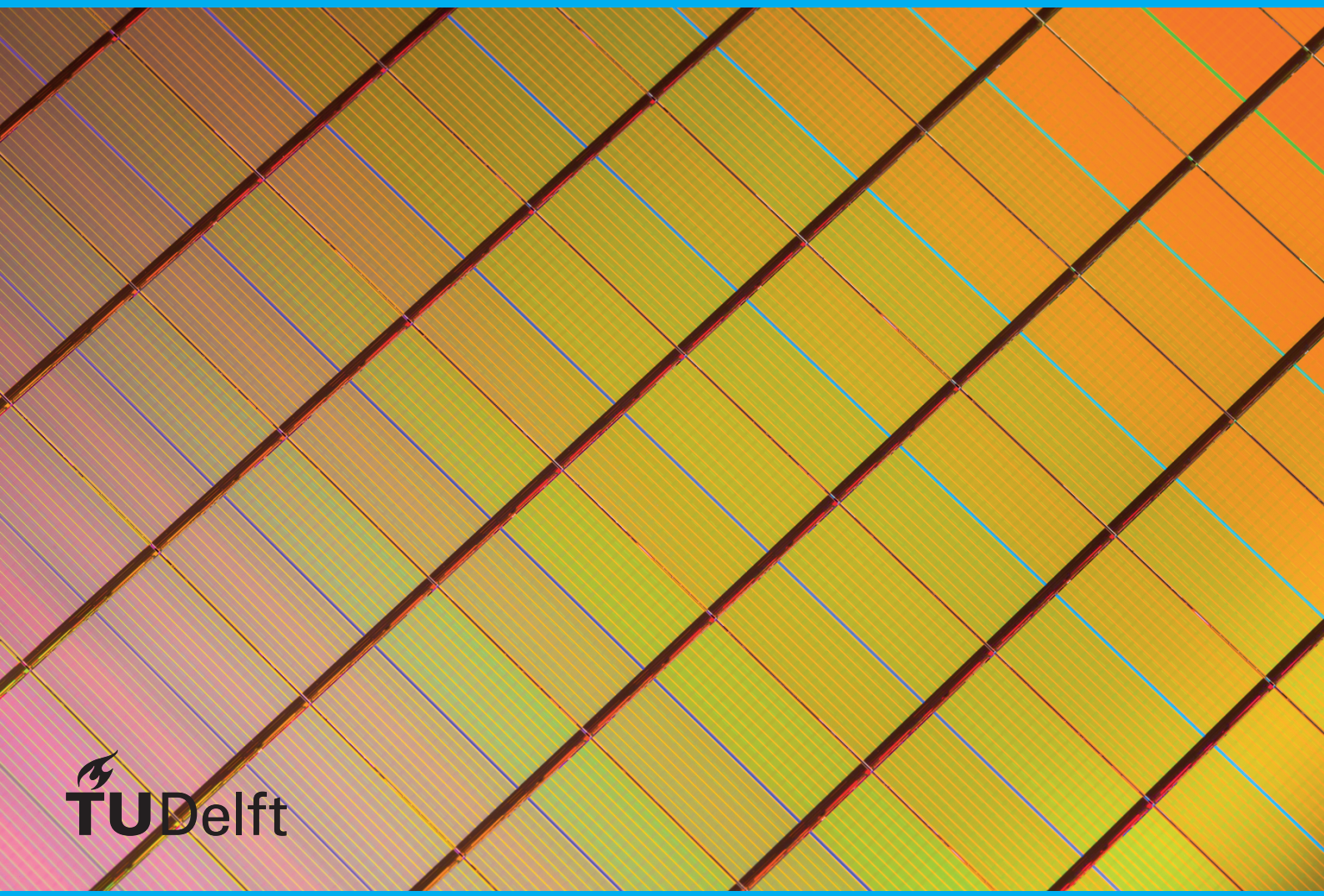
DRAM Reliability

Aging Analysis and Reliability Prediction Model

M.C.R. Fieback



TUDelft

DRAM Reliability

Aging Analysis and Reliability Prediction Model

by

M.C.R. Fieback

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday December 20, 2017 at 13:30

Student number:	4163850	
Thesis number:	CE-MS-2017-20	
Project duration:	November 1, 2016 – December 20, 2017	
Thesis committee:	Prof. dr. S. Hamdioui,	TU Delft, supervisor
	Dr. ir. T.G.R.M. van Leuken	TU Delft, member
	Dr. ir. J.S.S.M. Wong	TU Delft, member
	Dr. ir. M. Taouil	TU Delft, member

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Abstract

An increasing amount of critical applications use DRAM as main memory in its computing systems. It it therefore extremely important that these memories function correctly during their lifetime in order to prevent catastrophic failures. Already during the design phase, the reliability of the circuit needs to be predicted so that a reasonable lifetime expectation can be given. Although the importance of reliability analysis is clear, in literature not much research on DRAM reliability is available to designers. This thesis proposes a two phased DRAM reliability prediction model that can be used in the circuit design phase. During the first phase, the circuit performance is analyzed for different wear-out mechanisms affecting different subcomponents in the design. In the second phase, the results of the first phase are then used to determine the reliability of the the circuit.

In the first phase, the wear-out effects of Bias Temperature Instability (BTI) Hot Carrier Injection (HCI) and radiation trapping as well as transistor mismatch are examined. BTI is modeled using the RD-model, HCI with the lucky electron model, transistor mismatch with Pelgrom's model. Wear-out caused by radiation trapping is modeled as Gate Induced Drain Leakage (GIDL), the data for which are derived from retention time degradation measurements of an irradiated commercial DRAM. The circuit performance is analyzed per subcomponent of the DRAM design in a range of different metrics. Furthermore, the aging effects on a dwonscaled version of the circuit are investigated.

In the second phase, reliability functions are derived from the results from phase one per wear-out mechanism and per subcomponent. These reliability functions are used in an analytical reliability model which yields the overall circuit reliability.

The results from the first phase show degradation of the retention time as well as degradation of sensing delay metrics. Due to the relative low duty factor of memory cells, BTI and HCI have minor impact on the memory cell circuit performance. Radiation however, renders the circuit useless once the Total Ionizing Dose (TID) becomes more than 126 krad. On other subcomponents than the memory cells, BTI and HCI shift the reference voltage which results in an increase of retention time. BTI and HCI stressing of the sense amplifier also slightly increases retention time but mainly increases sensing delay. For both reference cells and the sense amplifier it holds that higher radiation doses break down the circuit completely. The same effects hold for the downscaled circuit, although the observed effects are more severe than in the unscaled device.

The system reliability prediction in the second phase shows the importance of individual reliability prediction of the subcircuits and wear-out mechanisms. Via the individual analysis, it becomes clear that the system reliability is mostly impacted by the degradation of the sense amplifier delay due to BTI and HCI. Other metric variations, like the increase in retention time caused by the reference cells, have less impact. It was found that the system reliability decreases to 0.84 after $1 \cdot 10^8$ s at a stressing temperature of 300 K.

Acknowledgements

Writing this thesis has been an intensive task which I could never have accomplished by myself. I therefore would like to thank all people who have helped me achieve this and some people especially!

First, I would like to thank my supervising professor, Said Hamdioui. I highly value his sharp eye for detail which steered this research in the right directions. Second, I want to thank Mottaqiallah Taouil, who guided me on daily basis. His tips when writing this thesis as well as the academic insights he shared during my research, really helped making this project a success. Then, I also would like to thank Marco Rovatti from ESA for his feedback during the implementation of the work.

Next, I thank my friends without whom this graduation project would have been impossible to complete. The beers I drank with Complex, the vroeger-was-alles-beter stories from my board members, the long nights with my roommates, the food hunts with Yikun, as well as the support from Dörthe all mean a great deal to me.

Finally, I wish to thank my family. Without their support and motivation during my studies, I would have been nowhere near this point.

Thank you!

Moritz Fieback
Delft, December 2017

Contents

List of Figures

List of Tables

Nomenclature

$\Delta V_{th,\,BTI}$	BTI threshold voltage shift
$\Delta V_{th,\,HCI}$	HCI threshold voltage shift
$\Delta V_{th,\,mis}$	Transistor mismatch threshold voltage shift
D	Radiation dose rate in rad/s
E_a	Activation energy in eV
$F(t)$	Failure distribution
I_D	Drain current
$I_{Rad\ GIDL}$	Radiation Gate Induced Drain Leakage current
k	Boltzmann constant
L	MOSFET length
q	Elementary charge
$R(t)$	Reliability function
T	Temperature in K
t_{ret}	Retention time
V_{DS}	Drain-source voltage
V_{GS}	Gate-source voltage
V_{SB}	Source-bulk voltage
V_{th}	Transistor threshold voltage
W	MOSFET width
$z(t)$	Failure or hazard rate
dXd	A memory operation where *d* is a logical value and *X* is an operation, i.e., Read (*R*) or Write (*W*)
B	Transistor Bulk
D	Transistor Drain
G	Transistor Gate
GND	Ground
S	Transistor Source
V_{DD}	Supply voltage
V_{ref}	Reference voltage
WL	word line

Glossary

AR Aspect Ratio.

BERT Berkely Reliability Tools.

BTI Bias Temperature Instability.

DDR Double Data Rate.

DRAM Dynamic Random Access Memory.

EDO Extended Data Out.

EM Electromigration.

FaRBS Failure Rate Based SPICE Prediction Method.

FPM Fast page Mode.

GIDL Gate Induced Drain Leakage.

HCI Hot Carrier Injection.

IC Integrated Circuit.

lifetime is the typical age at which a circuit fails.

MaCRO Maryland Circuit-Reliability-Oriented.

MDL Measurement Description Language.

MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor.

MTTF Mean-Time-To-Failure.

NBTI Negative Bias Temperature Instability.

NMOS n-type Metal-Oxide-Semiconductor Field-Effect Transistor.

PBTI Positive Bias temperature Instability.

PMOS p-type Metal-Oxide-Semiconductor Field-Effect Transistor.

RAMP Reliability Aware Microprocessor Model.

RD Reaction-Diffusion.

reliability describes the circuit population that works conform its specification over time.

retention time the time it takes for a DRAM memory cell to discharge until a bit flip occurs.

RSM Response Surface Model.

RTN Random Telegraph Noise.

SDRAM Synchronous Dynamic Random Access Memory.

SPICE Simulation Program with Integrated Circuit Emphasis.

SRAM Static Random Access Memory.

STI Shallow Trench Isolation.

SyRA System Reliability Analyzer.

TDDB Time Dependent Dielectric Breakdown.

TID Total Ionizing Dose.

1 Introduction

This chapter introduces this thesis. First, Section 1.1 presents the motivation and relevance of DRAM reliability. Then, Section 1.2 presents the state of the art in circuit reliability simulation. This section also presents the shortcomings and limitations that need to be overcome in order to enable accurate and precise DRAM reliability simulation. Section 1.3 presents the academic contributions of this thesis. Finally, Section 1.4 presents the outline of this thesis.

1.1. Motivation

Integrated circuits (ICs) age [1], causing degradation of circuit performance over time and may even lead to chip failure. Depending on the application of the device, a failure may lead to significant losses or even dangerous situations [2]. Consider for example the effects of a malfunctioning chip in an autonomous car. The manufacturer of the chip will be held responsible for the damage and the negative publicity it brings could even result in bankruptcy. To prevent this, manufacturers design their chips to meet a certain standard of robustness. They specify operating conditions and a period of time in which the chip is said to work reliable.

The expected time a circuit is able to operate according to its specifications is called lifetime. Manufacturers want their products to have a sufficiently high reliability to fulfill the requirements of their customers. An insufficient reliability will lead to failures as described above, while a product reliability that is higher than needed results in unnecessary development costs. The definition of reliability is closely related to lifetime and is described for electronics by Gielen *et al.*: *"Reliability is defined as the ability of a circuit to conform to its specifications over a specified period of time under specified conditions."* [1]. The reliability of a circuit or chip can be found in multiple ways. During the design phase it is possible to simulate the aging effects on a circuit and predict the circuit degradation. Once a chip is manufactured, the reliability of a chip can be determined by accelerated testing [3]. Once the reliability of a product is found, the expected lifetime of it can be derived [4], [5].

Dynamic Random Access Memory (DRAM) is commonly used as the main memory of a computing system and is therefore widely used in a large variety of appliances. To illustrate this impact, Figure 1.1 shows the quarterly worldwide DRAM revenue from 2013 to 2017 which was over 40 billion USD in 2016 [6]. DRAM sales account approximately for 13 % of the global semiconductor revenue [7]. Although DRAM is such an important component in the semiconductor industry, not much data about it is accessible for the scientific community. This is probably caused by the fact that the DRAM market is highly competitive and therefore manufacturers do not share their findings.

Space agencies like the European Space Agency (ESA) or National Aeronautics and Space Administration (NASA) demand components with the highest possible reliability for their spacecrafts [8]. A failing component in space can not be repaired and may even lead to complete failure of the spacecraft [2]. Next to these locational issues, radiation is a prime wear-out mechanism for ICs in space [9]. The highly energized space particles can get trapped in silicon and may generate leakage currents or even break down the transistor oxide [9], [10]. Therefore, the components in a spacecraft thus need an extremely high reliability. To guarantee this high reliability, space agencies usually resort to the use of older and mature technologies with high reliability [8], [11]. However, space agencies would like to be able to use more sophisticated technologies in their spacecrafts to reduce costs and increase the performance of their technology. However, this requires a new

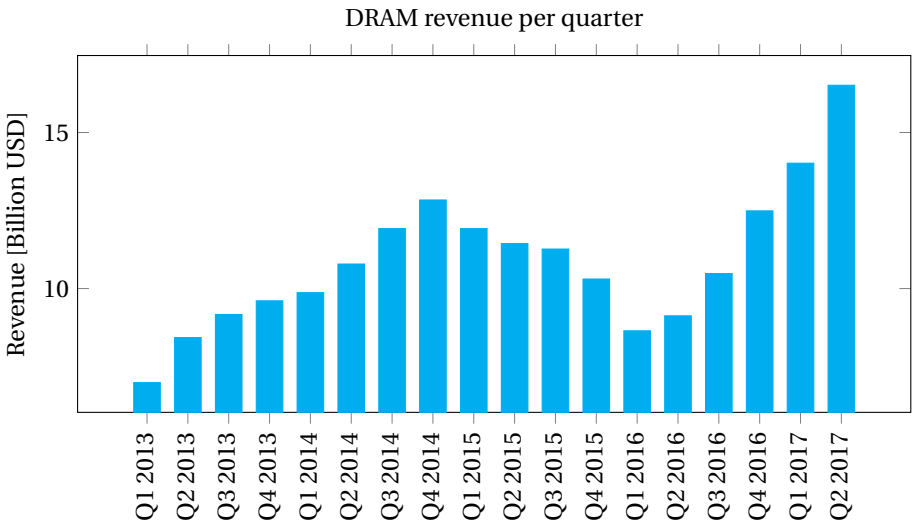


Figure 1.1: Worldwide DRAM revenue per quarter. Total revenue in 2016 is slightly more than 40 billion USD [6]

methodology where the reliability can be determined without waiting years and analyze how a technology performs.

The combination of a lack of DRAM reliability data and the desire of space agencies and other electronic manufacturers to have an accurate and fast method to determine the reliability during the design phase, is the motivation behind this research. This research investigates the effects of commonly seen wear-out phenomena like bias temperature instability and hot carrier injection on DRAM performance, but also includes total ionizing dose effects which are especially relevant to space applications. The results from this aging investigation are used in a reliability prediction model. The model predicts the reliability based on different component metrics and operating conditions.

1.2. State of the Art

Several methods have been proposed to estimate the aging effects of integrated circuits. Section 1.2.1 introduces them and briefly describes how they are implemented. Thereafter, Section 1.2.2 presents the limitations of these simulation methods for this research.

1.2.1. Reliability Prediction Models

Two distinctions in reliability prediction models can be made. One on the level at which circuits or systems are simulated and the other one on the prediction outcome they deliver. The simulation levels can be distinguished in simulations on individual transistor level, on gate level and on a behavioral level. The prediction outcome can be divided into two different types of outcomes being reliability and lifetime. Lifetime is defined as the mean-time-to-failure (MTTF) of a circuit while reliability also includes information on the circuit failures over time. Section 3.1 describes both terms in depth. Below a short introduction of all simulation methods is presented. A more in depth analysis of the methods presented below can be found in Section 3.4.3.

Berkely Reliability Tools The Berkely Reliability Tools (BERT) [12] were one of the first reliability predictors. The predictor calculates the degradation of individual transistors in the circuit based on their individual workloads. The results of these calculations are used to make an extrapolation which estimates the degradation of the circuit for a given aging time and subsequently predicts the lifetime. However, it has several limitations such as the inability to combine different failure mechanisms and inclusion of temperature effects. Next to a lifetime prediction, the BERT can predict reliability as well for oxide breakdown phenomena only.

Reliability Aware Microprocessor Model RAMP v1 [5] and its successor RAMP v2 [13] simulate aging effects on the behavioral level. It simulates the effects of aging on a complete processor design by taking basic as-

sumptions on the workload and other circuit parameters. These parameters are applied to generalized blocks constituting the complete processor. The result of this is a reliability prediction. The drawback is given by the fact that the prediction is made based on system assumptions rather than individual components. This neglects the correlating effects of individual component degradation.

Maryland Circuit-Reliability-Oriented MaCRO [14] is inspired by the BERT. It also evaluates circuit aging in multiple iterations where first a clean reference circuit is simulated before several aged circuits are evaluated. Based on the circuit workload and temperature, aging models are applied to the transistors which calculate the relative degradation of the transistors. Then the most degraded transistors are replaced with an aged equivalent in the circuit. The result of this analysis is a reliability prediction. Drawbacks are the ignorance of marginally degenerated transistors and the accuracy of the prediction models, which is low in order to allow for fast circuit simulation.

AgeGate Lorenz *et al.* proposed AgeGate [15] to simulate the aging effects of digital circuits on gate level. It defines the rise and fall times and delays of the gate models by compensating for aging. This allows to make a prediction on the lifetime of the circuit. For larger circuits a method to find critical paths and possible critical paths after aging is presented. A drawback of this method is that is only applicable to digital circuits only.

Response Surface Model Based Simulation Technique The RSM [16] is a reliability simulation prediction method which performs simulations at transistor level and predicts reliability of the circuit for given inputs like the workload and operating temperature. The method allows to simulate the degradation of ICs due to both aging effects and process variability. A Monte-Carlo method is used to simulate the circuit under test for various process variabilities and aging effects. The outcomes are then used to derive functions which describe the reliability of a circuit. With these functions it becomes possible to predict the reliability of a circuit faster than with a bold Monte-Carlo approach and it is possible to highlight weak spots in the design. The failure mechanisms in this approach are independent from each other, which could easily lead to over or under estimation of their effects.

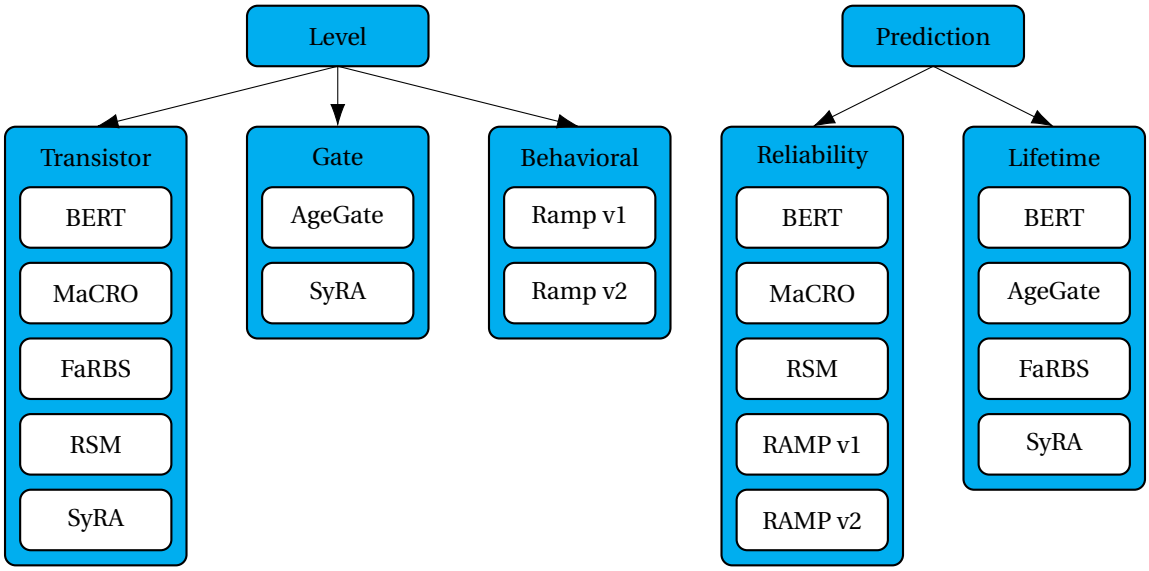
Failure Rate Based SPICE Prediction Method FaRBS [17] works on transistor level and predicts the lifetime. Based on the simulation of an unaged circuit the lifetimes per transistor using waveform information and per failure mode are evaluated. From these lifetimes the general MTTF for the complete circuit is calculated which can then be used to derive the lifetime of the circuit. The outcomes of this approach form a drawback as it remains unclear how the reliability will develop over time as only a lifetime prediction is made. Also the failure mechanisms are independent and are added in a weighted sum. This might over or under estimate the actual impact as correlating effects in the wear-out phenomena are not included.

System Reliability Analyzer The final method which is able to predict aging effects on transistor level is SyRA [18]. It simulates aging effects on analog circuits in an iterative way with a predefined timestep. At every timestep the newly acquired degradation of the transistor is set for the new simulation. This iterative fashion allows for example to see drift in bias circuits which is not possible if a direct calculation would be used. Besides transistor level simulation, it is also possible to simulate the effects of aging on digital gates. This part of the method focuses solely on delays of the gates in the circuit. Drawbacks of this method are the potential long simulation times if many timesteps are required and again, the independence of the failure mechanisms.

1.2.2. Limitations

Figure 1.2 presents an overview of all previously discussed aging simulation methods. The figure splits these based on simulation level and on the prediction outcome.

The previous section illustrates the benefits and drawbacks of several state-of-art methods in simulating aging and predicting reliability in integrated circuits. A common property is the applicability to straight forward digital or analog circuits. More complex circuits, containing for example feedback loops like the voltage boost circuits in a DRAM, can hardly be simulated accurately. Next to that, none of the above mentioned simulation approaches is able to simulate both digital and analog aging degradation in an efficient way. This is a desirable property when analyzing DRAM as some parts, like the control logic, work digitally while the memory array is made from cells working in an analog way. In order to make an accurate estimation of the



(a) Simulation level (b) Prediction outcome

Figure 1.2: State-of-the-art aging simulation methods comparison on simulation level and prediction outcome

circuit wear-out, it is important that failure mechanisms are dependent on each other as different effects may be caused by the same physical phenomenon. The capacitor in the memory cells has a complex mechanical structure that is different from the structures usually seen in ICs. None of the reliability predictors described above is able to include wear-out analysis of these devices. Finally, none of the presented approaches allows for the simulation of radiation effects which is an important wear-out characteristic for space agencies as was described in Section 1.1.

The following list summarizes the desired features of a DRAM reliability predictor. It should:

- Be able to simulate general wear-out phenomena on transistor level depending from on each other
- Be able to simulate DRAM capacitor wear-out
- Include workload and operating temperature or the operating conditions
- Include radiation effects
- Efficient simultaneous digital and analog simulation allowing dynamic simulation level
- Analyze both digital and analog simulation efficiently, allowing for a dynamic prediction level
- Predict both reliability and lifetime

1.3. Contributions

The wear-out effects of BTI and HCI on the performance of a DRAM are evaluated as well as the effects of space radiation trapping. Besides these time dependent degradation mechanisms also transistor variability is included in the model by the use of Pelgrom's model. All the beforementioned time dependent models are applied to the DRAM netlist that was operated under different temperatures and workloads. Furthermore the DRAM circuit was downscaled so that the effects of technology could be examined. Wear-out effects BTI and HCI were scaled accordingly.
The main contributions of this thesis are:

- Development of a DRAM space radiation model based on irradiated DRAM measurements
- Application of the BTI, HCI and space radiation models to a DRAM netlist for different temperatures and workloads
- Downscaling the DRAM netlist with a factor $\sqrt{\kappa}$ and applying the wear-out models to it

- A framework that is able to automatically simulate aging effects on the DRAM circuit
- A dynamic prediction model which predicts the reliability and lifetime of a DRAM circuit under various circumstances
- A paper submission to the 2018 IEEE Latin-American Test Symposium (LATS2018), included as Appendix B:
  M. Fieback, M. Taouil, S. Hamdioui and M. Rovatti, "Ionizing Radiation Modeling in DRAM", in *2018 IEEE Latin-American Test Symposium*

1.4. Outline

The remainder of this thesis is organized as follows. Chapter 2 introduces DRAM on architectural, circuit and technology level. Then, Chapter 3 introduces the terms and definitions of reliability, failure modes seen in CMOS as well as in DRAM specific and presents the circuit model, an in-depth overview of aging predictors as well as the implemented wear-out models. Chapter 4 describes the framework that was used in this research, containing the simulation platform, measurement metrics and the experiments to be performed. After this, Chapter 5 presents the results of these experiments and presents the reliability prediction model. Finally Chapter 6 presents a summary and a conclusion, discusses this research and draws some suggestions for future work.

2

DRAM Design and Technology

This chapter introduces computer memory and more specifically DRAM. A general introduction in computer memory is presented in Section 2.1. After this, Section 2.2 introduces the DRAM classifications which describe different types of DRAM. These classifications will then be used in Section 2.3 to show what differences in architecture constitute these classes. Section 2.4 presents the components that are present in an architecture and ends with an overview presenting the timing and interactions between these components. Then Section 2.5 presents a general introduction in IC fabrication technology before presenting the specifics for DRAM fabrication technology.

2.1. Computer Memory

In this section first a general introduction of different memory technologies will be given. The presented technologies will be ranked according to their speed, cost and storage capacity. After this the DRAM technology will be explored in more depth.

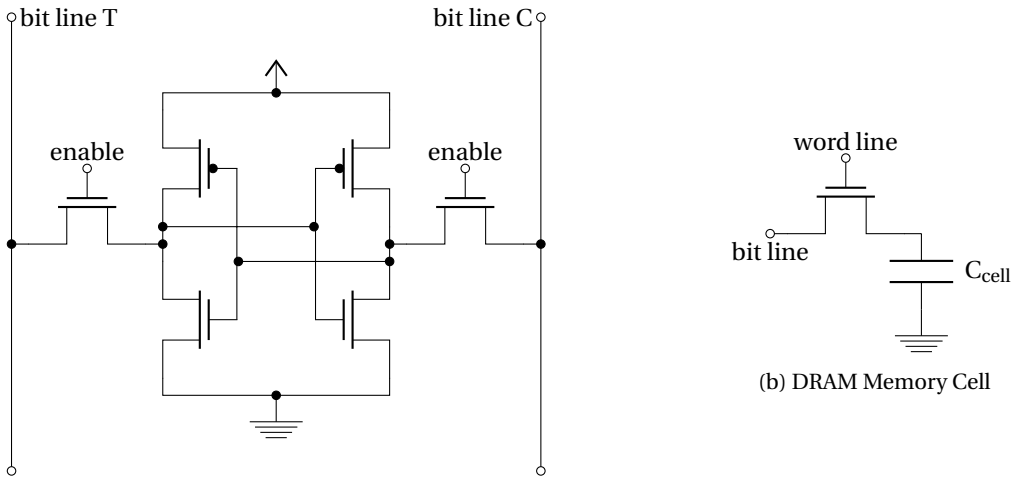
Memory is used by a computer to store data for later use. Depending on the size of the data and the demand for it while computing, different types of memory are used. Close to the processor fast cache memory is used in order to perform operations on it without large latencies. On the other side of the spectrum, one finds magnetic hard drives storing the bulk of the data [19]. Table 2.1 lists the different kinds of memory and ranks them according to speed, cost and capacity [20].

Another classification of memory could be based on the ability to hold data after the supply power is disconnected. Memory that loses its data after the power is removed is called volatile while one that keeps its data is called non-volatile. Close to the processor, e.g. in caches and main memory, usually volatile memory is used. This is due to the fact that these memories are faster in reading and writing than the non-volatile ones. Farther away from the processor in the memory hierarchy non-volatile memory is used.

In current commercially available technology the fastest type of memory is called Static Random Access Memory (SRAM) [21]. This memory is used closest to the processor as cache memory. A typical SRAM cell is made from two cross-coupled inverters storing the data and two access transistors which allow access to the cell. A typical SRAM circuit is presented in Figure 2.1a. The total of six transistors per bit makes it an expensive memory cell in comparison with other volatile memory types that use less transistors per bit.

One memory layer below the caches the main memory is found, this is typically Dynamic Random Access Memory (DRAM). In current technology DRAM is made using one capacitor to store charge and one transistor to access the capacitor [21]. Figure 2.1b shows the circuit of a DRAM cell. Section 2.4 will go deeper into the actual design and physical lay-out. DRAM is slower than SRAM and volatile as well but the smaller cell size allows for a denser memory which is thus cheaper per bit. The capacitor that stores the charge is leaky. This means that after a certain time the stored data will be lost and therefore a refresh operation is needed to keep the correct data in the memory. The time a cell can keep the correct data is called the retention time (t_{ret}).

Memory cells, both SRAM and DRAM, are placed in an array where individual cells can be accessed by a combination of a row and a column address. These addresses are supplied externally from the controlling processor. A schematic overview of a memory array is presented in Figure 2.2. The row addresses are translated to word lines (WL), the column addresses to bit lines (BL).



(a) SRAM Memory Cell

(b) DRAM Memory Cell

Figure 2.1: SRAM and DRAM memory cell circuits

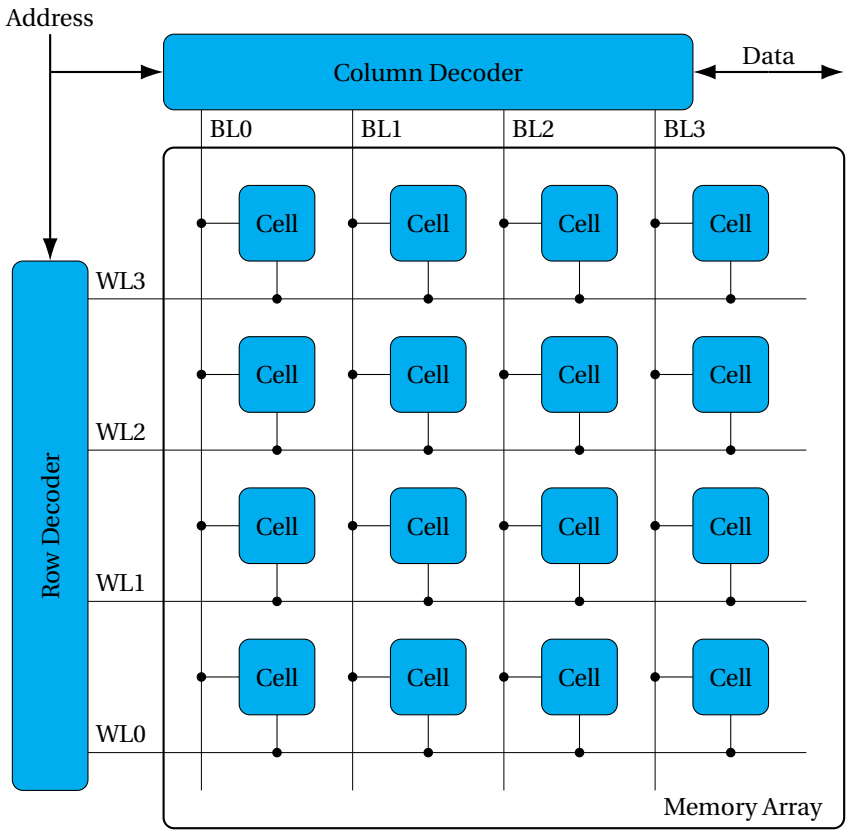


Figure 2.2: Memory array and address decoders

Below the main memory layer bulk memory is found. This memory is non-volatile and is cheaper per stored bit than DRAM is. This type of memory can be everything ranging from flash memory to hard disks.

