A LOW-POWER MEMORY CMOS INTEGRATED CIRCUIT FOR IMAGE SENSORS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MITHAT CEM BOREYDA ÜSTÜNDAĞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENG.

DECEMBER 2015

Approval of the thesis:

A LOW-POWER MEMORY CMOS INTEGRATED CIRCUIT FOR IMAGE SENSORS

submitted by **MITHAT CEM BOREYDA ÜSTÜNDAĞ** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Eng. Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Gönül Turhan Sayan _____
Head of Department, **Electrical and Electronics Eng.**

Prof. Dr. Tayfun Akın _____
Supervisor, **Electrical and Electronics Eng. METU**

Examining Committee Members:

Prof. Dr. Haluk Külah
Electrical and Electronics Engineering, METU _____

Prof. Dr. Tayfun Akın
Electrical and Electronics Engineering, METU _____

Assoc. Prof. Dr. Ali Ziya Alkar
Electrical and Electronics Engineering,Hacettepe Univ. _____

Assoc. Prof. Dr. Barış Bayram
Electrical and Electronics Engineering, METU _____

Assoc. Prof. Dr.Cüneyt Bazlamaçcı
Electrical and Electronics Engineering, METU _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MITHAT CEM BOREYDA ÜSTÜNDAĞ

Signature :

ABSTRACT

A LOW-POWER MEMORY CMOS INTEGRATED CIRCUIT FOR IMAGE SENSORS

Üstündağ, Mithat Cem Boreyda
M.S., Department of Electrical and Electronics Eng.
Supervisor : Prof. Dr. Tayfun Akın

December 2015, 65 pages

This thesis presents a low power SRAM block implemented in a $0.35\,\mu\mathrm{m}$ CMOS technology for imaging applications to be used inside a digital image processor ASIC (Application Specific Integrated Circuit). The SRAM structure is designed to be fast enough to store all the image data fed by a large format readout circuitry such as VGA (640x512), while requiring low power consumption. The low power consumption is a very critical requirement of such circuit, as the circuit will eventually be used in an embedded platform, which is generally battery operated. The circuitry is implemented with standard six transistor bitcells, write buffers, sense amplifiers, and a timing generator, while each sub-unit is designed very carefully to reduce the overall power consumption of the circuit. All interior signals are created by the timing generator, by asserting two control signals (enable and read/write) and a system clock. Sense amplifiers are selected to be current-type, which helps to improve both area and power consumption without reduced speed performance. The minimum achieved power consumption of the design is 1.28 mW for the read operation and 0.58 mW for the write operation. These numbers are comparable with the state of the art SRAM devices implemented in 14 nm CMOS node, if the capacitance values are scaled for a realistic comparison. This low power SRAM design also aims to be scalable and allows implementing from 4Kbit to 2Mbit storage areas, which correspond to array sizes of 16x256 to 8192x256. This scalable design also utilizes the local word line

assertion technique in addition to pulsed synchronous operation to reduce the power consumption of the SRAM further. The local word line assertion also enables multi-port operations with minimal additions to structure, increasing speed performance for imaging applications. In summary, the SRAM presented in this thesis not only satisfies the requirements spatial image processing of VGA image sensors at its maximum frequency for current setup with one-port, but also competes with the state of the art SRAMs in the literature in terms of power consumption.

Keywords: Low power, SRAM, image sensors, current sense amplifers, pulsed synchronous operation, local word line assertion, scalability

ÖZ

GÖRÜNTÜLEME SENSÖRLERİ İÇİN DÜŞÜK GÜÇ TÜKETEN CMOS HAFIZA ENTEGRE DEVRESİ

Üstündağ, Mithat Cem Boreyda
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Tayfun Akın

Aralık 2015 , 65 sayfa

Bu tezde görüntüleme uygulamalarında kullanılmak üzere $0.35\,\mu\mathrm{m}$ CMOS teknolojisinde tasarlanmış düşük güç tüketimli bir SRAM aygıtı anlatılmaktadır. Bu aygıtın öncelikli kullanımı bir sayısal görüntü işleme ASIC (uygulamaya özel tüm devre) içerisinde olacaktır. VGA çözünürlüğü gibi (640x512) büyük format okuma devresinden gelen veriyle beslenecek olup, bu verilerin tamamının kaydedilebilecek kadar hızlı bir yapı olması beklenirken, güç tüketiminin de az olması gerekmektedir. Az güç tüketimi devre için çok önemli bir gerekliliktir, nihai olarak, gömülü bir sistemde kullanılacak, büyük ihtimalle de gücünü bataryadan alacaktır. Devre standart altı transistörlü bit hücreleri, yazma tamponları, algı yükselticileri ve zamanlama üreteci alt devrelerini içerir. Bütün içsel işaretler zamanlama üreteci tarafından iki işaret (etkinleştime ve okuma/yazma) ve bir sistem saati işareti aracılığıyla üretmektedir. Algı yükselticileri, akım algılayan türdendir. Bu özelliği yükselticinin kapladığı yerin ve güç tüketiminin azalmasını sağlarken, hız performansını düşürmemektedir. Elde edilen minimum güç tüketim verileri yazma için 1.28 mW, okuma için ise 0.58 mW'dır. Eğer devrenin kapasite değerleri ölçeklendirilip hesaplanırsa en güncel 14 nm SRAM tasarımlarıyla karşılaştırılabilir enerji sonuçları elde edilir. Düşük güç tüketimli SRAM üretilecekken aynı zamanda kolayca yönetilebilecek bir ölçeklendirilebilirlik de hedeflenmiştir. Bu hedef sonucunda 4Kbit ila 2Mbit arasındaki boyutlarda, yani 16x256 ila 8192x256 arasındaki büyüklüklerde öbekler ile SRAM devreleri inşa edilebilir. Bu ölçeklendiri-

lebilir tasarım güç tüketimini düşürmek üzere atımlı eş zamanlı işlemler ve ek olarak da yerel kelime yolu seçimi tekniğini kullanarak SRAM devresinin güc tüketimini daha da düşürmektedir. Ek olarak yerel kelime yolu seçimi tekniği yardımıyla, çoklu giriş-çıkış özelliğine sahip olan bir devre oluşturmak, yapıya çok ufak miktarda bir ekleme yaparak mümkün olabilmektedir. Çoklu giriş-çıkış özelliği, görüntüleme uygulamalarının performansını artırmak için çok önemli bir imkandır. Sonuç olarak, bu tezde sunulan SRAM devresi mevcut durumunda, bir giriş ve çıkış portu barındıran halinde, en yüksek frekansında VGA görüntüleme sensörlerinin uzaysal görüntü işleme gerekliliklerini sağlamakla kalmaz, aynı zamanda güç tüketimi konusunda literatürdeki en gelişmiş SRAM devreleriyle yarışabilir durumdadır.

Anahtar Kelimeler: Düşük güç tüketimi, SRAM, görüntü sensörleri, akım temelli algı yükseltici, atımlı eşzamanlı işlemler, yerel kelime yolu seçimi, ölçeklendirilebilirlik

To my family

ACKNOWLEDGMENTS

I would like to express my gratitude towards my advisor Prof. Dr. Tayfun Akın for his supervision and guidance throughout this study. I would like to thank Dr. Selim Eminoğlu for his endless support and trust through my graduate life.

I would like to thank Mehmet Ali Gülden, Osman Samet İncedere, Nusret Bayhan and Murat Işıkhan for their technical guidance and sharing their knowledge and experience at every stage of this study.

I also would like to thank to my manager Umut Ekşi for his support during my thesis period and all of the colleagues at Mikro-Tasarım, Adnan Gündel, Suhip Tuncer Soyer, Cem Yalçın, Mustafa Karadayı, Serhat Koçak, Özge Turan, Ramazan Yıldırım, Alper Karakuzulu, Mert Can Ergün, Sercan Şahin, Aslı Doğan, Oğuz Kazan, Hakan Çeçen, Mehmet Akdemir, Murat Aktürk and Emine Aytekin for their invaluable friendship and support during this work.

I also would like to thank Mikro-Tasarım San. ve Tic. Ltd. Şti. for providing access to the IC design software, for fabricating the test chip implemented in this thesis and for supporting the test board fabrication.

I would like to thank my loving wife for all the emotional and technical support throughout the thesis period, and through my life.

Last but not the least, I would like to give my special thanks to my mother and my father for their endless trust, encouragement and patience in every moment of my life. I would like to thank my grandmother and grandfather for their endless love and interest on me throughout my life. Without their invaluable support, I would not achieve this success.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

In today's world, imaging covers a very important percentage of the people's life. It is the most powerful method to prove phenomenon or events. In a presentation, in a catalogue or in an online sales place, pictures and videos gave users trust. Social media is almost completely based on pictures and videos instead of written sources. Nearly every single person has their own personal imaging device. With the rapid evolution in the power electronics market, inexpensive imaging systems can fit into pockets of everyone. This evolution forced every manufacturer to create smaller and smaller devices, and also to achieve less power consumption for more performance. In an imaging system, every single piece of the system should consume less power and give better performance with every new generation. Thus, a digital ASIC chip and its sub-systems for supporting the imaging system also needs these prerequisites. Especially, the memory device inside the digital ASIC should be both frugal and powerful, since the conducted operation may change, but still every operation should read from and write to the same memory. To understand the motivation of this thesis, one need a general knowledge about both the imaging systems and the memory devices, in order to understand the fundamental idea properly.

1.1 Imaging Systems

An imaging system is a system that converts incoming light packages to the preliminary electronic signals in the first place. Generally, raw output of the sensors came out to be analog data. This analog data needs to be calibrated. Calibrated data may be manipulated, in order to eliminate deficiencies on the sensor or the readout circuitry

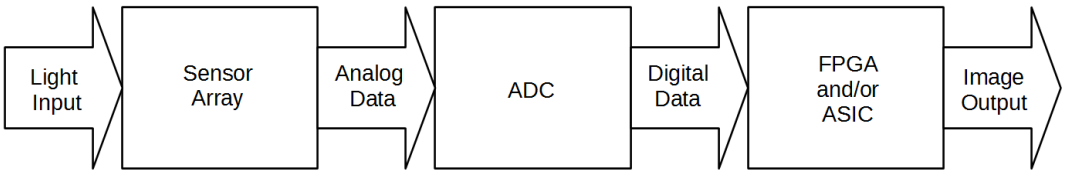

Figure 1.1: General structure of imaging systems

or to improve the image quality. Calibration and manipulation of the image data can be done by the supporting electronics of the imaging systems. Then, based on the purpose of the system electronic signals can have beamed on a screen or stored in a memory. A simple abstraction of an imaging system can be seen on Figure 1.1

1.1.1 Sensors

Sensor is the main part of the imaging devices. Indeed, sensors are used as arrays. As coupled with nature, their output is essentially analog data. Sensors take the photon energy, converts it to electrical equivalent value. Different kinds of sensors convert different parts of the electromagnetic spectrum. Current semiconductor technology allows inexpensive manufacturing of various kinds of sensors. For general imaging purposes, mostly visible light and near infrared (NIR) spectrum of the radiation is used. However, for different imaging applications, i.e. medical imaging different parts of the spectrum may be used. Especially group 3A-5A compound semiconductor technology allows huge advances with their optical properties. Also CMOS monolithic sensors came to a very decent place in usage. However, all the raw data collected form sensor arrays needed to be organized. These organization operations are done by both the readout circuitry and the supporting electronics. An example sensor and readout construct can be seen on Figure 1.2

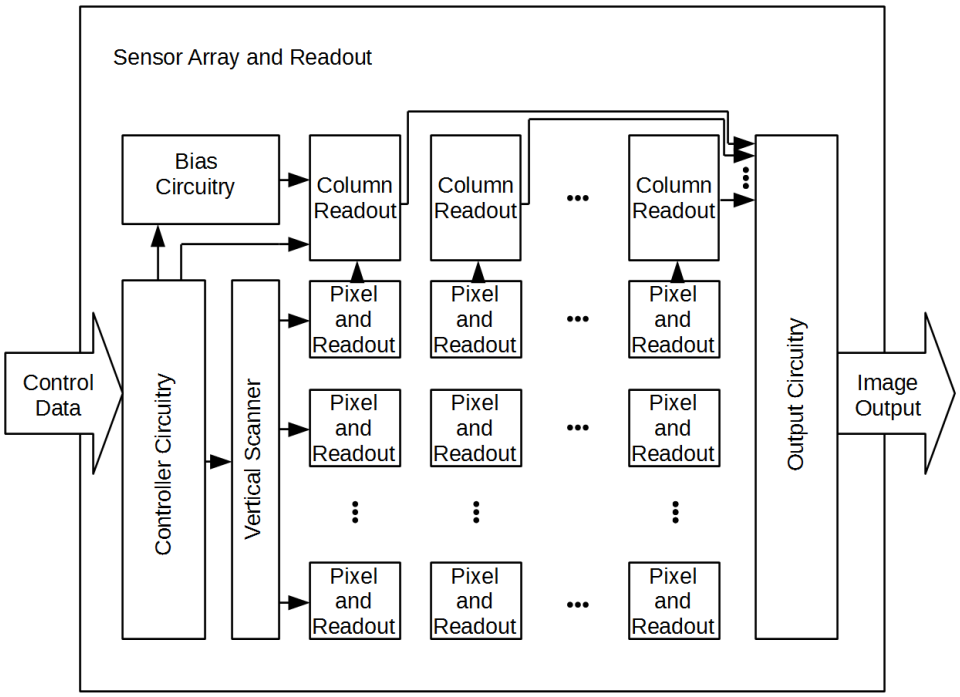

Figure 1.2: Abstraction of an example sensor readout structure

1.1.2 Supporting Electronics

Generally, if electronics are inside a sensor chip, we name the electronics as readout circuitry, but as all of the functionality cannot fit into one chip, and the cost of fitting everything into one chip is not generally feasible, there exists supporting electronics outside the chip to achieve needed operations. Some of the supporting features can be implemented inside the readout circuitry easily, but this time the area considerations and design time budget may create cold feet for a certain design. Thus, it is inevitable to use complementary circuitry if designing a complicated system. Some of the supporting electronics example can be seen in the following part.

1.1.2.1 ADC

ADC takes the analog data from readout circuitry, converts it to its digital equivalent. It feeds the digital data to the digital ASIC or FPGA, for being enhanced by digital image processors or by the software. ADC is a critical component of the various signal processing systems[24]. Thus, it is also very important for imaging systems. Resolution and accuracy of the ADC device impact performance of the imaging system significantly. There exist plenty of ADC types, available for commercial usage. Prerequisites of the application determine the kind of ADC that is suitable for the application. Also, if ADC will be integrated into an ASIC or readout circuitry, more restriction on design constraints occur. ADC varieties include fundamental ones that are, flash ADC, successive approximation ADC and its derivatives, and also ramp compare and its derivatives. In addition to these basic types, also there exists new approaches and improvements, but it's out of the scope for here.

1.1.2.2 FPGA

For rapid prototyping and eliminating manufacturing costs, FPGAs can take several different tasks in an imaging system. It can do any digital operation as its resources are enough to implement. It consists of programmable tables, which are named as look up tables (LUTs). LUTs can mimic any logical operation that is mapped onto them. They are basically a kind of memory which takes logic input as an address and at the relevant place there exists the result of the logical operation. In addition to LUTs, also there is a need for local memory which are flip flops. Using these structures, a meaningful sub-part of an FPGA can be constructed. Some of the companies as Xilinx Inc. name this structure as slice, and an example of it can be seen on Figure 1.3. In current FPGA designs, a lot more building blocks exists for different purposes. The reason behind these specialized blocks is to create an environment that can be comparable with a custom ASIC design with a lot more design space. FPGA initializes the readout circuitry, feeds the configuration data, (in some cases) reads the converted output data, and acts as a bridge between software and the readout circuitry. Also, FPGAs enable new features, on the fly with the same device [2], and a great amount of flexibility.

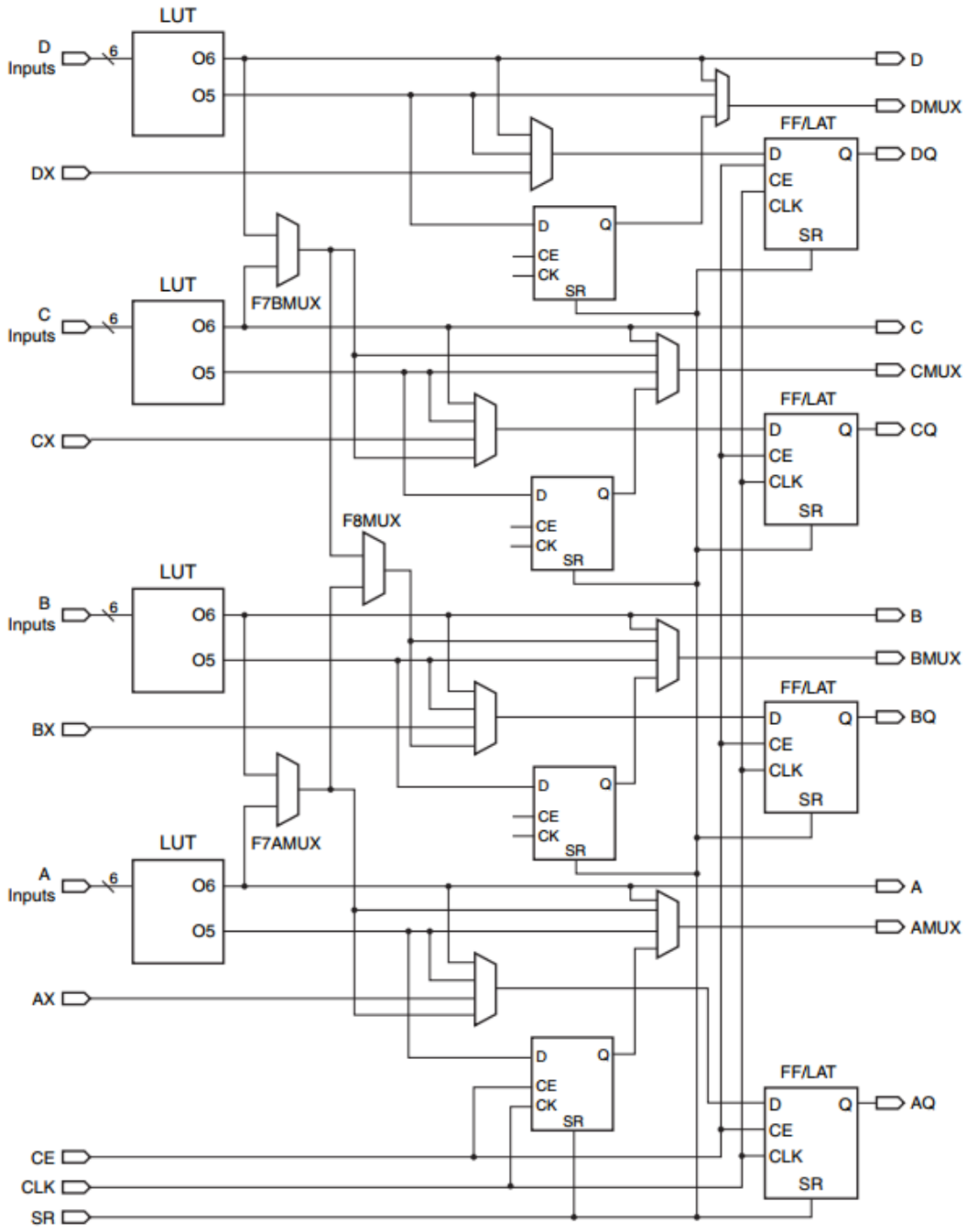

Figure 1.3: Building block of an FPGA (A Slice of Xilinx Spartan 6 [26])

1.1.2.3 ASIC

ASIC devices can do anything in its application context. As an example, it can combine an ADC and a basic multiplexer, helping to scale down the total PCB area. For some other cases, it can do digital image processing, on converted data, or can do both. It is the designers' choice, and also a matter of the cost in the end. Designing an ASIC chip kills the flexibility, but gives more opportunities in the design phase. It is costlier to start a new ASIC design, however with the less power consumption, and less complexity with respect to FPGA prototyping gives ASIC design an edge. In addition to latter advantages, also evolution of the imaging system design eventually forces ASIC usage at a point. Implementing brand new designs without the limited prototyping environment of FPGA forces this development. Fundamental sub-parts can be seen on Figure 1.4 for a digital ASIC. If design of the ASIC chip includes digital image processing, a decent memory device needed to be used inside. For implementing spatial image processing, there is a need for small and fast memory structure, in addition to the low power consumption needs. However, if there is a need for time domain image processing, i.e. frame by frame processing, then needs for memory devices change significantly. Size of the needed memory increases, and a more parallel architecture may be needed. In that context, power consumption could be increased with respect to former small memory. Moreover, interfacing the bigger memory creates more challenges, when more write and read ports implemented to obtain a more parallel architecture. Luckily in this work, the purpose is to implement a memory for spatial image processing, and this memory thought as an integral part of the core ASIC system. Bigger memory may have advantages, but implementing it with plain CMOS does not create a positive effect on cost and power consumption of the core ASIC circuitry. It could be connected at PCB sytem level.

