gate

Fin

20nm

Gate
These pictures show transistor gates and active regions. The metal interconnect is above the transistors and is not shown.
LAYOUT AFFECTS LITHOGRAPHY
REFERENCES

3. Dr. Fuller’s Lecture Notes, http://people.rit.edu/lffeee
1. Use SPICE to illustrate the operation of the SRAM cell. Show write of 0 and 1, show read.

2. Investigate the speed of operation of the SRAM cell using SPICE.

3. Design the sense amplifier and a data latch to read the cell.
SRAM cell plus sense amplifier and data latch.

Zach Allen, 2016
SRAM cell plus sense amplifier and data latch.

Zach Allen, 2016