Table 2.1 compares the different types of memory with each other and ranks them on speed, cost and storage capacity [20]. It can be seen that in general cheaper memories are slower and have a larger storage capacity while faster ones are more expensive and have smaller capacities.

Table 2.1: Comparison of different memory technologies and their relative ranking. An 1 indicates best performance in the given category, a 4 indicates worst performance [20]

Memory technology	Used as	Latency from Processor	Cost per bit	Storage Capacity
SRAM	Processor cache	1	4	4
DRAM	Main memory	2	3	3
Flash	Bulk memory	3	2	2
Hard disk	Bulk memory	4	1	1

2.2. DRAM Classifications

In 1970 Intel presented the first commercial DRAM [22]. It uses three transistors to access the memory cell and store charge. From that moment on new types of DRAM were invented, where the new types aimed for higher data bandwidth (the amount of bits transferred per second) and higher memory density (the amount of bits per area) than the previous ones had. Figure 2.3 gives an overview of all different classifications of DRAM that have existed or still exist. From this figure it becomes clear that there are two main branches in DRAM technology; asynchronous designs and synchronous designs. The difference between the two branches is that asynchronous designs are, as the name implies, not synchronized by a clocking signal while synchronous ones are. The increase in processor and memory operating speed decreased the timing margins between the two components. In order to keep increasing the operating speeds it became necessary to synchronize memory with the processor and thus the synchronous branch came into existence [21]. In [21] an extensive introduction of DRAM architectures is presented. Below, a summary of this information will be presented.

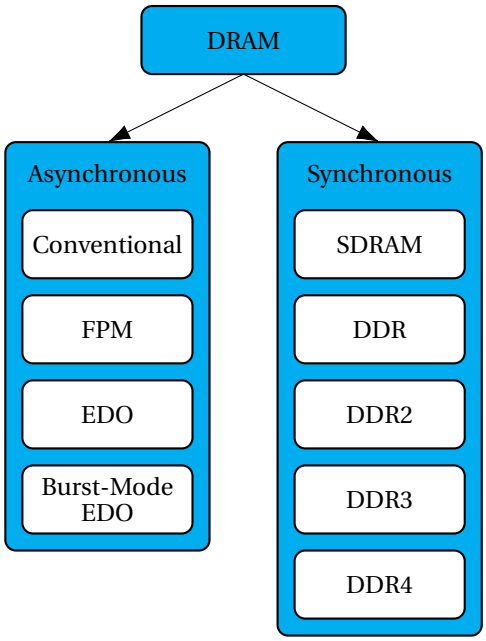


Figure 2.3: Different classifications of DRAM

The asynchronous sub-categories differ in the total memory bandwidth they can achieve. Figure 2.4 shows the row and column address timing and valid dataout timing for the asynchronous architectures. In conventional DRAM, the controlling processor must supply a row address and a column address to read one page from the memory array. A page is the amount of bits that is accessed when one row is operated. When

this page is read, a new address needs to be supplied before the next page can be accessed. Fast Page Mode (FPM) DRAM makes use of the locality of data principle [19] and allows to read from one row on multiple columns. The columns are then individually fed to the output by simply increasing the column address internally which removes the overhead of externally supplying new column addresses. Extended Data Out (EDO) DRAM latches the output of the DRAM which keeps the output data valid until the next column address is supplied and allows to start fetching the data earlier in a cycle, effectively increasing the data throughput. The EDO technology evolved in Burst-Mode EDO. In Burst-Mode one row and one column address are supplied and the column address is automatically increased. This removes the overhead of supplying addresses completely for locally spaced data [19], [21].

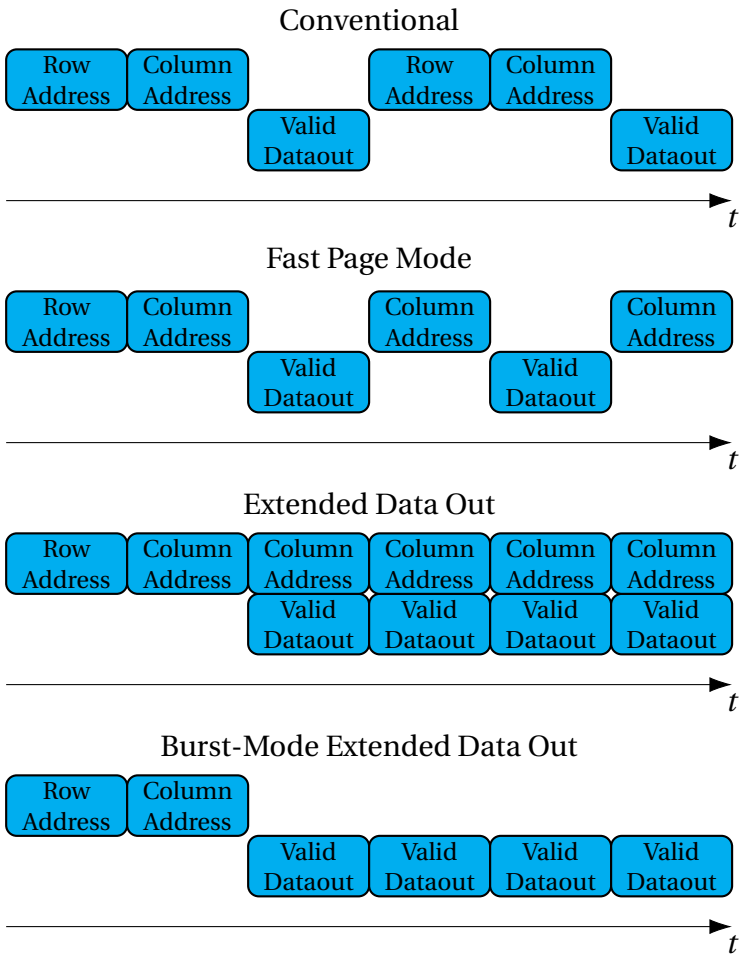


Figure 2.4: Asynchronous DRAM timing [21]

Applying a clock to the memories allows for higher data rates, as was stated above, which lead to Synchronous DRAM (SDRAM). In Figure 2.4 the address and data timing for SDRAM and DDR DRAM is shown. The use of a clocking signal controlling the internals of the DRAM allowed for higher operating frequencies and thus an increase in memory bandwidth. It is also possible to operate a SDRAM in burst-mode, which is shown in the figure. To further increase the bandwidth, data were also fed to the output at the falling edge of the clock. This type of SDRAM is called Double Data Rate (DDR) DRAM. The internal frequency does not need to be increased but the output data clock has to double in frequency [21]. DDR2 up to DDR4 apply this same method but increase the total amount of data words that is fetched per cycle. DDR2 fetches four data words while DDR3 and DDR4 fetch eight data words. DDR4 increases the memory bandwidth even further by dividing the memory into multiple banks which can be addressed simultaneously [23].

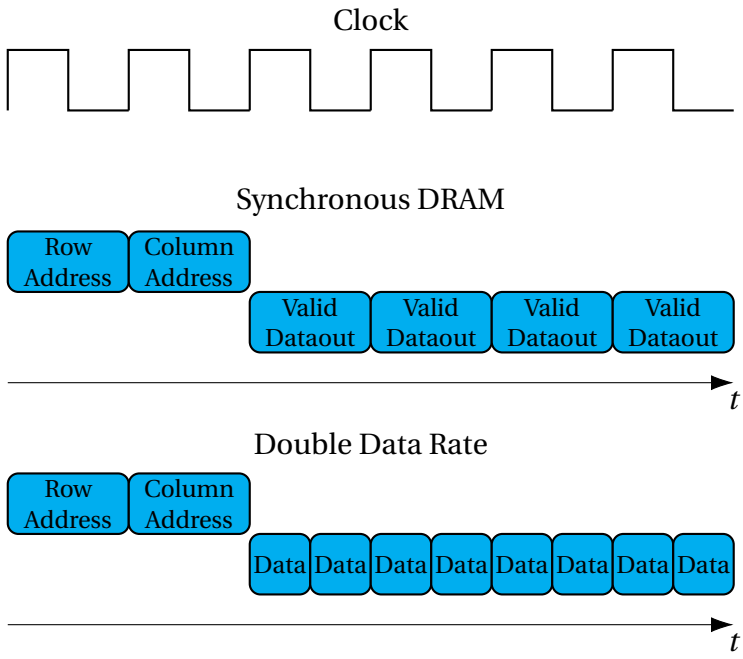


Figure 2.5: Synchronous DRAM timing [21]

2.3. Architectures

This section will introduce the main building blocks of a DRAM, the combination of which is called the architecture. The detailed explanation of the architectural subcomponents will be presented in Section 2.4.

Figure 2.6 shows a basic architecture for a SDRAM. The memory cell array stores the data and consumes the most of the chip area. To access the array, the input address must be decoded. The address will be split in the column and row that should be accessed. The memory cells in the array can then be accessed through word lines and bit lines. Word lines are connected to a complete row in the matrix whereas bit lines are connected to a complete column. Data is read out from the array using sense amplifiers to amplify the small signal coming from the array. By setting one word line and sensing a bit line, a cell can be read out. Writing to a cell can be done by setting the bit line to the value that has to be written and enable the word line. The access devices allow the in- and output buffer to read and write to and from the array. Because the data are stored as charge on a capacitor, refresh of the data is needed. The refresh counter keeps track of the rows that have to be refreshed. The whole process is governed by control logic which also receives commands from the processor. The cell array can be split in multiple smaller sub arrays called banks. The smaller arrays decrease the distance from the sense amplifiers to the farthest cell which decreases the delay from accessing a cell and having a valid output. This can increase the memory operating speed and with that the memory bandwidth.

Asynchronous designs differ from the architecture presented in Figure 2.6 in the fact that no external clock is supplied to it and slight adaptations are made in the hardware to some of the subcomponents. FPM DRAM is nearly equal to conventional DRAM, only in the control logic some small alterations had to be made. As was stated in Section 2.2, EDO DRAM adds a latch to the data output buffers only. Burst-mode was added by adding a multiplexer to the column decoder input that would switch between the column address supplied from outside the chip by the processor or to an internal counter which provides the increasing column addresses.

To supply two data words per clock cycle in synchronous DDR DRAM the cell array, sense amplifiers and access devices are doubled. The row and column decoders need to be altered as well in order to be able to read from the two separated arrays. In one read operation now two independent arrays can be accessed which will respond at the same time. The data output buffer uses a multiplexer and a latch to select one data word to output from the two arrays on a clock edge. DDR2 and DDR3 repeat this procedure of doubling the memory cell array to increase the amount of fetched data words. This structure for a DDR SDRAM design is presented schematically in Figure 2.7. DDR4 DRAM increases the amount of banks in the cell array and has some other modifications to the control logic and data buffers which increase the operating speed.

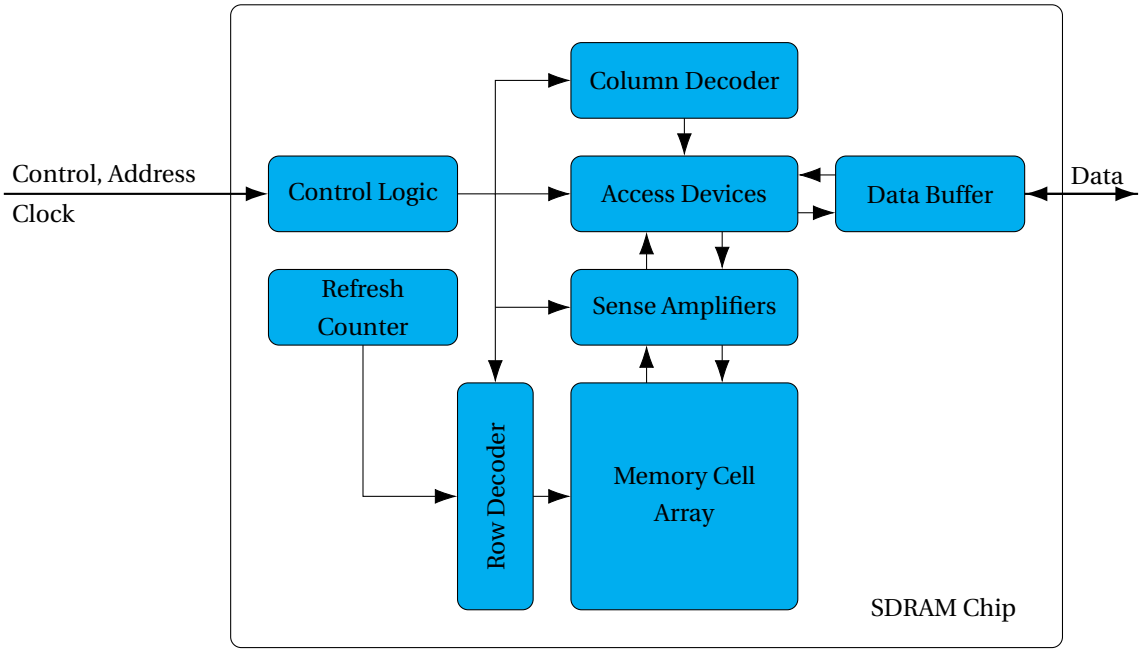


Figure 2.6: SDRAM Architecture [24]

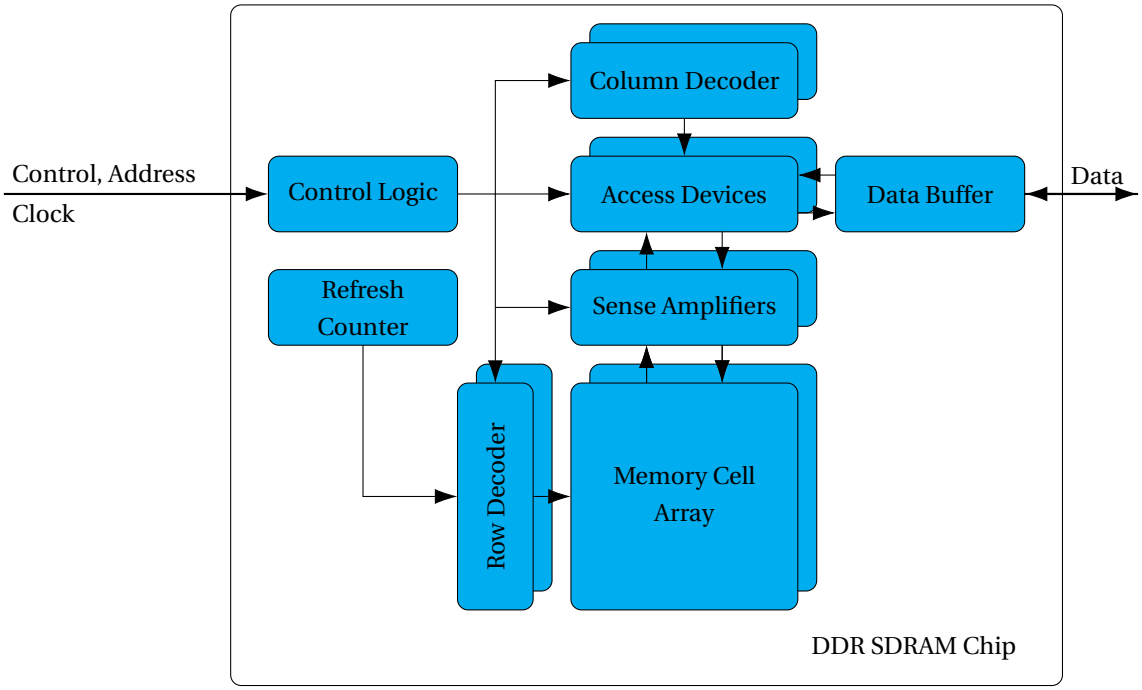


Figure 2.7: DDR SDRAM Architecture

2.4. Design

The different building blocks from which the architecture is made will be described in this section. After the circuit design has become clear to the reader, a general timing diagram will be presented that describes how the different components interact with each other.

2.4.1. DRAM Subcomponents

This section describes the circuit topology for the several subcomponents of a DRAM based on the work in [21]. Section 2.4.2 will introduce the operations that can be performed on a DRAM and Section 2.4.3 presents the timing and waveforms that constitute these operations.

Memory Cell

The memory cell stores data as charge on a capacitor. Figure 2.8 shows a typical cell design. The capacitor can be accessed through a single transistor where the gate is connected to a word line and the drain to a bit line. The cell can be written by setting the desired voltage on the bit line and enabling the word line. Reading from the cell is done by setting the bit line to half the maximal voltage and enabling the word line. The small amount of charge stored on the capacitor will cause a minor voltage change on the bit line. This voltage difference is then sensed by the sense amplifier, which will be discussed below. The physical construction of the memory cell will be discussed in Section 2.5.2.

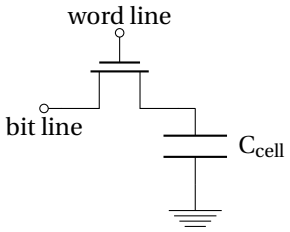


Figure 2.8: A DRAM memory cell

Precharge Circuit

The precharge circuit charges the bit lines to half the supply voltage, $V_{DD}/2$, before a read operation is executed. The structure of this circuit can be found in Figure 2.9. Its operation is easy to understand, the two transistors at the side of the circuit, T_T and T_C, charge the bit lines to the desired voltage and T_{eq} equalizes the bit lines. The latter is necessary as process fluctuations or transistor aging might cause different voltage drops over the two side transistors. After the bit lines are precharged, they get disconnected from the precharge voltage.

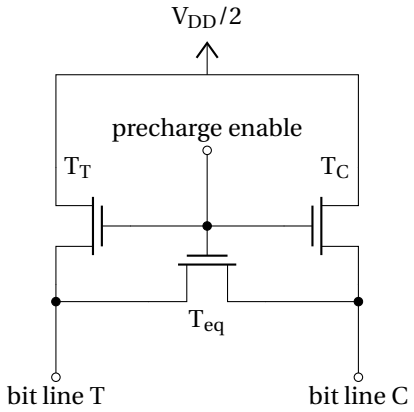


Figure 2.9: Precharge circuit

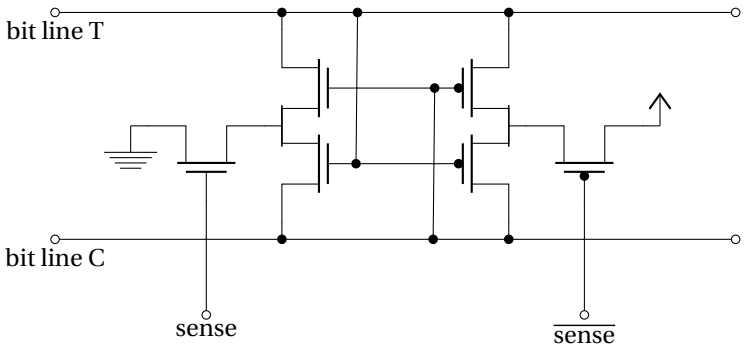


Figure 2.10: Sense amplifier for a DRAM. The bar indicates active when low behaviour.

Sense Amplifier
Figure 2.10 shows the sense amplifier circuit. The sense amplifier is made from two cross-coupled inverters. The amplifier senses small voltage differences between the two connected bit lines and amplifies this difference. The cross-coupling creates a feedback network that allows for very fast sensing operation. I.e., a lower voltage on one node forces the other node to a higher voltage via the inverter. The circuit will charge the bit line with the highest voltage to V_{DD} and the lower one to GND. During the amplification the word line of the cell is still active, so the sensing operation will restore the cell's contents to its original value. The resulting voltages on the bit lines are fed to the output of the DRAM.

Reference Cells
To every bit line a reference or dummy cell is connected. These cells do not store data but are used to provide an accurate reference voltage to the sense amplifier. Mismatch phenomena make it difficult to generate a good reference source that could be used in the sensing operation. To deal with this, dummy or reference cells are used. When the bit lines are precharged, the dummy cels are opened and the precharge voltage is stored on their capacitors. Now, when a data cell on the complementary bit line is accessed the reference cell is accessed at the same moment, providing an accurate reference voltage to the sense amplifier. The sense amplifier senses the voltage differences, restores the cell's data and provides data to the output of the circuit [25]. The circuit topology of a dummy cell is equal to that of a regular memory cell and can thus also be described by Figure 2.8.

Memory Array
All the memory cells are placed in an array. An example of this array can be seen in Figure 2.11. The horizontal lines connecting to the gates of the memory cells are the word lines, indicated with WL. The vertical connections are the bit lines, indicated with BT and BC. BT stands for the true bit line while BC stands for the complementary bit line. At the end of the bit lines the sense amplifiers and reference cells are present. The precharge devices are found at the other end of the bit lines. From the sense amplifiers the data are moved to the access devices that lead the data further out of the memory.

2.4.2. Read and Write Operations
Al-Ars describes five operations available in a DRAM [24]. These five operations are:

1. **Act** Activate a word line and move data to the sense amplifiers
2. **Rd** Read data from the sense amplifiers and transfer to output buffers
3. **Wr** Write data from the input buffers to the bit lines
4. **Pre** Precharge the bit lines and unassert any word line
5. **Nop** No operation is performed. Instead the previous command is extended in time.

Using these operations it becomes possible to specify the order of use of the previously described components. The combination of these operations forms complete read and write operations. Below the operations needed to read, write and refresh are described. Note that all the operations assume the bit lines to be precharged before the operation starts.

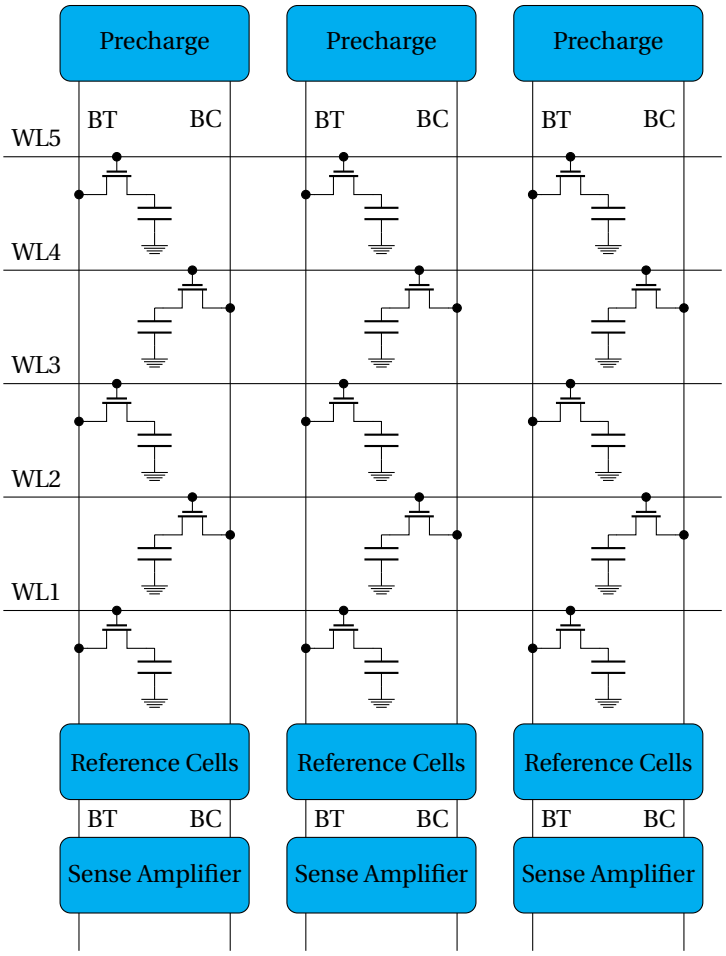


Figure 2.11: Memory array with precharge devices, reference cells and sense amplifiers

Reading:

1. **Act**
2. **Rd**
3. **Pre**

Note that the **Act** command also recharges the capacitor as the word lines are still asserted.
Writing:

1. **Act**
2. **Wr**
3. **Pre**

Refreshing:

1. **Act**
2. **Pre**

The operations described above are abbreviated in the remainder of this work. An operation on a memory cell can be described as: *dXd* where *d* is a logical value, being either '1' or '0', and *X* describes a reading or writing operation, denoted with *R* or *W* respectively. The first logical value describes the cell contents before the operation took place and the last one describes the contents off the cell after the operation was completed. For example: assume that a cell holds a logical '0' and an '1' is then written to it. The corresponding operation would then be described as: *0W1*. As was stated above, the **Act** operation refreshes the cell contents. A refreshing operation can then for example be described as: *1R1* where the '1' in the cell is restored.

2.4.3. Timing Diagrams and Waveforms

With the hardware description from Section 2.4.1 and the description of operations from Section 2.4.2 it becomes possible to draw the according waveforms. In Figure 2.12 a read operation is shown. At $t = 0$ s the word line of a cell is opened as well as the word line to the reference cell on the other bit line. After 4 ns the sense amplifier is enabled. The cross-coupled structure of the sense amplifier starts to discharge both the bit lines until the voltage difference between them is large enough. Then the PMOS transistors will charge the highest bit line to V_{DD} and the NMOS transitors will discharge the bit line with the lowest voltage top GND. The data can now be fed out of the DRAM to the processor. After the operation is completed the bit lines are precharged again, which can be seen in Figure 2.13. Both the bit lines are precharged back to $V_{DD}/2$ [24].

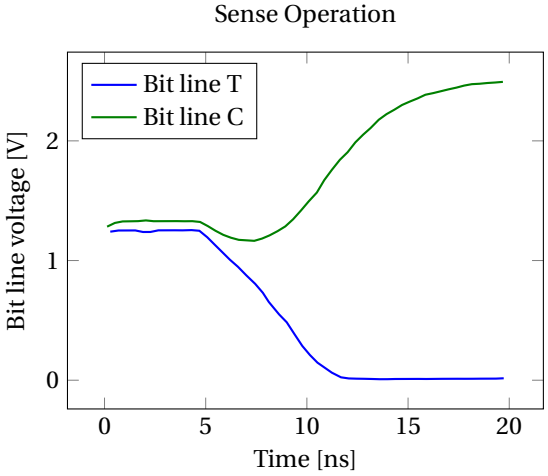


Figure 2.12: Sense amplifier operation [24]

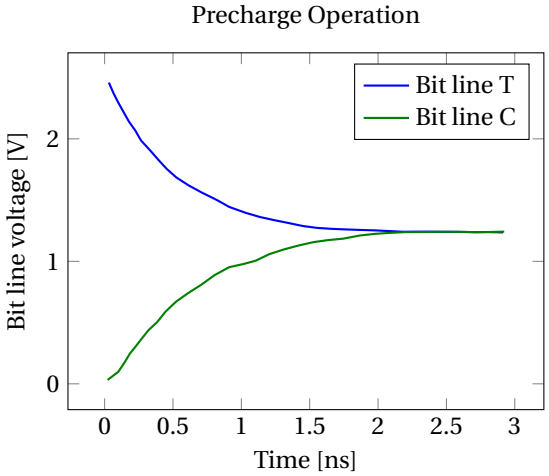
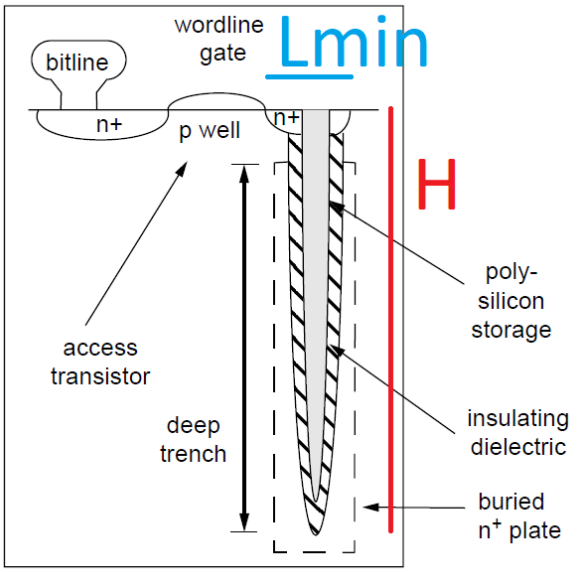
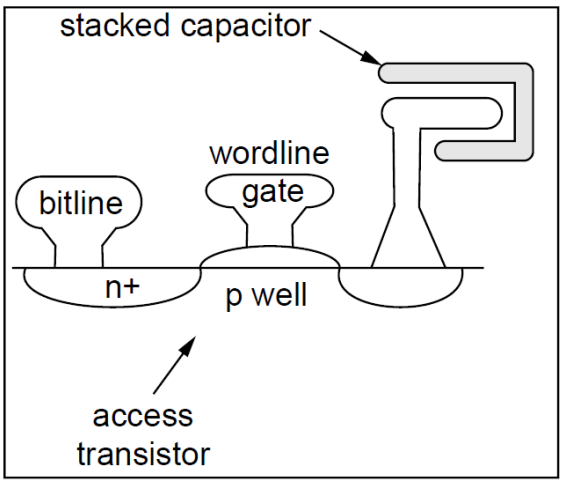


Figure 2.13: Precharge operation [24]

Figure 2.14: Sense and precharge operation of a DRAM [24]



(a) A DRAM trench capacitor with indicated aspect ratio [21]



(b) A DRAM stacked capacitor [21]

Figure 2.15: Two types of DRAM capacitors; trench in the substrate and stacked over the substrate

2.5. Technology

This section will go deeper into the technology that is needed to make a DRAM in silicon. As the storage capacitor is quite uncommon in general IC designs, this will be specially addressed in Section 2.5.2.

2.5.1. IC Technology

The requirements for a DRAM chip are different from those for a generic digital IC. The differences originate from the fact that in a DRAM low leakage, in order to keep a high retention time, is one of the most important properties a transistor needs to have while CMOS circuits are usually developed for the highest operating speeds possible [21], [25]. Lower leakage can be ensured by setting the threshold voltage of the transistor to higher values while speed can be increased by lowering this threshold voltage. This will be presented in more detail in the next chapter [21].

As most of the chip area is used for memory cells, the lay-out of the cells can be made repetitive. The remaining subcircuits on the chip are small in measures of area and their wiring is not complex either. This leads to chip designs that require less metal layers than for example processors from the same period do [21]. In comparison, in 2007 Intel presented their 45 nm technology which contains up to nine metal layers, while a DRAM from Micron in the same period was manufactured with only three metal layers [26], [27].

2.5.2. Capacitor Technology

Every memory cell needs a capacitor to store data. In order to keep a high memory density, a special design is needed for these capacitors to minimize the area of a cell. There are two ways to do this; One solution is to make a trench in the silicon wafer where the capacitor is buried, the other is to make a capacitor in the metal layers over the active area. Figures 2.15a and 2.15b show these designs respectively. The ratio between the capacitor height and the minimum feature size is called the aspect ratio (AR) of a storage node, AR = h_{cap}/L_{min} and is indicated in Figure 2.15a [21]. In order to keep increasing the memory density of chips while still meeting the performance requirements, the AR increases. This means that with every more modern technology the capacitor becomes taller while its footprint area decreases [21].

In order to have a capacitance as high as possible while maintaining a small footprint, manufacturers started to use dielectrics with a higher dielectric constant than SiO_2. These so-called high-k dielectrics have more defects per unit volume than SiO_2 has [28]. This increases the probability of dielectric breakdowns as the defects may form a conducting path between the electrodes. More on this dielectric breakdown can be found in Section 3.2.4.

