DESIGN AND FABRICATION OF FIVE MICRON NMOS SRAM

Robert Chizmadia
Senior Microelectronic Engineering Student
Rochester Institute of Technology

ABSTRACT

A 128 bit by one bit NMOS static RAM was designed following design rules for RIT's standard four layer NMOS process. Verification of working devices was done using the SPICE circuit simulator, but some concerns exist with this because of assumptions made in model parameters. Fabrication was an intended goal of this project, but time restraints allowed only masks to be made.

INTRODUCTION

High density Random Access Memory (RAM) devices are an integral part of any computer system. These devices can be built in two forms, static or dynamic. Dynamic RAM's (DRAM's) store data on a capacitor requiring refresh circuitry. To perform a read operation, this charge is removed giving a destructive read cycle. Static RAM's (SRAM's) store charge on the gates of cross coupled inverters as shown in Figure 1. Because of this feedback path, no refresh is necessary. When performing a read operation, the charge stored on a data line is discharged through one of the pull down transistors, while the other inverter continues to hold the state. Therefore, the read cycle is non-destructive.

Figure 1 shows a schematic of the base static memory cell. Transistors one and two are constantly saturated. Because of the small width to length ratio the current is low. This allows the pull down to sink more current when activated than the pull up can supply, giving a the low logic level. Transistors three and four are the pull down transistors, which when paired with transistors one and two respectively give the NMOS inverters. The drain of transistors one and three are connected to the gate of transistor four. When transistor three is turned on, it will pull node two to ground. Since node two is connected to the gate of transistor four, this will turn transistor four off, allowing node three to charge to Vdd minus the threshold voltage, or logic high. This will allow transistor three to turn on, fulfilling the original assumption. Therefore, the state is stored without any refresh being necessary. Data is placed in the cell by activating transistors five and six. The distinction