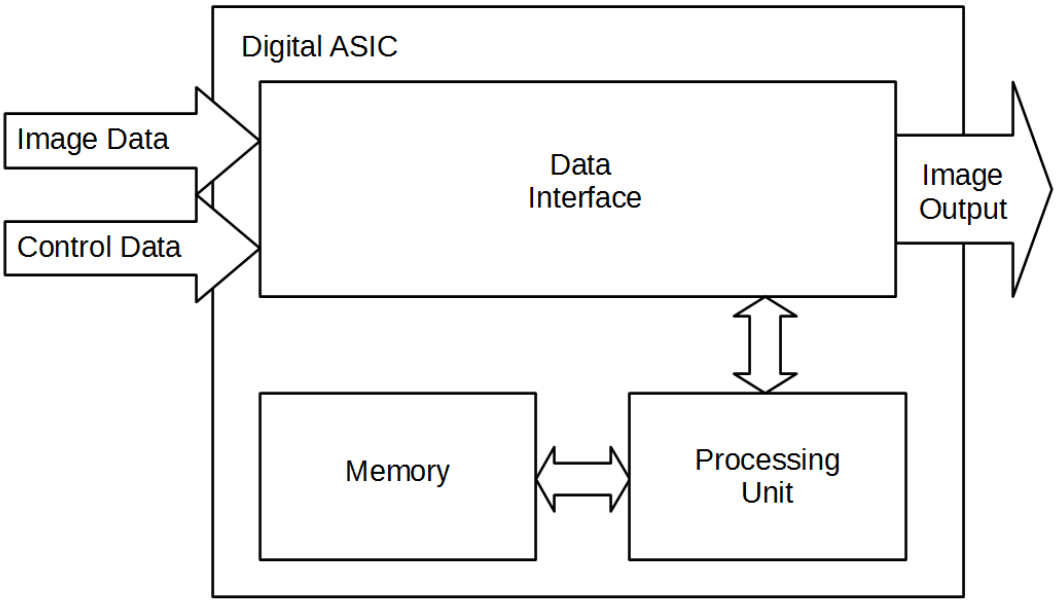

Figure 1.4: Sub-parts of the ASIC system

1.2 Memory Devices

In modern computing systems and electronic devices, there has always been a need for structures that aim to remember a previous state that were used before. For fulfilling this requirement, people suggested plenty of concepts throughout the era of the electronics. They asserted new concepts with a resemblance of the paper and the pen or paper and the pencil. In abstract part, modeling was easy to conduct, but blending them with the electronics concepts was the real problem. Ideal models, however only work in certain circumstances for the real world. For different requirements for different systems, myriads of solutions created as memory with the year over year work. One of the best solutions to answer these requirements, and reduce performance problems between processing units and different memory kinds is to use several levels of memory in a hierarchical method [18]. In addition to application area necessities, also power consumption and cost considerations came into forefront with passing time. Currently, there exists a classification for choosing the proper structure for the needed application within its power and financial budget. The ones that pave the way to this thesis will be mentioned in the following parts.

1.2.1 Non-Volatile Memories

The term non-volatile means, if the device is powered down, nothing changes the bits it stored. Non-volatile memories are useful, especially for firmware-like structures, where change of the stored data is not needed rapidly, or the data do not change at all. In recent history also non-volatile types of memory have been used for the mass distribution of data. Indeed, usage continues, but with increasing speeds of internet, quantities that were produced decreases, though do not become extinct. Endurance of these kind of media is generally high, unless of course rewriting occurs frequently. Storing or reading may be sequential or random access; it depends on the structure of the medium and cost-area preferences in certain structures. Commonly used ones in today's world are explained briefly below.

1.2.1.1 Magnetic and Optical Media

After a long term of paper usage for memory, with emerging electronic computing devices, early media for memory was also paper, but in the context, they do not value as a device, so it can be confidently said that, magnetic media was the first devices to be used as a memory device in the early computing devices.

The first example of the magnetic device was the drum memory. It was created as a big cylinder, coated with magnetic material. There exists one head for read-write for one track of data. For some time, this kind of memory was used as both primary and secondary memory. However, its speed was determined by the speed of the cylinder, so it was mechanically limited. Then, these were evolved into different kinds of memories, for different requirements.

For the primary memory, core memory was used afterwards the drum memory. It was indeed a random access media without mechanical motion which allows faster read and write times with respect to mechanical media with sequential access.

Also for the secondary memory, again a mechanical structure was used with magnetization. As performance needs were not high as primary memory, but density of data is more important, a sequential media named hard disk proposed. It has stacked

disks, and for every disk it has write/read magnetic heads. Disks were revolving for read and write operations. Although the structure seems ancient as a drum memory, it can still find place in the modern computers; as the packing density of them and enormous rewrite cycle count cannot be achieved. Nevertheless, they are not perfect. Mechanical structure makes them prone to physical shocks, which can damage data, or device. Stationary use is safe, but mobile usage is dangerous, despite the precautions. Moreover, for increasing their performances, again there exists mechanical limitations. As [9] asserts for example, disturbances in the servo control systems that is directly related to operation speed, have to be eliminated with an observer, and the design of the observer could lead to system stability problems.

In addition to these media, also several types of the cassettes were used for distribution purposes. They came into place as audio recording media first, then also used for computing and for the video. Before the advent of the optical discs, magnetic cassettes dominated all the market.

After the mass usage of cassettes, optical media came into place. It was invented, when data needs were increased and magnetic media cannot answer these needs. First commercial product was Compact Disc. It was used for audio again, at least firstly. Then, also usage for video and computing were spread. It was very successful, as it settled into all kind of media needs. The thing replaced it was also an optical disc, which is DVD, and currently Blu-Ray discs as the 3rd iteration of the optical discs was used as a distribution media. However, demand for these kind of media were diminishing, since internet serve as a mass pool of media.

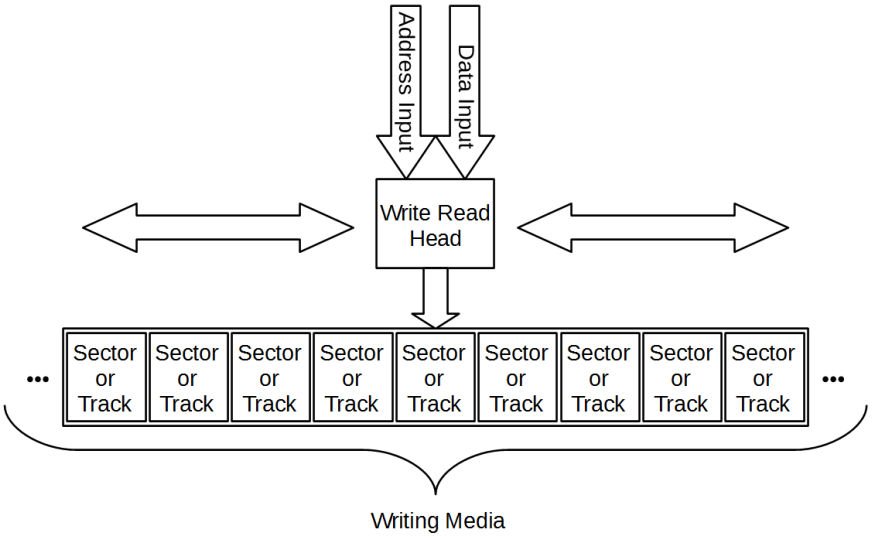

Figure 1.5: General abstract structure of magnetic and optical sequential media

1.2.1.2 ROM

Rom means read only memory, but today, the synonymous of rom is not only imply that. Of course the main purpose of the memory is reading, in modern times, though it has rewriting possibilities available, as they will be explained in latter part of the section.

The need for static configuration and static data lead the way to create first ROM devices. Firstly, diode arrays or similar structures were used as read only memory. According to arrangement, whether there exists a diode or not the reading value is "0" or "1" respectively. This matrix structure also used with the transistors, and practically no big difference exists between two.

These structures lead to more densely packed memories with the advent of the integrated circuitry. Also, new opportunities evolved. The first example is the mask memory or mask rom. It is completely read only semiconductor memory that its bit values were defined at the mask level from the manufacturer and it cannot change afterwards. It may be a good solution for the semiconductor manufacturers, but users need a little more flexibility. Also cost of production is high, as for every new memory, user must have gone through extensive design and manufacturing cycle. Then,

Programmable ROMs came into place (PROM).

PROMs gave the users opportunity that, they can write their bit sequences once without thinking design and manufacturing processes, only focusing to their code. Writing process involved a high voltage operation, if user wants a "1" and burns the selected part. Otherwise, remaining parts were "0" where the transistors do not burn and still conduct. Shortcomings of this structure, made manufacturers improve the design to allow re-writability, and Erasable PROM (EPROM) was born. Floating gate structure allowed threshold voltage changes with application of high voltage to the control gate by trapping the electrons to the gate. Also, this kind of memory can be erased with some overhead in time and with usage of UV light with specified intensity, angle and wavelength. After that a simpler method to remove bits were invented. Applying another electrical signal with different properties removed the trapped electrons inside the floating gate. They named this type of ROM as Electrically Erasable PROM (EEPROM). Though write cycle was limited to nearly thousands of erase and write operations. Also, erasing consumed more power with respect to writing in early devices. Improvements in the integrated circuit technology, enabled improved write cycle, improved erasing speed, and increased packing density. Indeed, in modern world they can easily be used as secondary memory, as named flash memory. Only a small conceptual difference exists between the commercialized flash memory and the EEPROM, which is, one can erase data on EEPROM word by word, however in flash memory for the sake of packing more devices, erasing takes place as block by block. Although write cycles were improved, still, nowhere near to competing mechanical magnetized media, so when designing, special care needed to erase and write to different physical areas of the device, so that wear effects can be minimized. One very important advantage is however, speed of the media. As no mechanical motion exists, whether sequential or random access, speed is much higher than the conventional hard disks. In addition to that, the cost of the flash memory decreases and the packing density increases with passing time, made this kind suitable for the handheld devices [23].

1.2.2 Volatile Memories

In addition to the non-volatile counterparts, also there exists volatile kind of memory devices, which cannot preserve their bit values after cutting off the supply. The purpose of this kind of memory is to access repeating parts of data quicker than secondary memory, hence system performance is increased. As a consequence, immense amount of data was written and erased back and forth in these kind of memories. There exist two kinds of volatile memory devices, and as they have both used for the fast operation, they are both random access type of devices.

1.2.2.1 DRAM

The dynamic random access memory takes its name from the need of refreshing its cells. When cells are idle, leakage current makes their voltage diminished with passing time. Between certain time intervals, DRAM bitcells needs refreshing, i.e. dynamic operation. Also, read cells need to be rewritten, as capacitive charge sharing effectively destroys the data in read cell during reading operation. Fundamentally DRAM bitcell is a charge storing device with one transistor and a capacitor. As bitcells are such basic devices that do not take much space in physical world, packing density is very high. It is the most critical advantage of the DRAM against SRAM. In addition to their packing density, also access speed of them is another advantage. Although they cannot achieve SRAM performance, they are much quicker than secondary memory and their predecessors. Especially, with the introduction of the new generations of the double data rate architectures, every generation increased the throughput by multiples of their predecessors. The increase in performance can be attributed to technology nodes, and buffer improvements. However, power consumption is not in line with the SRAM counterparts; in other words, due to nature of its memory cells, power consumption is high in DRAM structures [14]. They are essential for current computer systems as the main memory in order to buffer virtual memory and/or hard drives, as any miss in the main memory means much slower read operation from these sequentially written, mechanical media. They can be written again and again, virtually infinitely many times, which is one of the another important advantage of DRAM. Especially with improvements in current flash technology,

read performance became comparable with DRAM, nevertheless write cycles of the flash memories are not comparable that of the DRAM architectures (at least tens of thousands of difference still exists)

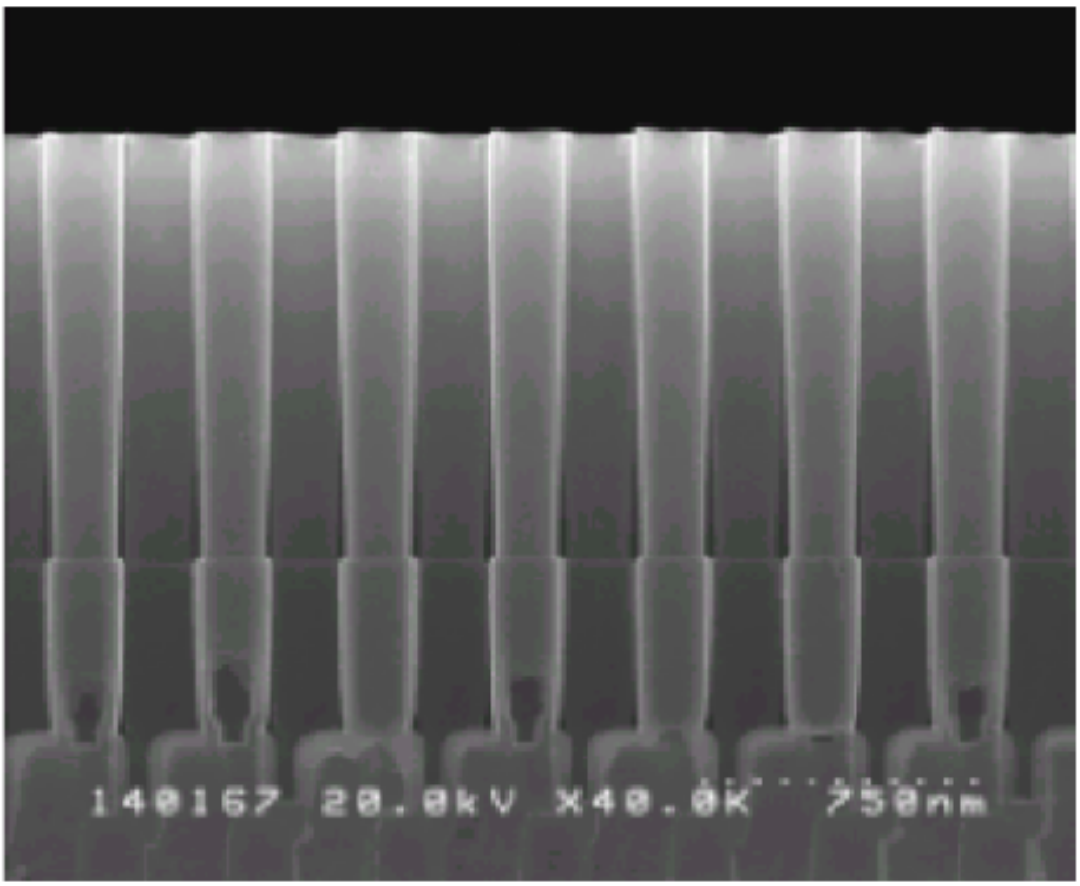

Figure 1.6: Cross-sectional SEM (spectral electron microscopy) image of novel DRAM capacitors implemented on a 90 nm CMOS technology [17]

1.2.2.2 SRAM

When it comes to the SRAM, it is the least dense counterpart of the memories. It is volatile in traditional sense, but in different external conditions, it seems that it have data remanence, which is it can protect electric charge on it; for example, in low temperature environments [22]. Its storage elements need at least four but generally more than that number of transistors. Its advantage is the operation performance, and the reduced latency, as small arrays means smaller length, area, resulting in smaller parasitic capacitances. This type of memory occupies the uppermost part of the memory hierarchy, which have least abstraction and if used as processor register, write and read operations are completely direct. Operation can be realized as writing directly, or reading directly, with given input address; no apparent waiting exists, or no elaborate structure that complicates the operation is expected. Inside the structure there

may be complicated circuitry or timing exists, but for the outside, it is better to protect direct approach, without giving a hint about the complexity. As an example, in the cutting edge technology, where physical limits of the silicon also tested, sleep modes for the bitcell ocean were implemented. Fairly complex reading operation with the help of the sense amplifiers exists, a more hierarchical address handling, in order to decrease input capacitances and achieve less latency were some of these techniques. Nevertheless, in the end, all operations are wrapped as write and read, no information needed about inside of the SRAM. Efficiency of the memory access in today's multi-core processors determines the overall system performance [25]. It is very important also because currently, a great percentage of the processing unit circuit area is used for the different levels of caches, which are essentially SRAM counterparts. Indeed, [8] asserts that some of the applications need an allocation of chip area of 90 percent for the cache memory. In addition to this asserted value, it can be seen on the Figure 1.7, SRAM Cache (probably shared L3 cache) of the circuitry takes nearly a half area of the CPU cores of the design, and it should be noted that, if only CPU and cache were squeezed into one separate chip, at least one third of the chip would be this SRAM cache, careful eyes can see though, inside the CPU cores, there exists L1 and L2 caches which are also SRAM structures. In addition to area considerations, power budget of the SoC applications significantly depends on the cache memories [20]. Using an architecture level optimization of number of memory access combined with reduction of array power in each access can reduce the power dissipation significantly [5]. Architectural improvements are important, but using cutting edge technologies as High-K FinFET [4] helps to improve performance or RRAM-based nonvolatile SRAM [13] helps to improve energy usage according to needed application.

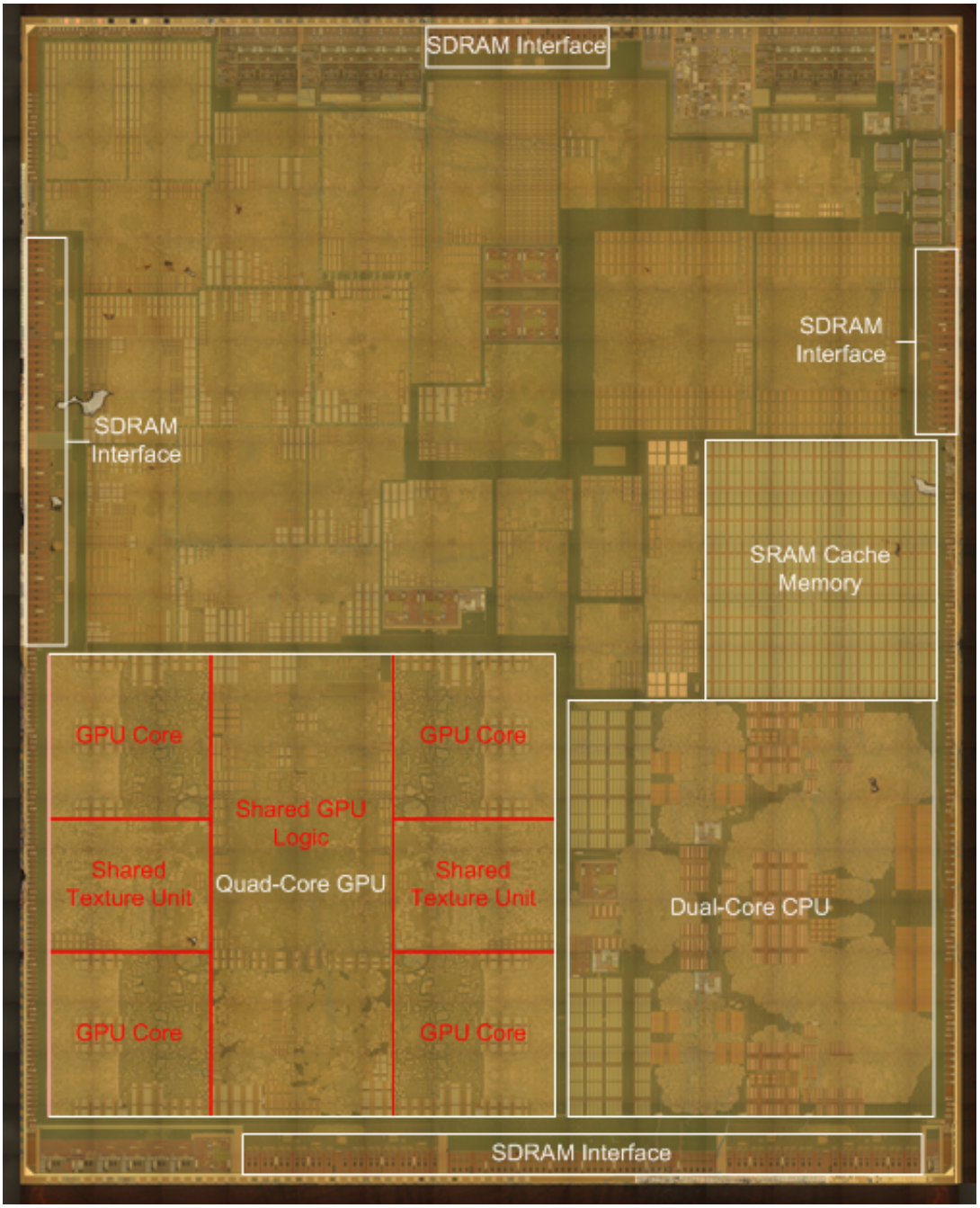

Figure 1.7: SEM image of an Apple A8 SoC with certain parts marked [11]

1.3 Motivation of Thesis and Thesis Organization

As digital image processing is a must for the current imaging devices, there exists a need for image processors. Currently there exists no exclusive memory device for imaging purposes. However, all kinds of processors, also image processors make extensive use of a memory structure, which have the image data and also the manipulation data. Image data is volatile data, which needs to be held at least line after line in order to make calculations on all of the set. It needs to be saved rapidly and can be read rapidly in order to process frame data on the fly. Otherwise, it cannot be used in an imaging system, rather it can be used in a render environment which processes the imaging data after recording it. Another very important constraint is the power draw of the structure. As a camera is generally used with a battery, or a charge pack, it has a limited amount of current for a limited amount of time. Thus, an image processor and its subsystems need to be balanced in their performance and the power usage to operate properly. As it can be seen in Table 1.1 within commercially available memory types, SRAM is the most appropriate choice to be implemented, also it has to be noted, erasing one flash cell consumes immense amount of energy with respect to writing energy, and in the table, value of erasing does not been specified. In order to use in a low power image processor ASIC, a low power SRAM needs to be designed.