3

DRAM Reliability

This chapter introduces DRAM reliability. First an introduction in the terminology and math related to reliability engineering will be given in Section 3.1. After this Section 3.2 will introduce the basic operations of the transistor and its related wear-out mechanisms. Then Section 3.3 will present DRAM specific failures in more detail. Finally, in Section 3.4 the DRAM reliability model used in this research will be described. The circuit as well as the wear-out mechanism models will be presented.

3.1. Reliability Engineering

This section describes the basic terminology and mathematics of reliability engineering.

3.1.1. Reliability Mathematics

Reliability is defined as the amount of correct functioning devices, given as a percentage, at a certain point in time. Adapted to circuits, Gielen, Wit, Maricau, *et al.* define reliability in circuits as *"Reliability is defined as the ability of a circuit to conform to its specifications over a specified period of time under specified conditions."* [1]. Reliability can be described with the reliability function, $R(t)$, expressed in Equation 3.1. $R(t)$ describes the relative amount of devices that function to their designed performance over time. $F(t)$ denotes the failure distribution, or the amount of devices that do not meet the design criteria. Devices that do not meet the criteria have failed.

$$R(t) = 1 - F(t) \tag{3.1}$$

The amount of failures over time is given by the hazard or failure rate $z(t)$ described in Equation 3.2. It indicates the lifetime expectation for the remaining population of correct functioning devices.

$$z(t) = -\frac{1}{R(t)} \cdot \frac{\mathrm{d}\,R(t)}{\mathrm{d}\,t} \tag{3.2}$$

Finally the mean-time-to-failure (MTTF) or lifetime is defined in Equation 3.3. The MTTF is the expected outcome of the reliability distribution, $R(t)$. If the system has a constant hazard rate, $z(t) = \lambda$, the MTTF is then thus given by MTTF $= \frac{1}{\lambda}$.

$$\mathrm{MTTF} = \int_0^\infty R(t)\,\mathrm{d}\,t \tag{3.3}$$

Summarizing, reliability is a continuous function that describes how a population of circuits degrades over time while the lifetime describes the typical or expected time at which a circuit will fail.

Reliability of a system can be derived in multiple ways, all varying in complexity and assumptions that are made. A first order approach in finding the reliability of a system is by multiplication of the system's individual components reliability [4]. The reliability of a system consisting of n parts can be described by Equation 3.4 in which W_k denotes a weighting factor for every subcomponent's reliability function and $R_k(t)$ is the subcomponent's reliability function.

$$R_{system}(t) = \prod_{k=1}^{n} W_k R_k(t) \tag{3.4}$$

This first order approximation assumes that all components' reliability is uncorrelated, which is generally untrue. In the case of DRAM degradation for example, the supply voltage regulator might degrade and decrease supply voltage. The maximum cell voltage is then decreased as well which causes the retention time to drop, degrading the reliability of the cell. Examples of reliability prediction methods which take these influences into account can be found in [13].

3.1.2. Reliability in Electronics

Reliability in electronics can generally be described by the bathtub curve, which can be seen in Figure 3.1 [29]. The curve presents the failure rate over time and shows three important regions. Failures in the first region, where the failure rate decreases, are called 'infant mortality'. The devices that fail during this period usually suffer from some kind of production failure. Manufacturers do not want to sell devices that have a high chance of failure early in the product lifetime so they filter these weak devices in the production facilities [3]. A form of filtering can be done by so called burn-in testing in which a device is stressed at elevated temperatures and voltages with the goal to fail and find the weak devices. Devices that do not fail this test are considered to perform according to the designed requirements. The next period depicted in the figure is the normal operation period of a device. In this period the devices operate as they should and only random failures define the failure rate, which is at its minimum. When the devices start to reach their designed lifetime, the hazard rate starts to increase again and the reliability of the device will drop fast. It is said that the circuits are wearing-out [29]. The next section will go deeper into these wear-out or failure mechanisms.

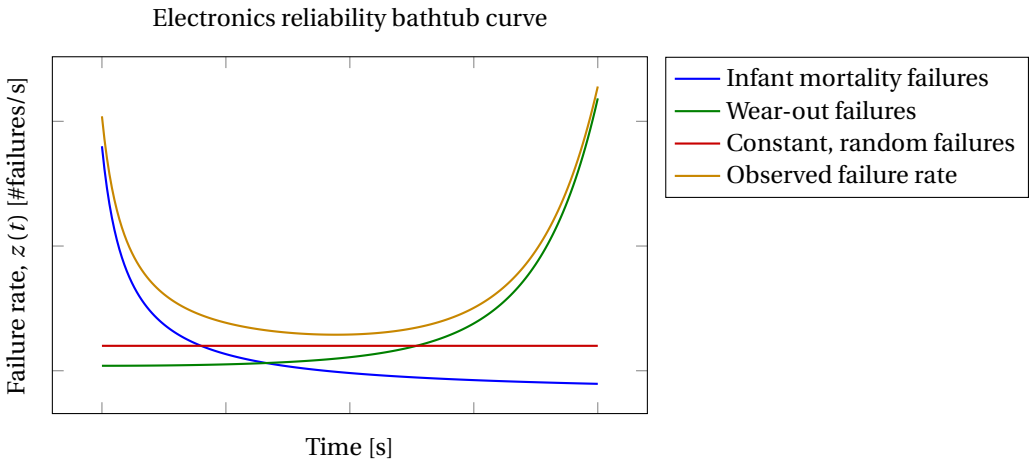


Figure 3.1: Reliability bathtub commonly seen in electronic wear-out [29]

3.2. CMOS Failures

In order to understand the factors influencing the reliability of DRAM, it is important to get familiar with the basic failures that occur in regular CMOS integrated circuits. This section will introduce five commonly observed failure mechanisms: bias temperature instability, hot carrier injection, time dependent dielectric breakdown, radiation failure and electromigration. First, a basic introduction of the metal-oxide-semiconductor field-effect transistor (MOSFET) is presented as four of the aging effects mentioned above occur in this device. Subsequently, the five wear-out phenomena will be described.

3.2.1. Transistor

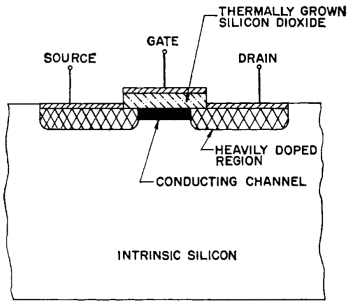
In this section an introduction of the MOS transistor is presented. The basics of the physical construction as well as its operating behavior will be described first, to be concluded by a specific leakage mechanism, gate induced drain leakage.

Structure

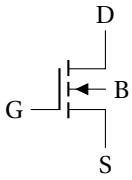
The MOSFET is the most used type of transistor in electronic circuits to date. Figure 3.2a shows the basic physical structure of a n-type MOSFET (NMOS) [30]. A MOSFET is made on a wafer. This wafer is usually charged with p-type carriers and forms the substrate of the transistor. When a PMOS transistor needs to be

made, a n-well can be made with n-type carriers in the substrate. To make a transistor first an isolating oxide is grown over the wafer. Afterwards the gate (G) made from either polysilicon or metal is placed over this oxide. Then the oxide that is not covered by the gate is removed. In the uncovered areas doping ions (n-type carriers for NMOS, p-type for PMOS transistors) are implanted which will form the drain (D) and the source (S) of the MOSFET [31]. When a gate voltage is applied, charges are attracted under the gate which form a conducting channel between the drain and the source [32].

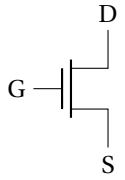
A MOSFET is a symmetric device, e.g. drain and source can be swapped without consequences. A MOSFET has four electrical connections; one to gate, two to drain and source and one to the substrate, also called bulk (B). Figure 3.2b and 3.2c show respectively the circuit symbols for the four terminal device and the digital device. In digital circuits the bulk is typically connected to ground in NMOS devices and to V_{DD} in PMOS devices, therefore this connection is not drawn in the figure.



(a) Basic MOSFET structure. Copied from [30]



(b) Four terminal MOSFET representation



(c) Three terminal MOSFET representation used in digital representations

Figure 3.2: Physical and circuit representations of a n-type MOSFET

MOSFET Operation

The following text holds for NMOS transistors. To apply the information to PMOS transistors, voltages and currents should be inverted and doping types should be switched. E.g. the bulk is made with n-type carriers in stead of p-type, current flows in the reverse direction. Current flows from drain to source when a voltage larger than a certain threshold voltage, V_{th}, is applied from gate to source, $V_{th} \leq V_{GS}$ [32], where V_{GS} is the gate-to-source voltage. If $V_{GS} < V_{th}$, the transistor is in cut-off and ideally no current flows. The region between drain and source is called the channel. It has a length, L, and a width, W. When the source is connected to the ground of the circuit, Equations 3.5 and 3.6 hold for the drain current, I_D [25]. Equation 3.5 applies when the device is in so called linear or triode region and Equation 3.6 holds when the device is in saturation. In Figure 3.3 the cut-off, triode and saturation region of a transistor are indicated.

$$I_D = k_n^{'} \frac{W}{L} \left((V_{GS} - V_{th}) V_{DS} - \left(\frac{V_{DS}}{2} \right)^2 \right) \quad \text{when} \quad V_{GS} > V_{th} \quad \text{and} \quad V_{DS} < (V_{GS} - V_{th}) \tag{3.5}$$

$$I_D = \frac{k_n^{'}}{2} \frac{W}{L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \quad \text{when} \quad V_{GS} > V_{th} \quad \text{and} \quad V_{DS} \geq V_{GS} - V_{th} \tag{3.6}$$

In these Equations V_{DS} is the drain-to-source voltage, $k_n^{'} = \frac{\mu_n \epsilon_{ox}}{t_{ox}}$ is called the transconductance parameter depending on the oxide thickness, t_{ox}, mobility of the carrier, μ_n and the oxide permittivity, ϵ_{ox}. When the device length of the channel is shortand the drain source voltage is high, thus resulting in a strong electric field in the channel, the channel can come in pinch-off. The channel becomes shorter than the physical channel length. This increases the current through the device and its influence is given by λ [25], [32]. In Figure 3.3 the influence of λ can be seen in the increase of the drain current in the saturation region.

Both Equations only apply when $V_{GS} > V_{th}$. If this is not the case, the transistor is in its off-state. In the off-state a small current still flows through the transistor, called the sub-threshold current. This current is given by Equation 3.7 [25].

$$I_D = I_S \exp\left(\frac{V_{GS}}{nkT/q} \right) \left(1 - \exp\left(-\frac{V_{DS}}{kT/q} \right) \right) (1 + \lambda V_{DS}) \tag{3.7}$$

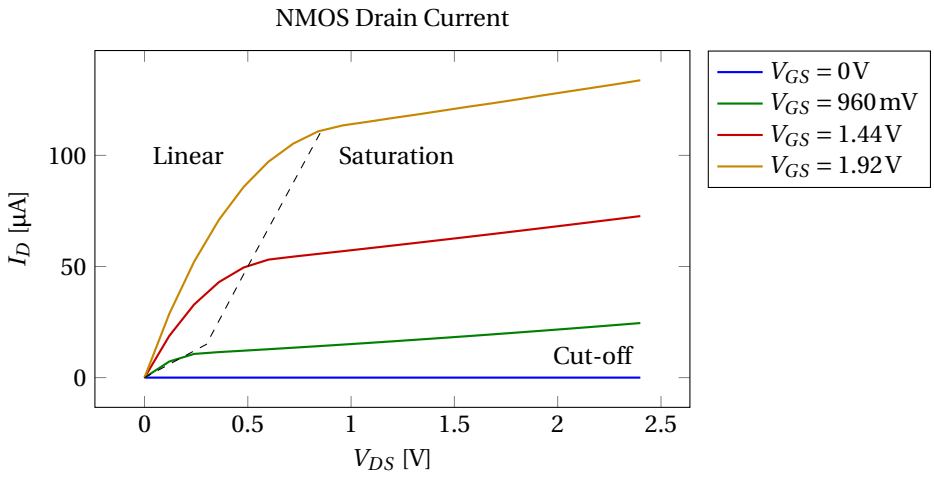


Figure 3.3: Linear, cut-off and saturation NMOS operating regions

In this Equation I_S and n are empirical defined parameters; k is the Boltzmann constant, q is the elementary charge and T is the device temperature.

The threshold voltage of a MOSFET is given by Equation 3.8 [25].

$$V_{th} = V_{th,0} + \gamma\left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right) \tag{3.8}$$

In this Equation V_{SB} denotes the voltage from source to bulk. γ is the body-effect coefficient, indicating the influence of changes in V_{SB} on the threshold voltage and ϕ_F is the Fermi potential. $V_{th,0}$ finally is the threshold voltage if no substrate bias voltage is applied. From this equation it follows that the threshold voltage of a MOSFET is dependent on its bulk bias and physical parameters. Equation 3.9 shows the calculation of $V_{th,0}$ [32]. In this equation V_{FB} is the flat band voltage, ϕ_{fp} is the difference between the Fermi level and the intrinsic Fermi level, ϵ_s is the permittivity of the semiconductor, N_a is the acceptor donor concentration and C_{ox} is the oxide capacitance. The threshold voltage thus depends on the oxide thickness, the oxide material and doping concentrations in the device [32].

$$V_{th,0} = V_{FB} + 2\phi_{fp} + \frac{\sqrt{2\epsilon_s q N_a (2\phi_{fp})}}{C_{ox}} \tag{3.9}$$

In the linear regime a transistor can be used as an amplifier. Small input voltage changes will results in larger current swing through the device. The relation between the output current and input voltage is called transconductance and is for a MOSFET described in Equation 3.10 [32].

$$g_m = \frac{2I_D}{V_{GS} - V_{th}} \tag{3.10}$$

The conditions in Equation 3.5 and 3.6 have the consequence that a n-type device is unable to pass the full drain source voltage. A drop equal to the threshold voltage of the device will occur. Following the same reasoning, a p-type device is unable to bring a voltage down to zero volt, again the threshold voltage will remain.

When $V_{GS} < V_{th}$ the transistor ideally should turn off and not conduct any carriers. However, in real devices this is not the case and a small current will flow as described in Equation 3.7. In Figure 3.4 the subthreshold current of a MOSFET is plotted [33] for negative V_{GS}. It follows that the current drops very fast when $V_{GS} < V_{th}$ until $V_{GS} \approx -0.5\,V$. From this point it increases again and Equation 3.7 no longer holds. This leakage current is now mainly caused by gate induced drain leakage.

Gate Induced Drain Leakage

Gate Induced Drain Leakage (GIDL) occurs in the overlapping area between the gate and the drain of a transistor [34]. If the gate of a NMOS device is biased at a low voltage, while the gate is at a high voltage, a deep-depletion region is formed in the overlapping region. For a PMOS device, this occurs when the gate is biased

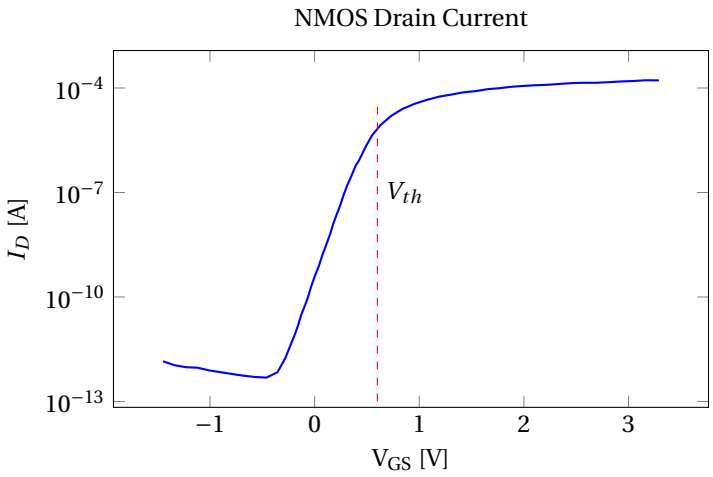


Figure 3.4: Subthreshold current in a MOSFET [33]

at high voltage while the drain is biased at a low voltage. Carriers can tunnel through the drain-to-bulk band gap and are then swept away [34]. This process is schematically presented in Figure 3.5 [34]. GIDL is an important leakage mechanism in DRAMs. Since the cells are accessed infrequently, the charge stored on the capacitor creates a strong electric field between in the gate and drain; this gives rise to GIDL currents. Hence, GIDL decreases the retention time of a memory cell.

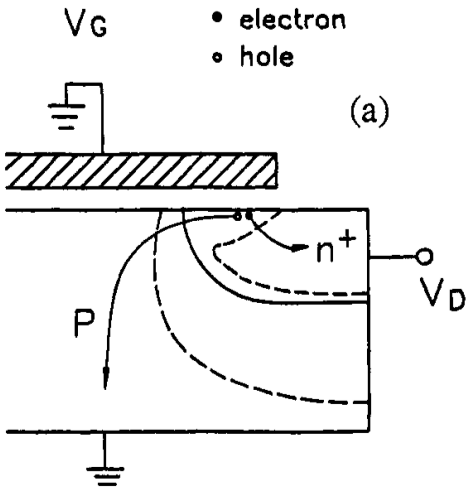


Figure 3.5: Deep-depletion in the gate drain overlapping region causes GIDL [34]

When there are traps present in the gate drain overlapping region, GIDL current increases. The traps form energy states in the band gap that can be occupied by the carriers. From these states, the carriers may either continue to cross the band gap or fall back in their previous energy band. Since these energy states are within the band gap a carrier needs less energy to cross the complete band gap [32].

3.2.2. Bias Temperature Instability

Bias Temperature Instability (BTI) shifts the transistor threshold voltage due to bond breaking in the Si-SiO_2-interface. The crystalline structure of pure silicon, as used in the substrate, and of silicon dioxide, the gate insulator, do not match. This means that some of the bonds of the oxide are not connected to any of the silicon atoms from the bulk [31]. In the production process a so called annealing step is used to bind the dangling bonds with hydrogen atoms. If this would not be done, the dangling bonds act like trapped charges at the silicon interface [32]. This results in an increase of the threshold value as these charges then add to N_a in Equation 3.9 [31], [32]. A schematic representation of the Si-SiO_2-interface is presented in Figure 3.6 [35].

The bonds between the hydrogen (H) and silicon (Si) atoms can break due to high electric fields and

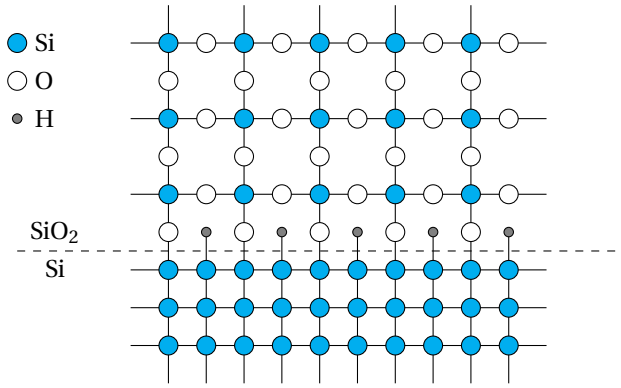


Figure 3.6: Schematic representation of Si-SiO_2-interface [35]

increased temperatures [1]. If a gate bias voltage is applied to a transistor, a high electric field will form through the gate oxide. This will break the bonds of the hydrogen atoms. The freed hydrogen atoms will start to diffuse through the gate insulator until they finally are swept away in the gate material. When the hydrogen atoms diffuse they leave an unbonded trap in the Si-SiO_2-interface which increases the threshold voltage. If the electric field is removed, some of the hydrogen atoms may diffuse back to the silicon and silicon dioxide interface thus decreasing the threshold voltage again. The effect of generation and recovery of traps on the threshold voltage is depicted in Figure 3.8 [36]. Because a transistor can partially recover from BTI degradation, the long-term threshold voltage shift depends on the duty factor at which the transistor operates. Figure 3.7 shows the relative impact of the duty factor for several sources in literature.

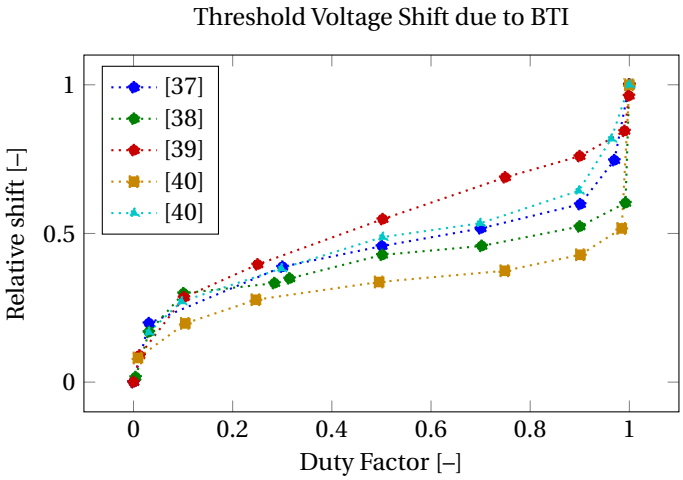


Figure 3.7: Impact of duty factor on long-term threshold voltage shift caused by BTI. All shifts are plotted relative to DC long-term stress

With increasing temperature more hydrogen atoms will diffuse away from the boundary and thus the threshold voltage will increase faster with increasing temperature. The effect described above is called bias temperature instability BTI [1].

Traditionally PMOS transistors suffer the most from BTI effects when their gates are biased negatively [41], [42]. The resulting BTI is then called Negative Bias Temperature Instability (NBTI). NMOS transistors generally suffer less from BTI effects than PMOS transistors, but the most noticeable degradation occurs when their gates are biased positively. This is then called Positive Bias Temperature Instability (PBTI) [42]. PMOS transistors are less prone to PBTI effects and NMOS transistors are less prone to NBTI effects than to the other two variants [42]. Modern transistors have a gate that is made form a high-k material which has a higher permittivity than SiO_2. These high-k materials have a higher defect density than SiO_2 which results in an increment of PBTI effects in NMOS transistors [28], [42], [43] relative to transistors with a polysilicon gate.

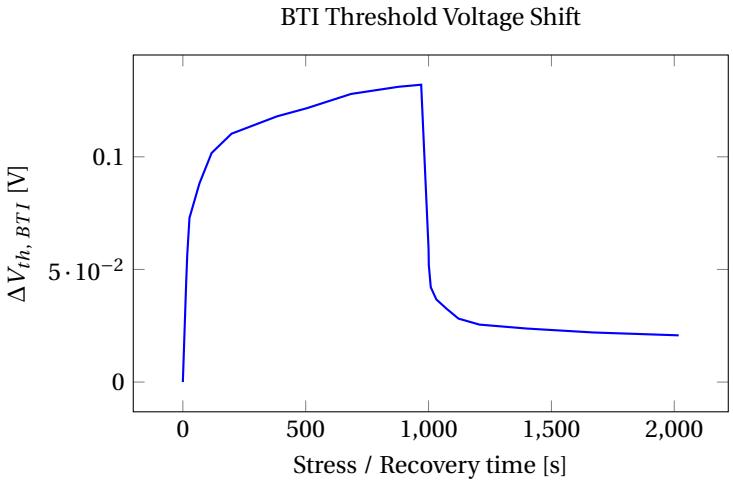


Figure 3.8: Generation and recovery of BTI wear-out [36]

3.2.3. Hot Carrier Injection

Another important wear-out factor in transistors is called Hot Carrier Injection (HCI) [44]. The strong electric field near the drain edge in the channel energizes carriers making them hot carriers. These hot carriers then break silicon-hydrogen bonds, just like in BTI, or occupy trap states in the gate oxide thus increasing the threshold voltage. Since the carriers only become hot near the end of the channel, the degradation takes place in this region. The bond braking phenomenon is presented graphically in Figure 3.9 [45]. The carrier can also be swept in the bulk of the transistor. If this happens, or the carrier is swept into the gate of the transistor, the net current flow is decreased and thus g_m decreases [1], [29], [44]. NMOS transistors are more prone to HCI than PMOS because electrons, the carriers in n-type devices, have a lower energy barrier to become hot electrons than holes, the p-type carriers, have to become hot holes [18], [44], [46]. The barrier for an electron is approximately equal to 3.2 eV and for a hole 4.7 eV [46]. Recovery as seen in BTI does not occur in HCI [44]. The threshold voltage can only increase with time.

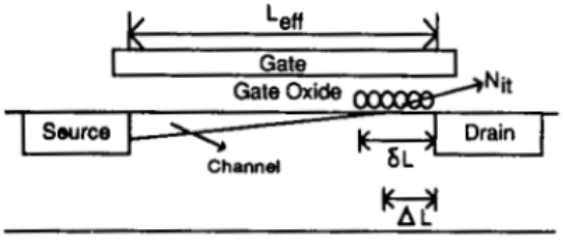


Figure 3.9: Hot carrier injection bond breaking at the end of the channel [45]

3.2.4. Time Dependent Dielectric Breakdown

The previously described wear-out phenomena, BTI and HCI, can be characterized by the gradual degradation of transistor properties over time. Time Dependent Dielectric Breakdown (TDDB) is a more direct wear-out phenomenon [1], [47]. Under influence of the strong electric field over the oxide, defects can be formed in the oxide. These defect states can be occupied by carriers, as mentioned in the sections above, that might start to form a gate current via these traps. Once a certain threshold of traps is reached, a conducting path is formed where large currents can flow. Figure 3.10 presents this schematically [48]. The oxide heats up and finally the transistor breaks down completely [1], [47].

3.2.5. Radiation Failure

Ionizing radiation can create electron-hole (eh) pairs in transistors that degrade of even disrupt the their performance [10]. The damage that occurs depends strongly on the dose rate (D) of the radiation given in rad/s. The total amount of ionizing radiation that accumulates in a transistor is called Total Ionizing Dose

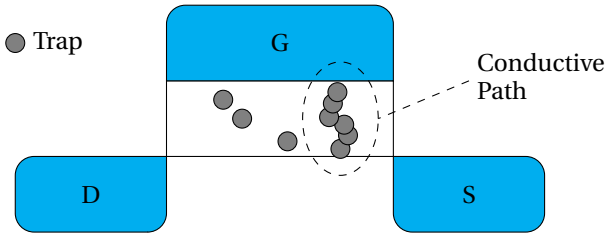


Figure 3.10: Time dependent dielectric breakdown caused by a conducting path of defects in the gate oxide [48]

(TID) and is measured in rad. In the case of low dose rates, the eh-pairs can get trapped in the oxide or break a Si-H bond at the Si-SiO_2-interface which both results in an increase of the threshold voltage, as was described earlier [10], [32]. The trapped charges in the oxide can lead to TDDB effects where a leakage current flows through the oxide and a breakdown of the oxide may occur. Higher dose rates can result in latch-up or even complete destruction of the transistor. Latch-up is the phenomenon where a parasitic bipolar transistor in the MOSFET is activated by the large amount of energy released by the ionizing particles [32]. A large current starts to flow making a short circuit. Even higher dose rates or energized particles can destroy the material properties of the transistor and render it useless [10].

3.2.6. Electromigration

Electromigration (EM) is a wear-out phenomenon that occurs in the metal in a chip [1]. Both mechanical and electrical stress can cause a drift of ions in a metal [1], [5]. For example, if a strong current were to flow always in one direction, the electric field causes metal ions to diffuse through the metal. At certain points more ions will start to diffuse, thus decreasing the overall metal width and further increasing the electric field in that part, increasing the diffusion even more. This finally can lead to a gap where metal is missing and an open circuit is formed. On the other hand the ions will start to move to certain points in the metal. Here a build-up of ions, a hillock, occurs which can lead to short-circuit situations [1]. In Figure 3.11 both the formation of hillocks and voids is presented [49].

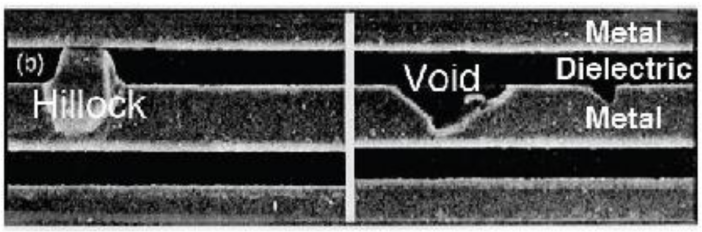


Figure 3.11: Formation of hillocks and voids due to electromigration [49]

3.3. DRAM Specific Failures

This section introduces DRAM specific failures. The failures are split in soft and hard faults [24].

3.3.1. Soft Faults

A soft fault happens when the data is written to a cell and has changed when reading it after some time which is shorter than the retention time.[24]. Soft faults are non-destructive and correctable after a write operation. The non-destructive property of this type of fault refers to the DRAM hardware, it does not suffer from any damage when this fault occurs. This also implies that the fault is correctable after a write operation because if this was not the case the circuit would have been damaged [24].

Causes of soft faults can be ionizing radiation or capacitive coupling with other cells [24], [50]. Ionizing radiation can release large amounts of electron hole-pairs that can add charge to the storage capacitor which might result in flipping a logical bit [50]. Cells can get capacitively coupled with other cells via densely placed bit lines. Operations on one cell influence a coupled cell which might result in a soft fault [24], [51]. It is found that the generation of soft faults depends on the data background of neighbouring cells [52], [53].

3.3.2. Hard Faults

In contrast to soft faults hard faults are uncorrectable and lead to direct circuit failure [24]. Examples of hard faults are broken connections in the cell or in the wiring which make it unable to write and read to that cell or stuck-at faults in which a cell is always set to a zero or an one [24], [54]. In order to keep the yield of a production process high, replacement rows are made on the chips. If a cell suffering from a hard fault is found during the testing of the product, the row containing it is disconnected and replaced by one of the replacement rows [21].

Soft faults may become hard faults. E.g., due to TDDB, the cell might have an increased leakage current, reducing the retention time [54], [55]. If the dielectric breaks down, no more data can be stored in the cell and the soft fault became a hard fault.

3.4. Reliability Simulation Model Description

In this section the model used to simulate the wear-out and aging of a DRAM will be described. First the circuit topology used for these simulations is presented in detail. The second part of this section consists of an overview of commonly used wear-out simulation methods and the description of the aging models used in this thesis.

3.4.1. Circuit

Due to the secretive and highly competitive nature of the DRAM industry it was impossible to use a modern day DRAM circuit. In order to keep the results realistic, a commercial DRAM netlist from the 1990s was used. Transistor simulation is performed on SPICE simulation level 3.

The circuit can be divided into two sections. The first section consists of the memory cells and their connecting bit lines which lead to a local sense amplifier. The sense amplifier is connected to column selectors which form the boundary between the two sections. After the column selectors the global precharge and sense amplifiers, which form the second section, are located. Other parts of the circuit, like the control logic, are included as ideal controlled voltage sources, not as hardware. Aging of digital circuits is investigated for example in [5], [13], [15].

The next sections will describe the circuit following the read path up, down from a single memory cell to the output buffers of the circuit. After this, the timing and the waveforms of this netlist will be discussed.

Memory Cell

In this design 1T1C memory cells are used meaning that one capacitor is accessed by one access transistor. Figure 3.12 shows the cell simulation model. The polysilicon that is used to connect the transistor with the capacitor has a higher resistivity than metal has. This is modeled as a resistance between drain and capacitor.

Note that the access transistor is a PMOS transistor. The bulk is biased at V_{DD}. This means that leakage in the transistor will charge the capacitor towards V_{DD}, instead of discharge it. Table 3.1 lists the important design details of the components in the memory cell.

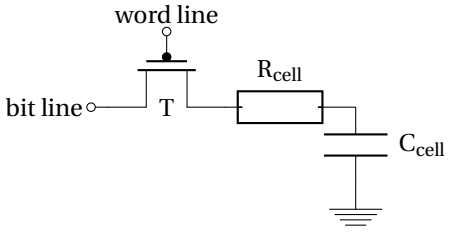


Figure 3.12: Memory cell

Table 3.1: Memory cell details

Parameter	Value	Description
T_L		Transistor length
T_W		Transistor width
R_{cell}		Drain-Capacitor resistance
C_{cell}		Cell capacitance

Reference Cells

The reference cells are also 1T1C cells. Figure 3.13 shows the circuit for two reference cells, both connected to their own bit line. T_{eq} is the equalization transistor. It is used to bring the reference cells to equal reference voltage, V_{ref}, before a new read operation is performed. Again, the higher resistivity of the polysilicon connection is modeled as a resistor. In Table 3.2 the specifications of this subcomponent are presented.