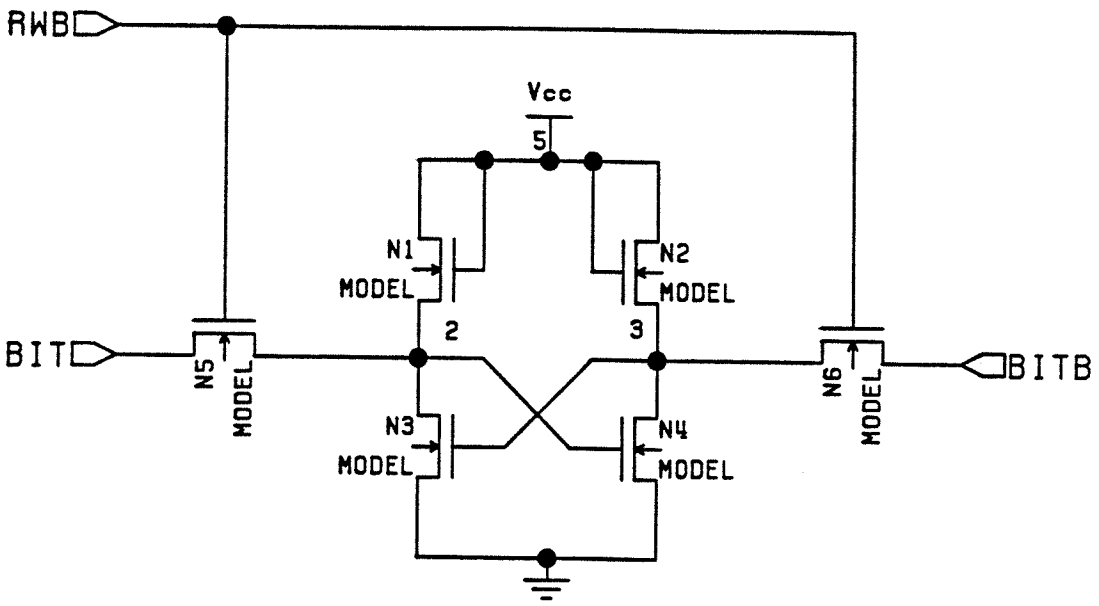

Figure 1: Transistor Schematic of Base Memory Cell

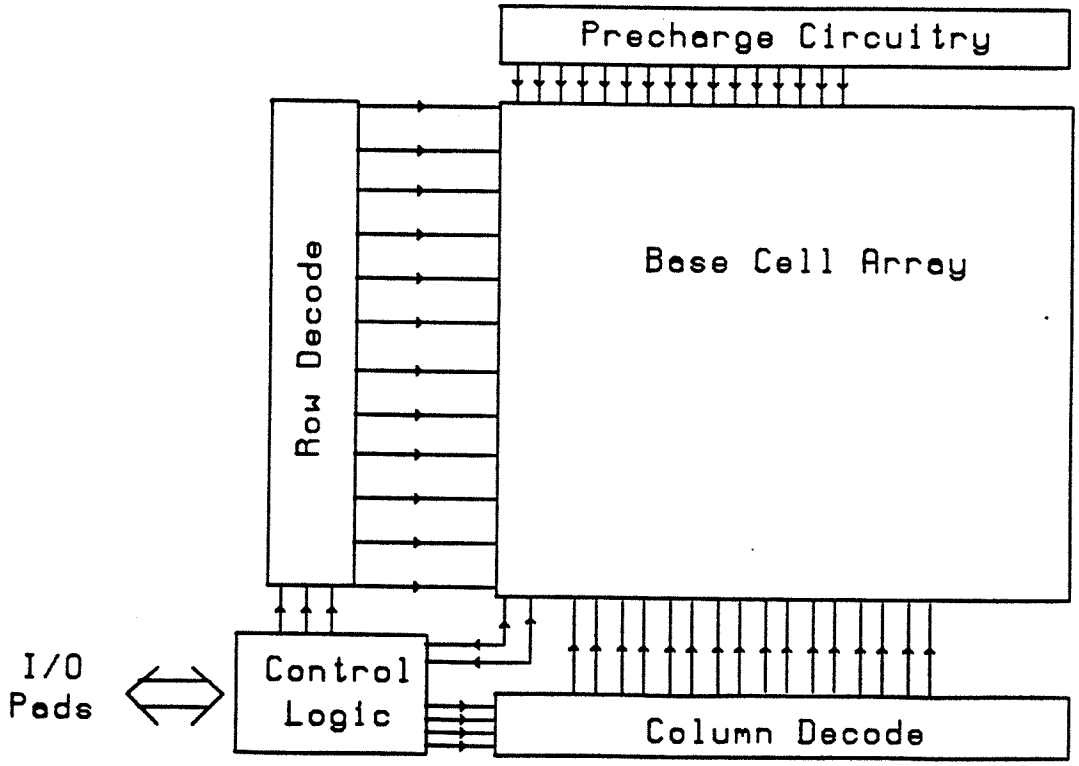

Figure 2: Block Diagram of SRAM Architecture

between read and write is dependent on the state of the **bit** and **bitb**, where **bitb** is the compliment to **bit**. For a write operation, the **bit** line is charge to the state to be stored while the **bitb** lines is charge to the compliment. Transistors five and six are then activated, allowing nodes two and three to be overridden with the new state. For a read operation, both **bit** and **bitb** are precharged to Vdd. When transistors five and six are turned on, one of the lines will be discharged while the other remains at Vdd. The state on the **bit** line is the state that was read. Simple NAND and NOR logic controls the precharging and the charging of the **bit** lines. Pass transistors trees were used for row and column decode.

DESIGN

This memory was designed using enhancement mode NMOS transistors only, reducing the number of lithography layers and fabrication steps. Poly was used for the gate material because the process was self aligning allowing for smaller gate lengths. Minimum width of poly was five microns, while all other spacing and width rules were ten microns. The architecture of the SRAM was 128 bit by one bit, requiring seven address lines, a read/writebar line, chip enable, in, out, power and ground. A block diagram of the design can be seen in Figure 2. Four layers were used, active, polysilicon, contacts, and metal.

The design was performed on Apollo workstations using Mentor Graphics layout tool Chipgraph. Precharging techniques, as described in the introduction section, were employed to reduce the size of the base cell. By using precharging, the drive capability of the transistors in the base cells is very small, allowing for small width to length ratios. Row and column select was accomplished by using pass transistor networks. Simulation of transistor sizes and connectivity was done using SPICE. The major concern with this was the models used which were untested and incomplete. The only parameters given were threshold voltage, which was set to one volt, and gate oxide thickness, which was set to 700 angstroms. It can be noted though than many other parameters are calculated from these in SPICE.

RESULTS/DISCUSSION

As this was a design project there are no distinct results as in an experiment, but some discussions can be made on problems encountered and future work that could be performed. The base cell, which contained the two inverters and two pass transistors, occupied an area of 170 microns by 270 microns. A layout of the base cell can be seen in Figure 3. This was placed in an array of 16 by 8. Adjoining cells were "mirror images" of each other, or in other words, one was flipped along it's vertical axis. This allowed power and ground lines