Organization of this thesis is as folows:

Chapter 2 explains the important sub-circuits of designed SRAM structure such as bitcell, address decoders, sense amplifier, timing generator, and column structure. In the explanations, also implemented crucial methods for timing generator (pulsed synchronous operation), and column structure (local word line assertion) and abstract overlooks of all the sub-circuitry exist.

Chapter 3 tells about the test chip that is to be manufactured, in order to confirm the operation of designed SRAM structure. In this part, auxiliary sub-circuits, along with their placements inside the test chip are given. Also, pad placement of the test chip was indicated in this part.

Chapter 4 gives information about various simulation results, and also proposed test setup for the test chip. Simulation results of the bitcell, address decoders, timing

Table 1.1: Commercially available memory device feature comparison [27]

	SRAM	DRAM	Flash	HDD
Reciprocal Density(F^2)	140	6-12	1-4	2/3
Energy/Bit(pJ)	$5x10^{-5}$	$5x10^{-3}$	$2x10^{-5}$	$1\text{-}10x10^{9}$
Read Time(ns)	0.1-0.3	10	10^{5}	$5\text{-}8x10^{6}$
Write time(ns)	0.1-0.3	10	10^{5}	$5\text{-}8x10^{6}$
Retention	supply on	«second	years	years
Endurance(cycles)	$>10^{16}$	$>10^{16}$	10^{4}	10^{4}

generator, column structure, and top level resides in that part. Also, comparison of the work with previous works can be found in this part. Lastly, an abstract view of test setup and constructed GUI for test operations can be found in this part.

Chapter 5 summarizes the thesis work, gives achievements, and suggests improvements for future works.

CHAPTER 2

DESIGN OF THE SRAM

2.1 Bitcell

Design of the bitcell is nearly the most critical part for achieving a great packing density. It is the most abundant structure in percentage as it has to be. Thus, smaller bitcell means more memory in the same area as a result. There exists plenty of evolved bitcell topologies that are derived from the standard six transistor bitcell structure. Although there exists evolved bitcell structures with 7,8,9 or even with 10 transistors [21], mere motivation of these approaches is very low read SNM values that makes read operations harder in deep-submicron technologies. Lowering supply voltages for reducing the leakage current, oblige designers to make compromises from the precious layout area. In our technology node however, area is much more important than the leakage current, as total leakage current is very low with respect to the state of the art CMOS SRAMs, but also as the minimum transistor sizes are huge, and any improvement in the bitcell area is more appreciated. In the light of these reminders, minimum width and length of the transistors were used for the bitcell. For clarity, these minimum sizes were for the boneless structures (contacts do not create a visual representation of a "bone" on both sides of the polysilicon gate) in order not to waste area from the horizontal length of the bitcell. In the rated supply voltage, our technology specifies at least 1.8 Volts. However, threshold voltages of the MOSes let lower supply voltages to be used as well. Using a deep submicron technology, increases concern of read reliability, and more read margin has to be given [1]. Also, in technologies beyond 130nm, there exist problems of intra-die variations and leakage which severely complicate low power SRAM design [6]. Moreover, according to

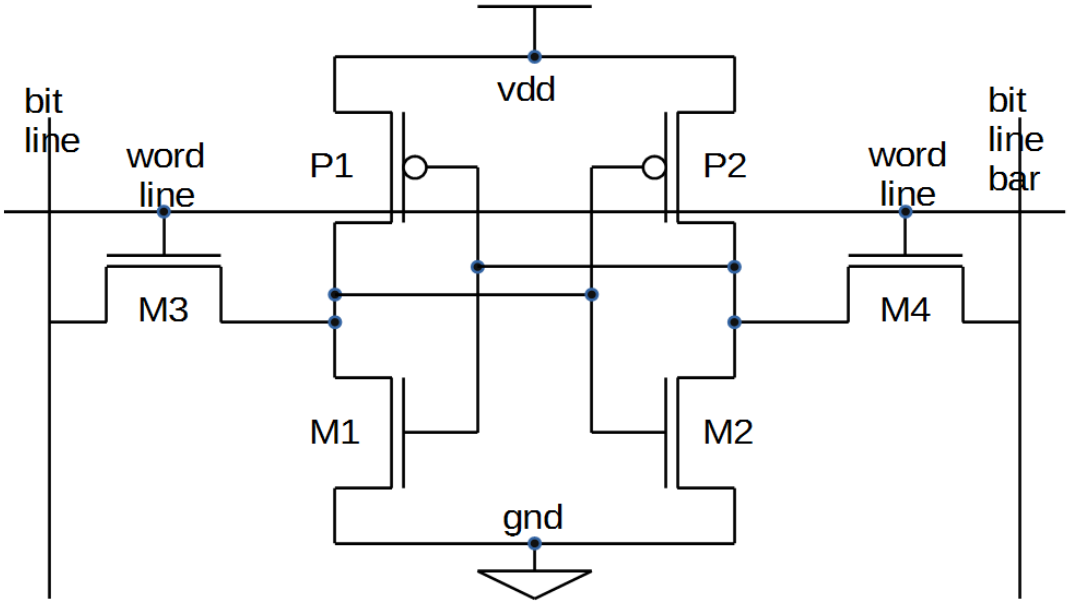

Figure 2.1: Schematic of the basic six transistor bitcell

[10], soft error rate per device increases with the decreasing size of the technology node. Fortunately, (as can be seen in SNM simulation results) cell SNM voltages have a margin and technology node allows more relaxed design. The advantage of the lower supply voltage comes from two sources; one is obvious as voltage itself, in addition to that also current drawn from the supply decreases with lower voltage applied. Thus, there exists nearly a quadratic advantage of power with respect to voltage. For the access transistors, also minimum boneless n type MOSes were used. Schematic of the bitcell can be seen on the Figure 2.1, and layout can be seen on Figure 2.2 . Also characteristics of the bitcell can be found in the simulations part.

2.2 Address Decoders

Generally, address decoder design is not essentially a critical part of the SRAM design, however with the lower supply voltage in mind, and also the need of, at least decent performance at the low supply, considerations must be done with some balancing of the implementation area and performance. The real problem is inevitable, as whatever is done, p type structures have less current drive (in this case 2.5 to 3

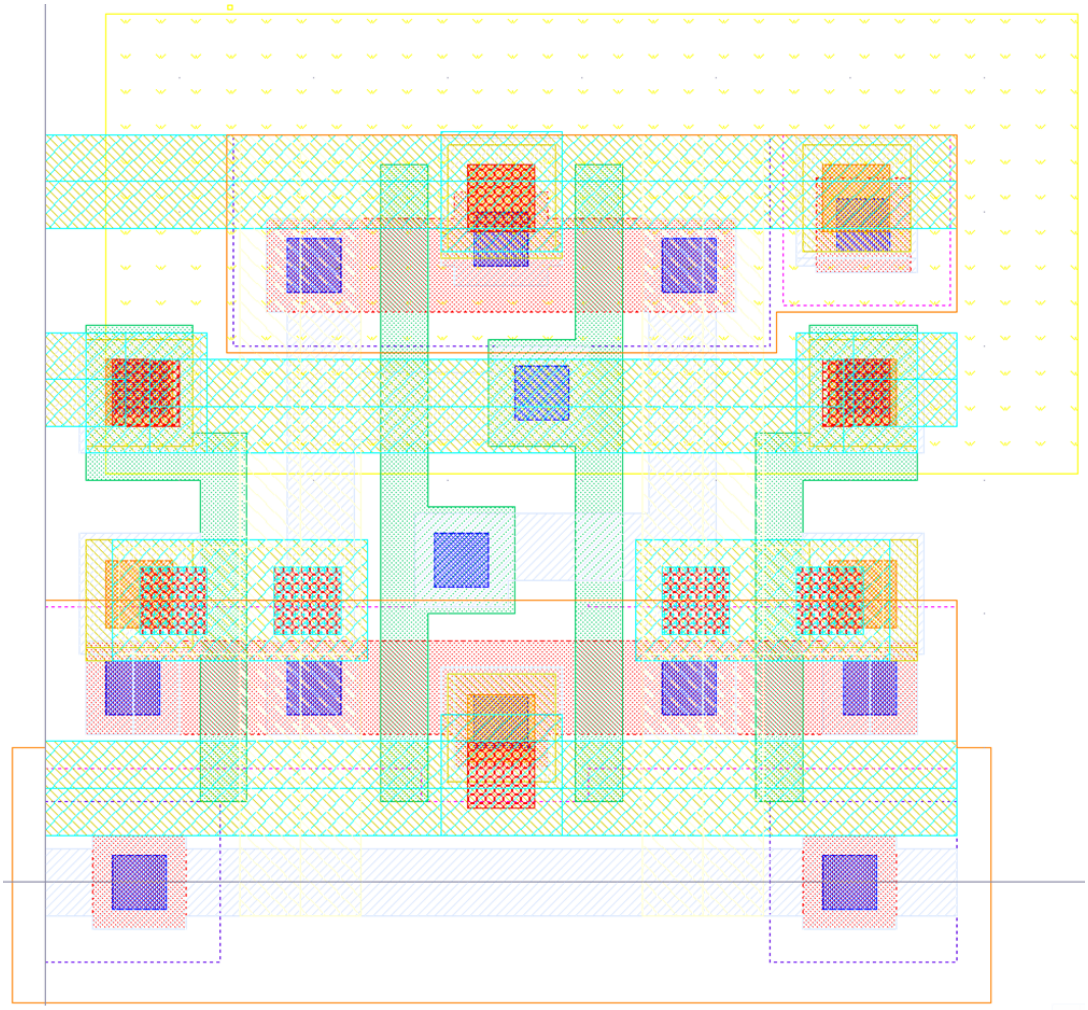

Figure 2.2: Layout of the basic six transistor bitcell

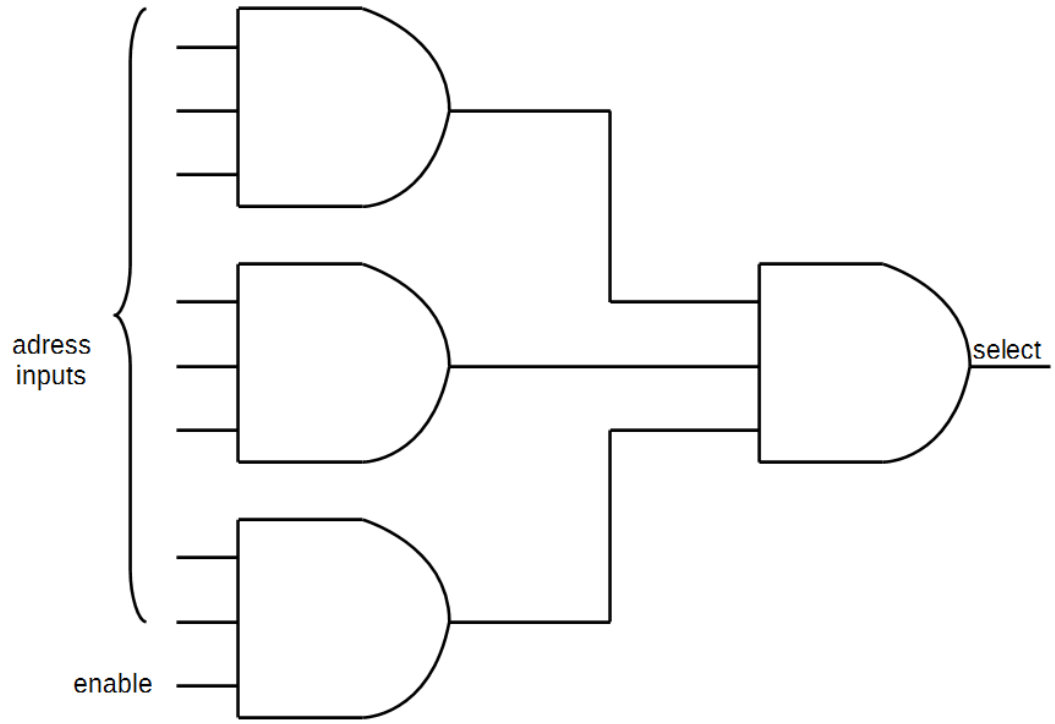

Figure 2.3: Schematic of the row address decoder structure

times less) which leads to using series of pull-up devices, getting even less current, and in some cases, weak p type MOSes stuck at the linear region. As a result, in the decoders, nand gates with inverters are used. Also in the remaining logic operations, the same issue addressed, and solved by same approach. The schematics of the address decoder and the column decoder were made up of three AND gates which takes three inputs and one NAND gate which takes four input to unify one level below and also adds one enable input. Schematic can be seen on the Figure 2.4, layouts of the column and row address decoders can be seen on Figure 2.5 and Figure 2.6. Also, simulation results of the basic circuitry can be seen in the its own part in the simulations chapter.

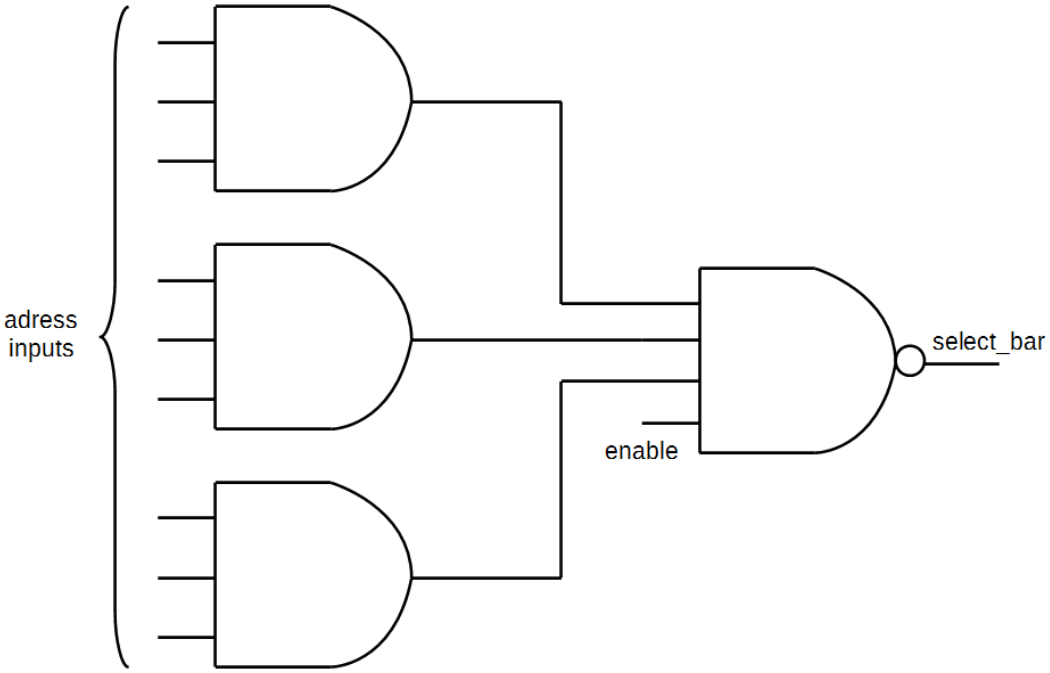

Figure 2.4: Schematic of the column address decoder structure

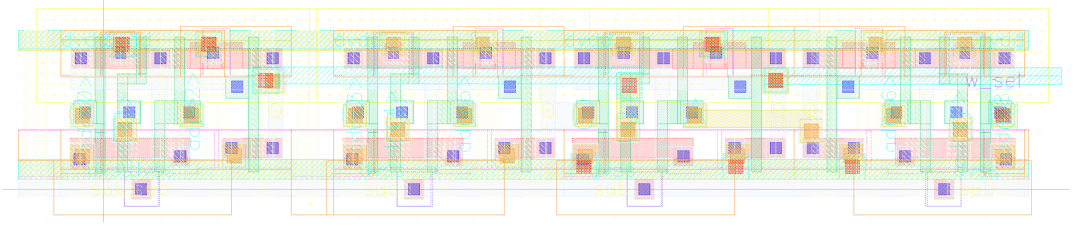

Figure 2.5: Layout of the row address decoder structure

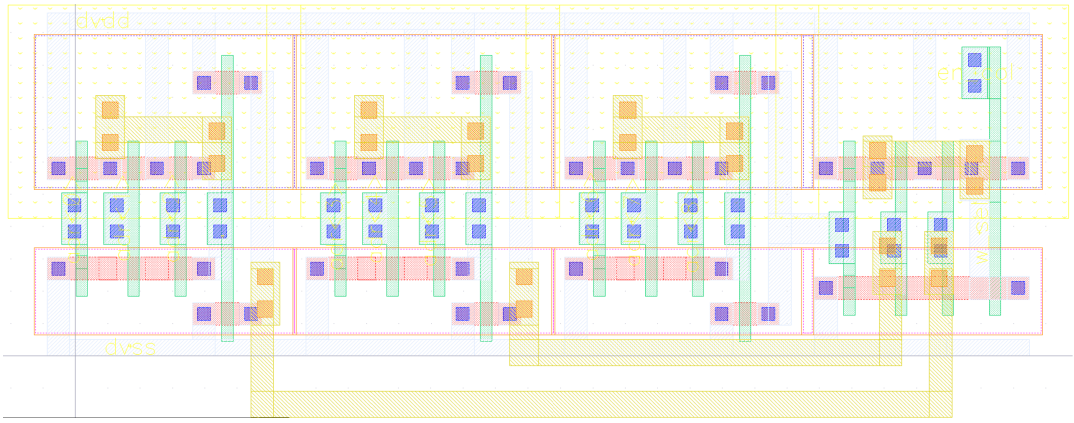

Figure 2.6: Layout of the column address decoder structure

2.3 Sense Amplifiers

A very important part of the SRAM is the sense amplifier. Sense amplifiers help to read differential voltage and current changes on the bit lines of the memory circuitry, and then amplifies the differential input with a high gain positive feedback amplifier to obtain values inside memory cell. If a sense amplifier not been used on SRAM memory design, then read times of the circuitry became very high compared to the write times. The reasons for that is the weak transistors of the memory cells (generally minimum transistor sizes were used) that are needed to drive the highly capacitive bitlines. All modern SRAM designs use sense amplifiers to have decent performance in both write and read operation. In this work, a current sense amplifier was used. Current sense amplifiers are widely used as one of the most effective ways to reduce both sensing delay and power consumption of the SRAM [7]. Topology inspired by [16]. It has both the advantage of the power and area with respect to standard voltage sense amplifiers. It works with the current difference coming from the bitcells, as either one of the bitcell or bitcell bar gives zero as output and other gives one. With a precharged bit line, one part which have zero output, sends less DC current to sense amplifier, and this current creates more voltage in the receiving end, leading positive feedback loop to rapidly give the output value. Moreover, in this topology, sense amplifiers do not drive the bit lines, thus smaller transistors inside this topology are capable of sensing. Schematic can be seen on Figure 2.7. Working principles of the amplifier in a simplified fashion can be seen on Figure 2.8 and Figure 2.9.