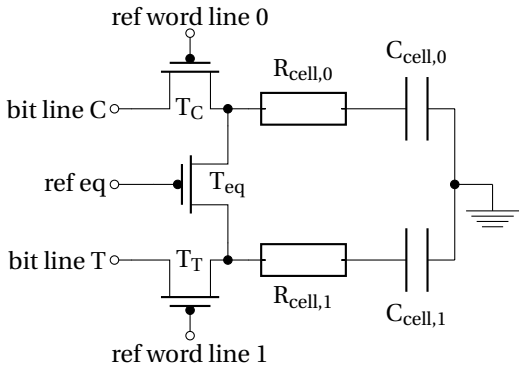


Figure 3.13: Reference cells

Table 3.2: Reference cells details

Parameter	Value	Description
$T_{T,C,eqn\ L}$		Transistor length
$T_{T,C,eqn\ W}$		Transistor width
$R_{cell\ 0,1}$		Drain-Capacitor resistance
$C_{cell\ 0,1}$		Cell capacitance

Sense Amplifiers

Figure 3.14 shows the sense amplifier circuit. The local sense amplifiers are smaller than the global ones but the circuit lay-out is nearly equal. Because the bit lines are always charged to V_{DD}, it is more important for a fast reading operation that the sense amplifier is able to discharge the bit lines fast. To allow for this, the faster NMOS transistors are designed wider than the charging PMOS transistors.

Operation of the local sense amplifier is controlled with the *setN* node in the circuit. This node is controlled by large transistors connecting the sense amplifier to GND, which enables fast discharging of the bit lines. The global sense amplifier is driven with a signal directly to one large discharge transistor connecting the sense amplifier to GND. Table 3.3 lists the transistor sizes used in both the local and global sense amplifiers.

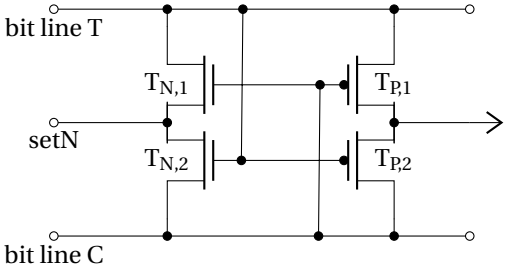


Figure 3.14: Sense amplifier

Table 3.3: Sense amplifier details

Parameter	Value	Description
$T_{N,1,2,local\ L}$		Transistor length
$T_{N,1,2,local\ W}$		Transistor width
$T_{P,1,2,local\ L}$		Transistor length
$T_{P,1,2,local\ W}$		Transistor width
$T_{N,1,2,global\ L}$		Transistor length
$T_{N,1,2,global\ W}$		Transistor width
$T_{P,1,2,global\ L}$		Transistor length
$T_{P,1,2,global\ W}$		Transistor width

Precharge Devices

Figure 3.15 shows the precharge circuit. Both the local and global equivalents use the same topology, but differ in transistor size. Table 3.4 presents the sizes of these transistors. When an operation is completed, the bit lines are precharged back to V_{DD} by setting the *precharge* node to a low voltage.

Timing

Now that the subcomponents are presented, their internal relationship and timing can be discussed. In order to give a clear illustration of all the events that happen during one clock cycle, a *0W0* operation is split into four pieces and explained in detail. One clock cycle in this DRAM circuit takes $T_{clk} = 60$ ns.

The start of the operation is presented in Figure 3.16 and is equal to the start of a read operation. The bit lines are precharged to V_{DD}, the reference cells to V_{ref} and in the memory cell a logical '0' is stored. When the word line is activated at $t = 6$ ns, both the memory cell on bit line T and reference cell 0 on the bit line C, share their charge with their bit lines. Since both the bit lines are at a higher potential than the cells, their internal capacitor voltages increase and the bit line potentials drop slightly. Because the voltage difference between the reference cell and its bit line, bit line C, is less than that of the memory cell and bit line T, bit line T is discharged to a lower voltage. Note that reference cell 1 is not activated yet and still holds V_{ref}. At $t = 11$ ns the local sense amplifier is enabled and the sensing operation starts. Since bit line T has a lower voltage than bit line C, the sense amplifier will discharge bit line T to GND and restore bit line C to V_{DD}.

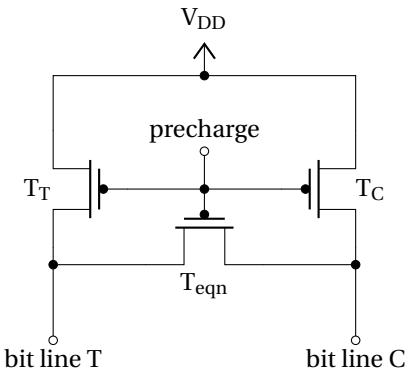


Figure 3.15: Precharge circuit

Table 3.4: Precharge circuit details

Parameter	Value	Description
$T_{T,C,local\ L}$		Transistor length
$T_{T,C,local\ W}$		Transistor width
$T_{eqn,local\ L}$		Transistor length
$T_{eqn,local\ W}$		Transistor width
$T_{T,C,eqn,global\ L}$		Transistor length
$T_{T,C,eqn,global\ W}$		Transistor width

The memory and reference cells are still connected to these bit lines and thus their internal voltage is also discharged and charged respectively. The memory cell voltage does not decrease back to its original value due to the threshold voltage of the access transistor and the word line potential supplied to its gate.

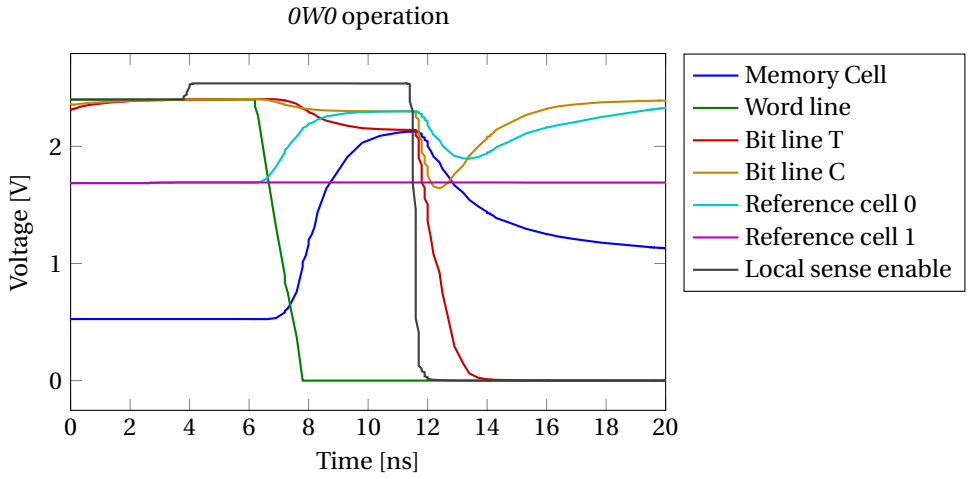


Figure 3.16: Sensing during a *0W0* operation

In Figure 3.17 the global circuitry sets its bit lines to the values that have to be written to the cell later in the cycle. With the activation of the write signal at $t = 20$ ns, global word line T is set to 0 V. This period of the cycle is only used for writing operations and gives room for previous reading operations to extend in time, for example, due to aging.

Figure 3.18 shows the period of the cycle where the bit lines are coupled to the global circuitry via the column selectors. The word line voltage drops below 0 V in order to mitigate the threshold voltage problem mentioned above. Besides that, now also reference cell 1 is connected to bit line T and discharged as well. Again the threshold voltage of the reference cell transistor limits how far the cell will discharge. 43 ns after the start of the cycle, the column selector becomes active and allows the data from the global bit lines to be fed into the memory cell or to read out the local bit lines with the global sense amplifier and forward the data to the output of the chip if a read operation would have been performed. After $t = 47$ ns, the global circuitry becomes disconnected again from the local circuitry after which the global bit lines are precharged again.

The last part of the cycle is presented in Figure 3.19. After 51 ns the cell word line is deasserted and the cell voltage is definitively set. The reference devices are equalized to the average voltage of both cells and finally the bit lines are both precharged back to V_{DD}.

3.4.2. Workload and Duty Factor

In order to find realistic degradation of DRAM, it is important to know how the transistors are stressed. From the workload the stress for every transistor can be found and thus a prediction of the degradation can be

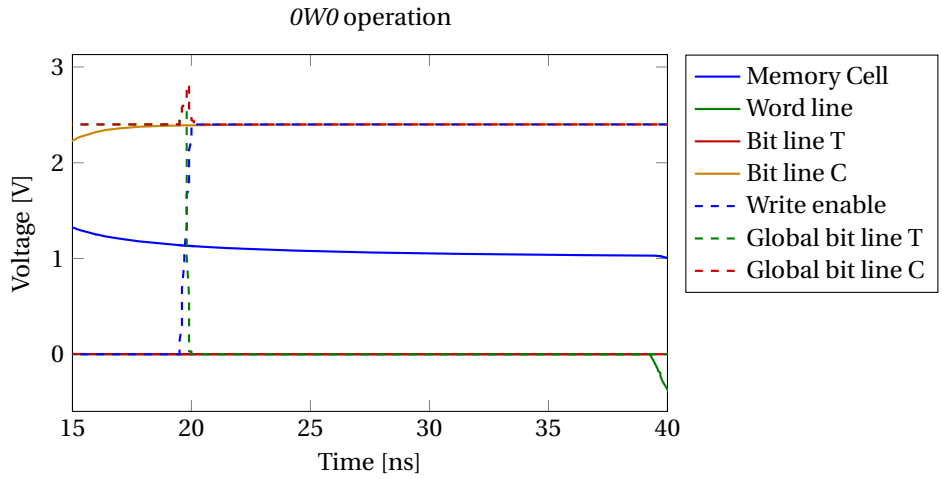


Figure 3.17: Sensing during a *0W0* operation

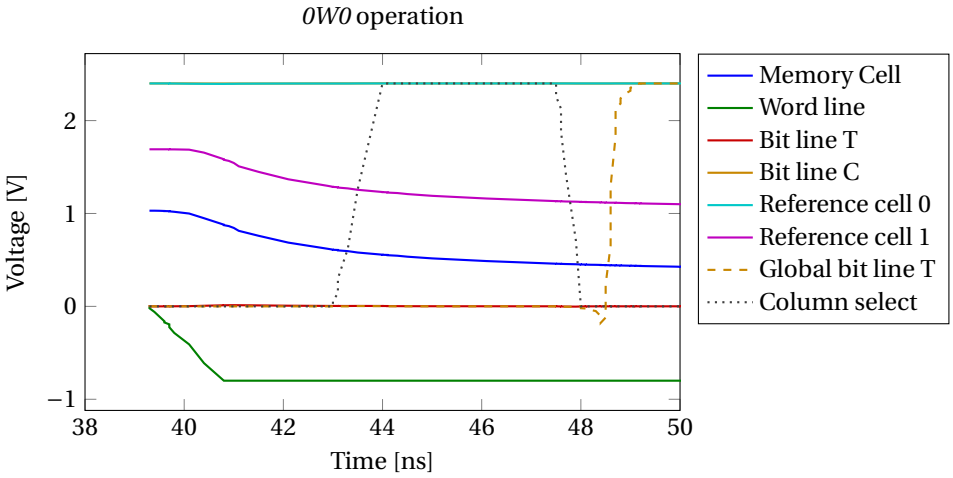


Figure 3.18: Sensing during a *0W0* operation

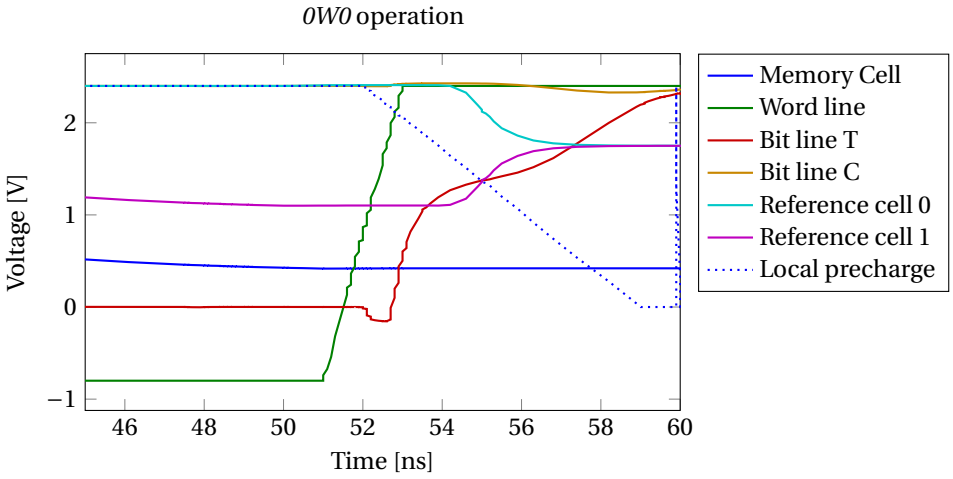


Figure 3.19: Sensing during a *0W0* operation

made more accurately. In order to do so, one must know the memory lay-out. Because it is unknown for which chip the circuit files are designed, a DRAM chip was selected with specifications that roughly match the properties of the circuit and its memory lay-out was assumed for this research. The selected reference SDRAM chip is manufactured by Siemens with type number: *HYB39S64400/800/160AT* [56]. Table 3.5 lists important parameters of this chip.

Table 3.5: Important parameters of selected reference DRAM chip, Siemens HYB39S64400/800/160AT [56]

Parameter	Value
# bits	64 Mbit
# banks	4
column address	8 bit
row address	12 bit
bits per data word	16
Technology	250 nm
V_{DD}	3.3 V
$f_{clk,\,max}$	125 MHz
$T_{refresh}$	64 ms

The workload needs to be split in the workload per column and the workload per cell to characterize the DRAM circuit. The column workload will be used to determine the degradation of sense amplifiers, reference cells and precharge devices while the cell workload will be applied to the cells. Since the circuit does not contain for example a row decoder, the row workload is not needed in this research but could be found using the same method that was used to find the other workloads as well.

To find the workload on transistors it, was assumed that accessing a data word happens randomly and independently from other accessing operations. This assumption would for example not hold in specific cases like row-hammering where a specific row is accessed repetetively with a high frequency [51]. The chance of accessing a single cell is given by the division of the total amount of bits by the amount of bits per word. Equation 3.11 shows this mathematically.

$$p_{cell} = \frac{1}{\#\,\text{bits/bits per word}} = \frac{1}{64\,\text{Mbit}/16\,\text{bit}} = 2.3842 \cdot 10^{-7} \tag{3.11}$$

The chance that a column will be accessed is given by the amount of columns multiplied by the amount of banks. Equation 3.12 summarizes this.

$$p_{column} = \frac{1}{\#\,\text{banks} \cdot \#\,\text{columns}} = \frac{1}{4 \cdot 2^8} = 9.7656 \cdot 10^{-4} \tag{3.12}$$

In order to see the differences in degradation due to the workload, two extreme cases were defined. In the best-case situation, only refreshes of the cell's contents stress the DRAM. The refresh rate forms a lower boundary on the workload on a cell transistor. One read-refresh cycle of the DRAM circuit takes $T_{clk} = 60\,\text{ns}$ while one refresh cycle of the complete DRAM takes $T_{refresh} = 64\,\text{ms}$ as stated in Table 3.5. The best-case duty factor for a cell is then given by Equation 3.13. DF_{WL} in this formula is the duty factor of the word line, WL, during one clock cycle, T_{clk}. The duty factor for a word line is the percentage of the complete clock cycle in which the line is active. From the timing graphs in Figures 3.16, 3.17, 3.18 and 3.19 it follows that the word line has a duty factor of $DF_{WL} = 0.718$.

$$DF_{cell,\,best-case} = DF_{WL} \cdot \frac{T_{clk}}{T_{refresh}} = DF_{WL} \cdot \frac{60\,\text{ns}}{64\,\text{ms}} \tag{3.13}$$

The best-case workload for a column device is given by Equation 3.14. The term $\frac{T_{refresh}}{\#\,\text{rows} \cdot T_{clk}}$ gives the amount of clock cycles per refresh operation. One of these cycles is used for the actual refresh operation, so 1 needs to be subtracted. $DF_{control}$ indicates again the duty factor of the driving source of the transistor.

$$DF_{col,\,best-case} = DF_{control} \cdot \left(\frac{T_{refresh}}{\#\,\text{rows} \cdot T_{clk}} - 1 \right)^{-1} = DF_{control} \cdot \left(\frac{64\,\text{ms}}{2^{12} \cdot 60\,\text{ns}} - 1 \right)^{-1} \tag{3.14}$$

The worst-case workload is based on the misses of the L2 cache in the SPEC CPU2006 benchmark [57] for a Intel Core 2 Duo Processor [58]. In [59] and [60] this benchmark was used to determine, among other

metrics, the L2 misses. The worst case L2 miss rate was reported in [59] to be $p_{miss,\,L2} = 50/1000$ for the 433.milc benchmark. If a L2 miss occurs, a complete memory block is replaced with one from the DRAM main memory, as was explained in Section 2.1. It is assumed in this research that one block has the size of 32 kB, which equals 16 kwords in the reference DRAM. The worst-case duty factor for a cell is shown in Equation 3.15 and for a column device in Equation 3.16.

$$DF_{cell,\,worst-case} = DF_{WL} \cdot p_{cell} \cdot p_{miss,\,L2} \cdot \text{words per block} \tag{3.15}$$

$$DF_{col,\,worst-case} = DF_{control} \cdot p_{column} \cdot p_{miss,\,L2} \cdot \text{words per block} \tag{3.16}$$

3.4.3. Aging Simulation Methods

This section goes into more detail in the state of art in simulating aging effects than the introduction in Section 1.2. The failure mechanisms covered in each simulation approach will be presented as well as an in-detail analysis of the aging modeling used. Every approach will be presented in the same order as they appeared in Section 1.2. Table 3.6 summarizes the details of all methods. In Section 3.4.3, the methods presented in this section will be used in the development of wear-out models for this research.

Berkely Reliability Tools

The BERT [12] were one of the first simulation methods to simulate the effects of aging on a circuit. It allows to simulate the effects of HCI, TDDB and EM although not combined. The BERT offer three different simulation modules which are independent from each other. In order to simulate HCI wear-out, first a fresh circuit is simulated using SPICE. Then from the operating conditions a so called AGE-parameter is calculated. The formulae to determine AGE for NMOS and PMOS devices are presented in Equations 3.17 and 3.18 respectively. H_s, H_g, m_s and m_gX are technology parameters, τ is the total stress time, I_{sub} is the substrate current, I_DS is the drain-source current, I_G the gate current and W the transistor width.

$$\text{AGE}_{NMOS} = \int_0^\tau \frac{I_{DS}}{W \cdot H_s} \cdot \left[\frac{I_{sub}}{I_{DS}}\right]^{m_s} \mathrm{d}\,t \tag{3.17}$$

$$\text{AGE}_{PMOS} = \int_0^\tau \frac{1}{H_g} \cdot \left[\frac{I_G}{W}\right]^{m_g} \mathrm{d}\,t \tag{3.18}$$

From Equations 3.17 and 3.18 and the further equations used to determine the actual aged parameter set [45], [61], [62], it becomes clear that these models do not take into account the temperature of the circuit in order to derive the aged operating conditions. The aged parameters are derived from an inter- or extrapolation on measured data points. E.g. given a calculated AGE, the BERT will try to interpolate the derived parameters from two points of the measurement data. The circuit is then simulated again but now with altered transistor models in order to see the effects of aging. In case of a NMOS device, a current source connected to drain and source is added to simulate the degradation of device current. For a PMOS device the SPICE library parameters I_D, V_{th} and η are altered. Figure 3.20 shows the altered model for the NMOS transistor. HCI simulations are thus done on the circuit level and the outcome is reliability since this method allows to see the degradation of the circuit over time.

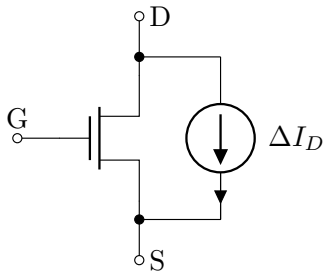


Figure 3.20: NMOS transistor model used to simulate aging effects due to HCI in the BERT [12], [62]

For TDDB and EM the approach is different. For these failure mechanisms the chip as a whole is evaluated and the outcomes are reliability for the complete system. For TDDB the needed oxide thickness that would

result in a breakdown after a user defined time is calculated. Taking into account the actual oxide thicknesses in the system from the previous calculation, the distribution of failing transistors can be derived. When EM is examined, the BERT simulate the fresh circuit and use the derived currents to find the time-to-failure for all the interconnects. This will result in a distribution of failures over time. Depending on the failure criteria a MTTF can also be found.

Reliability Aware MicroProcessor Model v1
The first Reliability Aware MicroProcessor model, RAMP v1 [5], is able to do so and predicts system reliability by performing an analysis on behavioral level. It estimates failure rates for the EM, stress migration (SM), TDDB and temperature cycling (TC) failure mechanisms per structure in the circuit. It is assumed that the failures are independent and have a constant failure rate. The MTTF for the complete system is given by summing all the failure rates, λ_{il}, for the different structures and failure mechanisms with a weighing function, W_{il}, as presented in Equation 3.19.

$$\text{MTTF}_{sys} = \frac{1}{\sum_{i=1}^{j}\sum_{l=1}^{k} W_{il} \cdot \lambda_{il}} \tag{3.19}$$

For some of the failure mechanisms, like EM, information about the circuit is needed to derive the MTTF. This implies that RAMP v1 can be partially aware of workload.

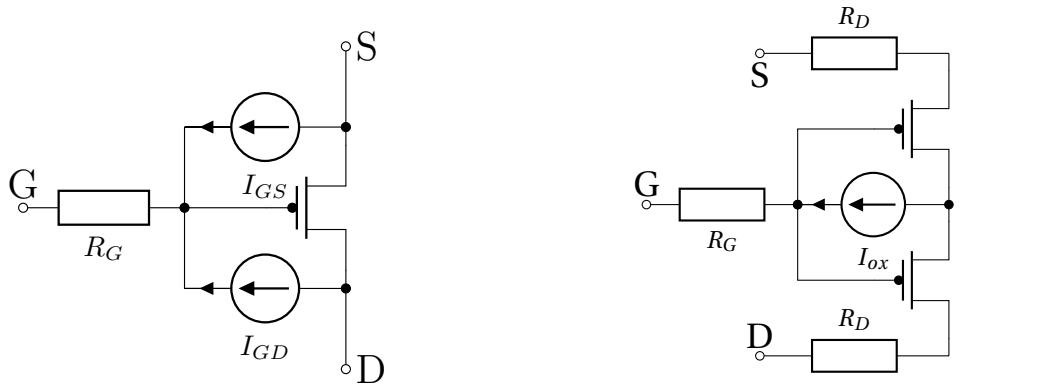
Reliability Aware MicroProcessor Model v2
A few years after the introduction of RAMP v1 the same researchers developed Reliability Aware MicroProcessor model v2, RAMP v2 [13]. The failure rate is not constant any more; in stead of an exponential distribution now a log-normal distribution of failures is used. This new distribution results in an increment of failure rate towards the end of life of a device. In other words, it becomes possible to simulate both the floor and the right wall of the bathtub curve as were presented in Figure 3.1. NBTI is also added as a failure mechanism in the evaluation. The third improvement over RAMP v1 is the ability to simulate parallel and redundant structures. To do so a Monte-Carlo simulation is performed to determine the reliability metrics of the parallel structures and their influence on the complete system. The Monte-Carlo approach results in a reliability distribution for the system.

Maryland Circuit-Reliability-Oriented
MaCRO [14], is able to simulate aging effects due to HCI, TDDB and NBTI. It first uses SPICE to find operating conditions of all the transistors. Once the operating conditions are known, the lifetimes of all the components for all different failure mechanisms are calculated. Equation 3.20 shows how the lifetime under HCI stress is calculated. In this Equation A_{HCI} is a model prefactor, n is a process parameter and $E_{a,\,HCI}$ is the activation energy, temperature is taken into account with parameter T, k is the Boltzmann constant and W is the device width.

$$t_{f,\,HCI} = A_{HCI}\left(\frac{I_{sub}}{W}\right)^{-n} \exp\left(\frac{E_{a,\,HCI}}{kT}\right) \tag{3.20}$$

When all the lifetimes are found the transistors are sorted based on these lifetimes. E.g. the most degraded transistors have the lowest lifetime prediction. The most degraded transistors are then replaced with aged equivalent circuits. It is possible to combine multiple failure mechanisms in one aged transistor model. Figure 3.21a shows the aged equivalent model used in MaCRO to simulate NBTI in a PMOS transistor and 3.21b shows how both NBTI and TDDB in a PMOS device are combined into one equivalent aged circuit model. The values of, for example, the current sources do not depend on the previously calculated lifetimes. In stead they are derived from operating conditions directly. The degraded circuit can be used to simulate further aging until circuit functionality is lost. With this information reliability and lifetimes of the circuit can be derived.

AgeGate
AgeGate [15] is able to simulate aging effects on digital gates. It predicts a lifetime for the circuit under test. AgeGate uses, like SyRA [18], a work profile to determine the load on the transistors in gates but also includes switching activity and temperature. The work profile is used to determine variations in gate delay and in the slopes of the signals. NBTI results in threshold voltage shifts while HCI will result in drain current variations. In order to derive the work profile, first a simulation is run without any aging factors involved. Then the



(a) PMOS transistor suffering from NBTI [63]

(b) PMOS transistor suffering both NBTI and TDDB [63]

Figure 3.21: Examples of aged equivalent transistor models used in MaCRO [14], [63]

added delay due to aging per gate is calculated as well as changes in the slopes of the signals. Equation 3.21 shows how the additional delay for a gate is derived. In this equation Δd denotes the delay, ΔI_{on} is the current through a transistor when it is active. GATE is the set of all transistors in a gate.

$$\Delta d = \sum_{m \in \text{GATE}} \left(\frac{\partial d}{\partial V_{th,m}} \cdot \Delta V_{th,m} + \frac{\partial d}{\partial I_{on,m}} \cdot \Delta I_{on,m} \right) \tag{3.21}$$

To enable fast simulation of aging effects on larger circuits, it is necessary to consider only a limited amount of data paths in the aging analysis. The presented approach in AgeGate is to make a timing graph of the network and remove all the nodes and links that do not contribute to the critical path or can become a critical path after aging. In three steps the graph is reduced and thus allows for faster simulation. The detailed reduction is not relevant for this thesis, but could prove useful when this research is extended to a DRAM netlist that includes digital circuitry. The aging analysis will result in an expected lifetime under given operating conditions. Also the data path on which this occurs will be known.

Response Surface Model Based Simulation Technique

The Response Surface Model Based Simulation Technique (RSM) [16] is able to predict reliability on a transistor level for both aging effects as well as process variability. Determining circuit behavior can be denoted mathematically in Equation 3.22 as:

$$\mathbf{P}^i = \vartheta(\mathbf{f}, t_i) \tag{3.22}$$

In this Equation $\mathbf{f} \in \mathscr{F}$ is the vector containing n parameters which have a process dependent spread, e.g. width and length of transistors. The function ϑ maps the function space of the circuit, $\mathscr{F}$, to the performance space, $\mathscr{P}$ at time t_i. The performance space contains vectors, $\mathbf{P}$, with m performance metrics for every timestep t_i. Performance can here for example be defined as the gain of the circuit, the consumed power or any other metric describing circuit functionality. The function ϑ needs to be found in order to enable fast reliability predictions for a circuit which is influenced by process variability. To evaluate ϑ, one could use a Monte-Carlo approach. This would require an enormous amount of simulations as for every sample a complete aging analysis needs to be performed. It becomes beneficial to select only those parameters in $\mathscr{F}$ which have a large impact on the outcome of ϑ, a response surface model (RSM). Screening design is used to find the parameters in this reduced set resulting in a mapping function $\widehat{\vartheta}$. The error made by $\widehat{\vartheta}$ is evaluated. If this error is small enough, the circuit reliability can be described by the derived function. If this is not the case, a second phase is needed, called regression design. In regression design more parameters can be taken into account and a more complex approach in mapping them is used to derive a suitable RSM.

The reliability for the complete circuit can be derived with a Monte-Carlo analysis. With the reduced set of influential parameters and modeling the circuit as an analytical expression this process is sped up significantly. As the influential parameters are known from the RSM, the designer is enabled to quickly identify weak spots in the design by doing a sensitivity analysis on the RSM.

The evaluation of aging on the circuit performance is done by deriving the workload on the transistors and substituting the transistor models with aged equivalents. Figure 3.22 shows the aged equivalent circuit used in the RSM approach. The circuit presented in the figure contains three failure mechanisms, BTI, HCI and TDDB. Both BTI and HCI are modeled in the voltage source in series with the gate of the transistor and two current sources connecting drain and source. Partial TDDB (increased leakage current before breakdown) is modeled using two resistors connecting the model gate to the source and the drain.

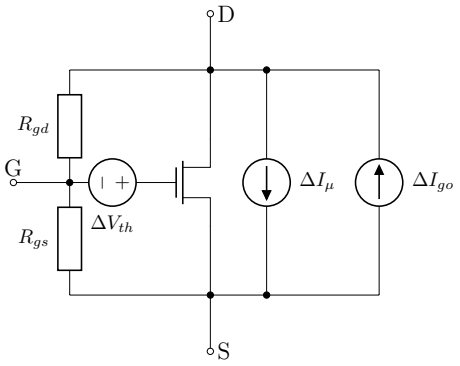


Figure 3.22: Aged equivalent transistor model used in the RSM [16]

Failure Rate Based SPICE Prediction Method

FaRBS [17] evaluates the lifetime of complete systems with simulations on single transistor level. It does this by simulating the full circuit and extracting the waveforms at certain points in time. If enough time samples are taken, the workload of the system can be taken into account accurately. Every sample j, the stress time, t_j, gate voltage, V_{gj}, drain voltage, V_{dj} and temperature, T_j, are sampled. Then the total stress, T_N, for a transistor for a single failure mechanism is calculated using for example Equation 3.23 to calculate stress due to TDDB. In this equation, A_N and A_T denote the area of the oxide in the device under stress and of a reference transistor, E_a is the activation energy, β and γ are shape and voltage parameters that are derived from accelerated testing respectively.

$$T_N = \sum_{j=1}^{J} t_j \left(\frac{A_N}{A_T}\right)^{\beta} \exp\left(\gamma\left(V_{gj} - V_{th}\right)\right) \exp\left(\frac{E_a}{k}\left(\frac{1}{T_T} - \frac{1}{T_j}\right)\right) \tag{3.23}$$

Once all the stresses for all the transistors are found, the reliability for the complete circuit can be found with a simple product of all reliabilities as shown in Equation 3.24.

$$R_S = \prod R_{TDDB} R_{EM} R_{HCI} R_{BTI} \tag{3.24}$$

It is important to mention that FaRBS assumes complete independence of failures where the first failure leads to a complete system failure and also assumes a constant failure rate. With the constant failure rate assumption the lifetime or MTTF for the system can be derived from the outcome of Equation 3.24 as was described in Section 3.1.1.