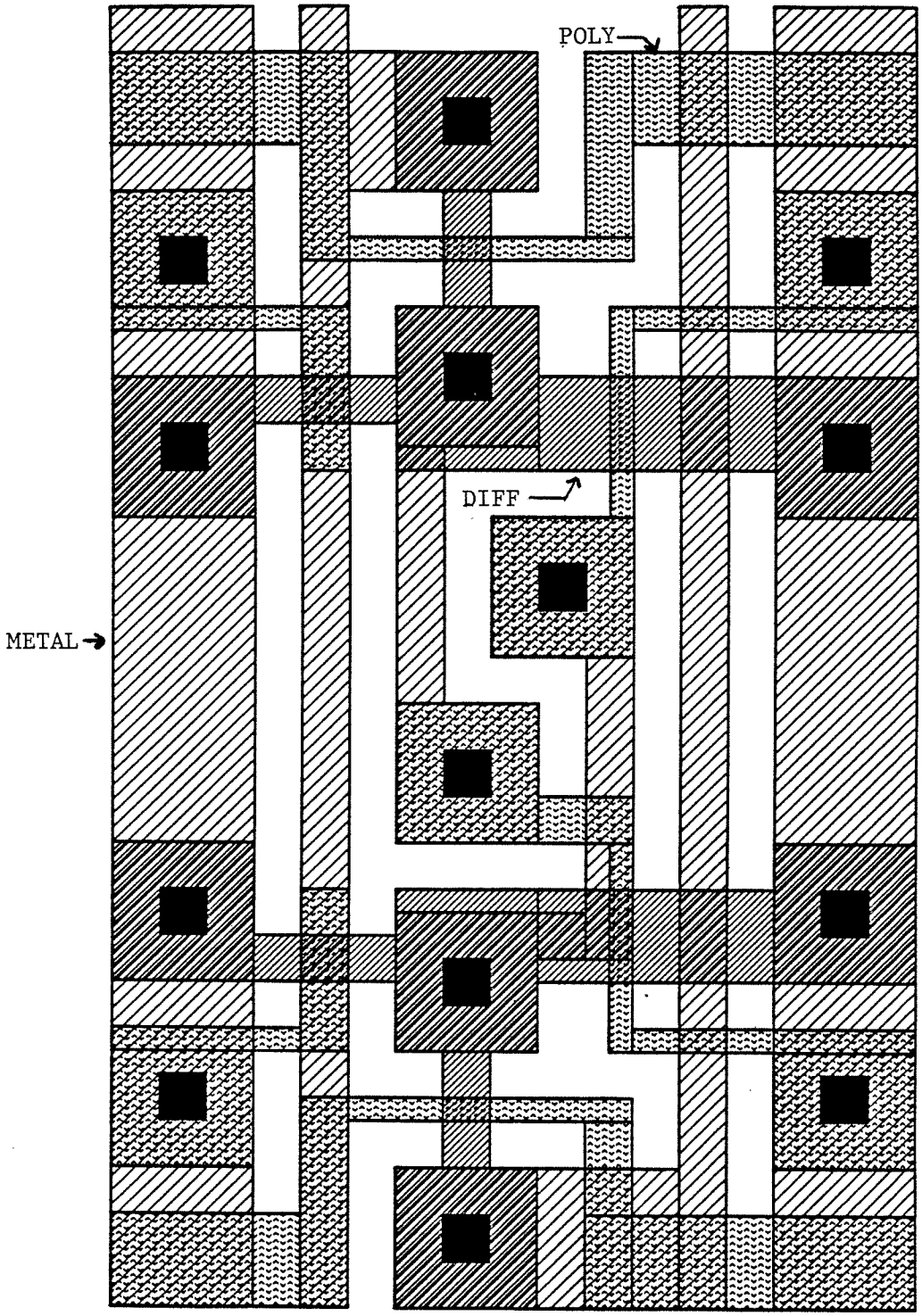

Figure 3: Layout of Base Memory Cell

within the array to be shared, saving 40 microns per every two cells, saving 320 microns allowing an extra two columns to be placed in the array. A large amount of time was spent laying out the base cell to allow each cell to connect to the other. This allowed a dense array because no routing lanes were needed in the internal array.

As stated in the Design section, the SPICE models used were incomplete, but circuit functionality could be verified as long as some care was taken not to believe all results as exact. The major problem concerning the SPICE simulations was a lack of true capacitance values for parasitic devices. This design relies heavily on capacitances due to the large number of pass transistors. In many situations where a pass transistors is turned off, the charge stored on the gate of another transistor is vital to maintain correct operation. The lack of parasitic capacitances in the SPICE models caused placement of capacitors in the input decks. The values for these capacitors was usually between 0.1 picofarad and 1.0 picofarad. Without any knowledge of test results from the process, these values seemed reasonable. It must also be remembered that timing was not a consideration during any of the design phase.

A fair amount of time was spent determining the exact method for testing the devices. A twelve pad structure was chosen even though the design required 13 inputs. It was determined that a additional pad could be added as long as it was much larger than an ordinary probe pad. Space was available so this option was used. This allowed testing of the device with available equipment at RIT.

Future work that could be performed on the device would be an update of the SPICE models allowing for more accurate simulation and a greater guarantee of working devices. Another way of determining if the design is functional would be to actually fabricate and test the devices. Contact, poly, and diffusion reticals have been made, as have poly and diffusion masks.

CONCLUSIONS

As stated previously, this was a design project, not an experiment. Since fabrication was not performed, it is difficult to determine if the design was functional. SPICE simulation showed working devices, but the models used for SPICE were not complete. Future work that could be performed would be the development of complete SPICE models for the NMOS process, and the fabrication and testing of devices. Some masks have been made helping this possible future work.