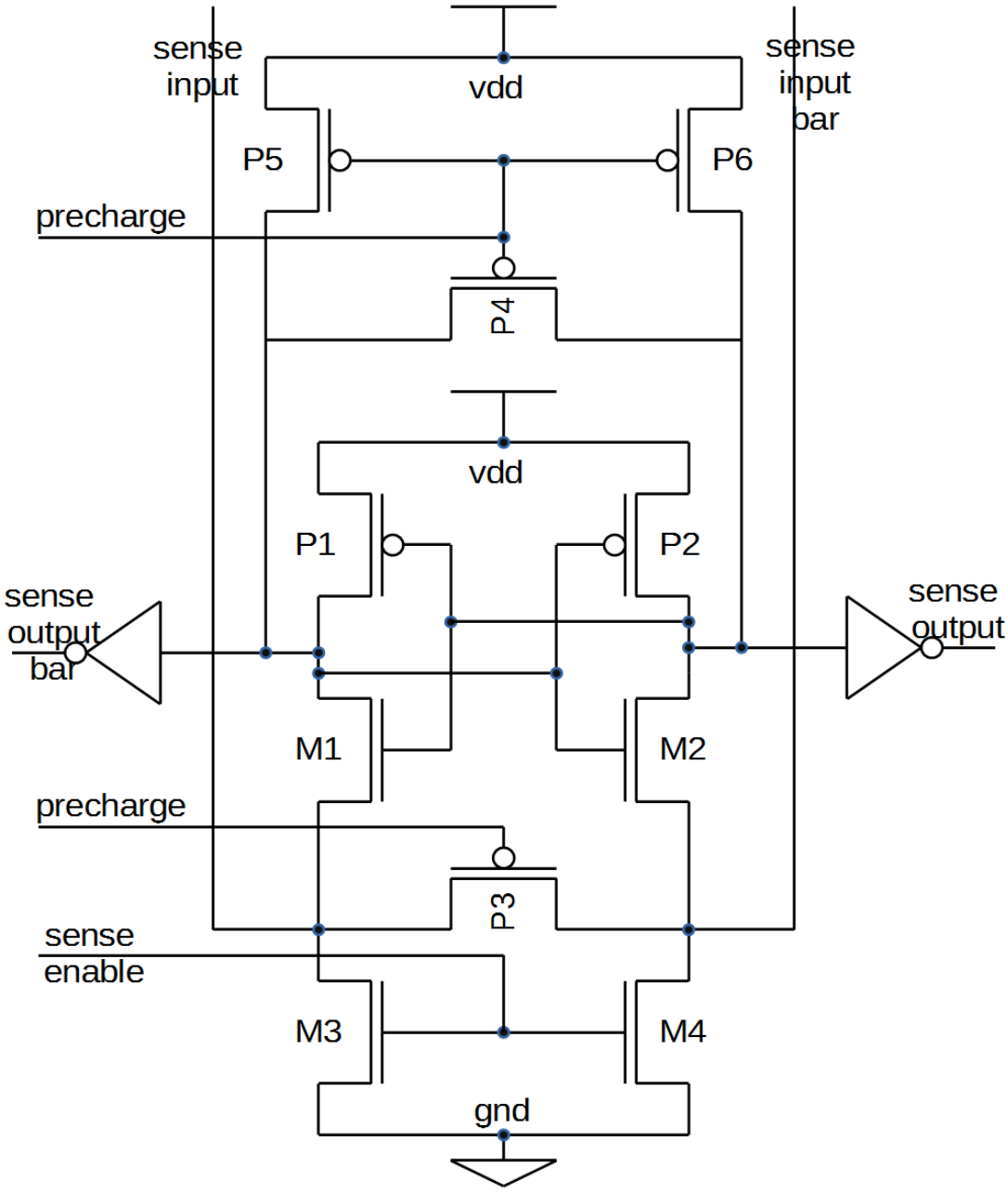

Figure 2.7: Schematic of the sense amplifier [16]

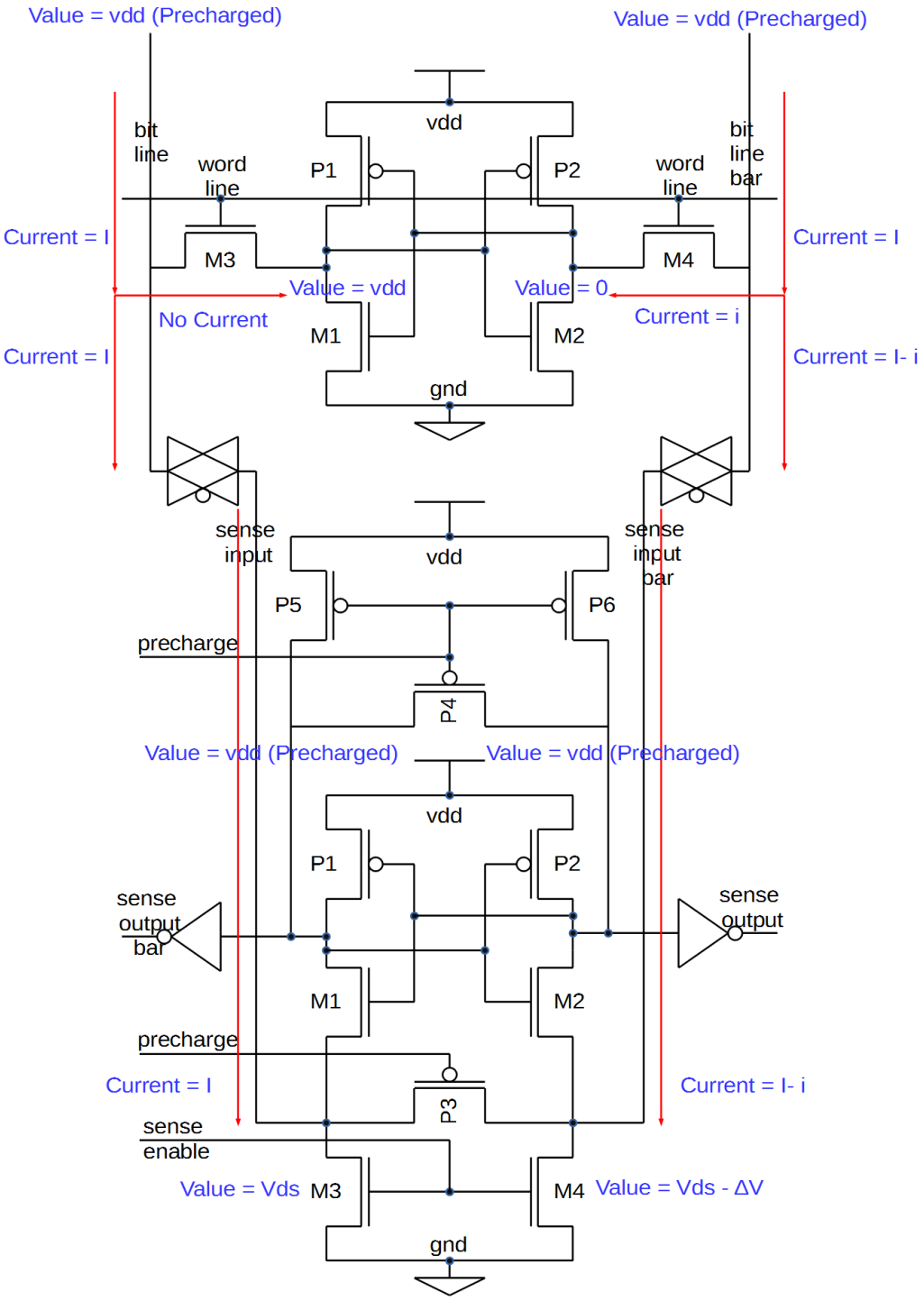

Figure 2.8: Sense amplifier operation illustration that shows the current difference occurring between bitlines

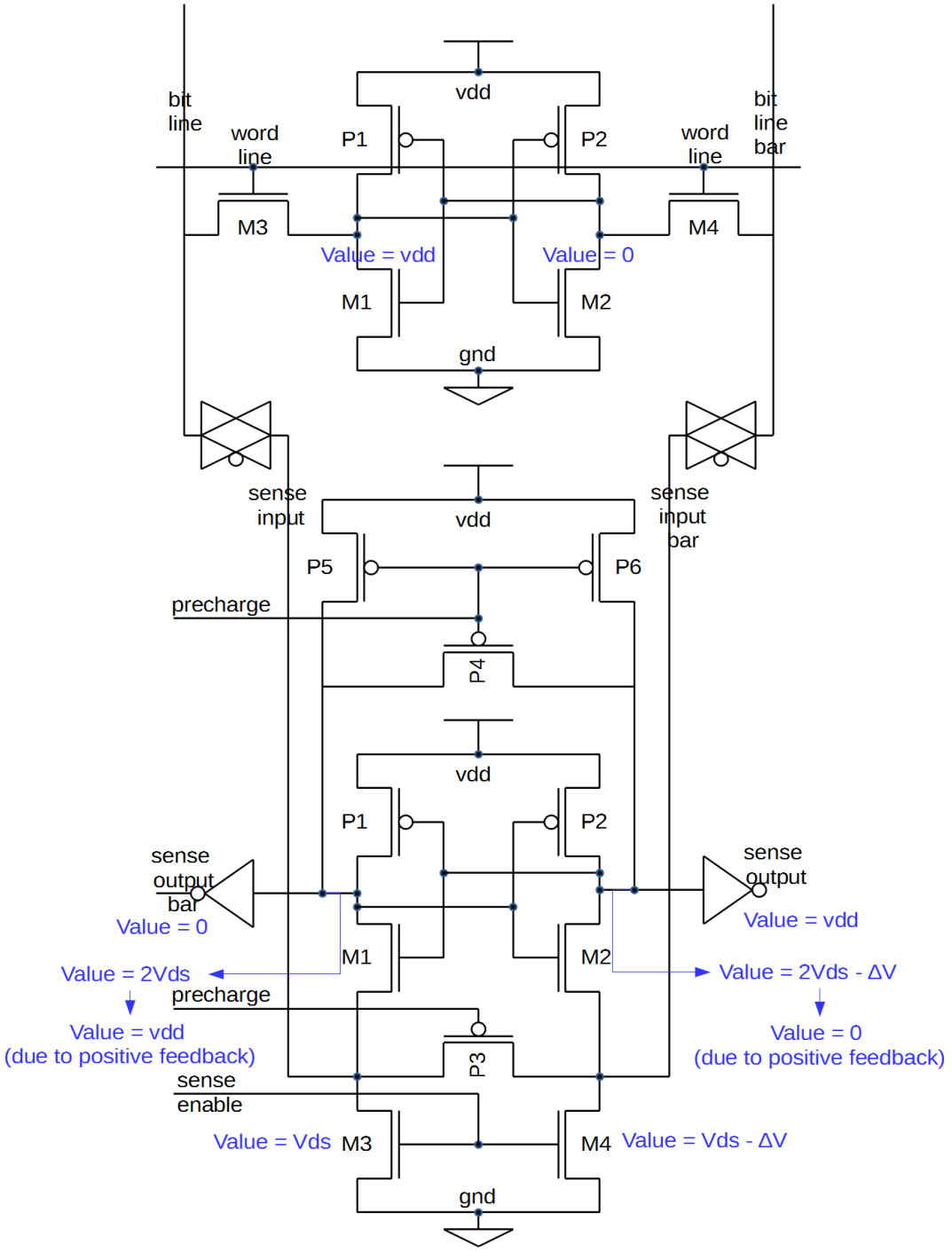

Figure 2.9: Sense amplifier operation illustration that shows the positive feedback and output creation

2.4 Timing Generator

Normally, control signals of the SRAM should be driven from the processor that uses the SRAM. However, when a standalone SRAM test chip is considered, programming all these discrete signals outside the chip creates a very important timing concern. Also, already crowded input and output space at the pad side takes hit from increased complexity. Thus, implementation of a timing generator helps to increase testability; at the same time reduces output pad circuitry complexity. Thus, although a timing generator, is not an essential part of an SRAM, it gives ease of use when implemented. It helps to close timing constraints, without external adjustments with complexity. Also, it helps to increase autonomy, thus allowing a scalable design. Once, critical timing was created to write and read data, it is easier to transport the same signals to everywhere, only with an attention to create needed buffer network to achieve same or very similar timing in everywhere. The state structure and signal generation can be seen, in the Figure 2.10. Designated output waveforms of the structure can be found in Figure 2.11 and Figure 2.12. The most important reason to use a timing generator is to synchronize everywhere, but as a secondary reason, to obtain two edges of the clock signal properly. The circuitry is aware of the clk signal, whether it is at high pulse or at the low pulse with the help of flip flops.

2.4.1 Pulsed Synchronous Operation

All working parts of the circuitry are triggered with an enable signal associated with them. In the first level there exists a column enable signal, asserted when address inputs are ready. All signals and dynamic changes inside column structure start with the column enable signal. It lets signals as precharge and word line enable to act without glitches. Otherwise, these glitches may lead to unnecessary dynamic changes, in addition to these changes, writing and reading errors may occur. Timing generator designed with pulsed synchronous structure allows glitch free operation, and resultantly less dynamic changes and less power consumption

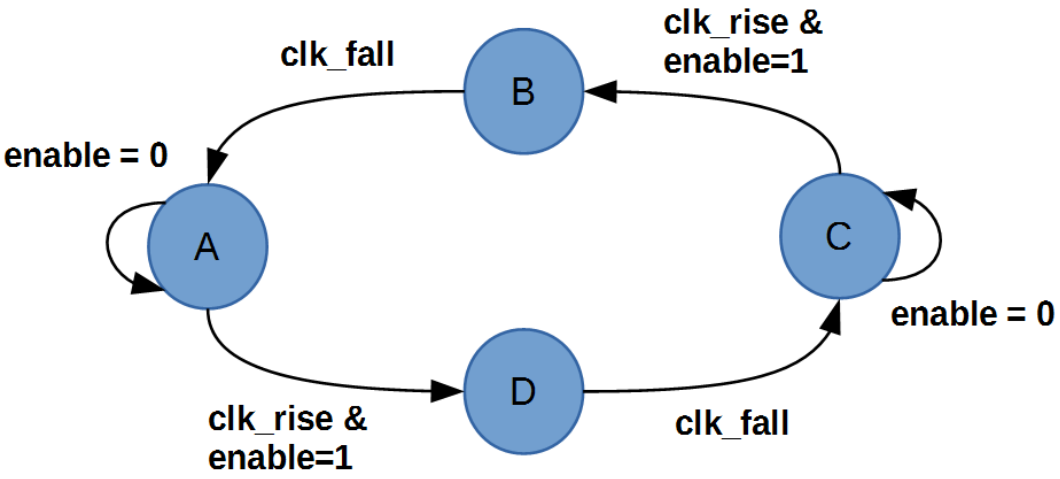

Figure 2.10: State machine of the timing generator

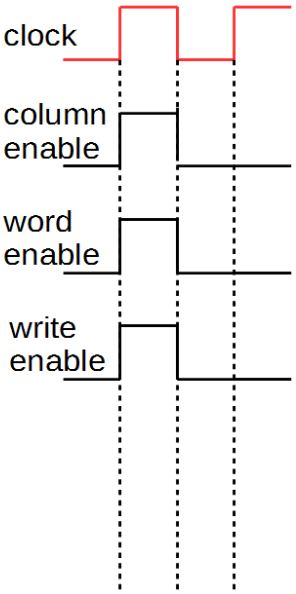

Figure 2.11: SRAM timing generator write signals

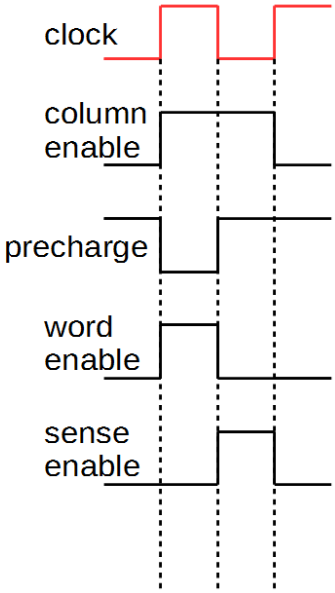

Figure 2.12: SRAM timing generator read signals

2.5 Column Structure

The column structure is a very important part of the design. Its creation purpose is to achieve a scalable architecture, yet with simple implementation. It has its local row address decoders, and one word of data, which is 16 bits in this case. In addition to these structures, also precharge, column decoder and column select transmission gate structures were present. Column have 256 rows of words. In Figure 3.1 basic assembly of the column structure can be seen. It has all the required decoders and buffers to implement an SRAM circuitry even with only one column. Moreover, with simple buffer organization from the timing generator also lets more and more columns to be stacked horizontally together for different sizes for different applications.

2.5.1 Local Word Line Assertion

In column structure, every row of the column is separated from other columns. Each column has its own row decoders, locally. Horizontal driving capacitances decrease significantly, enabling less power consumption. In addition to less consumption, this

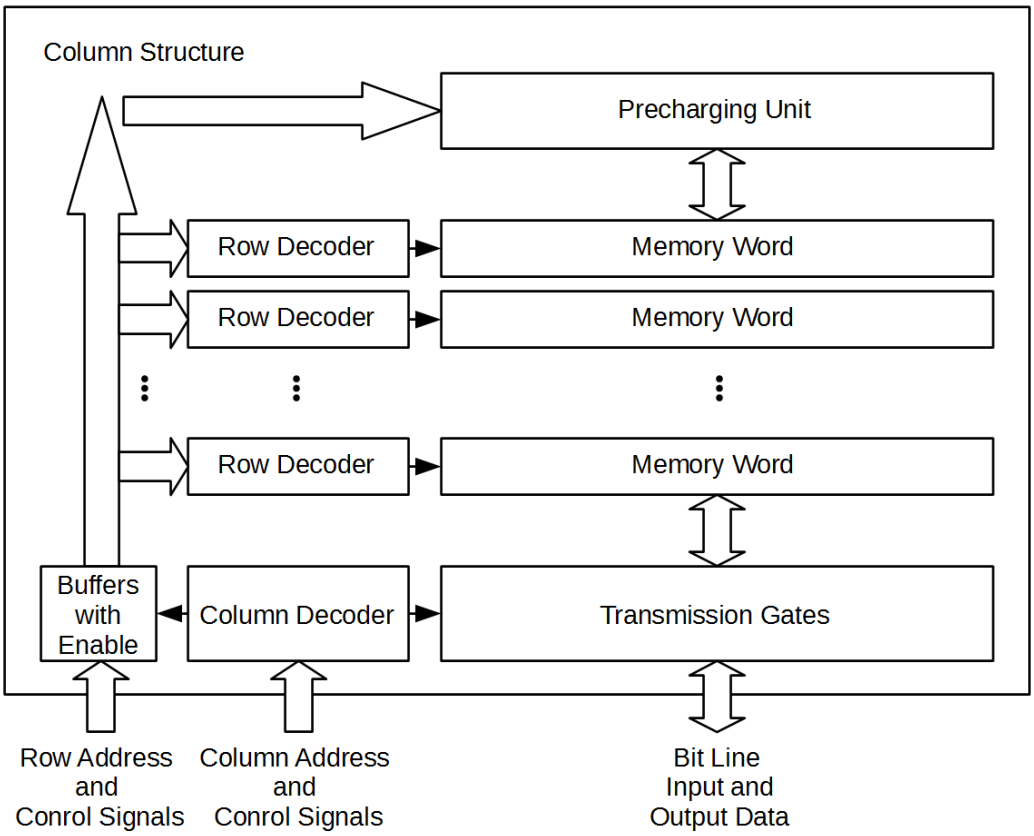

Figure 2.13: Structure of one column of the memory chip

structure also increases the achievable operating frequency. There exist coarser local word line assertion designs in literature [19], but structure used in this work enables another important possibility for imaging applications: multi-port operations. This work in current setup is not fully compatible for multi-port operations; however, with small changes in column structure as write lock and read lock, and multiple interconnected timing generators that are aware of these locks, implementation of multi-port operations is possible.

CHAPTER 3

TEST CHIP IMPLEMENTATION

For the purpose of verifying the simulation results and the design ideas, it is decided to fabricate a test chip. As the main blocks were explained in detail in the previous sections, no more explanation will be provided in this chapter. However, additional side blocks will be explained throughout this chapter.

3.1 Input Latches

In the test chip, for the sake of proper timing of data and address bits, input latches were used. Although naming them as latches, indeed these structures are negative edge triggered flip-flops with asynchronous reset capabilities. Input latches take address and data bits at negative clock edge beforehand the planned operation. This structure presents one clock cycle of latency at the desired operation, but throughput stays same, one operation per clock cycle. In an ASIC chip there exists no need to use this structure, however in a test chip it makes operations easier and eliminates most of the timing problems that might occur outside the test chip.

3.2 Output Latches

Output latches needed for both test chip and for the final implementation inside ASIC. The reason of this need is the output structure of the sense amplifiers. Sense amplifiers needed to precharge to supply voltage before operation, which make their outputs equal to ground level for at least half of the clock cycle. In order to make output

data available throughout a clock cycle, there exists an output latch structure at the output of the sense amplifiers. Output latches trigger with their own signal, derived from sense amplifier enable signal, in order to obtain proper timing. One drawback of input latched address and data combined with output latches is the two clock cycle latency when getting the first output data. Nevertheless, in that case also throughput still stays the same; one operation per clock cycle.