System Reliability Analyzer

SyRA [18] predicts the lifetime of circuits on single transistor and gate level. Simulations on gate level allow only digital circuits to be analyzed while the single transistor approach is more suited to analog circuits. SyRA shifts the threshold voltage for NBTI aging effects and changes drain current for HCI wear-out. To simulate aging in digital circuits, first the work profile is extracted by simulating the digital behavior of the circuit. With this work profile, the stress on each transistor in a gate can be found. Then using long-term formulae the change in threshold voltage or drain current is calculated. This outcome is converted to an additional gate delay as described in Equation 3.25 for the case of NBTI. In this equation $\Delta t_{d,\,Vdd}$ is the change in delay due to a shift in supply voltage, ΔV_{dd}, which can be found in the gate library. The factor η is the weighted contribution of a transistor in a gate. For example in a NOR-gate the PMOS transistor connected to the output can only degrade if the other one is activated.

$$\Delta t_{d,\,Vth} = -\frac{V_{dd}}{V_{th}} \cdot \left(\frac{\Delta t_{d,\,Vdd}}{\Delta V_{dd}} \right) \cdot \frac{\eta \cdot \Delta V_{th1} + \Delta V_{th2}}{2} \tag{3.25}$$

In the case of analog circuit examination, the transistors are replaced by aged equivalent models in the netlist. This is done by first simulating an unaged circuit and from this circuit the operating conditions are derived. The equivalent model for a PMOS device is given in Figure 3.23 for aging effects caused by NBTI. A voltage source in series with the gate simulates the degraded threshold voltage. The voltage source is driven by a Verilog-A script which calculates the voltage shift based on the long-term models. Because bias point in analog circuits may shift due to aging, it becomes necessary to use an iterative way to determine the effects of circuit aging. SyRA uses a constant timestep to allow bias point shifts to be predicted.

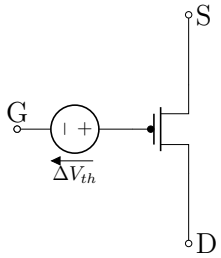


Figure 3.23: NBTI aged equivalent PMOS model used in SyRA [18]

In the case of digital circuit simulation the change in path delay of the 20 % longest paths is considered as the reliability metric for the whole circuit. If one of the path delays is longer than allowed, the circuit is considered failed. SyRA will find the moment this happens and thus predicts a lifetime. In the case of analog simulation the same procedure holds but now correct analog circuit behavior is the metric considered.

Ideal Reliabiltiy Prediction Method

Table 3.6 states that the ideal DRAM reliability simulation framework should be able to do. below the properties of such an ideal reliability prediction method are explained.

Inclusion of a netlist in the reliability investigation is important as this allows to see the degradation per transistor. With this high level of detail it becomes possible to pinpoint directly what causes most circuit degradation. Once the detailed analysis is done, the reliability can be evaluated at a higher level for these components, saving analysis time. Besides a netlist, also the operating conditions should be included in order to give a fair estimation of the wear-out the circuit will observe.

The faults that are analyzed need to be dependent on each other to represent the physical wear-out as realistic as possible. For example, both BTI and HCI can generate traps in the oxide of the transistor. These traps are indistinguishable once present and thus have the same impact on the transistor performance. Also the amount of traps that can be present in an oxide is limited and thus blindly adding all resulting wear-out phenomena could result in a wear-out estimation that is too high.

Finally, the method should allow to use variable metrics to determine the circuit reliability. For example, when analyzing the memory cell, one might want to include cell voltage as a metric to set the reliability while in the sense amplifier analysis the sensing delay might be the metric to be considered. The method then should be able to predict the reliability as well as the lifetime of the circuit.

3.4.4. Wear-out Models

This section describes the choices and implementation of the aging and wear-out models used to determine the aging of a DRAM in this research. Three wear-out phenomena and transistor mismatch are investigated. Transistor Mismatch are discussed in Section 3.4.4, Hot Carrier Injection in Section 3.4.4, Bias Temperature Instability in Section 3.4.4 and Radiation Induced Effects in Section 3.4.4.

In the previous section, an ideal reliability simulator configuration for DRAM was presented. It stated that faults should be dependent from each other in order to give a fair physical representation. In literature no consistent models that describe these mutual influences on the transistor for all mentioned wear-out phenomena are presented. This lead to the choice to include all aging mechanisms individually in this research instead of in a combined model.

Table 3.6: Detailed overview of aging and reliability prediction methods

Name	Netlist	Level	Failure mechanism	Workload aware	Fault dependencies	Prediction Outcome	Failure Metric
BERT [12]	Yes	Transistor	HCI, EM, TDDB	Nodal waveforms; No temperature for HCI	No, only individual failures	Reliability; Lifetime	TDDB: breakdown probability for points in time single devices; EM: time-to-failure for all wires; HCI: degradation of circuit properties
RAMP v1 [5]	No	System	EM, TDDB	Switching activity partially taken into account. Temperature included	Weighted sum; failures independent	Reliability	
RAMP v2 [13]	No	System	EM, TDDB, NBTI	Switching activity partially taken into account. Temperature included	Weighted sum; failures independent	Reliability	MTTF for components in system and complete system
MaCRO [14]	Yes	Transistor	HCI, TDDB, NBTI	Lifetime estimation: average voltages and temperature included; Functionality: waveforms and aged models	Weighted sum; failures independent	Reliability	Degradation of circuit properties resulting in loss of circuit functionality
AgeGate [15]	Yes	Gate	HCI, NBTI	Duty cycle and switching activity; Temperature included	Weighted sum; failures independent	Lifetime	Critical path delay
RSM [16]	Yes	Transistor	HCI, TDDB, BTI	Waveforms and Temperature included	Faults combined in one transistor model; failures independent	Reliability	Circuit parameter like gain, output current, etc.
FaRBS [17]	Yes	Transistor	HCI, TDDB, NBTI, EM	Samples over waveforms. Temperature included.	Weighted sum; failures independent	Lifetime	MTTF for the complete system
SyRA [18]	Yes	Transistor, Gate	HCI, NBTI	Digital: duty cycle; AMS: only generalized AC/DC waveforms; No temperature	Combination unclear; Failures independent	Lifetime	Digital: Path delay; AMS: degradation of circuit properties
Ideal	Yes	Dynamic	All	Actual voltages and currents; temperature included	Yes, dependent failures	Reliability & Lifetime	Different metrics per subcomponent, e.g. Cell: retention time, capacitor voltage; Sense amplifier: detection limit, sensing delay;

Inclusion of TDDB to this research was not possible because of its physical nature. Simulating the wear-out effects of TDDB would require to include the increase in gate leakage current caused by the conducting trap path in the oxide as well as complete breakdown of the transistor oxide. The latter would be more suited for simulation if a complete DRAM, containing all cells, were to be evaluated. TDDB in the cell capacitors was not included due to a lack of data on this wear-out mechanism. Application of transistor TDDB models to these capacitors was not considered suitable because of the geometric differences next to the behavior of a total breakdown. EM is left out of this research because it applies to the physical lay-out of the wires in the chip. As no lay-out of the circuit was available, this wear-out mechanisms was not implemented.

The NMOS transistor model that includes all wear-out and aging mechanisms used in this thesis is presented in Figure 3.24. It shows that the effects of transistor variations, HCI and BTI are included as threshold voltage shifts by putting voltage sources ($\Delta V_{th,\,mis}$, $\Delta V_{th,\,HCI}$, $\Delta V_{th,\,BTI}$, respectively) for these mechanisms in series with the gate. The radiation effects are modeled as a current source (I_{rad}) between the bulk and the drain of the transistor. The PMOS model is equal to the NMOS model except that a PMOS transistor is used and the current flows from the bulk to the source.

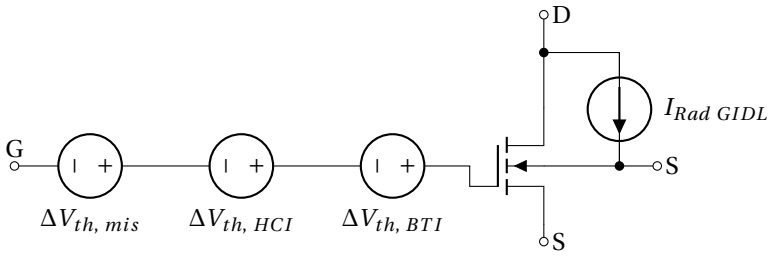


Figure 3.24: Transistor model used in this thesis. $\Delta V_{th,\,BTI}$ indicates the threshold voltage shift caused by BTI, $\Delta V_{th,\,HCI}$ for HCI and $\Delta V_{th,\,mis}$ for threshold voltage shifts caused by transistor variations. $I_{Rad\ GIDL}$ represents the radiation induced leakage current.

Transistor Variations

Transistors on a wafer or even on the same die do not have the same properties [64], [65]. This is caused by random variations in the manufacturing process and defects and impurities in the silicon wafer, called mismatch or process variations. The process variations might impact all transistors on a die of even a complete wafer. For example, all threshold voltages might shift in one direction. But even on one die, transistors are not equal due to the variations in doping and impurities [64], [65]. The mismatch variations are modelled using the commonly used Pelgrom model [64]. Pelgrom's model describes the variance in threshold voltage among transistors on a chip. The equation describing the variance of the normally distributed threshold voltage is presented in Equation 3.26.

$$\sigma^2\left(\Delta V_{th,\,mis}\right) = \frac{A_{VT}^2}{W \cdot L} \tag{3.26}$$

In this Equation A_{VT} is a technology node dependent parameter that decreases with decreasing feature size, also named Pelgrom's constant [65]. Table 3.7 summarizes the values for A_{VT} for different technology nodes.

Table 3.7: Pelgrom's constant A_{VT} for different technology nodes [65]

Technology Node	NMOS A_{VT} [mV·µm]	PMOS A_{VT} [mV·µm]
2.5 µm	30	35
1.2 µm	21	25
1.0 µm	13	16.5
0.7 µm	13	22
0.5 µm	11 ~ 12	13
0.35 µm	9	9
0.25 µm	6 ~ 7.5	6
0.18 µm	3.3 ~ 5	5.49

Process variations have an effect on the whole chip and result in a shift of all parameters in a certain direction, e.g. all NMOS threshold voltages increase. Investigation of these effects is out of the scope of this work.

Bias Temperature Instability
In this research two models were investigated to simulate the impact of BTI on the circuit. The first model, the atomistic model, is based on the work in [66], [67], the second is based on the commonly used Reaction-Diffusion model (RD-model) [68], [69].

Atomistic Model First the atomistic model was used. The model was in-house available and only needed an altering of parameters to fit the circuit used in this thesis. It turned out that the model is too modern and too computationally intensive to deal with the old DRAM design that is used in this work. Below the features of this model and the difficulties to use it with older technology are presented.

Kaczer, Grasser, Roussel, *et al.* showed in [66] that the nature of BTI resembles that of Random Telegraph Noise (RTN) for small devices. The shifts generated in threshold voltage for both effects are found to be due to the trapping and detrapping of single charges. The atomistic model uses defect densities to come up with a number of defects per device on a point in time and thus the related threshold voltage shift. Weckx, Kaczer, Toledano-Luque, *et al.* in [67] extend this model with densities for slower and faster traps, allowing for accurate short and long time predictions on the threshold voltage degradation. The model is workload aware in the sense that is takes duty factor, operating voltage and frequency, aging time and temperature in account when evaluating the generation of traps in a transistor. It was found that setting the parameters of the model to those of the transistors in the circuit (length, width, threshold voltage) does not result in correct estimations of the threshold voltage shift as described in [35], [70], [71] for this older technology. This is due to the fact that the gate oxide considered in [66], [67] is HfO_2 and not SiO_2 as used in the DRAM circuit. HfO_2 contains more defects per unit volume than SiO_2 does [28]. This difference makes the provided parameters invalid. Secondly, the larger oxide volumes in older transistors decrease the impact of single traps and carriers on the threshold voltage, although the total amount of traps increases. Simulating all the traps individually, as the atomistic model does, therefore consumes a lot of computation time without resulting in better results than the RD-model.

The model parameters were altered empirically to match with the data for older technologies, [35], [70], [71]. Figure 3.25 shows the shift in threshold voltage after several aging times. The results are matching the data in [35], [70], [71] reasonably well.

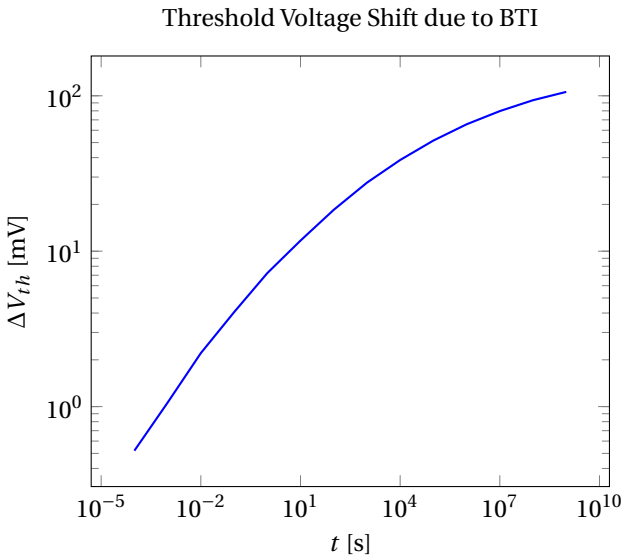


Figure 3.25: Threshold voltage shift due to BTI using the Atomistic Model. W=600 nm, L=450 nm, T=125 °C

Reaction-Diffusion Model After it became clear that the atomistic model was not suitable enough for this research, the RD-model was implemented [68], [69]. The RD-model was the first model to capture the thresh-

old voltage degradation caused by BTI. Alam and Mahapatra [69] extended the model to include the effects of temperature and voltage. Solving the differential equations in the RD-model gives a good prediction of the circuit degradation at every point in time. This is possible because the exact history of the circuit, being the voltages that were present and the previously calculated degradation, is known. Of course, keeping track of this greatly increases the simulation time. Since the waveforms in this DRAM circuit are repetitive and this research focuses on long-term degradation, a long-term approximation of the RD-model was used including the duty factor of the transistors. This approximation does not take into account historical alterations in the circuit operating conditions due to aging exactly but stills gives a fair approximation and thus speeds up simulation time signifiantly. Equation 3.27 shows the long-term approximation for a constant, DC, stress [72].

$$\Delta V_{th,\,BTI} = A \exp\left(-\frac{E_a}{kT}\right) \exp\left(\gamma V_{gs}\right) t^n \tag{3.27}$$

In this Equation t denotes the aging time, γ is a gate voltage dependent coefficient [72] and n is the BTI time exponent, which needs to be set to approximately $n \approx 0.25$ [68], [69], A is a scaling factor, E_a the activation energy, k is the Boltzmann constant and T is the temperature.

Stochastic variation in the threshold voltage for BTI was implemented as well. Equation 3.28 shows the standard deviation for the mean threshold voltage shift ($\mu\left(\Delta V_{th,\,BTI}\right)$) given by Equation 3.27 [73]. This deviation was used in to add a Gaussian distributed variation on the threshold voltage.

$$\sigma\left(\Delta V_{th,\,BTI}\right) = \sqrt{\frac{q \cdot t_{ox} \cdot \mu\left(\Delta V_{th,\,BTI}\right)}{\epsilon_{ox} \cdot W \cdot L}} \tag{3.28}$$

In this equation q is the elementary charge, t_{ox} is the oxide thickness, ϵ_{ox} is the oxide permeability, W is the transistor width and L is the transistor length.

To include the effects of different duty factors of transistors, the formula in Equation 3.27 was altered. As was presented in Section 3.2.2, the degradation of a transistor due to BTI is affected non-linearly by the duty factor at which it operates. As no specific data on BTI degradation of the transistors as used in the DRAM model are available, it was chosen to make a fitting curve based on data from other transistors available in literature. Figure 3.26 shows the relative degradation of different transistors for several workloads as was presented already in Figure 3.7. The measurements are all relative to constant DC stress. This figure also contains a fitted curve that was used in the BTI model in this research to include the workload of a transistor. Equation 3.29 shows the formula describing this fitting curve. In this equation DF denotes the duty factor of the transistor. As can be seen from the curve, influence of the duty factor varies a lot with transistor technology. The duty factor function was simply multiplied with the long-term degradation formula as showed in Equation 3.27 to represent the long-term BTI degradation with inclusion of duty factor. The complete description of the BTI threshold voltage shift is given in Equation 3.30.

$$f(DF) = \frac{(5/4) \cdot DF^{1/3} + (7/25) \cdot DF + DF^{30}}{253/100} \tag{3.29}$$

$$\Delta V_{th,\,BTI} = f(DF) \cdot A \exp\left(-\frac{E_a}{kT}\right) \exp\left(\gamma V_{gs}\right) t^n \tag{3.30}$$

The fact that n-type devices are less prone to BTI degradation than p-type devices is taken into account as well. A NMOS transistor suffering from PBTI has a threshold voltage shift approximately ten times less than a PMOS transistor [41], [42]. This effect was included in the model by dividing the scaling factor, A, by 10 for the NMOS model in comparison with the PMOS model. The influence of duty factor on BTI degradation in NMOS transistors is set equal to that of PMOS transistors [74], [75].

Hot Carrier Injection

Hot carrier injection is modeled with the lucky electron model that describes the threshold voltage shift [72], [76]. The change in threshold voltage, ΔV_{th}, is given by Equation 3.31 [72].

$$\Delta V_{th} = L_{eff}^{\alpha} \left[t \cdot \frac{I_d}{W} \cdot \left(\frac{I_{sub}}{I_d} \right)^m \right]^n \exp\left(\frac{-E_a}{kT}\right) \tag{3.31}$$

In this equation L_{eff} is the effective transistor length, which may become shorter than the real fabricated length, α is a technology related constant, the constant n is used to include the power law time exponent

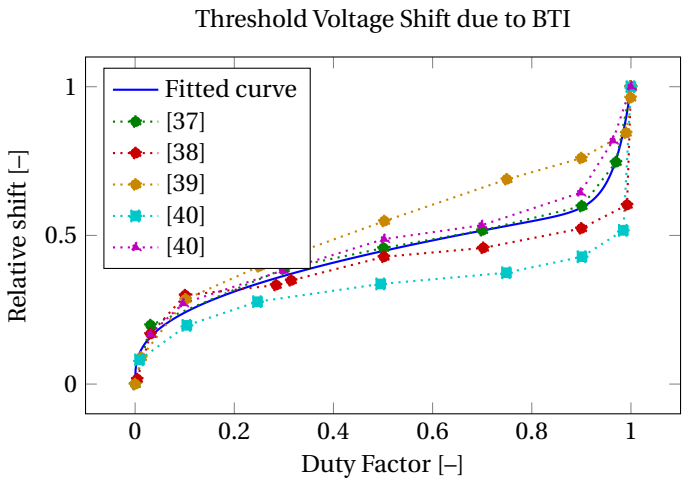


Figure 3.26: Impact of duty factor on long-term threshold voltage shift caused by BTI. All shifts are plotted relative to DC long-term stress

of the HCI degradation which can vary between 0.5 and 1.0 [77]. E_a is the HCI activation energy which is typically 0.05 eV [78]. The constant m is given by the ratio $m = \phi_{it}/\phi_{ii}$. ϕ_{it} is the critical energy to create an interface trap, which is 3.7 eV for travelling electrons (n-type devices) and is 4.6 eV for travelling holes (p-type devices) [46], [72], [79]. ϕ_{ii} is the impact ionization threshold energy, which is approximately 1.3 eV [72]. The values for the currents were derived from a clean circuit simulation and are an average of the peak currents observed for the cell transistors. This assumption may overestimate the HCI contribution to the transistors. Figure 3.27 shows the simulated threshold voltage shift of a NMOS transistor with a duty factor equal to 1.0.

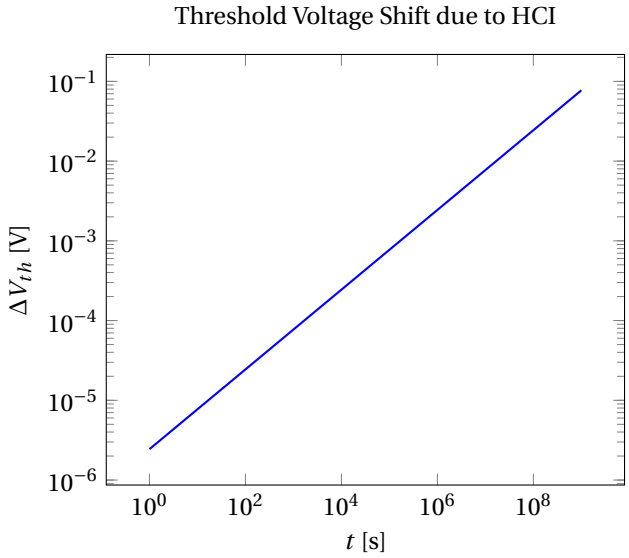


Figure 3.27: Threshold voltage shift due to HCI. Duty factor was set to 1.0

The model presented in Equation 3.31 was calibrated to fit data from transistor measurements [77], [80] and comply with theoretical model parameters [46], [72], [77], [79]. Duty factor was included in the model as well via a multiplier on the calculated threshold change. As described in Section 3.2.3 this is a valid approach, as there is practically no recovery from HCI effects observed [44]. Besides the duty factor also some statistical spread was added to the currents in the model, including the random fluctuations observed with HCI as well. The standard deviation of the gaussian spread on the shift in threshold voltage is given in Equation 3.32 [81].

$$\sigma\left(\Delta V_{th,\,HCI}\right) = \sqrt{\frac{K \cdot 2q \cdot \mu\left(\Delta V_{th,\,HCI}\right)}{C_{ox} \cdot W \cdot L}} \tag{3.32}$$

The standard deviation thus depends on geometries of the transistor (W, L, C_{ox}), the mean threshold voltage degradation ($\mu(\Delta V_{th,\,HCI})$) and a unknown factor K. In [81], experimental proof suggests a value of $K = 5.6$ in 45 nm-technology and $K = 5.9$ in 65 nm-technology. No further data were available so the assumption was made that K scales linearly with technology node. This sets $K \approx 7.5$ for the DRAM design used in this research. Figure 3.28 shows the impact of varying K on the retention time of the circuit for three different operating temperatures. As was expected from Equation 3.32 the spread in the retention time increases with increasing K. As the increase in spread is limited by the square root function, an error on the estimation that $K \approx 7.5$, is acceptable.

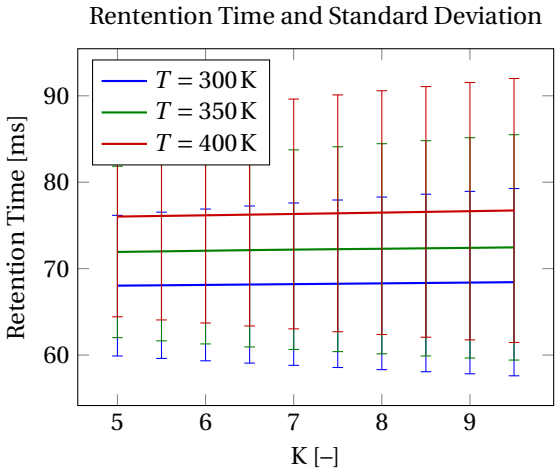


Figure 3.28: Mean retention time values and its standard deviation for varying K for three temperatures

Radiation Induced Effects

This section describes the implementation of two radiation wear-out models first and finally data on the space radiation environment are presented. The first model investigated was developed for this research specific and is based on fitting real-world measurements of a DRAM under radiation. The second model is based on the physics behind radiation trap forming. After these models are explained the space radiation environment is presented.

Implementing Radiation Trapping Effects as Gate Induced Drain Leakage An adapted version of the text in this section was submitted to the 2018 IEEE Latin-American Test Symposium (LATS2018).

In [50], the authors related the measurements of retention time in a DRAM under irradiation to GIDL leakage in transistors. The authors link the degradation of retention time to the generation of interface traps in the oxide of the transistor, which increases the gate induced drain leakage current. This model fits the degraded parameters to physical ones that can be used on transistor level SPICE simulations. It is important to mention that the data presented in [50] are all measurements from a single commercial DRAM chip. Therefore, the accuracy of the model could be improved if more chips are taken into account.

In order to model the effect of the increased leakage current on the retention time, first a general GIDL model was implemented based on the one used in the BSIM4 physics based transistor model [82]. Equation 3.33 shows this GIDL current. The parameters of this equation are described in Table 3.8.

$$I_{GIDL,\,normal} = \mathrm{AGIDL} \cdot W_{eff} \cdot \frac{V_{ds} - V_{gs} - \mathrm{EGIDL}}{3 \cdot t_{ox}} \cdot \exp\left(-\frac{3 \cdot t_{ox} \cdot \mathrm{BGIDL}}{V_{ds} - V_{gs} - \mathrm{EGIDL}}\right) \cdot \frac{V_{db}^3}{\mathrm{CGIDL} + V_{db}^3} \tag{3.33}$$

To allow modeling of trapped charge due to radiation, we add an extra multiplier to the GIDL current equation. Its value depends solely on the total amount of radiation trapped in the oxide (T_D) measured in krad, and the duty factor (DF) or activity factor of the transistor. The resulting GIDL current, $I_{Rad,\,GIDL}$, is presented in Equation 3.34.

$$I_{Rad\ GIDL} = I_{GIDL,\,normal} \cdot f(T_D, DF) \tag{3.34}$$

Table 3.8: GIDL Current Parameters

Parameter	Description
$I_{GIDL,\,normal}$	Gate Induced Drain Leakage Current
W_{effCJ}	Effective Source/Drain Diffusion Width
t_{ox}	Transistor Oxide Thickness
V_{db}	Drain-Bulk Voltage
V_{ds}	Drain-Source Voltage
V_{gs}	Gate-Source Voltage
AGIDL	Pre-exponential Coefficient
BGIDL	Exponential Coefficient
CGIDL	Parameter for Body-Bias Effect
EGIDL	Fitting Parameter for Band Bending

Figure 3.29 shows the amount of cells with a retention time less than 300 ms based on the data from [50], as well as a fitting curve through these points. The data suggests that the degradation of the circuit is exponentially dependent on TID.

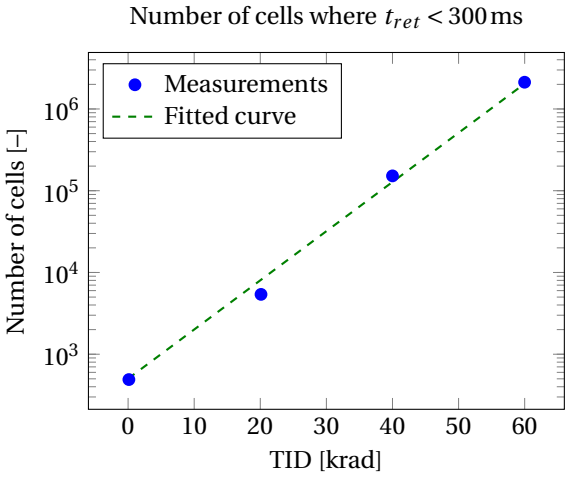


Figure 3.29: Cells with $t_{ret} < 300$ ms and fitted curve versus TID [50]

The multiplier function $f(T_D, DF)$ is further described in Equation 3.35. In this function, $f(DF)$ is a function that describes the impact of the duty factor on total radiation and X_1 is a dimensionless fitting parameter. E_a presents the activation energy in eV, k is the Boltzmann constant and T the temperature in K. The exponential term links the TID to the shift in retention time under radiation. The term -1 ensures that $f(T_D, DF) = 0$ when $T_D = 0$ krad. Below these terms will be explained in more detail.

$$f(T_D, DF) = f(DF) \cdot X_1 \left(\exp\left(\frac{E_a \cdot T_D}{kT} \right) - 1 \right) \tag{3.35}$$

The duty factor, $f(DF)$, of the transistor has an influence on the total ionizing dose that accumulates in the transistor. If the device is biased during irradiation more charges will get trapped than when the device is unbiased [33], [83]. It is assumed in the model that the amount of trapped charges increases linearly with the duty factor from the unbiased amount to the completely biased amount. From [33] it also follows that PMOS devices are more prone to radiation trapping effects than NMOS devices but the amount of trapped charge depends less on the duty factor. These differences are included in the model as well.

X_1 and E_a can be found by analyzing the discharge current graphs of a memory cell for different TID. In this work, we calcualte the discharge current from the retention time measurements in [50]. Figure 3.30 shows the change in the retention time density after a total dose of 60 krad is accumulated. The mean retention time shifts down from $\mu_{0\,\text{krad}} \approx 2800$ ms to $\mu_{60\,\text{krad}} \approx 1300$ ms, which is a factor $\frac{\mu_{0\,\text{krad}}}{\mu_{60\,\text{krad}}} \approx 2.15$.

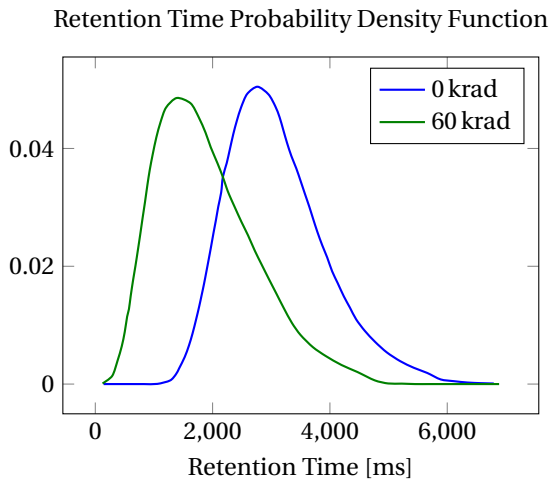


Figure 3.30: Change in retention time density function due to radiation trapping. Remodeled after [50]

The retention time of a cell is described by Equation 3.36 [24].

$$t_{ret} = \frac{(V_1 - V_2) \cdot C_{cell}}{I_{leak,\ average}} \tag{3.36}$$

In this Equation, V_1 denotes the high voltage (i.e., the voltage on the capacitor right after writing to it), V_2 the voltage where the logic cell value flips, C_{cell} is the cell capacitance and t_{ret} is the retention time of the cell. The values of V_1, V_2 and C_{cell} are design dependent. Finally, $I_{leak,\ average}$ is the average current that flows when the capacitor is discharging and consists of two components, $I_{leak,\ average} = I_{Rad\ GIDL,\ average} + I_{Other,\ average}$. $I_{Other,\ average}$ is the leakage current in the cell caused by sources different from radiation trapping effects, and $I_{Rad\ GIDL,\ average}$ is the current that is caused by radiation trapping effects. When no charges are trapped (TID = 0 krad), there is no additional leakage current and hence $I_{Rad\ GIDL,\ average} = 0$ A. The retention time of the cell is affected only by $I_{Other,\ average}$. As was seen in the previous paragraph, the mean retention time decreases 2.15 times once 60 krad of charge is accumulated in the transistor. This means that the average leakage current ($I_{leak,\ average}$) must have increased 2.15 times. Since $I_{Other,\ average}$ does not change (as we assume that this is unaffected by radiation), $I_{Rad\ GIDL,\ average}$ must have increased. Setting $t_{ret,1} = 2.15 \cdot t_{ret,2}$ and solving for $I_{Rad\ GIDL,\ average}$ yields $I_{Rad\ GIDL,\ average} = 1.15 \cdot I_{Other,\ average}$. Therefore, the exponential term in Equation 3.35 should have a slope of $\frac{E_a}{kT} = 0.0128$ in order to satisfy this relation. Calculating the average current is nearly impossible in an actual circuit because the current strongly depends on the bias voltages of the pass transistor. These voltages change over time as the charge leaks away from the capacitor and the voltage on the bit line may change depending on the performed operations. Therefore, X_1 is fitted based on retention time simulations.

Figure 3.30 clearly shows not only a degradation of the mean retention time but also an increase in standard deviation. These statistical properties are not taken into account in this model due to a lack of detailed measurement data. It is for example impossible to know how a single cell degrades as only complete chip distributions are presented. This makes it harder to derive how radiation will affect the tail distribution, i.e., the cells with the lowest retention times, as they have the most impact on the complete DRAM reliability [84].

The extra GIDL current generated due to trapped charges was implemented as a Verilog-A controlled current source to allow easy simulation of the radiated circuit. The current source discharges either from drain to bulk for NMOS transistors or from bulk to source for PMOS transistors. Figure 3.31 shows the NMOS radiation circuit. The code for this model can be found in Appendix A.

Correct functioning of the model was verified by simulation of a single transistor suffering from radiation damage and as an memory cell transistor in the DRAM circuit. In Figure 3.32 the resulting additional leakage current due to radiation for a single NMOS and PMOS transistor is shown. From the figure it becomes clear that the total current depends strongly on the total amount of trapped charges, as was expected from Equation 3.35 and the results presented in [83]. Also the current increases fast with increasing drain voltage, which is of course due to the increasing voltage difference between the gate and the drain. The minimal current is set to $2 \cdot 10^{-20}$ A for ease of comparison when plotted together with other larger currents and mostly to prevent potential simulation problems caused by the limited precision of floating point numbers.

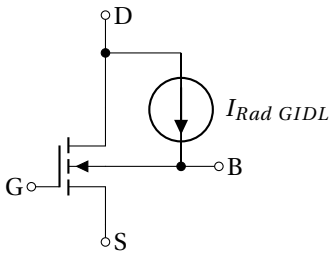
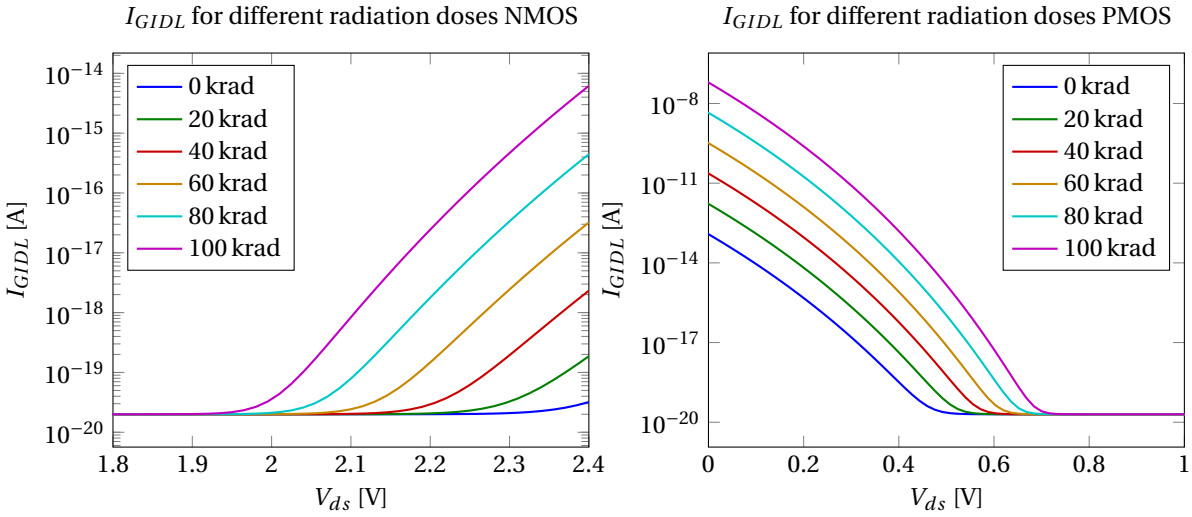


Figure 3.31: NMOS transistor trapped radiation GIDL



(a) Additional GIDL current due to charge trapping in a single NMOS transistor. Source, Gate and Bulk connection all biased at 0 V,

(b) Additional GIDL current due to charge trapping in a single PMOS transistor. Source, Gate and Bulk connection all biased at 2.4 V,

Figure 3.32: I_{GIDL} for NMOS and PMOS transistors for different TID

After the model was applied to a single transistor, it was applied to the access transistor of a DRAM memory cell. The retention time of the cell was measured for different doses of trapped charges. Figure 3.33 shows the results and a fitted curve through the measurement points. The retention time measurement method will be described in Section 4.2.1.

The retention time starts to drop almost linearly after a total dose of 40 krad is accumulated. The small impact of small doses on the retention time can be explained by the DRAM design used. As was stated in Section 3.4.1 the access device is a PMOS transistor which cannot pass a lower voltage than its own threshold voltage, given that the word line is set to 0 V. The threshold voltage in the technology used is approximately 700 mV. The headroom, the voltage difference between the threshold voltage and the voltage at which the additional GIDL current turns off, is less when a smaller dose is trapped in comparison with a larger trapped dose as can be seen in Figure 3.32b for PMOS devices. Similarly in NMOS devices, the voltage at which the GIDL current becomes important drops with increasing accumulated dose. This in combination with the smaller GIDL currents limits the influence of small doses of radiation on the retention time. The increase of the leakage current results in faster discharging of the capacitor and thus faster reaching the point where the influence of GIDL becomes negligible. The capacitor discharges then via other leakage mechanisms. Figure 3.34 shows the increase in cell voltage (leakage is towards V_{dd} in a PMOS device) directly after writing a low voltage to the cell. When the GIDL discharge current becomes negligible in comparison with other leakage mechanisms one sees that the voltage plots become parallel to each other, independent of dose.

Trap Based Radiation Model A defect based model able to predict degradation due to interface and oxide traps is presented by Esqueda, Barnaby, and King in [85]. As was presented in Section 3.2, BTI, HCI and

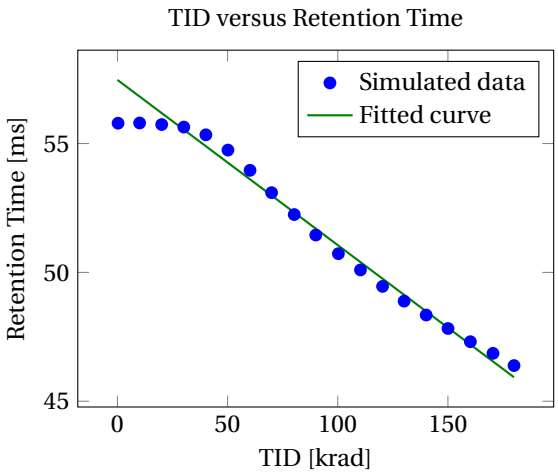


Figure 3.33: Retention time degradation in a DRAM due to radiation trapping with a fitting curve fitted in the region 50 krad ≤ TID ≤ 180 krad

radiation effects generate traps in a transistor that alter the threshold voltage. The presented model uses calculations of the surface potential to find the threshold voltage shift generated by these traps. The model was used in this research only to model the effects of radiation degradation. This was done to verify the correctness of the GIDL model and because BTI and HCI models were available before development of the radiation model, so there was no need for this combined model. In future work, combination of the three mentioned wear-out mechanisms via this defect based model could be a promising step.

The alteration in the threshold voltage is captured in the so called trap potential (ϕ_{nt},) which can be found via Equation 3.37 [85]–[87]. In this Equation ψ_s denotes the surface potential of the transistor, ϕ_b is the bulk potential. N_{ot} is the areal density of oxide trapped charges in cm^{-2} and D_{it} is the density of interface traps in $eV^{-1}cm^{-2}$ [85]–[87].

$$\phi_{nt} = \frac{q}{C_{ox}} \left[N_{ot} - D_{it} \left(\psi_s - \phi_b \right) \right] \tag{3.37}$$

To find the defect potential, the surface potential and the trap densities must be known. The surface potential can be calculated non-iteratively with the algorithm presented in [87].

Barnaby, McLain, Esqueda, *et al.* presented a model that links dose rates to trap densities in shallow trench isolation (STI) [86]. STIs isolate two different transistors from each other to reduce the effect of the parasistic MOSFET that is formed between them. These oxides are much thicker than gate oxides and thus the presented model cannot be applied directly to the gate oxides. Besides that, without a physical lay-out of the DRAM circuit, it is impossible to incorporate STI effects in this research. Barnaby, McLain, Esqueda, *et al.* describe a method that derives the change in oxide trap density (ΔN_{ot}) iteratively in [86].

The lack of radiation trapping data for the older technology used in this research, rendered the presented method useless. Next to that the equation depends on τ_{ot}, the annealing time for trapped holes. The value of this parameter strongly affects the total amount of generated interface traps and should therefore be set very accurately. Unfortunately, no suited value for the technology that was used in this work could be found in literature.

Space Radiation Environment The space radiation environment describes the type of radiation and the amounts of radiation a chip will face at certain points in space for certain moments in time. Figure 3.35 shows the space radiation environment around the earth [88]. The environment in space varies a lot with distance from earth, distance to the sun, the thickness and material of a shielding layer and even position above the earth's surface [89]. At lower altitudes (low earth orbit to medium earth orbit) protons are the main source of radiation damage to a chip, while at higher altitudes (medium earth orbit to geostationary orbit) electrons are more severe degrading factors. Besides the electrons and holes there is, for both aforementioned altitudes also bremsstrahlung present [88].

The aforementioned parameters make it difficult to find out what the total ionizing dose will be for a certain space mission. SPENVIS (Space Environment Information System) [90] is a tool developed by the Royal

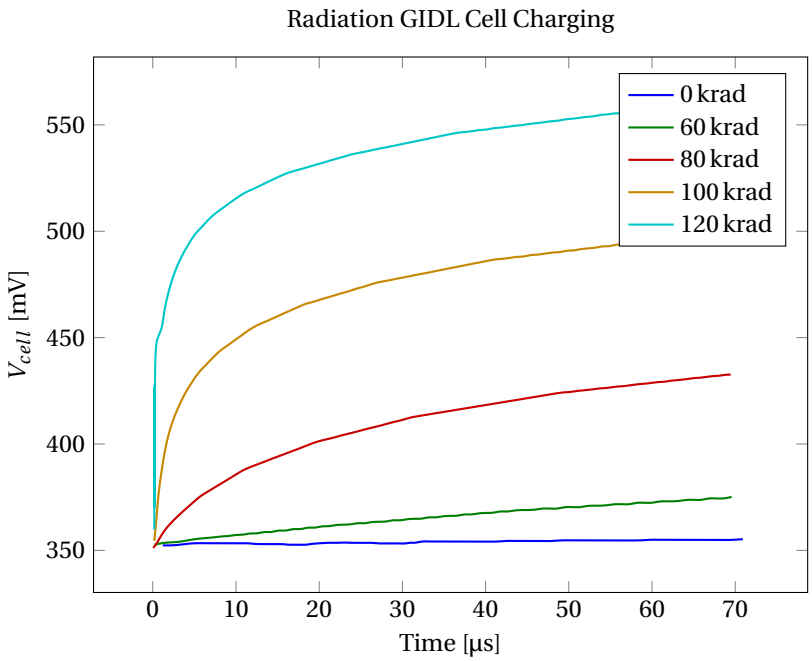


Figure 3.34: Increase of memory cell voltage directly after a write operation for varying TID

Belgian Institute for Space Aeronomy for the European Space Agency (ESA). The tool is able to calculate the TID for a user programmed mission and show how shielding and the planned traject influence the TID. For this research it was assumed that the chip would be used in a mission comparable to the ESA's Sentinel-1 mission [91]. For this mission a report about the space radiation environment, meteoroids and debris was made that contains Figure 3.36 [92]. The figure shows the TID in silicon for varying thickness of shielding for the complete mission. With the data in the graph combined with the mission duration the average accumulated TID per second can be calculated and is used in this research. In this work, two thicknesses of shielding aluminum layers are investigated. The thin shielding layer being 0.05 mm and the thick one 10 mm. The first shielding thickness would relate to the chip being placed closed to a satellite's outer surface, while the thicker shielding layer would place the chip in a well covered part of the satellite. The corresponding dose rates are then: $D = 29.59\,\mu\text{rad/s}$ for 10 mm thick aluminum shielding and $D = 1.26\,\text{mrad/s}$ for 0.5 mm thick aluminum shielding.

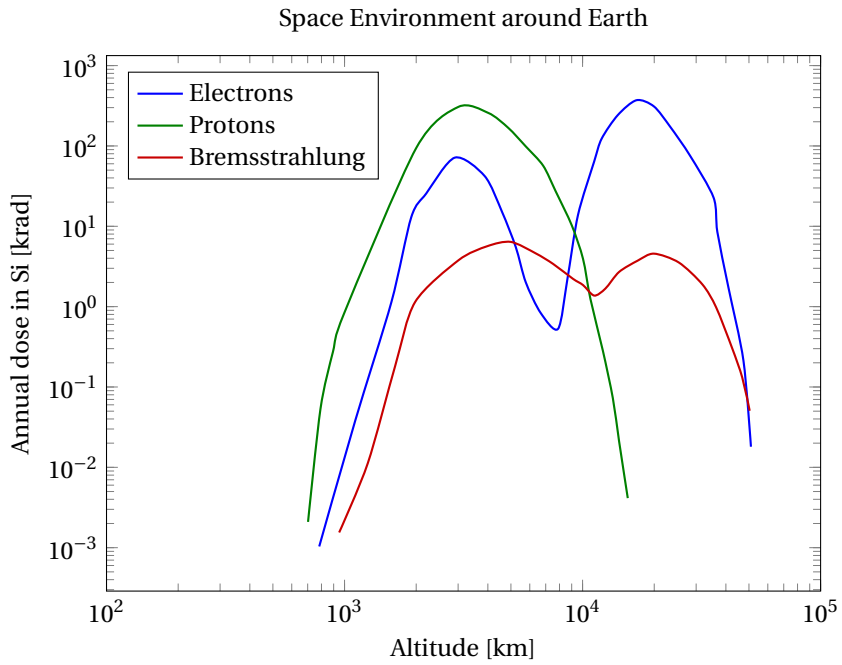


Figure 3.35: Space Radiation Environment around earth in a 4 mm diameter aluminium sphere [88]

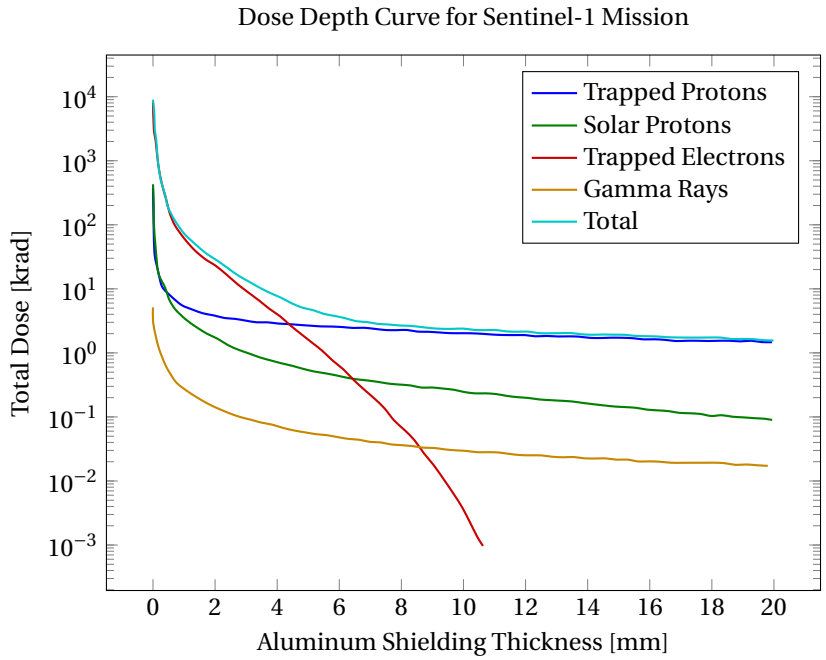


Figure 3.36: Total Ionizing Dose for varying shielding thickness for the Sentinel-1 Mission [92]

4

Framework

In this chapter, the framework of this research will be explained. The control and execution of the simulation, the simulation platform, will be described in Section 4.1. After this, the experimental metrics will be described in Section 4.2. In the following Section 4.3 the operating conditions of the circuit will be presented that are used in the experiments as described in Section 4.4. To finalize this chapter, the initial circuit performance which serves as a reference will be presented in Section 4.5.

4.1. Platform

Figure 4.1 presents the simulation platform schematically. Simulations were performed by the Virtuoso Spectre circuit simulator, version 14.1.0 64bit [93]. Spectre is able to simulate both circuits described in the SPICE language as well as in its native Spectre language. Next to circuit simulation abilities, Spectre also includes two features important for this research: a Measurement Description Language (MDL) [94] and the inclusion of Verilog-A modules [95]. MDL can be used to set up more extensive and automated measurements than would be possible with a SPICE *.measure* statement. Verilog-A modules allow the user to describe the analog behavior of a component without designing an actual circuit in the Verilog language. In order to be able to fully use the simulator's capabilities, the original SPICE circuit was rewritten to the Spectre language. This also allowed for easier automated model selection and alterations of the circuit by the controlling framework.

The simulations were all controlled by a Matlab script running on version R2015b [96], that altered circuit parameters and wear-out mechanisms for every simulation and combined and analyzed the results. To ease the process of result generation, control of the Monte-Carlo simulation was split between the Matlab controlling script and the MDL-scripts. Matlab seeds the MDL-script that runs one Monte-Carlo iteration. Matlab finally, combines all data points from all the iterations and presents the results.

4.2. Metrics

This section describes which metrics are used to evaluate the degradation of a DRAM circuit. The five metrics are: Retention time, Cell Voltage, Sensing Delay, Writing Delay and Energy Consumption. In the subsequent sections the metrics will be discussed in detail. The nominal values of these metrics, can be found in Section 4.5.

4.2.1. Retention Time

The definition of retention time of a DRAM was first presented in Section 2.1 and is repeated here for sake of clarity. The retention time of a DRAM cell is the time a cell is able to maintain its stored data correctly. This practically means that the retention time is the time it takes for the cell capacitor to discharge to a level where it becomes impossible to read out the previously written data. Because the performance of all hardware needed to read a cell might change due to wear-out and aging effects, the retention time is found by doing an actual read operation on the memory. The retention time is measured by doing a binary search in multiple simulation steps. Figure 4.2 shows a flowchart of the binary search algorithm used to find the retention time. An initial guess is made for the retention time of the circuit and then the circuit is simulated. If the guess was too high, meaning the data are lost, the next step sets the guess to half the maximum retention time minus

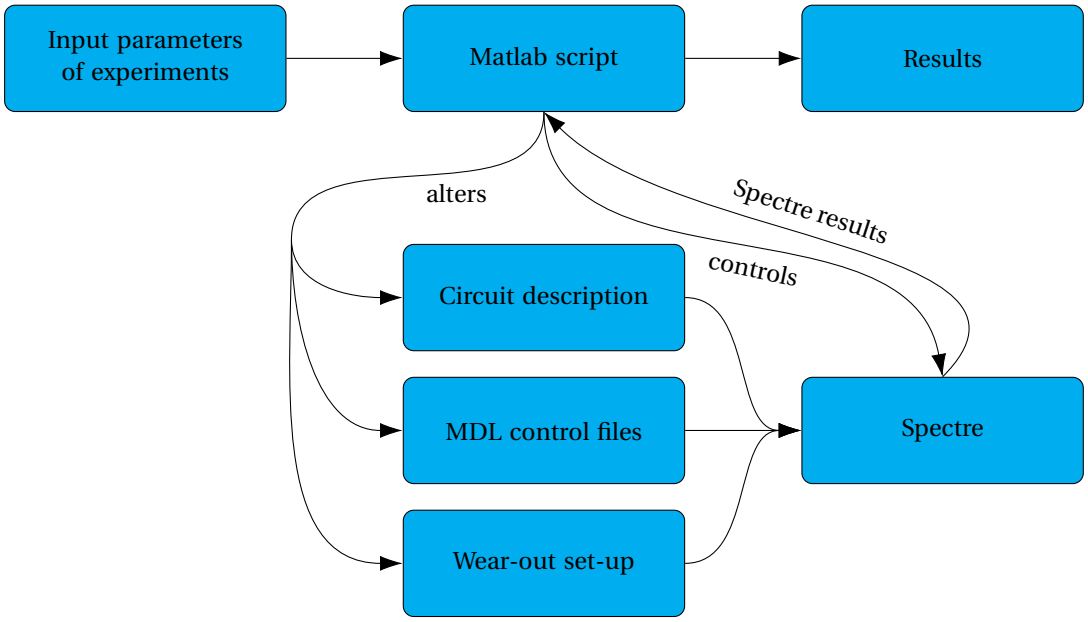


Figure 4.1: Schematic description of the simulation platform

a quarter of the maximum. If the guess was too low, so the data are still valid, the quarter is added to the half value of the previous guess. This process is repeated for n times resulting in an answer in n bits. Each extra step increases the resolution of the measurement until the resolution of extra steps becomes less than the duration of one operation cycle, which fundamentally limits this searching approach. In this research $n = 12$ and $t_{ret,\,max} = 120\,\text{ms}$ which leads to a resolution, of $r = \frac{t_{ret,\,max}}{2^n} = 29.29\,\mu\text{s}$. Figure 4.3 illustrates the binary search with two read operations, one where the estimated retention time is lower than the cell's actual retention time and one where the estimated retention time is too high. If the investigated retention time is too low, the read operation will restore the data stored in the cell. If it were too high and the data are lost, the read operation restores the flipped logical value in the cell.

4.2.2. Cell Voltage

The retention time depends on the voltage written to the cell. If the written voltage after a *0R0*-operation is higher and the overall leakage is constant, the cell will keep its contents for a shorter time. The threshold voltage of the pass transistor of a cell limits the maximum, in case of a NMOS transistor, or the minimal, in case of a PMOS transistor, voltage that can be written to a cell as discussed in Section 3.2.1. Wear-out and aging of a transistor may increase the threshold voltage and thus limit the voltage swing possible in the cell. The impact of aging on the cell voltage is measured 51 ns after the start of a cycle. Figure 4.4 shows a *1W0W0* operation on a cell and the moment the cell voltage is measured. From the Figure one can see that the measurement is always performed after the column access devices and the word line are disabled and before the next cycle starts. This ensures that the cell voltage has settled.

0R0, *1R1*, *0W0* and *1W1* operations are repeated three times before the measurement is performed to ensure a stable voltage is present in the cell. *0W1* and *1W0* operations are performed only once since applying the operations multiple times would result in another operation (e.g. *0W1* would become *1W1* after the first operation).

4.2.3. Sensing Delay

Wear-out can slow down the circuit which may lead to timing requirement violations causing the chip to fail [15]. In the design used in this research, an important metric is the sensing delay of the local sense amplifiers and therefore a measurement was defined for it. In Figure 4.5 the waveform corresponding to a *1R1* operation is shown. The sensing delay is defined as the time difference between the moments the *setn* voltage (the signal that enables the sensing operation) and *bit line C* voltage cross $50\% \cdot V_{DD} = 1.2\,\text{V}$. In the figure this difference is indicated with an arrow and ΔS_{1R1}. In the case of measuring the *0R0* sensing delay, the measured bit line would be *bit line T* instead of bit line C.

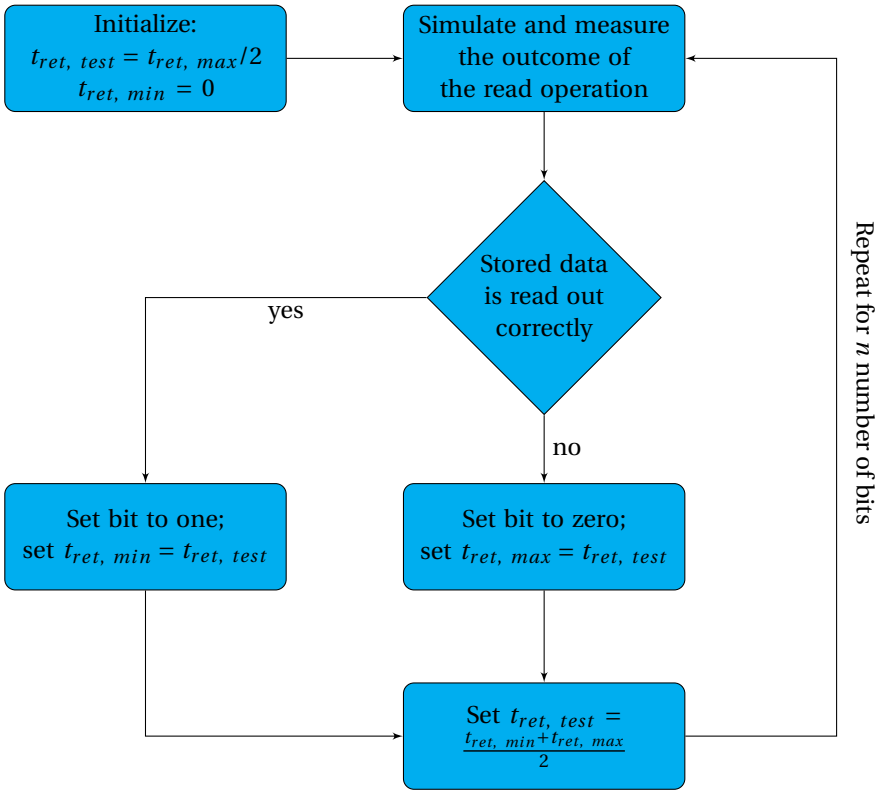


Figure 4.2: Binary search algorithm to find the retention time of a DRAM cell. The number of binary steps corresponds to the n number of bits in the final answer

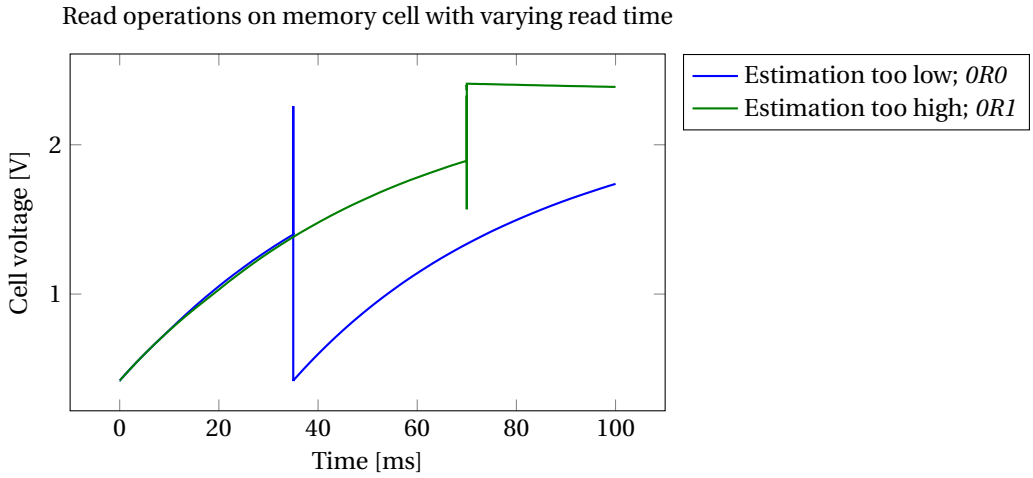


Figure 4.3: Example of retention time guesses that are too low and result in a $0R0$ operation and guesses that are too high that result in a $0R1$ operation. The spikes at $t = 35$ ms and $t = 70$ ms are caused by the read operation.

4.2.4. Writing Delay

Wear-out also affects the writing delay. In Figure 4.6 the measurement set-up for the *0W1* operation for this metric is shown. The writing delay is defined as the difference in time between the moment the write enable voltage rises above 1.2 V and the moment the internal cell voltage reaches this voltage level as well. In the figure, this time is denoted by ΔW_{0W1} and an arrow. For operations *0W0* and *1W0* the time difference is defined at the moment the internal cell voltage drops below 1.2 V. Measuring the delay for a *1W1* operation is impossible as the cell voltage does not change after the sense amplifier is enabled during the rest of the cycle.

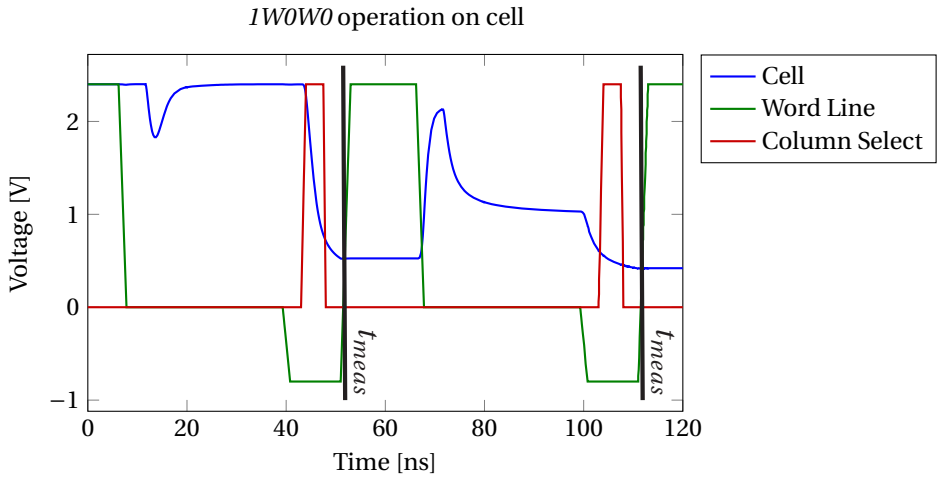


Figure 4.4: *1W0W0* operation on a cell and moment of cell voltage measurement

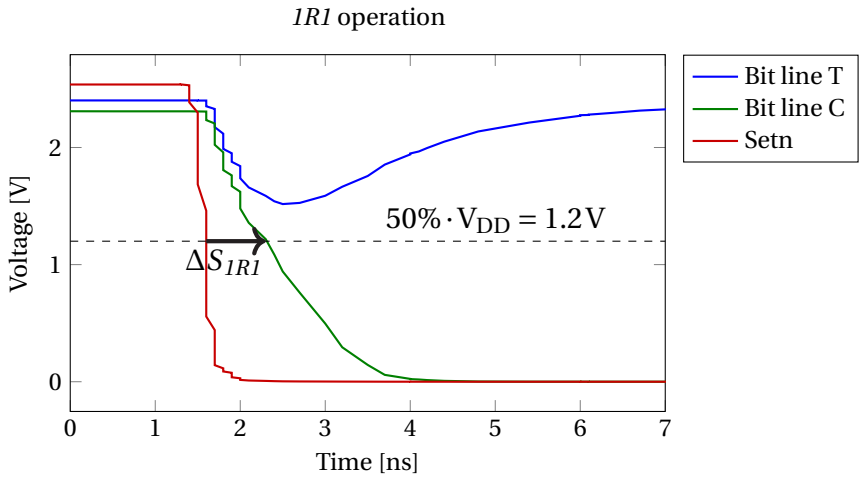


Figure 4.5: Bit line swing after an *1R1* operation on a cell. Local sensing delay, ΔS_{1R1}, is indicated with an arrow.

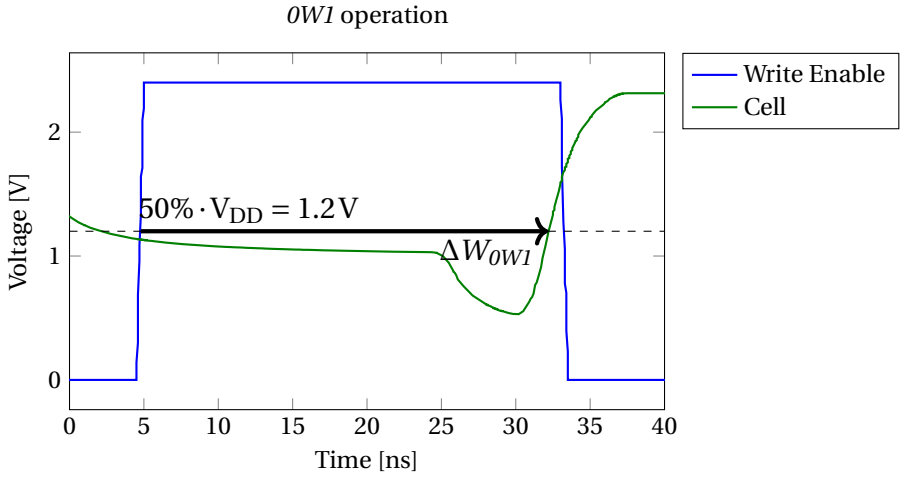


Figure 4.6: Definition of the writing delay for a *0W1* operation. The measured delay is indicated with ΔW_{0W1}.