3.3 Write Buffer

These are tristate buffers for input data driving. However, as there exist transmission gates on the column structure, usage of the buffers need extreme caution. However, pulsed synchronous operation also helps to stabilize the input states before column and word line enabled to write. No discrepancies created, at the same time power consumption reduced, as buffers are disabled when there is no need.

3.4 Address and Control Signals Buffers

These are standard buffers that could change according to size of the total array. There exist differently sized buffers for different signals, as the load differs for every different signal.

3.5 Top Level Integration

For top level integration, there is a need for a timing generator, and one words of write buffer along with one words of sense amplifier initially. Then with the needed number of columns, and buffers that would be driving the total load of these columns needed. In this case it is 16 columns, which equals to 64 Kb of SRAM. The size of the total structure is not enough to store a whole frame, but it is enough to store more than three lines of 1024 horizontal pixel coded with 16 bits (indeed 4 lines can be stored). As having three consecutive lines enable spatial image processing for a pixel, this size can be useful in spatial image processing tasks. The illustration

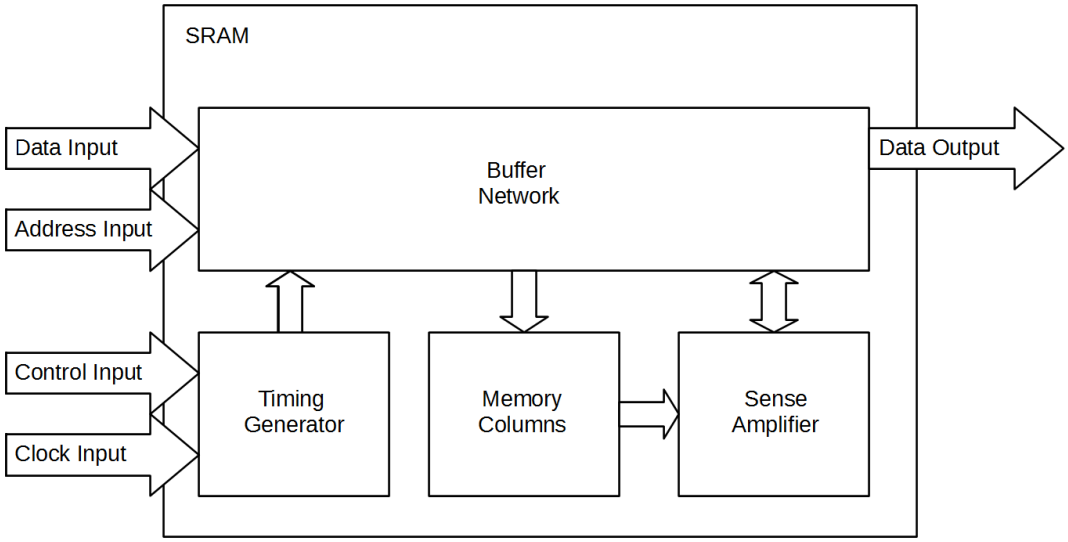

Figure 3.1: Structure of the SRAM memory chip

of a pixel neighborhood can be seen in the Figure 3.2. Also for the sake of easy synchronization, input address and data and output data are latched (not shown on Figure 3.1 as these are generic blocks, could be thought as latches are inside buffer network). This approach gives an overall one clock cycle latency, but guarantees proper operations as a result. Also, giving the data and address latched did not change the throughput of the circuitry, still operation per clock is realized. The schematics are ready for the implementation, simulations were done and layout is finished. General structure of the memory chip can be seen as in the Figure 3.1. Finished layout of the memory can be seen on Figure 3.3 for internal placements, and also Figure 3.4 for pad placements, . Pad names with respect to their numbers can be found in Table 3.1.

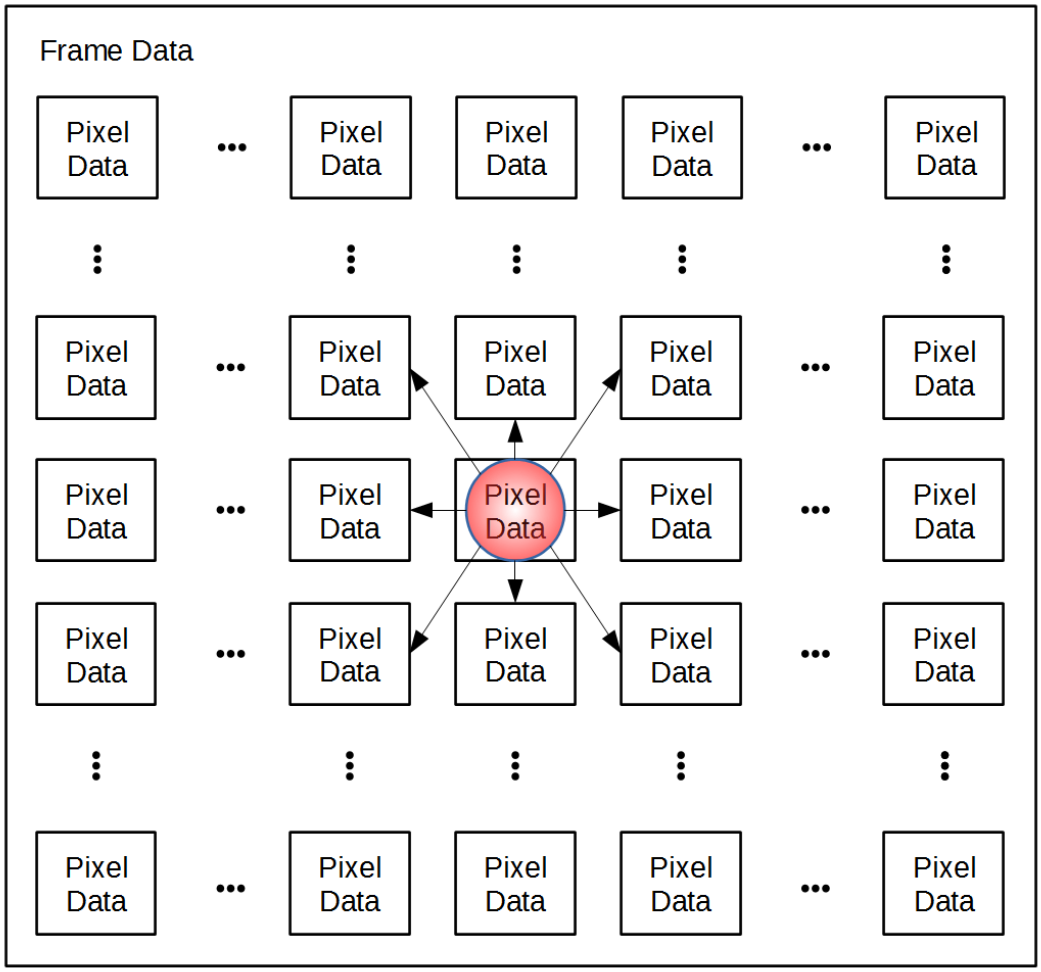

Figure 3.2: Eight connected neighborhood of a pixel inside a frame

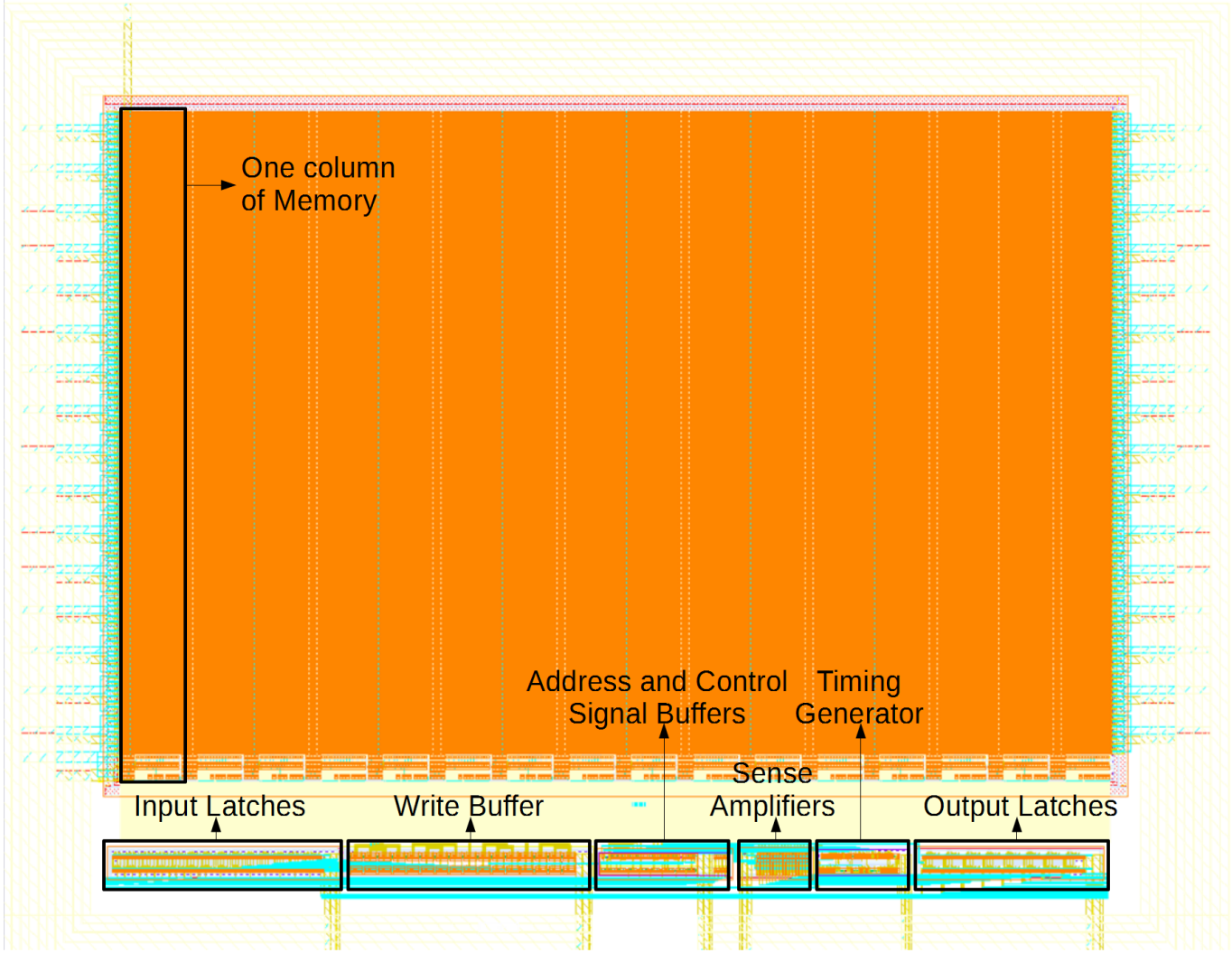

Figure 3.3: Core layout of the SRAM memory chip with internal placements (x=2850 µm, y=2230 µm)

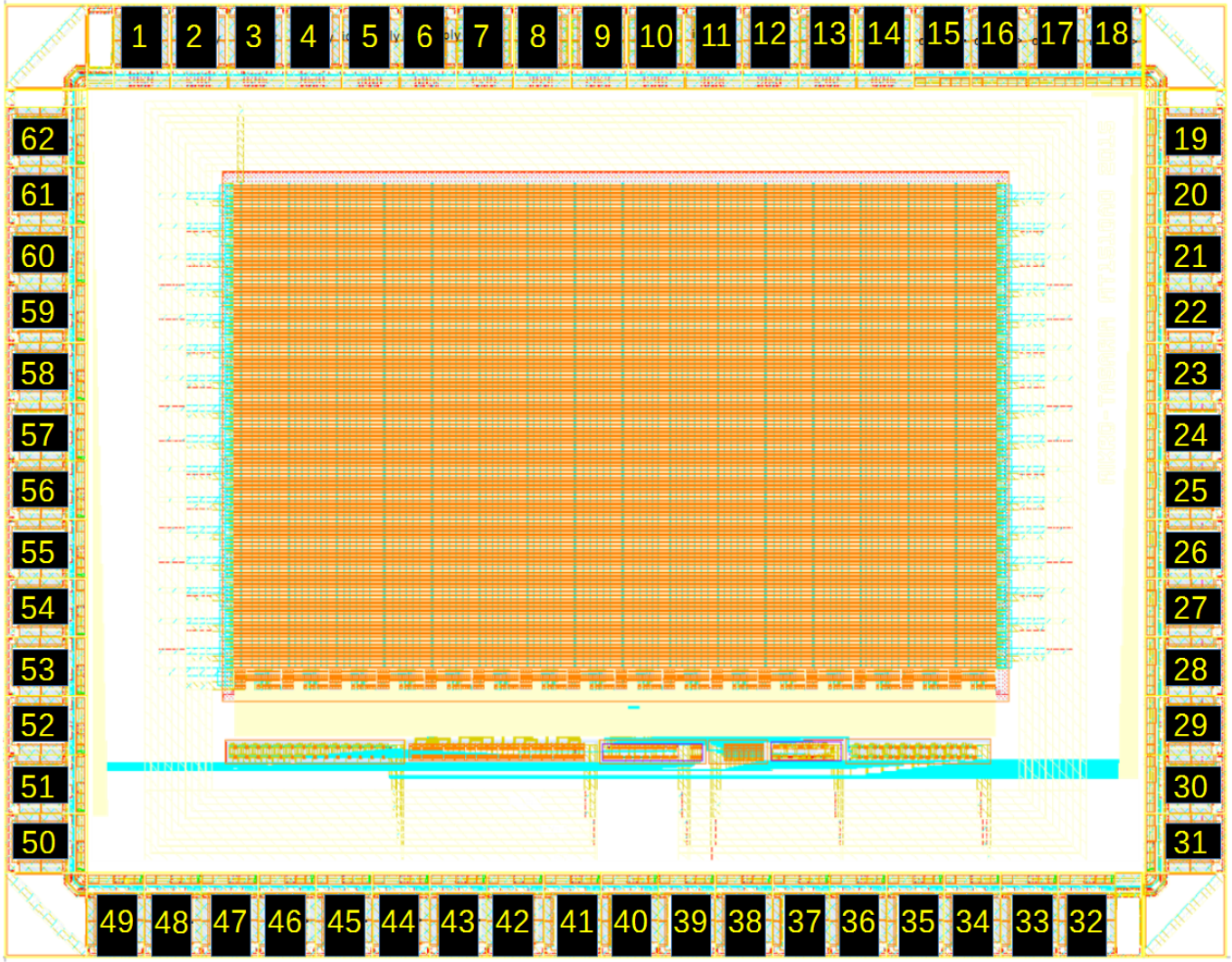

Figure 3.4: Top layout of the SRAM memory chip with pad placement (x=3700 µm, y=2800 µm)

Table 3.1: Pad placement of the test chip in clockwise order

Pad No	Pad Name	Pad No	Pad Name	Pad No	Pad Name	Pad No	Pad Name
1	Supply	17	Output(13)	33	Input data(0)	49	Operation
2	Supply	18	Output(12)	34	Input data(1)	50	Enable
3	Supply	19	Output(11)	35	Input data(2)	51	Adress(0)
4	Supply	20	Output(10)	36	Input data(3)	52	Adress(1)
5	IO supply	21	Output(9)	37	Input data(4)	53	Adress(2)
6	IO supply	22	Output(8)	38	Input data(5)	54	Adress(3)
7	Ground	23	Output(7)	39	Input data(6)	55	Adress(4)
8	Ground	24	Output(6)	40	Input data(7)	56	Adress(5)
9	Ground	25	Output(5)	41	Input data(8)	57	Adress(6)
10	Ground	26	Output(4)	42	Input data(9)	58	Adress(7)
11	IO ground	27	Output(3)	43	Input data(10)	59	Adress(8)
12	IO ground	28	Output(2)	44	Input data(11)	60	Adress(9)
13	ESD	29	Output(1)	45	Input data(12)	61	Adress(10)
14	ESD	30	Output(0)	46	Input data(13)	62	Adress(11)
15	Output(15)	31	Clock in	47	Input data(14)		
16	Output(14)	32	Resetb in	48	Input data(15)		

CHAPTER 4

SIMULATION RESULTS AND TEST SETUP

4.1 Bitcell Characteristics

Simulation results for bitcell shows the leakage current, and write times of the bitcell within various supply voltage levels. They can be seen in Table 4.1. Also read SNM illustrations for the selected supply voltages can be seen on the Figure 4.1 and Figure 4.2. The variability of the threshold voltages is a great problem in the small channel area [3], so the simulations in this part are especially performed using 6 sigma variations. SNM values for supply at 1.2 Volts is 0.21 Volts and it is 0.31 Volts for the 1.8 Volts case.

4.2 Address Decoder Characteristics

For the address decoders, there was a need for 9-input nand gate, and the possible solution can be 3 NANDs and one NOR, but minimum sized NOR3 is not good for its very weak pull up, so it has to be at least 2 times the minimum size of the p

Table 4.1: Bitcell write time and leakage current values at various supply voltages

Voltage	Write Time	Leakage Current
1 V	2 ns	3.11 pA
1.2 V	0.82 ns	3.72 pA
1.8 V	0.38 ns	5.55 pA
2.2 V	0.32 ns	6.78 pA