4.2.5. Energy Consumption per Operation

The final metric that is evaluated in this research, is the energy consumption of a cycle for a specific operation, e.g. *1R1* or *0W1*. This metric is measured by multiplying the output current and voltage of every source in

the circuit and taking the integral of this over the period of one operation cycle. Equation 4.1 illustrates this procedure mathematically. In the equation, J denotes the total number of power sources in the circuit, t_0 the start time of an operation cycle and Δt the total cycle time. $V_k(t)$ is the voltage of source k and $I_k(t)$ is its current.

$$E_{cycle} = \sum_{k=1}^{J} \left(\int_{t_0}^{t_0+\Delta t} V_k(t) \cdot I_k(t) \right) \mathrm{d}t \tag{4.1}$$

4.3. Operating Conditions

The circuit was operated under different conditions to see the effects of workload and temperature on the wear-out of the circuit. Below these conditions are explained in more detail.

4.3.1. Workload Dependence

The best-case and worst-case workload that were defined in Section 3.4.2 were applied to the circuit. Per subcomponent the workload for every transistor was set by the duty factors as presented in Section 3.4.2 and the duty factor of the controlling source. E.g. in the reference devices the equalizing transistor has a different workload than the cell transistors. It was assumed that the workload between two bit lines was shared equally, meaning that every bit line is pulled down as often as pulled high.

4.3.2. Temperature Dependence

As can be seen from the model descriptions in Section 3.4.4 the impact of HCI and BTI changes with temperature. Besides this, temperature also alters the circuit performance; e.g. leakage currents and junction voltages increase. In order to make a fair comparison, the circuit was aged at elevated temperatures but verified at a constant circuit temperature of 300 K. The evaluated temperatures are: 300 K, 350 K and 400 K.

4.4. Performed Experiments

In this research three experiments, individual impact of aging models, combined impact of aging models and transistor downscaling, are performed. Table 4.1 summarizes these experiments. In the following subsections the experiments will be described in more detail.

Table 4.1: Performed experiments in this research. In the column 'Wear-out', the letters B, H and R mean BTI, HCI and radiation respectively. In column 'Subcomponents', C denotes memory cells, R reference cells, S the sense amplifier and P the precharge transistor

Individual Impact		Combined Impact		Transistor scaling	
Wear-out	Subcomponents	Wear-out	Subcomponents	Wear-out	Subcomponents
B	C	BH	C	B	CRSP
H	CR	BHR	CR	H	
R	CSP		CSP	R	
	CRSP		CRSP	BH	
				BHR	

4.4.1. Individual Impact of Aging Models

First the individual impact of the aging models is investigated. The wear-out models are applied individually to a varying set of subcomponents in order to include their individual contributions to the combined wear-out of the circuit. The set of subcomponents follows the read path of the DRAM up from the cell to the sense amplifier. The combinations therefore are: memory cells only; memory and reference cells; memory cells, local precharge devices and local sense amplifier; memory and reference cells, local precharge devices and local sense amplifier.

4.4.2. Combined Impact of Aging Models

Once the effects of the individual aging models are examined, combinations with them can be made. These two combinations are BTI combined with HCI and the combination of all three aging mechanisms, BTI, HCI

and radiation. Again, the experiments will be performed on the varying subset of components as presented above.

4.4.3. Impact of Transistor Scaling

Next to the impact of the subcomponents wear-out, the influence of scaling on the DRAM performance was also examined.

Circuit Scaling

The complete circuit (also containing the pass transistors, global sense amplifier and global precharge devices) and transistor library were scaled manually based on standard CMOS scaling to the next technology node with scaling factor $\kappa = \sqrt{2}$ [97]–[99]. Because no modern DRAM transistor libraries were available to this research, the scaling had to be done manually. As presented by Baccarani, Wordeman, and Dennard in [99] voltages do not scale with factor κ, but rather with a factor $\lambda = \kappa \frac{V_{dd,\,new}}{V_{dd,\,old}}$. The new supply voltage was set to 2.2 V instead of 2.4 V setting $\lambda = \sqrt{2} \cdot \frac{2.4\text{V}}{2.2\text{V}} = 1.2964$. This was done because the circuit lay-out was not altered with the scaling and thus not lowering the supply voltage would increase energy consumption tremendously. Lowering the supply voltage further down leads to malfunctioning of the circuit. An overview of all SPICE transistor parameters and their scaling coefficient can be found in Table 4.2. The third column gives the reasoning behind the chosen scaling factor. For some parameters no adequate scaling function could be found or derived; these rows contain the text '*scaling unknown*' and the scaling factor is set to 1.

Table 4.2: Spectre transistor model parameters with their scaling factor and a descriptive reasoning of the scaling parameter. If an adequate scaling function could not be found '*scaling unknown*' is entered, the scaling factor is then set to 1. Scaling follows observartions in [98], [99]

Spectre Parameter	Scaling	Description
tox	$1/\kappa$	Oxide thickness
xj	$1/\kappa$	Junction depth
tpg	1	Gate material
delta	*scaling unknown*	Width effect on threshold
ld	*scaling unknown*	Lateral length diffusion
kp	*scaling unknown*	Transconductance parameter
theta	*scaling unknown*	Mobility modulation
rsh	$1/(\lambda \cdot \kappa)$	Diffusion sheet resistance
nsub	$\lambda \cdot \kappa$	Substrate doping
vmax	λ/κ	Maximum drift velocity of carriers
eta	λ/κ	Static feedback
kappa	*scaling unknown*	Saturation field vector
cgdo	$1/\kappa$	Gate drain overlap capacitance
cgso	$1/\kappa$	Gate source overlap capacitance
cgbo	$1/\kappa$	Gate bulk overlap capacitance
cj	$1/\kappa$	Bulk junction bottom capacitance
mj	$1/\kappa$	Bulk junction bottom grading coefficient
cjsw	$1/\kappa$	Bulk junction sidewall capacitance
mjsw	$1/\kappa$	Bulk junction sidewall grading coefficient
pb	λ/κ	Bulk junction potential
vto	λ/κ	Zero bias threshold voltage

Wear-out Mechanism Scaling

The aging mechanisms HCI and BTI were also scaled down. Radiation effects were scaled down via the resized transistor parameters. For HCI the currents were again re-calibrated following the same procedure as was used to set up the original model. The lower threshold voltage in the scaled circuit increased the substrate leakage current to 178.8 nA. The drain current remained nearly equal at 2.1861 µA. Next to this, the smaller device sizes increased the statistical spread of the HCI effect (see Equation 3.32). The factor K in this same Equation was set to $K = 7.2$ to include the effects of the smaller technological feature size.

The main driving factors in BTI scaling are transistor dimensions and the operating voltages. The dimension scaling, length, width and oxide thickness, increases the Gaussian spread on the threshold voltage

variation, as can be seen from Equation 3.28. The oxide thickness also has an impact on the mean value of the threshold voltage shift. With equal bias conditions, transistors with thinner oxides suffer more from BTI than thicker ones [70], [71]. From the data presented in [70] it follows that halving the oxide thickness results in a tripled threshold voltage shift under same bias conditions. This factor was added to the down-sized model as well. However, the downscaled circuit will not suffer three times more BTI degradation, as the supply voltage was reduced as well.

Scaling Experiments
The down-sized circuit was simulated without any aging effects but with downscaled process variations to define a nominal measurement. The following operating conditions were applied to the downscaled circuit. The aging temperature was set to: 300 K, 350 K and 400 K, the workload was set to the worst-case workload. The experiment was performed on the combination of memory cells, reference cells, sense amplifier and precharge devices. The other components are scaled down as well, but do not suffer from aging effects.

4.5. Initial Circuit Performance

In order to evaluate the results from the experiments, it is needed to know how the circuit behaves originally. A reference was made by doing a Monte-Carlo simulation on the circuit with all wear-out mechanisms in place and the time set to $t = 0$ s. The time setting ensures that no wear-out effects will manifest themselves. The results thus only depend on the initial mismatch of the transistor given by Pelgrom's model. The references were made by performing 300 Monte-Carlo iterations from which the mean and standard deviation could be found for every metric. The results for the unaltered, normal circuit are presented in Table 4.3a and for the downscaled circuit in Table 4.3b. Next to the dynamic energy consumption (e.g. Energy *0W1* in the tables) also the static energy consumption was measured. The static energy consumption for one clock cycle in the unscaled circuit was 467 aJ.

Table 4.3: Nominal circuit measurements

(a) Unaltered circuit results for all metrics

Metric	Mean	Std. Deviation	Unit
Retention time	57.496	4.627	ms
Delay *1R1*	0.695	$1.210 \cdot 10^{-2}$	ns
Delay *0R0*	0.555	$6.587 \cdot 10^{-3}$	ns
Delay *1W0*	38.941	$1.724 \cdot 10^{-2}$	ns
Delay *0W1*	40.244	$1.430 \cdot 10^{-2}$	ns
Delay *0W0*	10.162	$4.167 \cdot 10^{-1}$	ns
Energy *1R1*	9.274	$1.922 \cdot 10^{-2}$	pJ
Energy *0R0*	9.029	$1.213 \cdot 10^{-2}$	pJ
Energy *1W1*	9.317	$2.110 \cdot 10^{-2}$	pJ
Energy *1W0*	14.016	$2.269 \cdot 10^{-2}$	pJ
Energy *0W1*	13.940	$1.825 \cdot 10^{-2}$	pJ
Energy *0W0*	9.084	$1.225 \cdot 10^{-2}$	pJ
Voltage *1R1*	2.401	$3.913 \cdot 10^{-6}$	V
Voltage *0R0*	0.418	$1.504 \cdot 10^{-2}$	V
Voltage *1W1*	2.401	$3.914 \cdot 10^{-6}$	V
Voltage *1W0*	0.525	$1.436 \cdot 10^{-2}$	V
Voltage *0W1*	2.236	$3.346 \cdot 10^{-3}$	V
Voltage *0W0*	0.418	$1.504 \cdot 10^{-2}$	V

(b) Downscaled circuit results for all metrics

Metric	Mean	Std. Deviation	Unit
Retention time	34.583	4.912	ms
Delay *1R1*	0.583	$1.204 \cdot 10^{-2}$	ns
Delay *0R0*	0.505	$7.567 \cdot 10^{-3}$	ns
Delay *1W0*	37.953	$1.077 \cdot 10^{-2}$	ns
Delay *0W1*	38.433	$1.117 \cdot 10^{-2}$	ns
Delay *0W0*	6.152	$4.465 \cdot 10^{-2}$	ns
Energy *1R1*	62.074	$4.837 \cdot 10^{-1}$	pJ
Energy *0R0*	61.815	$4.726 \cdot 10^{-1}$	pJ
Energy *1W1*	62.273	$4.837 \cdot 10^{-1}$	pJ
Energy *1W0*	65.945	$4.727 \cdot 10^{-1}$	pJ
Energy *0W1*	68.336	$4.835 \cdot 10^{-1}$	pJ
Energy *0W0*	61.864	$4.725 \cdot 10^{-1}$	pJ
Voltage *1R1*	2.201	$1.797 \cdot 10^{-5}$	V
Voltage *0R0*	0.175	$1.147 \cdot 10^{-2}$	V
Voltage *1W1*	2.201	$2.422 \cdot 10^{-5}$	V
Voltage *1W0*	0.219	$1.129 \cdot 10^{-2}$	V
Voltage *0W1*	2.178	$6.419 \cdot 10^{-4}$	V
Voltage *0W0*	0.175	$1.147 \cdot 10^{-2}$	V

5

Experimental Results and Reliability Prediction Model

Following the structure from Section 4.4, the results of the performed experiments are presented in this chapter. In Section 5.2 the individual impact of the aging models will be examined, while in Section 5.3 the results of the combination of aging models will be presented. Then, in Section 5.4 the impact of downscaling the circuit is presented. The findings of the previous sections will be summarized in Section 5.5. The chapter finishes with the a reliability prediction model based on the results of this research in Section 5.6.

5.1. Introduction

From the resulting measurements it turned out that the worst-case workload only increased the effects that can be seen under best-case workload situations. The same argument holds for the temperature dependence; increasing the temperature increases the effects seen on lower temperatures. It was therefore chosen to introduce the observed effects in detail the first time they appear in the text and summarize the effects in the following parts by showing only degradation under worst-case workload or for the highest and lowest temperature. The retention time figures are plotted as error bars. The error in this plot indicates 1σ of standard deviation.

Retention time is taken as the prime metric to evaluate the results. It was found that the delay and energy metrics do not vary as much as the retention time metric does. Therefore these metrics are described more briefly in a table. The tables present the absolute change in delay, and the relative delay and energy consumption changes. The voltage measurements are only included when certain behavior needs to be explained as these measurements do not directly result alter the circuit performance. The prediction model however, will give equal priority to the delay and energy metrics and retention time.

5.2. Individual Impact of Aging Models

In this section, the impact of the individual wear-out mechanisms will be described. First, the effects of BTI on the DRAM will be presented, then the HCI effects and finally the effects of radiation trapping. First the results of applying the aging models to the memory cells only are presented to be followed by the combinations memory cell and reference cells, memory cells, precharge devices, sense amplifier and memory cells, reference cells, precharge devices, sense amplifier.

5.2.1. Bias Temperature Instability

In this section the impact of BTI on the different arrangements of subcomponents will presented.

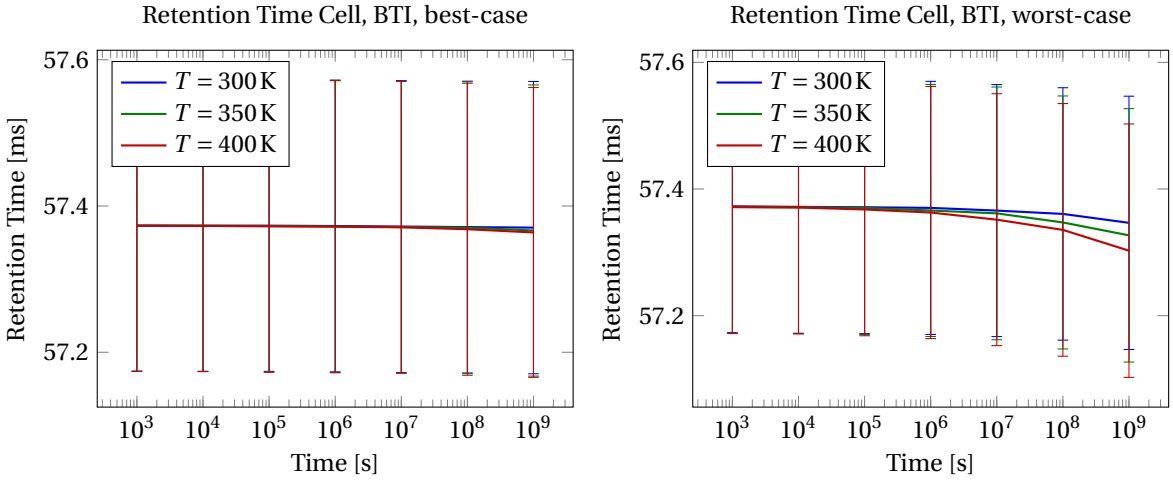
Impact on Memory Cells

Figure 5.1a and 5.1b show the impact of BTI on the memory cell only for the best-case workload for the worst-case, respectively.

From the figures, one clearly sees the impact of workload on the degradation of the DRAM; in the best-case scenario almost no degradation is observed, even not at higher ages or temperatures while in the worst-case

scenario the retention time drops with higher temperature and higher age. The decreasing retention time can be explained easily by the fact that a PMOS transistor is used as an access device to the cell capacitor, as was presented in Section 3.4.1. BTI increases the threshold voltage of the transistor and thus increases the minimum cell voltage after a logical '0' is written. With all the other components unaltered and not suffering from wear-out, the total amount of charge that needs to be dissipated in order to switch the stored value in the cell decreases and so the retention time of the cell decreases. The overall degradation is even in the worst possible setting relatively small. This is caused by the low duty factor of the cell, which limits the threshold voltage shift that occurs.

Table 5.1 summarizes the changes in delay and energy metrics for the worst-case workload at the highest temperature after $t = 1 \cdot 10^8$ s of aging. The most notable entry is given by the *0W0* delay, which increases by 0.759 %. This is caused by the increased threshold voltage of the cell transistor. Note that, as was defined explained in Section 4.2.3, no delay measurement can be made for *1W1* operations.



(a) Cell BTI best-case

(b) Cell BTI worst-case

Figure 5.1: BTI impact on memory cell

Table 5.1: Cell BTI worst-case workload, $T = 400$ K, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	0.216	0.039	0.001
1R1	-0.080	-0.011	0.021
0W0	77.083	0.759	-0.003
0W1	0.588	0.001	-0.004
1W0	3.029	0.008	-0.007
1W1	—	—	-0.005

Impact on Memory Cells and Reference Cells

The effects of BTI on the reference cells and the memory cells are presented in Figure 5.2 and in Table 5.2. Note that in Figure 5.2 only $T = 300$ K and $T = 400$ K are plotted, preventing cluttering of the figure.

From the figure it becomes clear that BTI effects increase the retention time. This is caused by the increase in absolute threshold voltage of the reference cell's access transistors. The minimum voltage on the lowest reference capacitor increases due to the increased threshold voltage. Then, when the reference cells are equalized, the reference voltage, V_{ref}, also increases, as was explained in Section 3.4.1. Figure 5.3 illustrates why a small change of the reference voltage can increase the retention time significantly. Because charging of a capacitor with constant series resistance is not linear, the time it takes to charge it from one voltage to another is also non-linear for the same voltage step. The small increase in reference voltage thus significantly improves the retention time of the circuit. The drawback of this is that the voltage range for a logical '1' decreases, i.e., the maximum voltage that is read as a logical '0' increases, therefore decreasing the range in which an '1' is read.

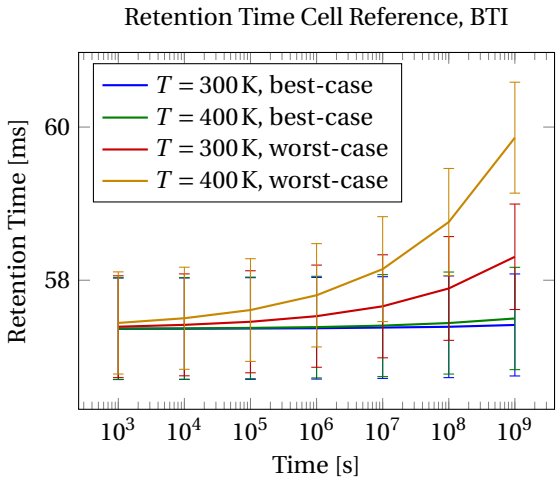


Figure 5.2: Cell Reference BTI

Table 5.2: Cell Reference BTI worst-case workload, $T = 400\,\mathrm{K}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-3.112	-0.561	-0.077
1R1	5.008	0.721	0.061
0W0	73.576	0.724	-0.077
0W1	-6.919	-0.017	-0.138
1W0	-0.378	-0.001	-0.057
1W1	—	—	0.048

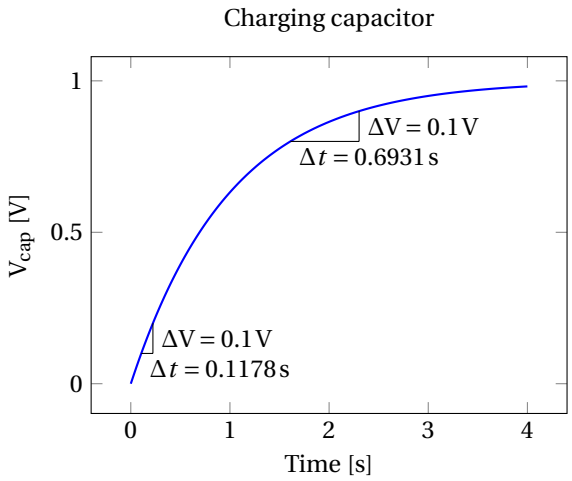


Figure 5.3: Differences in capacitor charging time for equal voltage increments

Impact on Memory Cells, Precharge Devices and Sense Amplifier
Figure 5.4 and Table 5.3 show the results of BTI wear-out on the memory cell, the precharge devices and the sense amplifier. From the figure it can be seen that the retention time and its spread increase slightly after longer periods of aging. This phenomenon will be described in more detail in Section 5.3.1. Furthermore, from Table 5.3 it can be seen that the sensing delays and the *0W0* writing delay increase slightly. The increase in writing delay is related to the increase in cell threshold voltage, as can be seen in the analysis for the cell only. The sensing delays increase because both the threshold voltage of NMOS and PMOS transistors increases, which slows down the sensing operation. PMOS transistors are more affected by BTI than by HCI, while NMOS transistors have it the other way around. Section 5.2.2 will show how this difference affects sense amplifier degradation.

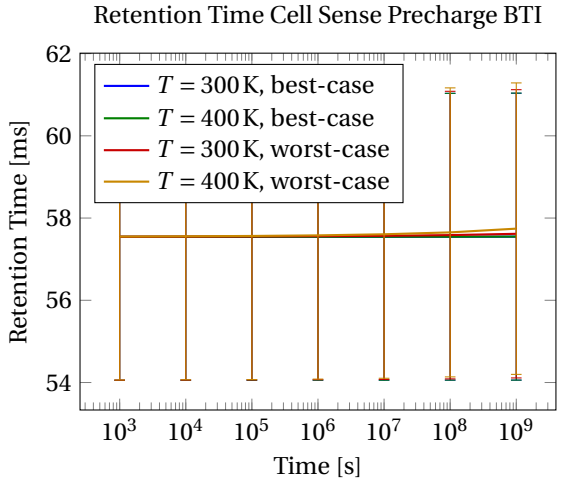


Figure 5.4: Cell Sense Precharge BTI

Table 5.3: Cell Sense Precharge BTI worst-case workload, $T = 400\,\mathrm{K}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	2.518	0.454	-0.028
1R1	3.733	0.537	-0.035
0W0	80.531	0.792	-0.030
0W1	18.648	0.046	-0.521
1W0	-18.115	-0.047	-0.516
1W1	—	—	-0.045

Impact on Memory Cells, Reference Cells, Precharge Devices and Sense Amplifier
The effects when the memory cells, the reference cells, the precharge devices and the sense amplifier are aged with BTI only can be seen in Figure 5.5 and Table 5.4. The results basically show an addition of the previous results; the retention time increases mainly due to the aging of the reference cells with a small contribution of the sense amplifier and the precharge devices, both compensating for the small reduction of retention time from the cell only. The sensing delays are affected by both the reference cells and the precharge devices resulting in a small *0R0* sensing delay decrease and relatively large increase of 1.273 % for *1R1*. The writing delay for *0W0* is again set by the memory cell degradation only. It also becomes clear that for all operations the energy consumption is slightly reduced, except for the *1W1*-operation.

5.2.2. Hot Carrier Injection

The effects of hot carrier degradation on all the individual subcomponents will be presented in this section.

Impact on Memory Cells
In Figure 5.6 and Table 5.5, the influence of HCI on the memory cell is shown. The degradation is less than when the cell is stressed by BTI which can be explained by the fact that PMOS transistors, used as the access device, are less prone to HCI than BTI. Furthermore, the duty factor is low which also leads to less degradation.

Impact on Memory Cells and Reference Cells
As was the case with BTI, V_{ref} again increases which leads to an increase in retention time and *1R1* sensing delay and a decrease of *0R0* sensing delay, as can be seen in Figure 5.7 and Table 5.6. The impact of HCI is larger than for BTI which can be explained by the different impact of duty factor on both wear-out mechanisms. As was described in Section 3.4.4, HCI stressed transistors do not experience any recovery and thus the duty factor linearly influences the threshold voltage shift. BTI, however, has a recovering component and thus the duty factor does not have a linear impact on the threshold voltage shift, as was stated in Section 3.4.4.

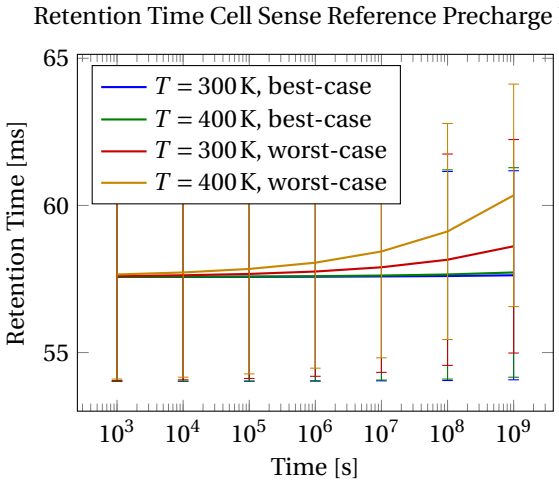


Figure 5.5: Cell Sense Reference Precharge BTI

Table 5.4: Cell Reference Sense Precharge BTI worst-case workload, $T = 400\,\mathrm{K}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.875	-0.158	-0.107
1R1	8.846	1.273	-0.001
0W0	76.909	0.757	-0.110
0W1	11.014	0.027	-0.658
1W0	-21.490	-0.055	-0.567
1W1	—	—	0.006

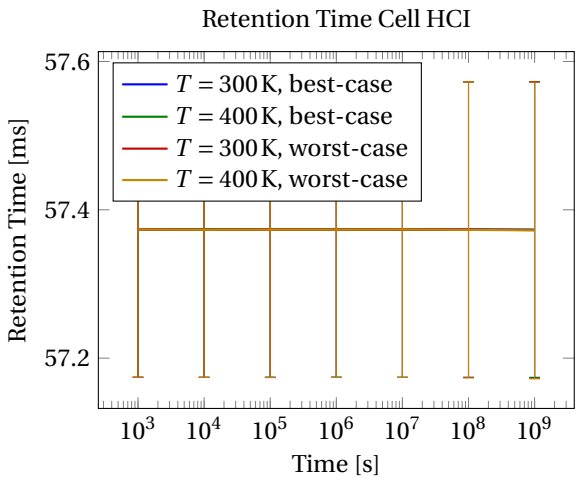


Figure 5.6: Cell HCI

Table 5.5: Cell HCI worst-case workload, $T = 400\,\mathrm{K}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.213	-0.038	-0.006
1R1	-0.118	-0.017	0.020
0W0	-6.033	-0.059	-0.010
0W1	0.041	0.000	-0.007
1W0	-0.288	-0.001	-0.008
1W1	—	—	-0.006

Figure 5.8 shows the fitted duty factor curve for BTI and the linear impact of duty factor for HCI. From this figure it can be seen that for a duty factor $DF \lesssim 0.4$ BTI has the highest relative impact and for a higher duty factor HCI has most impact. Under the worst-case workload, the duty factor for a reference cell is equal to $DF_{ref,\,wc} = DF_{col,\,wc} \cdot DF_{ref} = 0.5112$. The higher duty factor leads to a higher threshold voltage shift under HCI stress than BTI stress and thus a net higher reference voltage. Since the duty factor for the memory cell is also relatively low in the worst-case scenario, BTI is more important as a degrading factor which also comes back in the *0W0* delay measurement.

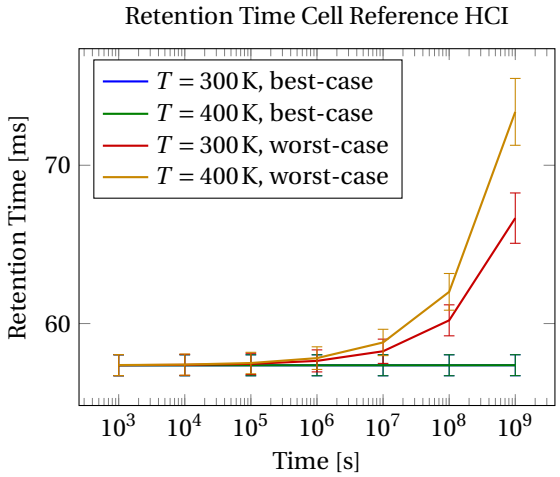


Figure 5.7: Cell Reference HCI

Table 5.6: Cell Reference HCI worst-case workload, $T = 400\,\mathrm{K}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-10.631	-1.915	-0.252
1R1	16.273	2.341	0.147
0W0	-16.969	-0.167	-0.243
0W1	-26.899	-0.067	-0.473
1W0	-12.743	-0.033	-0.178
1W1	—	—	0.169

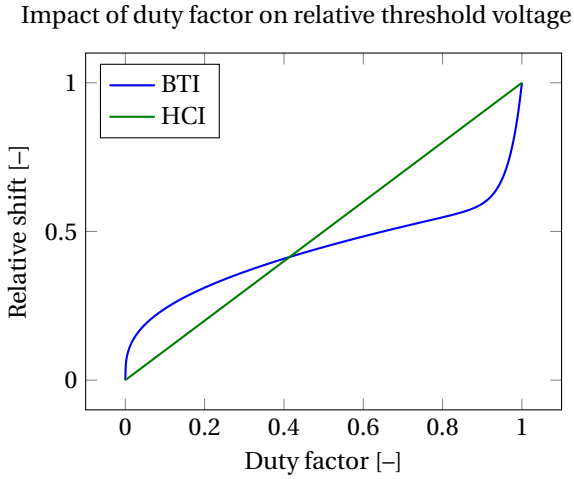


Figure 5.8: Impact of duty factor on relative threshold voltage shift for BTI and HCI

Impact on Memory Cells, Precharge Devices and Sense Amplifier

Stressing the memory cell, the precharge devices and the sense amplifier again slightly increases the mean retention time and increases its spread, as can be seen in Figure 5.9. The causes of this will be explained in Section 5.3.1. From Table 5.7 it follows that both sensing delays increase again; *0R0* 11.996 % and *1R1* 14.112 %. Because the bit lines are precharged to V_{DD}, the sensing delay is mainly dependent on the discharge delay of the bit line with the lowest voltage. Discharging the bit lines is done via the NMOS transistors in the sense amplifier. Since NMOS transistors suffer more from HCI than PMOS transistors and the workload for these column devices is high, as mentioned above, aging causes the sensing delays to increase. The delay for a *1R1* operation is higher because in this case one of two bit lines needs to be discharged almost from V_{DD} to ground, while for a *0R0* operation one bit line is already closer to ground level. From Table 5.7 it also follows that the writing delay for a *0W0* operation increases more than what would be expected of cell wear-out only.

An explanation for this can be found in the timing of the sense amplifier. When a value is written to a cell the local sense amplifier is still enabled and thus helps to drive the cell to the desired voltage. When the sense amplifier degrades, the writing delay will then increase. The overall energy consumption decreases for all operations slightly. This can be attributed by the increased threshold voltages which lowers leakage currents.