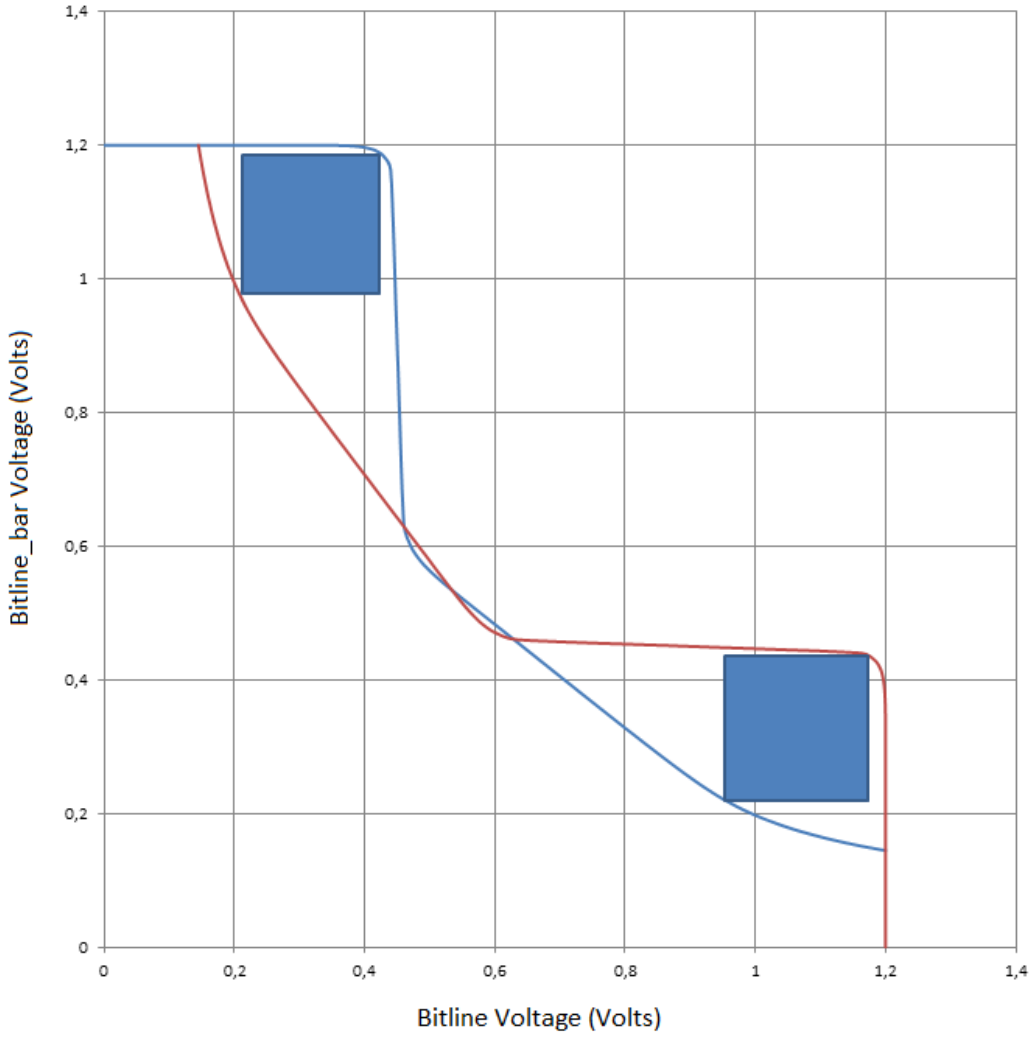

Figure 4.1: Read SNM Plot of the bitcell at 1.2 Volts supply voltage

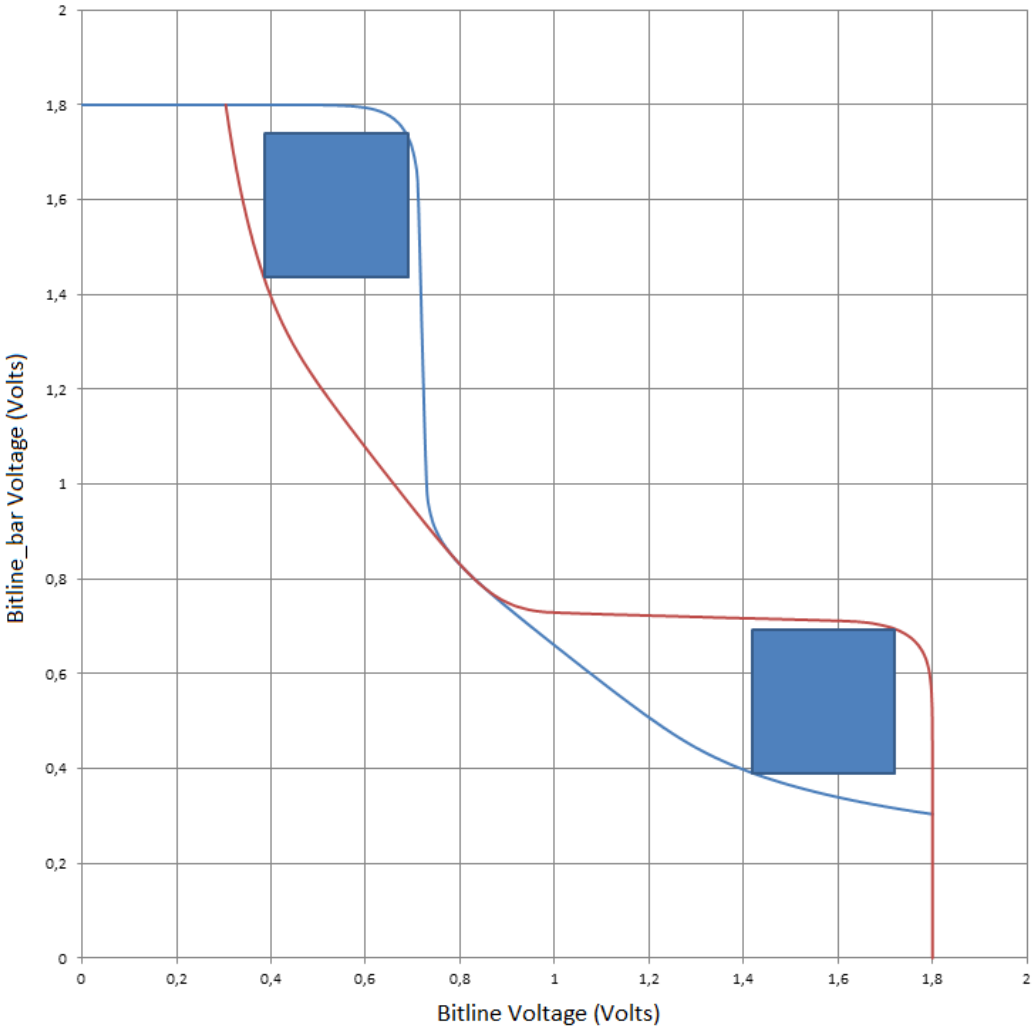

Figure 4.2: Read SNM Plot of the bitcell at 1.8 Volts supply voltage

Table 4.2: Rise and delay times of building block gates for address decoders at selected supply voltages

Voltage	AND3 Rise Time	AND3 Delay Time	NOR3 Rise Time	NOR3 Delay Time
1.2 V	3.34 ns	2.56 ns	5.90 ns	4.06 ns
1.8 V	1.25 ns	1.07 ns	2.20 ns	1.40 ns

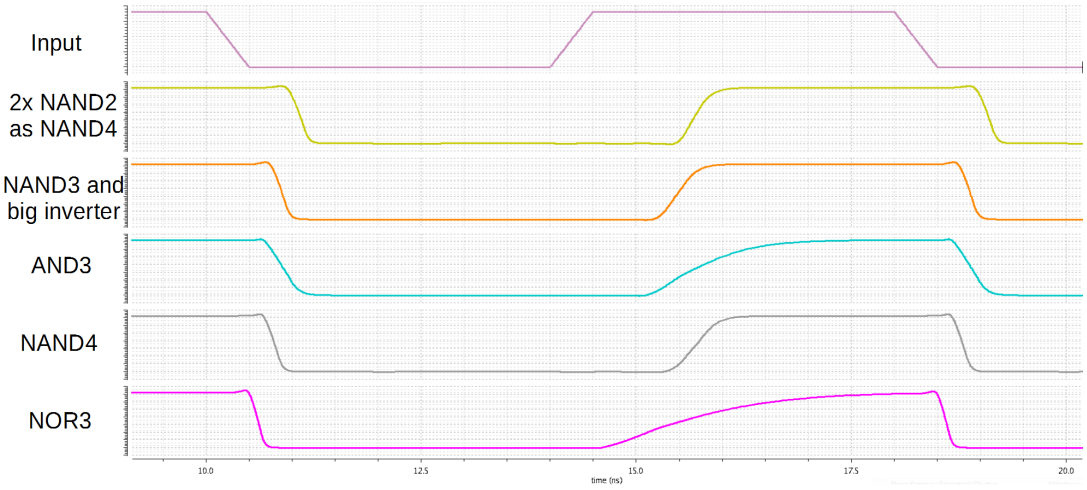

Figure 4.3: Simulations of address and row decoder possibilities at 1.8 Volts supply and 0.5 ns rise time

type transistor. Nevertheless, as it can be seen on the Table 4.2, it is not better than an AND3 created with one NAND3 and an inverter. As the simulated NOR3 uses double sized PMOS for pull-up, the area advantage of the NOR3 no longer exists, and it leads to choice of AND3 for the building block of the decoders. Various probabilities of the address decoders named as their second level gate simulated. End results and comparison of the signals can be seen on Figure 4.3 and Figure 4.4. Address decoders that have 4-inputs at the second level are candidates for the column address decoder, as their rise time fires more signals in the column and more address space is needed for scalability. However, decoders that have 3-input gates at the second level are candidates for row address decoder structures.

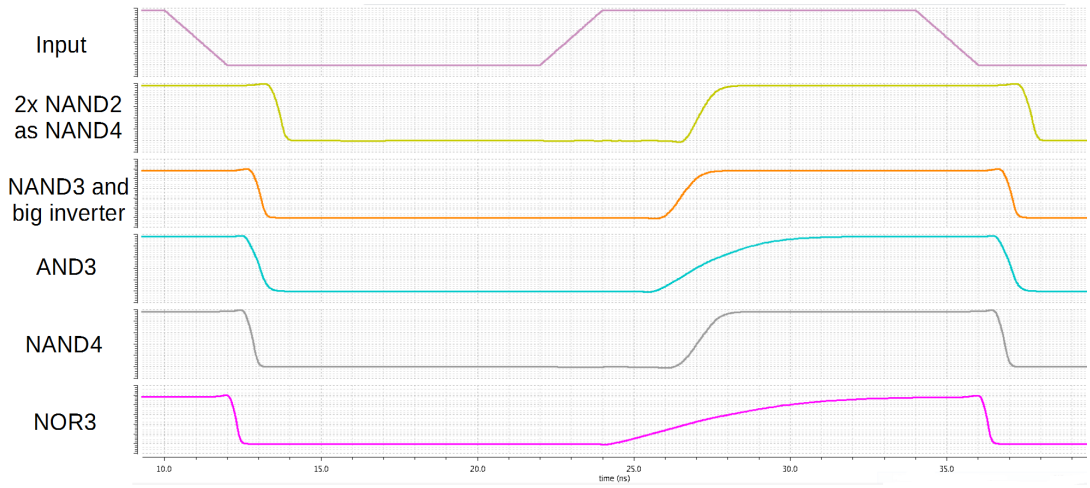

Figure 4.4: Simulations of address and row decoder possibilities at 1.2 Volts supply and 2 ns rise time

4.3 Timing Generator Characteristics

The signals created with the timing generator can be seen on the simulation results. In the simulations, the same signals were created, but with changing the frequency, and the supply voltages, there exists different rise and fall times. Results can be seen on Figure 4.5.

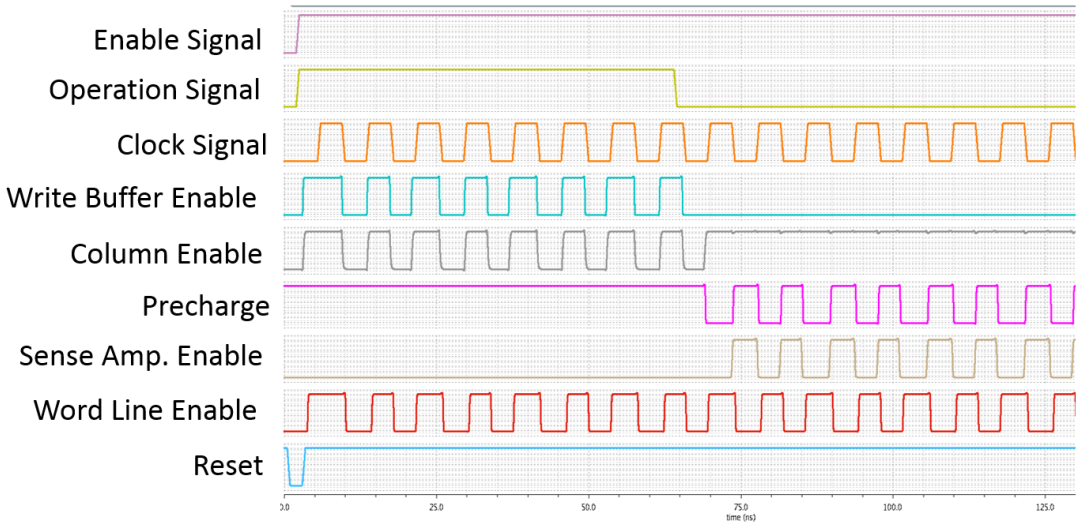

Figure 4.5: Outputs of the timing generator at 125 MHz of clock input and changing control inputs

4.4 Column Read and Write Simulations

In these simulations, three different frequency and two different supply voltages were used. Signals on the simulations results are clock, chip enable and operation signals, with changing order. Also at the bottom of results, digitalized input data, output data and address input can be seen with changing orders. As one column write and read simulation setup do not include input and output latches, no latency exists at the input side, but expectedly, output data still come at the end of clock pulse, but one clock less latency than the test chip implementation. Also, output data are available for only small time with respect to a clock cycle. Thus, in order to show read data at the second part of the simulations, magnified versions of the waveform at read operations added. The results can be seen in the Table 4.3. Also, in order to compare the topology with a standard voltage sense amplifier, there exist simulation results in the Table 4.4 in addition to the simulation results for current topology.

Table 4.3: Average power consumption of the column in various cases

Voltage	Frequency	Read Current	Read Power	Write Current	Write Power
1.2 V	42 MHz	1.07 mA	1.28 mW	0.48 mA	0.58 mW
1.8 V	125 MHz	4.75 mA	8.55 mW	2.55 mA	4.59 mW
	62.5 MHz	2.76 mA	4.97 mW	1.33 mA	2.39 mW

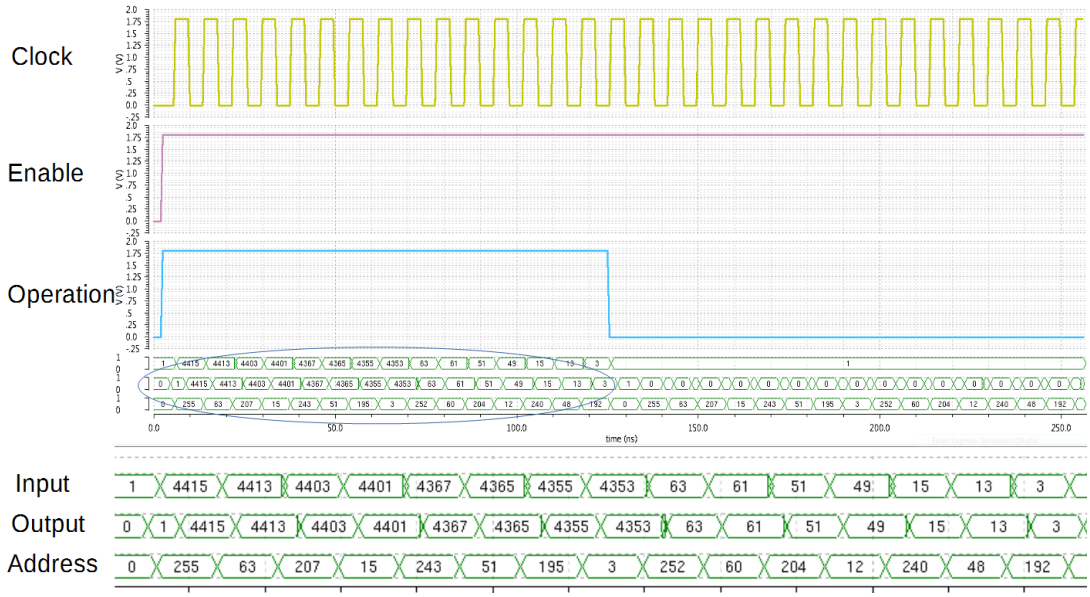

Figure 4.6: Waveforms and digitalized data for read and write phase simulations of one memory column at 125 MHz clock and 1.8 Volts supply

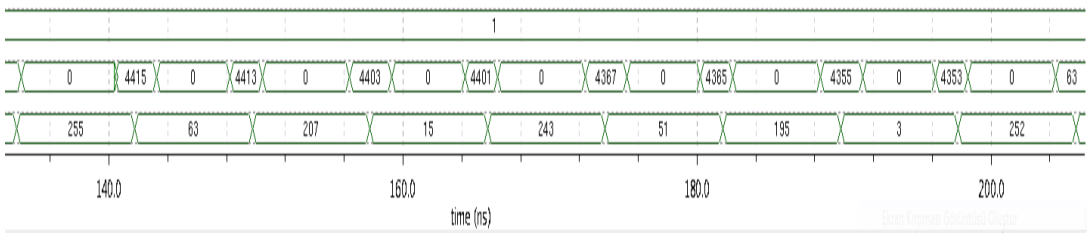

Figure 4.7: Waveforms and digitalized data for read and write phase simulations of one memory column at 125 MHz clock and 1.8 Volts supply, zoomed to first half of the batch reading

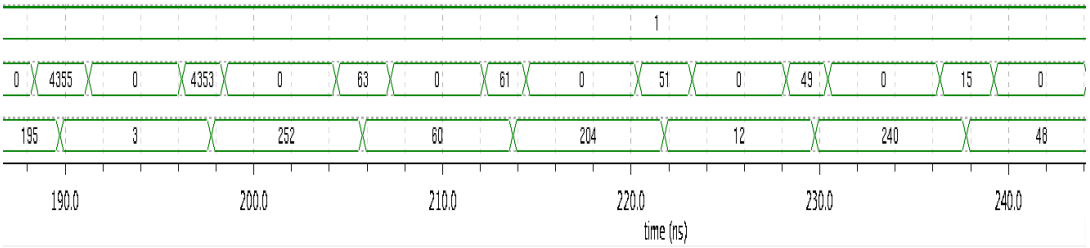

Figure 4.8: Waveforms and digitalized data for read and write phase simulations of one memory column at 125 MHz clock and 1.8 Volts supply, zoomed to second half of the batch reading

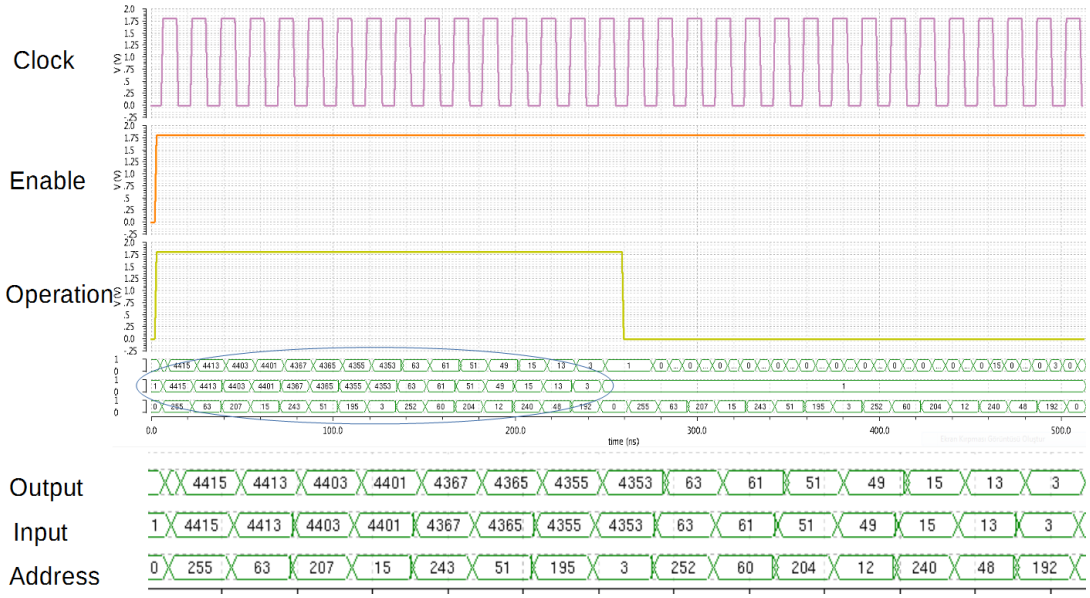

Figure 4.9: Waveforms and digitalized data for read and write phase simulations of one memory column at 62 MHz clock and 1.8 Volts supply

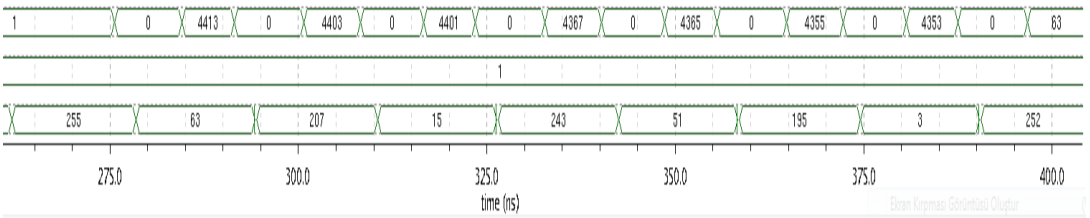

Figure 4.10: Waveforms and digitalized data for read and write phase simulations of one memory column at 62 MHz clock and 1.8 Volts supply, zoomed to first half of the batch reading

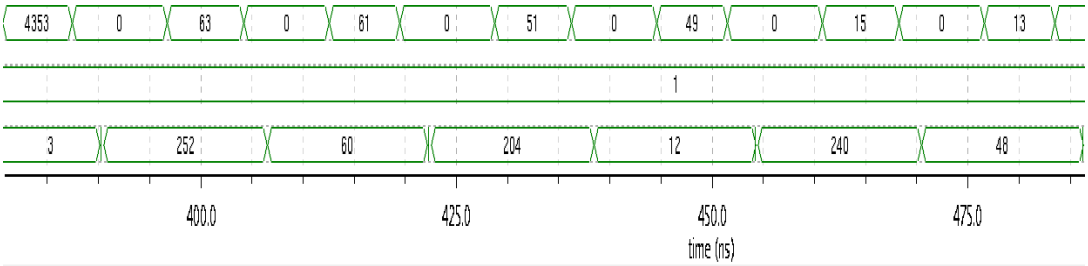

Figure 4.11: Waveforms and digitalized data for read and write phase simulations of one memory column at 62 MHz clock and 1.8 Volts supply, zoomed to second half of the batch reading

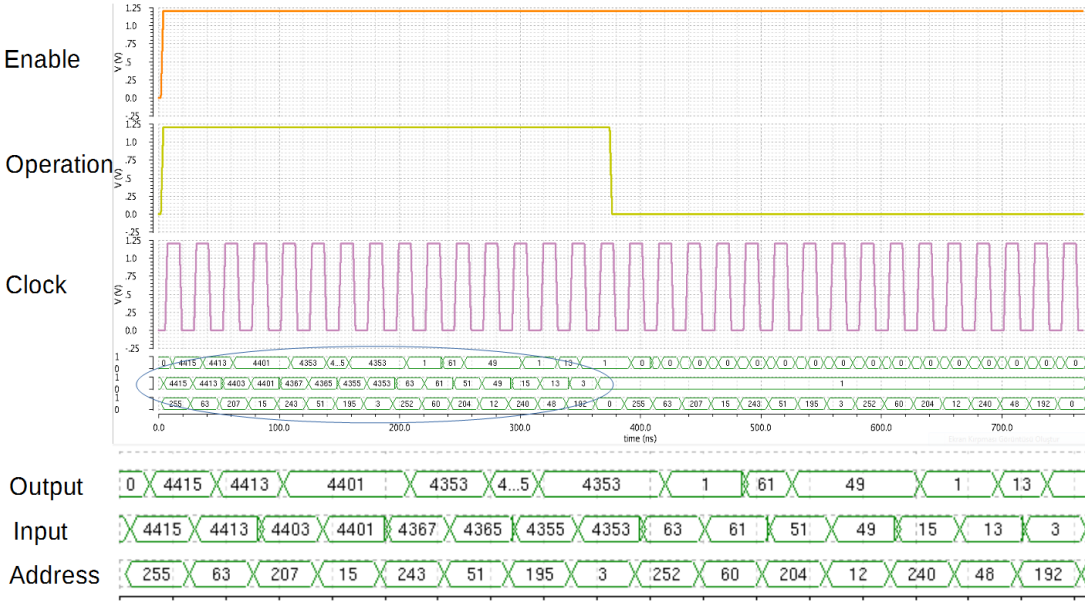

Figure 4.12: Waveforms and digitalized data for read and write phase simulations of one memory column at 42 MHz clock and 1.2 Volts supply

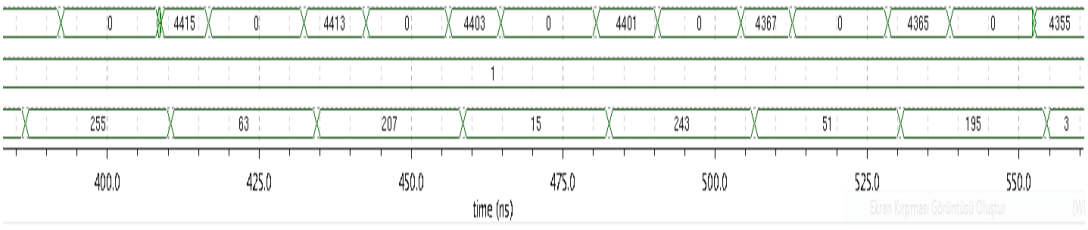

Figure 4.13: Waveforms and digitalized data for read and write phase simulations of one memory column at 42 MHz clock and 1.2 Volts supply, zoomed to first half of the batch reading

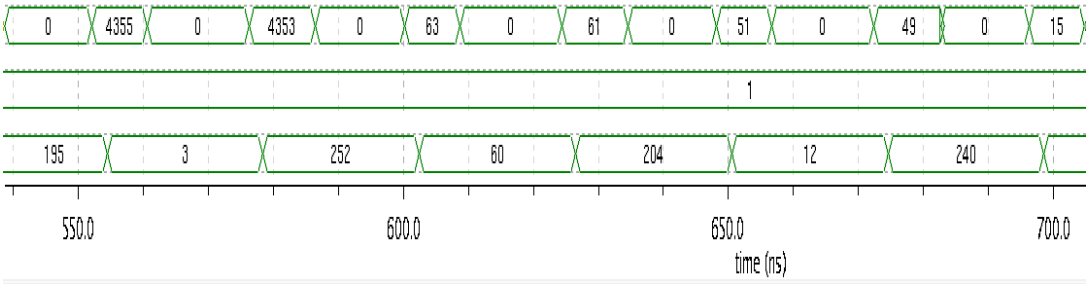

Figure 4.14: Waveforms and digitalized data for read and write phase simulations of one memory column at 42 MHz clock and 1.2 Volts supply, zoomed to second half of the batch reading

Table 4.4: Average power consumption of the column with different sense amplifier topologies

Voltage	Frequency	Read Current	Read Power	Write Current	Write Power
With Standard Voltage Sense Amplifier					
1.2 V	42 MHz(max)	1.27 mA	1.52 mW	0.49 mA	0.58 mW
1.8 V	125 MHz(max)	5.49 mA	9.88 mW	2.55 mA	4.59 mW
1.8 V	62.5 MHz	3.30 mA	5.94 mW	1.33 mA	2.39 mW
With Current Sense Amplifier					
1.2 V	42 MHz(max)	1.07 mA	1.28 mW	0.49 mA	0.58 mW
1.8 V	125 MHz(max)	4.75 mA	8.55 mW	2.55 mA	4.59 mW
1.8 V	62.5 MHz	2.76 mA	4.97 mW	1.33 mA	2.39 mW

Table 4.5: Average power consumption of the chip core compared with one column and 128 column array in various cases

Voltage	Frequency	Read Current	Read Power	Write Current	Write Power
One Column Data(16*256 bits)					
1.2 V	42 MHz(max)	1.27 mA	1.52 mW	0.49 mA	0.58 mW
1.8 V	125 MHz(max)	5.49 mA	9.88 mW	2.55 mA	4.59 mW
1.8 V	62.5 MHz	3.30 mA	5.94 mW	1.33 mA	2.39 mW
Test Chip Core (16 Columns, 256*256 bits)					
1.2 V	42 MHz(max)	1.07 mA	1.28 mW	0.49 mA	0.58 mW
1.8 V	125 MHz(max)	4.75 mA	8.55 mW	2.55 mA	4.59 mW
1.8 V	62.5 MHz	2.76 mA	4.97 mW	1.33 mA	2.39 mW
128 Column Array (2048*256 bit)					
1.8 V	125 MHz(max)	7.81 mA	14.06 mW	4.14 mA	7.45 mW
1.8 V	62.5 MHz	6.34 mA	11.41 mW	2.12 mA	3.82 mW

4.5 Top Level Simulations

In the top level simulations, input address, input data and two control signals were fed to the circuitry, then latched. Also outputs were latched, because when circuitry is working for the next output, previous output should not go away. The first part of the simulation shows writing to the memory, second part shows reading from these written places, showing proof of both writing and reading from the same place. Simulation setup and waveforms can be seen on Figures ranging from Figure 4.15 to Figure 4.17. Note that, when writing and reading, one clock pulse latency exists because of the input data and address latches combined with output latches, but throughput is the same. Also for easy comparison, and to show scalability, one column simulation data and 128 column array of simulation data added to test chip's simulation results. They can be seen on Table 4.7. Test chip at 125 MHz is able to make six operations for one pixel inside a VGA imaging sensor that have 16 bits per pixel. Total bandwidth required for this sensor at 60 Hz frame rate is 1.77 Gb/s for four read and two write operations. Design gives 2 Gb/s at 125 MHz and sufficient for the duty.

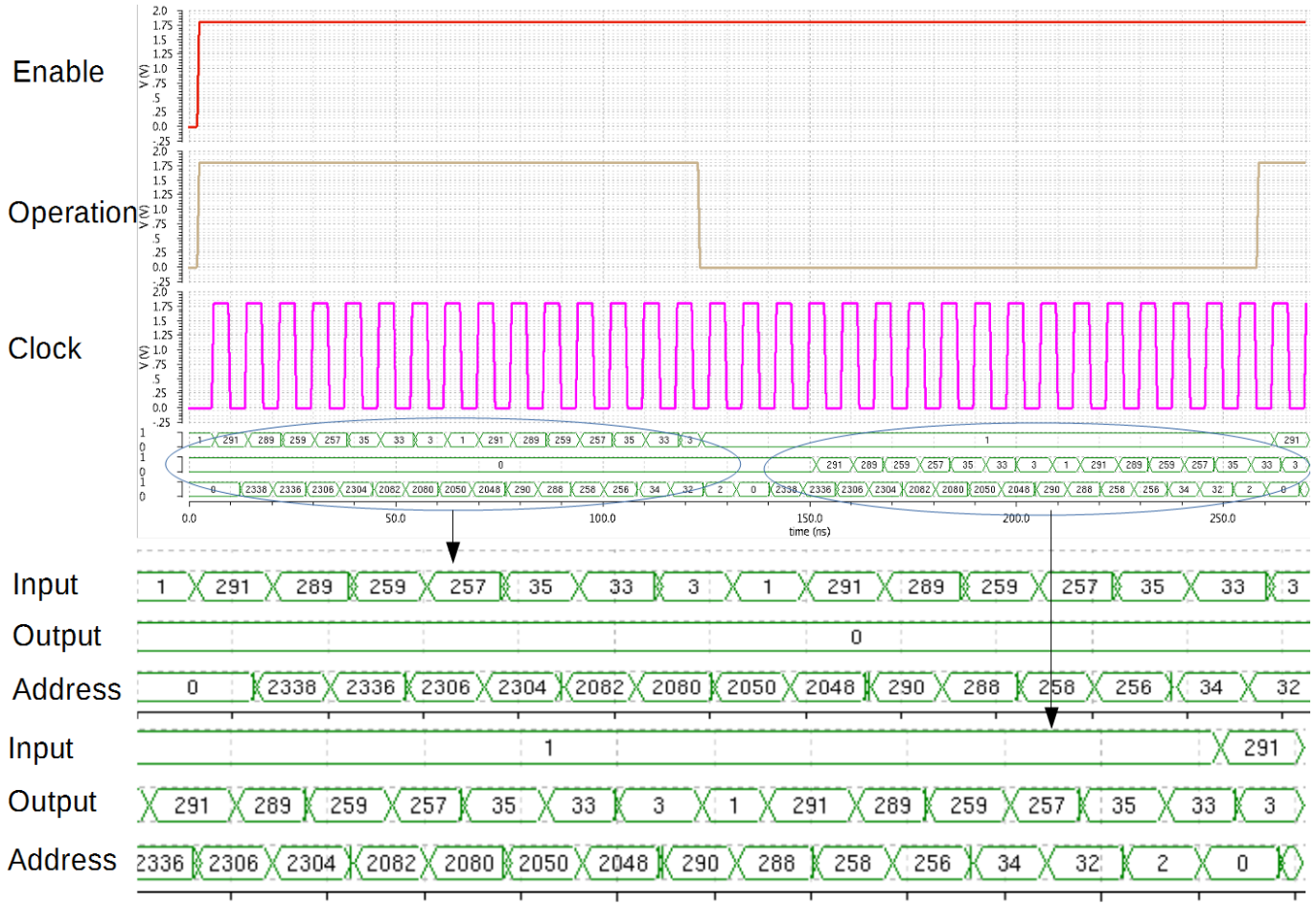

Figure 4.15: Waveforms and digitalized data for read and write phase simulations of top level memory core at 125 MHz, 1.8 Volts

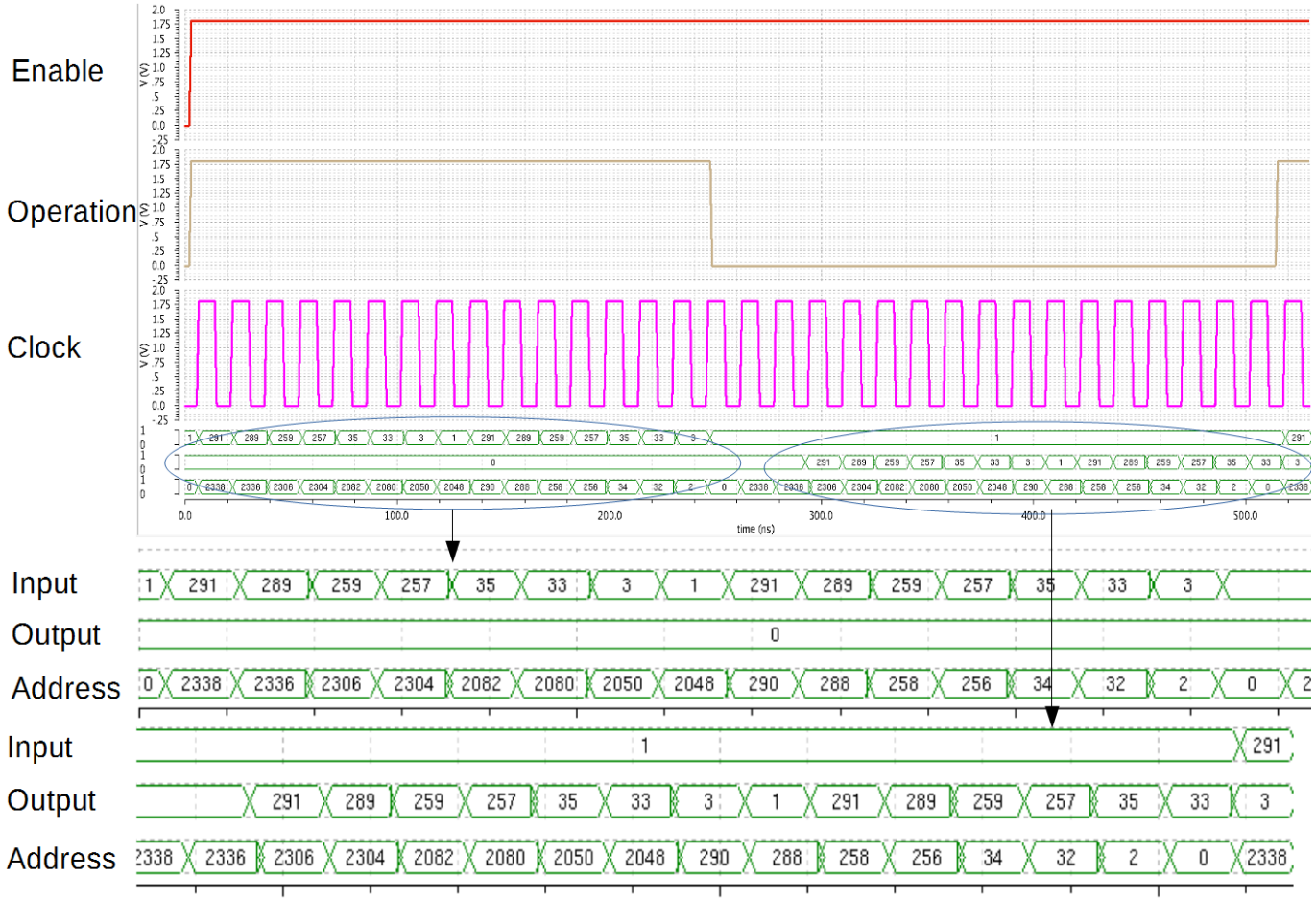

Figure 4.16: Waveforms and digitalized data for read and write phase simulations of top level memory core at 62 MHz, 1.8 Volts

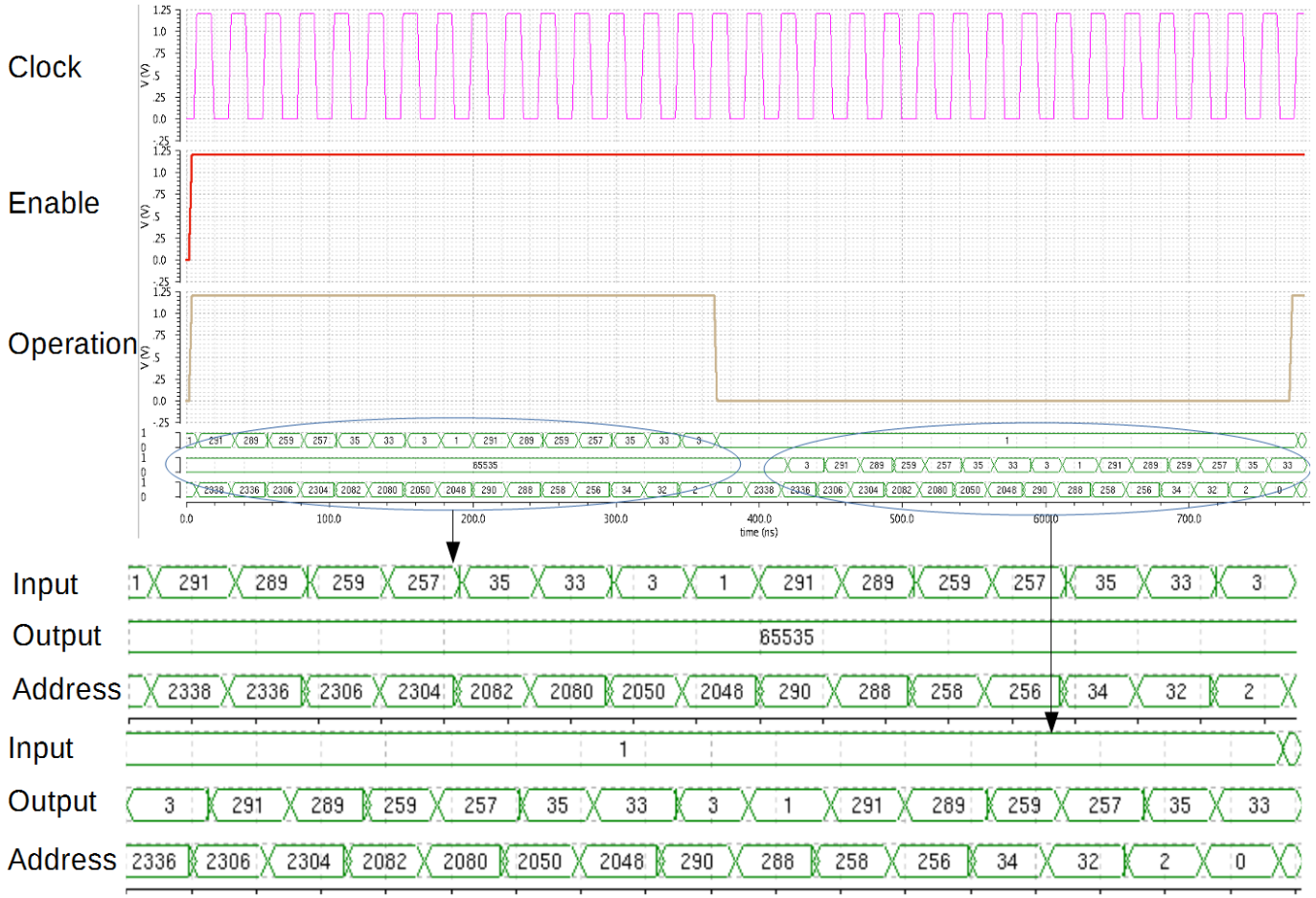

Figure 4.17: Waveforms and digitalized data for read and write phase simulations of top level memory core at 42 MHz, 1.2 Volts

4.5.1 Comparison with previous works

An example work carried out in the same node achieves 2.7mW of active power at 1V supply voltage at 10 MHz of operation [15]. It is clear that, at 1.2 Volts of supply voltage and greater frequency (42 MHz), this works' power consumption is superior. However, in order to fully understand the merits of the circuitry, there is a need for more recent works. For eliminating the scaling factor of capacitances in the more current works, and the state of the art technology, the original values and the scaled values of the newer circuitries would be supplied. As an example, in the work [6], an SRAM memory constructed with 90 nm technology was the subject. The original values for the energy consumption per access (read or write) at 240 MHz is nearly symmetrical, and similar (2638 fJ for read, 2748 fJ for write and 2675 fJ for average respectively). In order to eliminate different node barriers one could convert power data in 0.35 um technology to 90 nm technology. In that case, equivalent write and read energies of this work found to be 1058 fJ and 2447 fJ respectively. This work is clearly better in both operations. In another work, which was conducted with the cutting edge FinFET technology at 14nm node, there exists write and read energy values for its proposed circuitry [12]. Values for write and read operations are 28.7 fJ and 20.4 fJ respectively. Scaling again from 0.35 um to 14nm yields comparable results. In this work equivalent energies for writing and reading found to be 26 fJ and 59 fJ respectively. It can be seen clearly that, this work has better write energy, but improvements in read department could give the design an edge. Also one can conclude that, the most important player on power consumption game is the technology node that the design is on. All of the original energy, and their equivalents on the 14 nm technology can be found on Table 4.6, in order to compare easily each work

Table 4.6: Operation energy comparison with previous works, capacitance scaled values are compararable with cutting edge devices that designed at 14 nm node

Work	Original Write Energy	Original Read Energy	Scaled Write Energy	Scaled Read Energy
1998 ($0.35\,\mu\mathrm{m}$ node) [15]	270 pJ	270 pJ	432 fJ	432 fJ
2009 (90 nm node) [6]	2675 fJ	2675 fJ	64 fJ	64 fJ
2015 (14 nm node) [12]	28.7 fJ	20.4 fJ	28.7 fJ	20.4 fJ
This work	16 pJ	37 pJ	26 fJ	59 fJ

Table 4.7: Average power consumption of the chip with pad circuitry in various cases

Voltage	Frequency	Read Current	Read Power	Write Current	Write Power
1.8 V	125 MHz	9.00 mA	16.2 mW	4.45 mA	8.01 mW
	62.5 MHz	5.75 mA	10.35 mW	2.59 mA	4.66 mW

4.6 Top Level Simulations with Input Output Load Modeling and Pad Circuitry

In these simulations, a total test chip and card organization setup are to be tested. In order to be sure about test chip's functionality in real world, pad circuitry, FPGA input capacitances, core to pad capacitances inside circuitry, and wirebond model for all the circuit pads were added. Wirebond model used in simulations were straight wire inductance model.

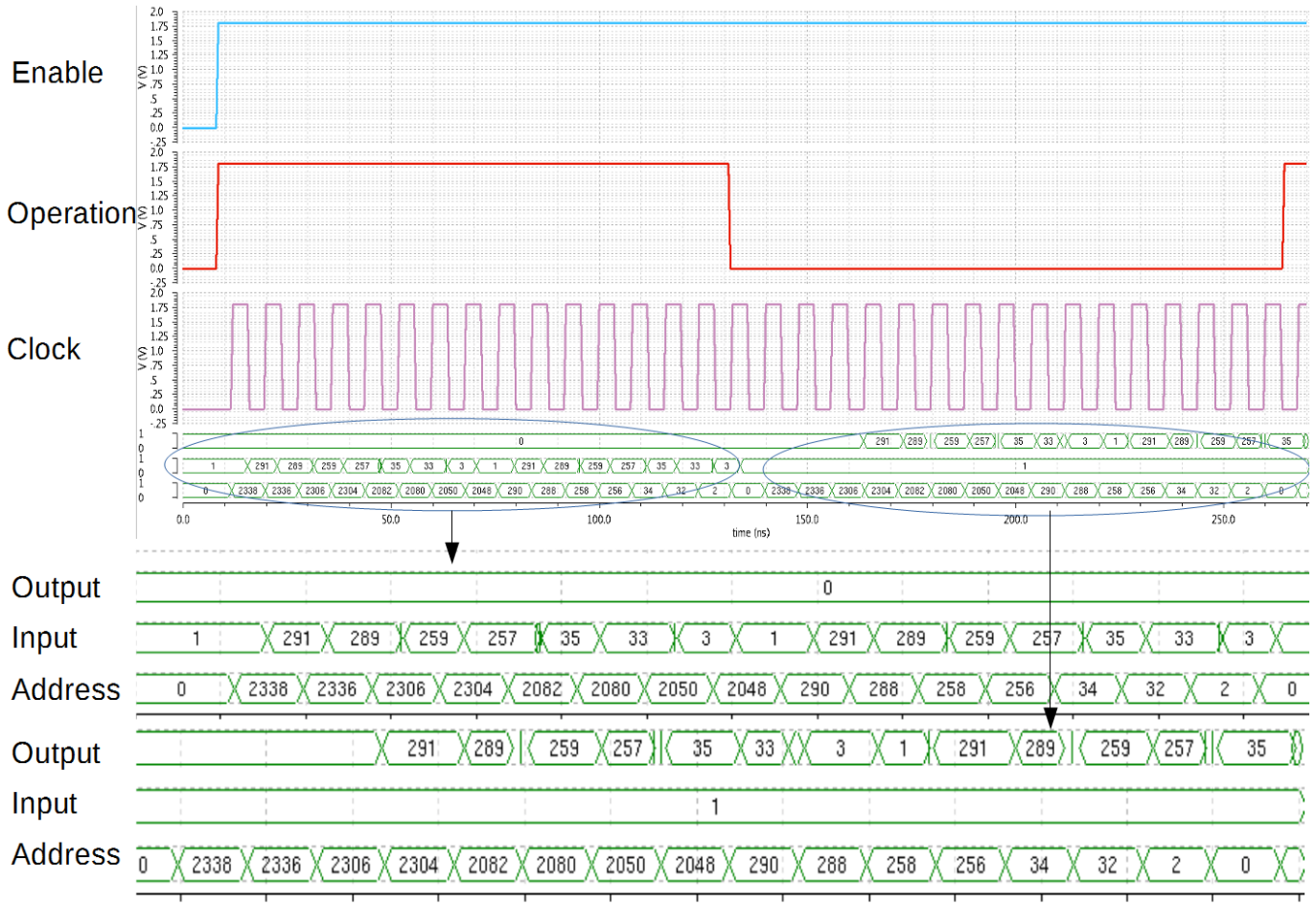

Figure 4.18: Waveforms and digitalized data for read and write phase simulations of memory with loads and straight wire wirebond model at 125 MHz, 1.8 Volts

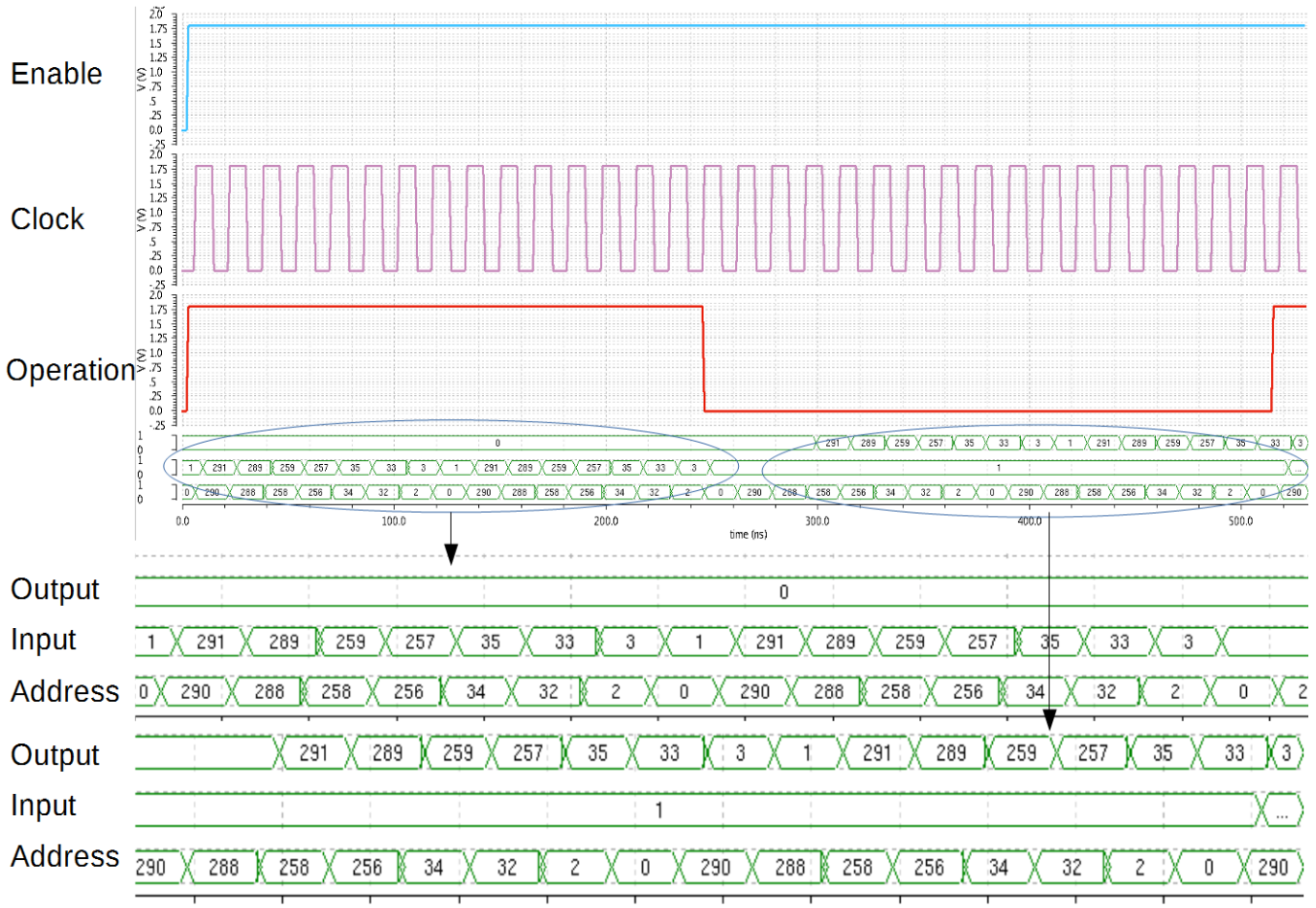

Figure 4.19: Waveforms and digitalized data for read and write phase simulations of memory with loads and straight wire wirebond model at 62 MHz, 1.8 Volts

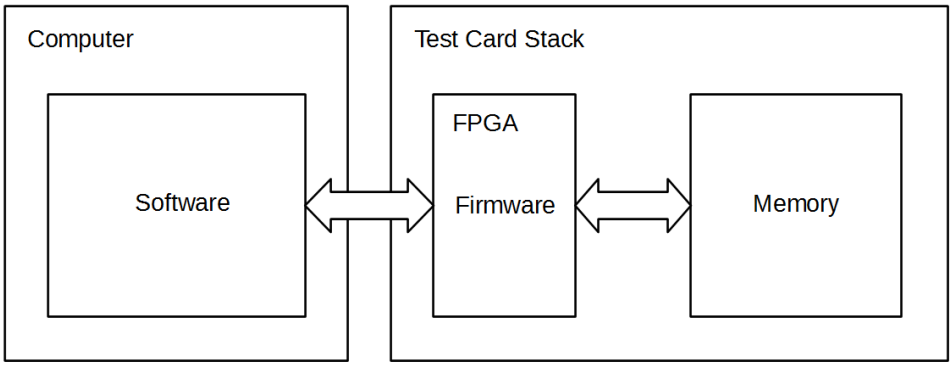

Figure 4.20: Proposed test setup for the memory chip

4.7 Test Setup

In this part, proposed test setup of the SRAM chip is discussed. Setup consists of a proximity card, test software and firmware. Details of the particular subject will be explained throughout following sections. Related illustration can be found on Figure 4.20

4.7.1 Test Card

Test card includes basic elements to allow test chip to operate. Voltage regulators and auxiliary parts are needed to make chip work. Also, as data, address and control inputs creates a wide array of data, an FPGA card also needed to drive the chip with the help of the firmware and software. The card should be designed with a PCB design tool. There should be one power input, and different voltage values will have generated from that input with the help of the voltage regulators.

4.7.2 Software and Firmware

Test chip is a memory device so, writing and reading is the fundamental operations to be done. Also, health of the particular chip can be noted to a file, in order to verify all rows and columns of the chip by means of address input and comparison of written

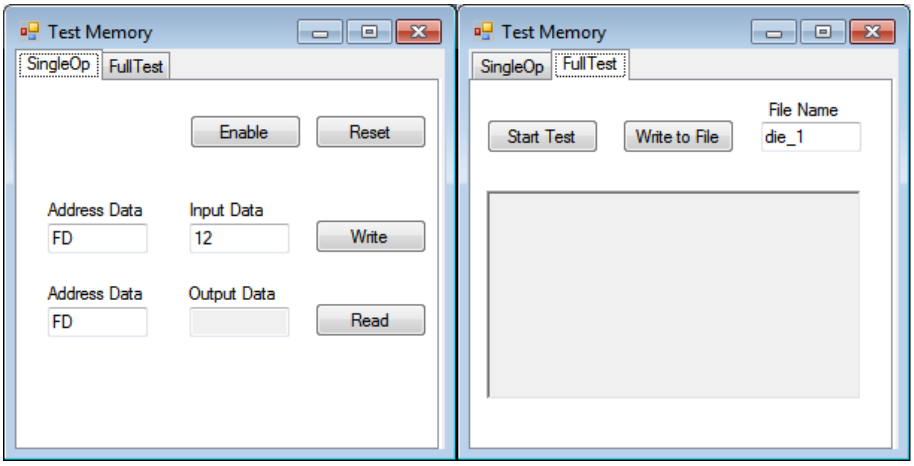

Figure 4.21: GUI elements of the test software, single test and full chip test respectively

data with read value for a particular address.

4.7.2.1 Firmware

Basic write and read operation should be realized on the FPGA environment. The reason of that is the atomic and sequential nature of the low level hardware. Inputs and outputs should be taken care clock by clock. After enough data accumulated, then for displaying and storing purposes, it should be transported to the software and properties of the chip, especially the errors should be stored, and evaluations of the chips should have made accordingly. All firmware programming will be done by means of Verilog HDL.

4.7.2.2 Software

Batch operations and accumulated data can be seen and stored on the software part of the test environment. Also, it should let one operation a time, in order to see the immediate result manually, if desired. Also non-operational row and column data should be saved to a file, which also includes the part number of the particular chip. The software part of the test setup is realized by Visual C# programming environment. GUI (Graphical User Interface) of the test software can be seen on Figure 4.21

CHAPTER 5

SUMMARY AND CONCLUSIONS

The research conducted in this thesis deals with an SRAM designed for spatial image processing in mind. There exists an easily scalable structure, constructible in sizes ranging from 4 Kbits to 2 Mbits. These capacity values equal to one column (16x256) and 512 columns (8192x256) respectively. In order to reduce power consumption in this design, some precautions taken as local word line assertion and pulsed synchronous operation. Local word line assertion technique reduces driving capacitances and allows unnecessary parts of the circuitry continue to sleep, i.e. no dynamic changes in these parts occur. Also, pulsed synchronous operation prevents circuitry from all kind of glitches that may cause false selections of addresses that increase dynamic power consumption, and that may lead to errors. Also one of the power saving technique, namely local word line assertion enables a very important opportunity for imaging purposes: multi-port operations. Although current setup of the design does not allow error-proof multi-port operations, small changes inside column and timing generator easily allow an error-free multi-port SRAM. A test chip with 64 Kbit capacity was sent to CMOS MPW run after extensive simulations. Simulations of the test chip give better converted operation energy values than previous works, except the most recent 14 nm node devices. Write operation of the design is better than the 14 nm node device in scaled energy data, but read energy is worse, improvements may needed in future works in that department. Obtained simulation power data at different supply and frequency values can be seen on Table 5.1. In order to test functionality of the circuitry, also a test setup will be prepared. A GUI is designed to show planned test operations on the manufactured memory chip.

Table 5.1: Power consumption simulation values for test chip

Voltage	Frequency	Read Current	Read Power	Write Current	Write Power
1.2 V	42 MHz(max)	1.07 mA	1.28 mW	0.49 mA	0.58 mW
1.8 V	125 MHz(max)	4.75 mA	8.55 mW	2.55 mA	4.59 mW
1.8 V	62.5 MHz	2.76 mA	4.97 mW	1.33 mA	2.39 mW

REFERENCES

[1] A. Azarpeyvand, M. Salehi, F. Firouzi, A. Yazdanbakhsh, and S. M. Fakhraie. Instruction reliability analysis for embedded processors. In *Design and Diagnostics of Electronic Circuits and Systems (DDECS), 2010 IEEE 13th International Symposium on*, pages 20–23, April 2010.

[2] V. Betz and J. Rose. VPR: a new packing, placement and routing tool for FPGA research. In W. Luk, P. Cheung, and M. Glesner, editors, *Field-Programmable Logic and Applications*, volume 1304 of *Lecture Notes in Computer Science*, pages 213–222. Springer Berlin Heidelberg, 1997.

[3] D. Burnett, K. Erington, C. Subramanian, and K. Baker. Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits. In *VLSI Technology, 1994. Digest of Technical Papers. 1994 Symposium on*, pages 15–16, June 1994.

[4] Y.-H. Chen, W.-M. Chan, W.-C. Wu, H.-J. Liao, K.-H. Pan, J.-J. Liaw, T.-H. Chung, Q. Li, C.-Y. Lin, M.-C. Chiang, S.-Y. Wu, and J. Chang. A 16 nm 128 mb SRAM in high-k metal-gate finFET technology with write-assist circuitry for low-VMIN applications. *Solid-State Circuits, IEEE Journal of*, 50(1):170–177, Jan 2015.

[5] M. Cho, J. Schlessman, W. Wolf, and S. Mukhopadhyay. Accuracy-aware sram: A reconfigurable low power sram architecture for mobile multimedia applications. In *Proceedings of the 2009 Asia and South Pacific Design Automation Conference*, ASP-DAC '09, pages 823–828, Piscataway, NJ, USA, 2009. IEEE Press.

[6] S. Cosemans, W. Dehaene, and F. Catthoor. A 3.6 pj/access 480 mhz, 128 kb on-chip SRAM with 850 mhz boost mode in 90 nm CMOS with tunable sense amplifiers. *Solid-State Circuits, IEEE Journal of*, 44(7):2065–2077, July 2009.

[7] A.-T. Do, Z.-H. Kong, K.-S. Yeo, and J. Y. S. Low. Design and sensitivity analysis of a new current-mode sense amplifier for low-power SRAM. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 19(2):196–204, Feb 2011.

[8] S. Huang, Z. Huang, A. Kurokawa, and Y. Inoue. A novel SRAM structure for leakage power suppression in 45nm technology. In *Communications, Circuits*

and Systems, 2008. ICCCAS 2008. International Conference on, pages 1070–1074, May 2008.

[9] Y. Huang and W. Messner. A novel disturbance observer design for magnetic hard drive servo system with a rotary actuator. *Magnetics, IEEE Transactions on*, 34(4):1892–1894, Jul 1998.

[10] E. Ibe, H. Taniguchi, Y. Yahagi, K. Shimbo, and T. Toba. Impact of scaling on neutron-induced soft error in SRAMs from a 250 nm to a 22 nm design rule. *Electron Devices, IEEE Transactions on*, 57(7):1527–1538, July 2010.

[11] D. James, D. Yang, and P. Gamache. Inside the iphone 6s. Technical report, Chipworks Inc., 1891 Robertson Road, Suite 500, Ottawa, ON K2H 5B7 Canada, 2015.

[12] H. Jeong, T. Kim, Y. Yang, T. Song, G. Kim, H. sig Won, and S.-O. Jung. Offset-compensated cross-coupled PFET bit-line conditioning and selective negative bit-line write assist for high-density low-power SRAM. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 62(4):1062–1070, April 2015.

[13] A. Lee, M.-F. Chang, C.-C. Lin, C.-F. Chen, M.-S. Ho, C.-C. Kuo, P.-L. Tseng, S.-S. Sheu, and T.-K. Ku. RRAM-based 7t1r nonvolatile SRAM with 2x reduction in store energy and 94x reduction in restore energy for frequent-off instant-on applications. In *VLSI Circuits (VLSI Circuits), 2015 Symposium on*, pages C76–C77, June 2015.

[14] S. Mittal. A survey of architectural techniques for DRAM power management. *IJHPSA*, 4(2), 2012.

[15] K. Nii, H. Makino, Y. Tujihashi, C. Morishima, Y. Hayakawa, H. Nunogami, T. Arakawa, and H. Hamano. A low power SRAM using auto-backgate-controlled MT-CMOS. In *Low Power Electronics and Design, 1998. Proceedings. 1998 International Symposium on*, pages 293–298, Aug 1998.

[16] S. A. Palatham-Veedu. Design and analysis of sense amplifier circuits used in high-performance and low-power SRAMs. Master's thesis, The University of Texas at Austin, 2011.

[17] Y. K. Park, Y. S. Ahn, S. B. Kim, K. H. Lee, C. H. Cho, T. Y. Chung, and K. Kim. Efective capacitance enhancement methods for 90-nm DRAM capacitors. *Journal of the Korean Physical Society*, 44(1):112–116, January 2004.

[18] G. Pasandi and S. Fakhraie. A new sub-threshold 7t SRAM cell design with capability of bit-interleaving in 90 nm CMOS. In *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, pages 1–6, May 2013.

[19] A. Pavlov and M. Sachdev. CMOS SRAM circuit design and parametric test in nano-scaled technologies. pages 31–33. 2008.

[20] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, and J. Rabaey. Standby supply voltage minimization for deep sub-micron {SRAM}. *Microelectronics Journal*, 36(9):789 – 800, 2005. {ISQED} 2004 International Symposium on Quality Electronic Design 2004.

[21] S. Singh, N. Arora, M. Suthar, and N. Gupta. Performance evaluation of different SRAM cell structures at different technologies. *International Journal of VLSI design & Communication Systems (VLSICS), pp97-109*, 2012.

[22] S. Skorobogatov. Low temperature data remanence in static RAM. *University of Cambridge, Computer Laboratory*, 2002.

[23] T. Tanaka, Y. Tanaka, H. Nakamura, K. Sakui, H. Oodaira, R. Shirota, K. Ohuchi, F. Masuoka, and H. Hara. A quick intelligent page-programming architecture and a shielded bitline sensing method for 3 v-only NAND flash memory. *Solid-State Circuits, IEEE Journal of*, 29(11):1366–1373, Nov 1994.

[24] R. Walden. Analog-to-digital converter survey and analysis. *Selected Areas in Communications, IEEE Journal on*, 17(4):539–550, Apr 1999.

[25] D.-P. Wang, H.-J. Lin, C.-T. Chuang, and W. Hwang. Low-power multiport SRAM with cross-point write word-lines, shared write bit-lines, and shared write row-access transistors. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 61(3):188–192, March 2014.

[26] Xilinx Inc. *Spartan-6 FPGA Configurable Logic Block User Guide*, February 2010.

[27] J. J. Yang, D. B. Strukov, and D. R. Stewart. Memristive devices for computing. *Nature Nano*, 8(1), 2013.