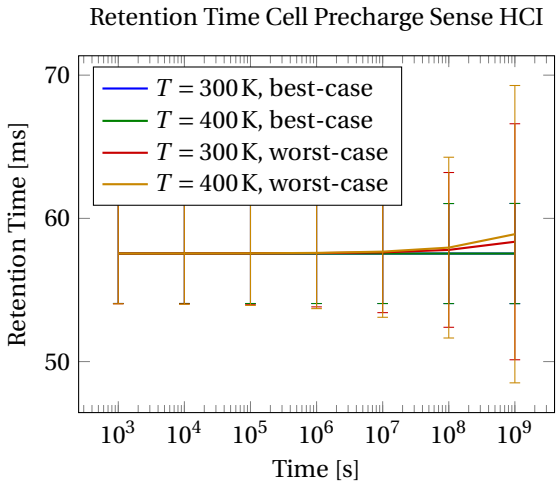


Figure 5.9: Cell Precharge Sense HCI

Table 5.7: Cell Sense Precharge HCI worst-case workload, $T = 400\,K$, $t = 1 \cdot 10^8\,s$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	66.581	11.996	-0.591
1R1	98.074	14.112	-0.751
0W0	55.659	0.548	-0.599
0W1	-65.360	-0.162	-2.067
1W0	-38.215	-0.098	-2.096
1W1	—	—	-0.631

Impact on Memory Cells, Reference Cells, Precharge Devices and Sense Amplifier

Application of HCI to all subcomponents again results in a summation of the previously seen results, presented in Figure 5.10 and Table 5.8. From the figure again the total increase in retention time can be clearly seen under the worst-case workload. Also the increase in spread caused by the sense amplifier degradation, is clearly visible. From the table it can be seen that the combination of sense amplifier and reference cell degradation causes the *1R1* sensing delay to increase more than 17 %. The *0R0* operation sensing delay caused by the sense amplifier degradation, is partially covered by the delay decrease caused by the reference cell degradation. Overall, the energy consumption decreases.

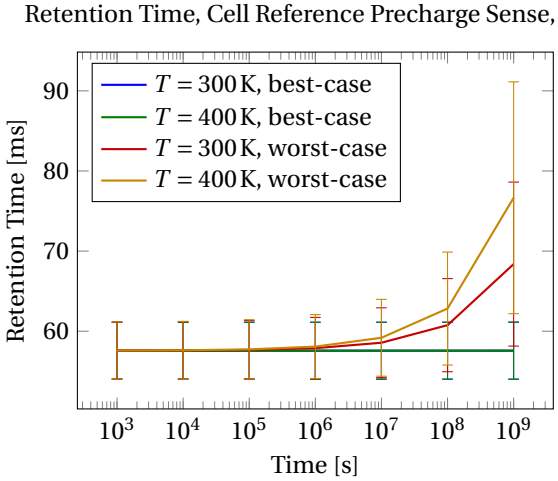


Figure 5.10: Cell Reference Precharge Sense HCI

Table 5.8: Cell Reference Sense Precharge HCI worst-case workload, $T = 400\,K$, $t = 1 \cdot 10^8\,s$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	53.540	9.647	-0.841
1R1	118.838	17.099	-0.599
0W0	42.184	0.415	-0.875
0W1	-86.231	-0.214	-2.467
1W0	-50.389	-0.129	-2.272
1W1	—	—	-0.481

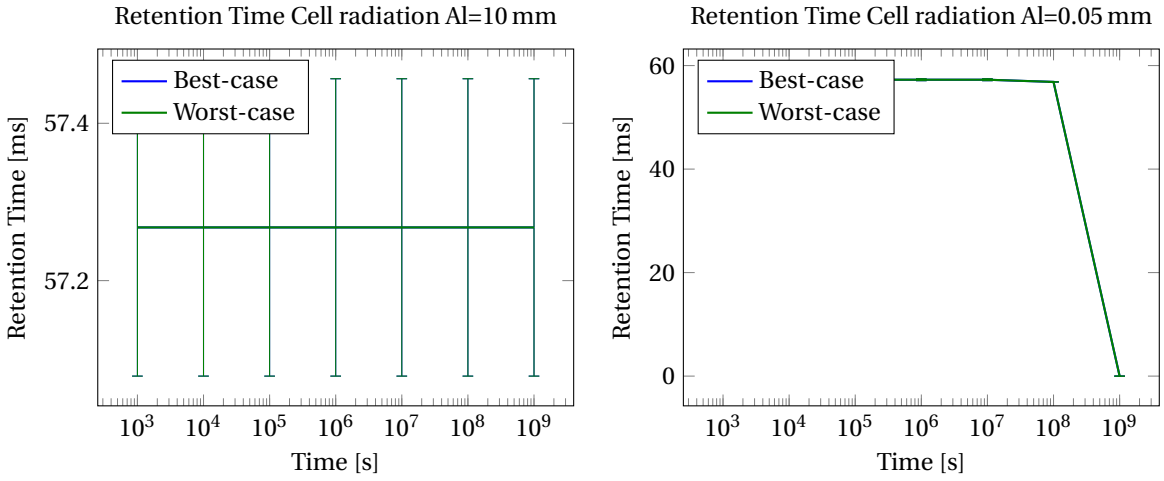
5.2.3. Radiation Effects

This section will describe the effects of radiation trapping on all subcomponents.

Impact on Memory Cells

The impact of radiation trapping on the memory cells only is depicted in Figures 5.11a and 5.11b respectively shielded by a 10 mm and 0.5 mm thick layer of aluminum. The corresponding dose rates are then: $D = 29.59\,\mu\text{rad/s}$ for 10 mm aluminum and 1.26 mrad/s for 0.5 mm aluminum. As was presented in Section 3.4.4, the radiation model is temperature independent and the total amount of trapped charge depends on the workload but charge trapping also occurs when a transistor is unbiased. From the figures it follows that the thickness of shielding, and thus the total accumulated dose, is the most important parameter describing the circuit degradation. The cell leakage currents start to increase notably with the thin layer of shielding after $t = 1 \cdot 10^7$ s and have increased so much at $t = 1 \cdot 10^9$ s that none of the cells are able to store data.

Table 5.9 shows that the delay and energy metrics do not vary much. An explanation for this is that the leakage currents involved with radiation trapping are relatively small and are only active for a short period of time, as was described in Section 3.4.4.



(a) Cell rad (b) Cell rad

Figure 5.11: Retention time of a memory cell suffering from radiation wear-out

Table 5.9: Cell, Radiation, worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.346	-0.062	-0.008
1R1	-0.130	-0.019	0.020
0W0	-35.106	-0.345	-0.012
0W1	-0.116	-0.000	-0.007
1W0	-1.353	-0.003	-0.008
1W1	—	—	-0.006

Impact on Memory Cells and Reference Cells

Figure 5.12 and Table 5.10 show the results when radiation trapping is applied to the memory cells and the reference cells. Again it becomes clear that when a total dose between 126 krad and 1.26 Mrad is accumulated, the circuit fails to store data. From the table oit becomes clear that delays and energy consumption are not affected significantly by radiation.

Impact on Memory Cells, Precharge Devices and Sense Amplifier

In Figure 5.13 and Table 5.11 the impact of radiation on the memory cell, sense amplifier and precharge devices is shown. There are no entries for $t > 1 \cdot 10^8$ s because the circuit degrades so much that it becomes impossible to perform a successful read operation, meaning the hardware is unable to provide a valid output signal. From the table follows again that no significant degradation occurs to the delay and energy consumption metrics.

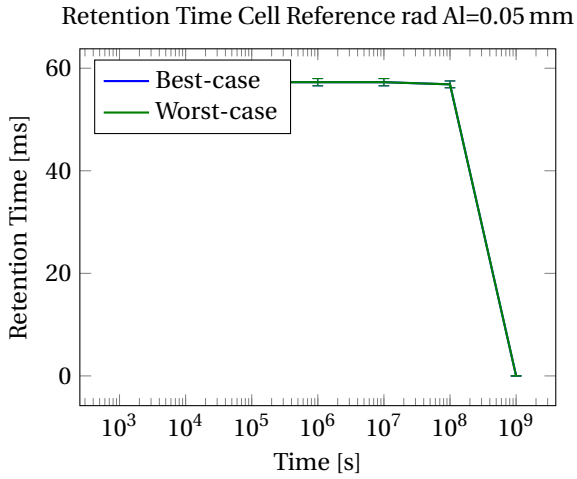


Figure 5.12: Cell Reference rad

Table 5.10: Cell Reference Radiation worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 1 \cdot 10^8\,\text{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.324	-0.058	-0.007
1R1	-0.245	-0.035	0.019
0W0	-35.089	-0.345	-0.010
0W1	0.026	0.000	-0.004
1W0	-1.336	-0.003	-0.007
1W1	—	—	-0.007

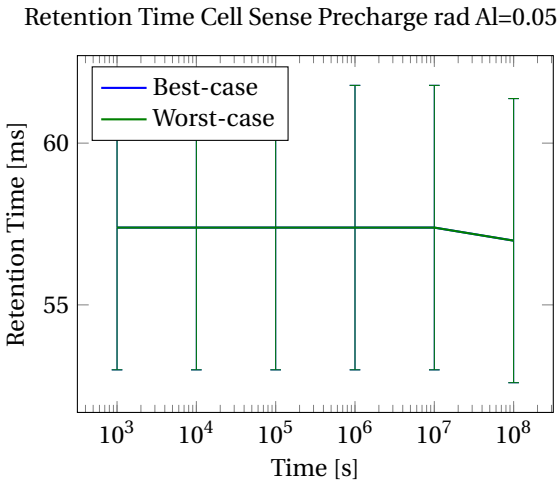


Figure 5.13: Cell Sense Precharge rad

Table 5.11: Cell Sense Precharge Radiation worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 1 \cdot 10^8\,\text{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.118	-0.021	-0.002
1R1	0.018	0.003	-0.001
0W0	-34.796	-0.342	-0.004
0W1	-0.443	-0.001	-0.002
1W0	-1.226	-0.003	-0.001
1W1	—	—	-0.001

Impact on Memory Cells, Reference Cells, Precharge Devices and Sense Amplifier
Again, the circuit fails when the TID exceeds 126 krad and the delay and energy consumption are nearly unaltered, as can be seen in Figure 5.14 and Table 5.12.

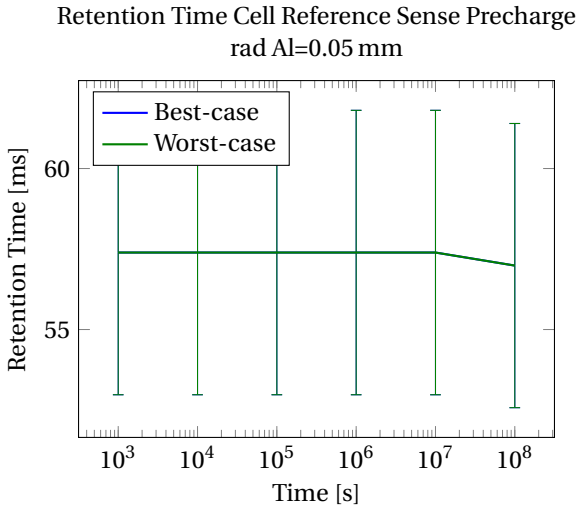


Figure 5.14: Cell Reference Sense Precharge rad

Table 5.12: Cell Reference Sense Precharge Radiation worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 1 \cdot 10^8\,\text{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.108	-0.019	-0.002
1R1	0.005	0.001	-0.001
0W0	-34.783	-0.342	-0.004
0W1	-0.473	-0.001	-0.005
1W0	-1.223	-0.003	-0.001
1W1	—	—	-0.002

5.3. Combined Impact of Aging Models

In the previous section the impact of all individual wear-out effects was presented. This section combines the BTI and HCI wear-out effects and then also adds the effects of radiation to this combination. The same order of subcomponents as was used in the previous section, is used here. First the results of applying the aging models to the memory cells only are presented to be followed by the combinations memory cell and reference cells, memory cells, precharge devices, sense amplifier and memory cells, reference cells, precharge devices, sense amplifier.

5.3.1. Bias Temperature Instability and Hot Carrier Injection

Impact on Memory Cellss
In Figure 5.15 and Table 5.13 the combination of BTI and HCI and its application to a single memory cell is shown. From the individual analysis of wear-out mechanisms it can be concluded that BTI is the most important factor for this degradation which is caused by its low duty factor. HCI effects are smaller than the effects of BTI but can not be neglected. From the table it becomes clear that the *0W0* writing delay increases slightly in comparison with BTI only degradation which must therefore be attributed to the threshold voltage increase caused by HCI.

Impact on Memory Cells and Reference Cells
Figure 5.16 presents the increase of retention time when both the reference and the memory cells are stressed with BTI and HCI. From Section 5.2 it followed that HCI has a larger impact than BTI has caused by the high duty factor of the reference cells. Next to that, Table 5.14 shows again the previously seen decrease in *0R0* and increase in *1R1* sensing delay. The slight decrease in delay of the *0W0* caused by HCI is not enough to compensate the increased delay caused by BTI but the increase is less than with BTI degradation only.

Impact on Memory Cells, Precharge Devices and Sense Amplifier
Figure 5.17 and Table 5.15 show the results for the sense amplifier and the precharge devices suffering from BTI and HCI wear-out. The figure shows again a slight increase in retention time as was expected from the individual contributions. Figure 5.18 splits the retention time degradation into two parts to explain the increase in retention time: retention time degradation caused by aging of the memory cell and precharge transistors and caused by degradation of the memory cell and sense amplifier. From the figure it becomes clear that sense amplifier degradation causes the increase in standard deviation. This is caused by the mismatch between the two cross-coupled inverters that form the sense amplifier. If the threshold voltages of the NMOS

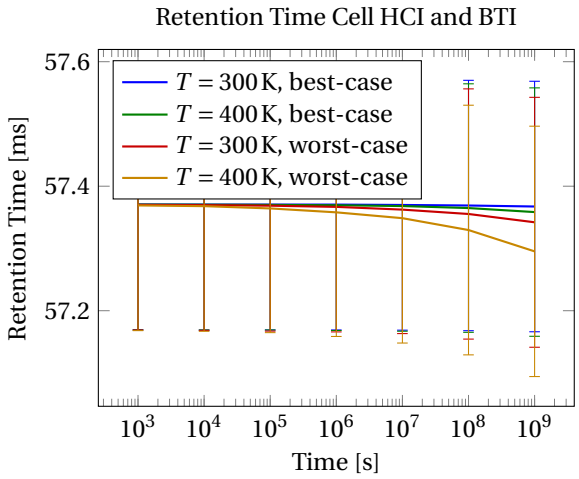


Figure 5.15: Cell HCI+BTI

Table 5.13: Cell BTI+HCI Worst-case workload, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	0.263	0.047	0.002
1R1	-0.075	-0.011	0.021
0W0	86.852	0.855	-0.002
0W1	0.645	0.002	-0.004
1W0	3.396	0.009	-0.008
1W1	—	—	-0.005

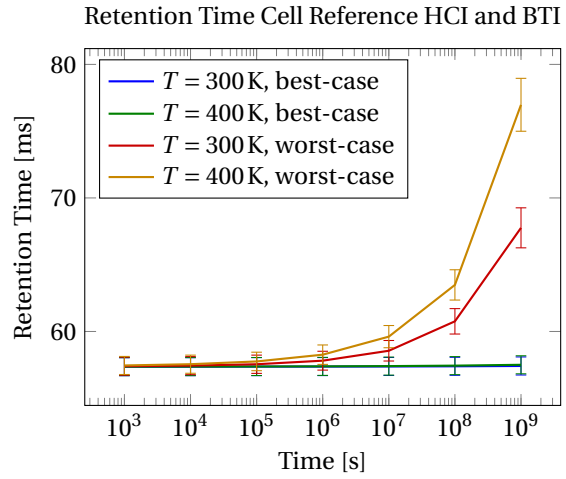


Figure 5.16: Cell Reference HCI+BTI

Table 5.14: Cell Reference BTI+HCI Worst-case workload, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-13.315	-2.399	-0.319
1R1	21.186	3.048	0.200
0W0	72.427	0.713	-0.278
0W1	-34.689	-0.086	-0.633
1W0	-16.229	-0.042	-0.271
1W1	—	—	0.225

transistors shift in different directions, increasing the flipping threshold voltage and decreasing the voltage on the bit line to be read, the retention time will increase. If the contrary is true however, the retention time will decrease. This mismatch is the cause of the slight increase in retention time that is observed. As was shown in Figure 5.3, an equal voltage increment when charging a capacitor results in different charging times for different starting voltages; i.e., the capacitor does not charge linearly. Therefore, the mismatch of the transistors causes unequal shifts of the retention time, effectively increasing it slightly.

Table 5.15 shows again the earlier observed increase in sensing delay caused by aging. The individual wear-out mechanisms show that HCI is the most important cause for this degradation. This is caused by the fact that HCI has a more severe impact on NMOS transistors which are the discharge components and thus are the bottleneck when it comes to sensing speed. Also, as explained before, the duty factor of the sense amplifier makes the HCI impact more pronounced than that of BTI. From the table it also becomes clear that the writing delay increases. Again this is mainly attributed to the HCI degradation with a small addition of the BTI.

Impact on Memory Cells, Reference Cells, Precharge Devices and Sense Amplifier

The effect of the combination of all subcomponents and both BTI and HCI results in the retention time development as shown in Figure 5.19. The delay and energy consumption changes are presented in Table 5.16. As was the case with the individual wear-out analysis, a summation of the previously described effects is seen. After longer aging times the retention time starts to increase and the spread of it increases as well. Next to that, as can be read from the table, the sensing delays increase by 9.665 % for the *0R0* operation and even

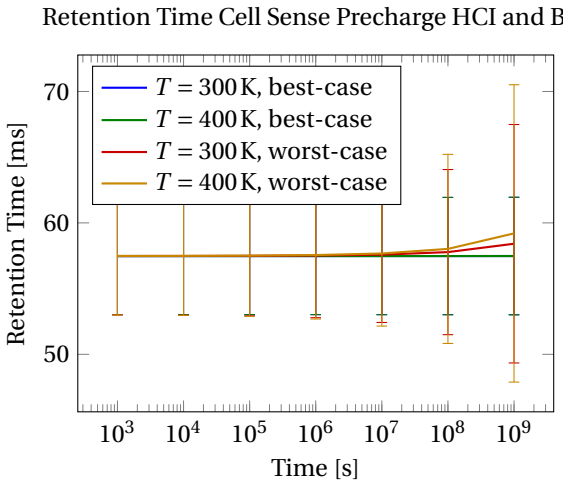


Figure 5.17: Cell Sense Precharge HCI+BTI

Table 5.15: Cell Sense Precharge BTI+HCI Worst-case workload, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	70.723	12.743	-0.615
1R1	102.744	14.783	-0.791
0W0	153.101	1.507	-0.623
0W1	-36.053	-0.090	-2.516
1W0	-54.424	-0.140	-2.558
1W1	—	—	-0.670

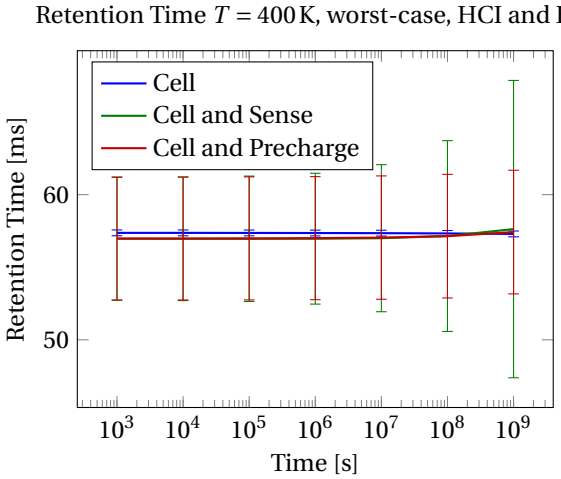


Figure 5.18: Individual impact of sense amplifier and precharge device on in combination with cell degradation HCI+BTI

by 18.773 % for a *1R1* operation. The difference between the two is caused by the fact that degradation of the reference cells decreases the *0R0* delay and increases the *1R1* delay. Degradation of the sense amplifier and precharge devices both increase these delay. The result of which is a net delay increase for both operations. From the table it also follows that the writing delay for a *0W0* operation increases as well, which was attributed to the memory cell transistor and sense amplifier degradation. The energy consumption decreases for all operations, again caused by the increase of the threshold voltage.

5.3.2. Bias Temperature Instability, Hot Carrier Injection and Radiation Effects

In this section the effect of both aging mechanisms and radiation trapping effects will be described.

Impact on Memory Cells

The effects on the memory cell when suffering both BTI, HCI and radiation effects are presented in Figure 5.20 and Table 5.17. The figure shows clearly that at higher ages total ionizing dose effects have more impact than BTI and HCI have. Also it becomes clear that with a 10 mm thick layer of shielding the radiation effects can be mitigated. In Table 5.17 the combination of the three effects is visible in the *0W0* writing delay. BTI and HCI have an increasing effect on this as was previously seen while radiation slightly reduces these effects.

Impact on Memory Cells and Reference Cells

When also the reference cells start to suffer from radiation effects, interesting behavior is shown by the circuit which can be seen in Figure 5.21. As was presented before, BTI and HCI increase the retention time of

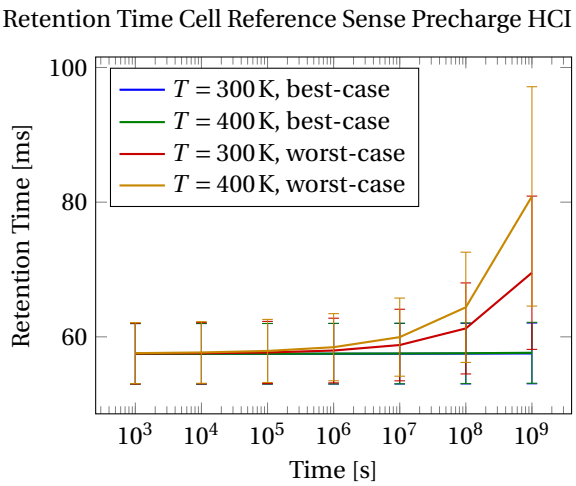


Figure 5.19: Cell Reference Sense Precharge HCI+BTI

Table 5.16: Cell Reference Sense Precharge BTI+HCI Worst-case workload, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	53.642	9.665	-0.933
1R1	130.472	18.773	-0.603
0W0	135.256	1.331	-0.945
0W1	-59.552	-0.148	-3.042
1W0	-73.660	-0.189	-2.835
1W1	—	—	-0.480

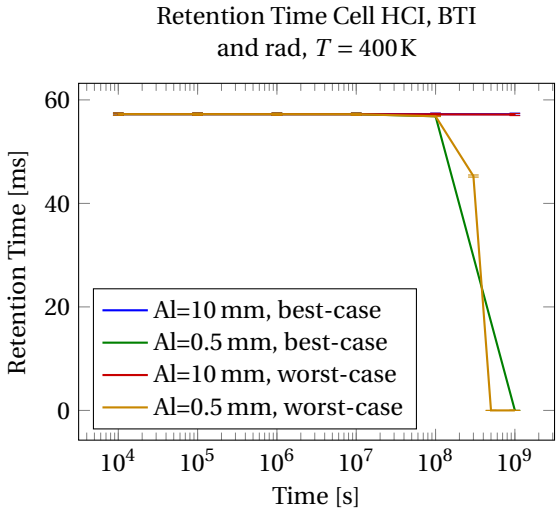


Figure 5.20: Cell HCI+BTI+Rad

Table 5.17: Cell BTI+HCI+Radiation Worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	0.281	0.051	0.002
1R1	-0.073	-0.011	0.021
0W0	91.751	0.903	-0.002
0W1	0.664	0.002	-0.001
1W0	3.556	0.009	-0.008
1W1	—	—	-0.005

the circuit and radiation will lower it at higher ages. The combination of these effects can be seen in the figure for the experiments where the aluminum shielding thickness was 0.5 mm and the worst-case workload is applied. The mean retention time first increases and suddenly starts to drop fast once TID = $D \cdot t = 1.26\,\text{mrad/s} \cdot 1 \cdot 10^8\,\text{s} = 126\,\text{krad}$ is accumulated. The data presented in Table 5.18 show a previously seen effect. The sensing delay for a *0R0* operation decreases while it increases for a *1R1* operation and the writing delay for a *0W0* operation increases as well.

Impact on Memory Cells, Precharge Devices and Sense Amplifier
Figure 5.22 and Table 5.19 present the results when the memory cell, precharge devices and sense amplifier are subjected to BTI, HCI and radiation effects. When the shielding layer is thick, again the slight increase in retention time manifests itself while thinner shielding clearly shows that radiation effects are the significant degradation factor. The table shows again that BTI and HCI have more impact on alterations in delay measurements and energy consumption than radiation effects have. This was expected from the individual contributions described previously.

Impact on Memory Cells, Reference Cells, Precharge Devices and Sense Amplifier
Finally the combination of all wear-out mechanisms applied to all subcomponents of the circuit can be evaluated. As was the case in previously seen analyses on all subcomponents, the effects are a combination of all subcomponents behaviors. From Figure 5.23, one can see the increase in spread of the retention time due

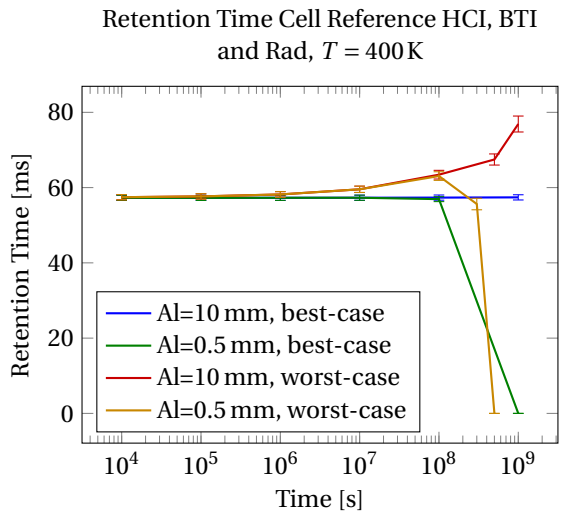


Figure 5.21: Cell Reference HCI+BTI+Rad

Table 5.18: Cell Reference BTI+HCI+Radiation Worst-case workload, $D = 1.26\,mrad/s$, $t = 1 \cdot 10^8\,s$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-13.295	-2.396	-0.318
1R1	21.251	3.058	0.200
0W0	77.380	0.761	-0.279
0W1	-34.742	-0.086	-0.632
1W0	-16.072	-0.041	-0.272
1W1	—	—	0.226

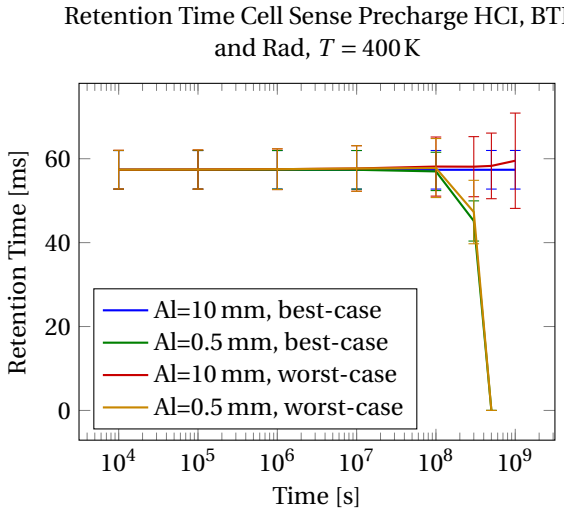


Figure 5.22: Cell Sense Precharge HCI+BTI+Rad

Table 5.19: Cell Sense Precharge BTI+HCI+Radiation Worst-case workload, $D = 1.26\,mrad/s$, $t = 1 \cdot 10^8\,s$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	70.091	12.629	-0.619
1R1	103.092	14.834	-0.780
0W0	157.078	1.546	-0.625
0W1	-36.695	-0.091	-2.512
1W0	-53.903	-0.138	-2.543
1W1	—	—	-0.658

to sense amplifier aging, the increase in retention time due to the aging of the reference cells as well as the decrease in retention time when higher total ionizing doses are accumulated. From Table 5.20 it follows again that the reference cells reduce the increase of the *0R0* sensing delay caused by aging of the sense amplifier while the *1R1* delay increases more. Also all operations consume less energy because the threshold voltages of the transistors increase.

5.4. Impact of Transistor Scaling

The effects of transistor scaling are presented in this section. As was mentioned in Section 4.4.3, the scaling was applied to all components in the circuit while the wear-out mechanisms are applied to the memory and reference cells, the local precharge devices and the local sense amplifier.

5.4.1. Bias Temperature Instability

In Figure 5.24 the retention time development for the downscaled circuit suffering from BTI is presented. As was the case in the unscaled circuit, the retention time increases over time which is again caused by the increase in reference voltage. It can also be seen that workload has more impact on the retention time degradation than temperature has, although a slight increase in retention time can be observed in the best-case workload at higher temperatures. Table 5.21 presents the variations in delays and energy consumption. As

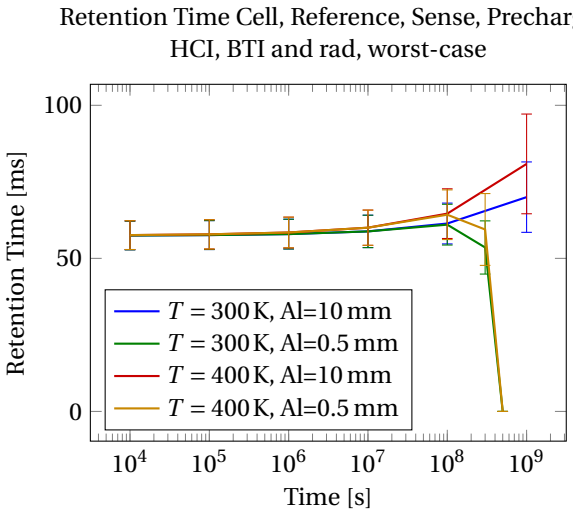


Figure 5.23: Cell, Reference, Sense, Precharge HCI+BTI+Rad

Table 5.20: Cell Reference Sense Precharge BTI+HCI+Radiation Worst-case workload, $D = 1.26\,\mathrm{mrad/s}$, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	52.912	9.534	-0.940
1R1	130.975	18.845	-0.583
0W0	139.072	1.369	-0.956
0W1	-60.308	-0.150	-3.043
1W0	-73.426	-0.189	-2.822
1W1	—	—	-0.466

was the case with the unscaled circuit, the *0R0* sensing delay decreases while the *1R1* delay increases. Also it can be seen from the table that the energy consumption decreases for all operations. This is caused by the increase in threshold voltage of the transistors which, as explained in Section 3.2, reduces the leakage currents.

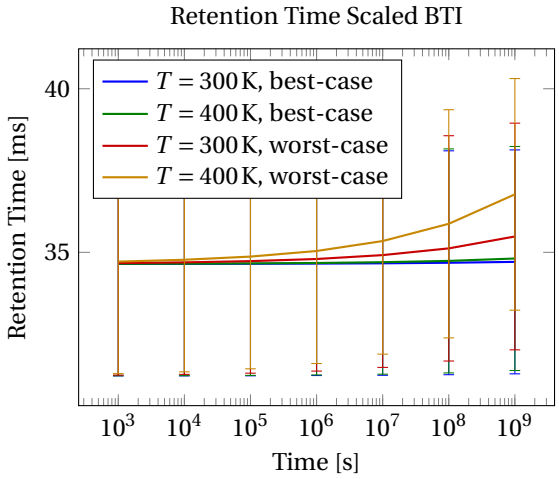


Figure 5.24: Scaled BTI

Table 5.21: Scaling BTI Worst-case workload, $t = 1 \cdot 10^8\,\mathrm{s}$

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	-0.881	-0.174	-0.456
1R1	4.994	0.857	-0.440
0W0	2.114	0.034	-0.457
0W1	13.379	0.035	-0.700
1W0	-16.891	-0.045	-0.556
1W1	—	—	-0.452

5.4.2. Hot Carrier Injection

The larger impact of HCI compared to BTI due to its high duty factor can be seen in Figure 5.25 and Table 5.22. As was expected the spread in retention time increases and the mean value increases as well with increasing aging time. From the table it becomes directly clear that the downscaled circuit is more susceptible to HCI than the unscaled one, the sensing delays increase by 21.522 % for a *0R0* operation and 31.653 % for an *1R1* operation and the *0W0* increases by 2.176 %. Again, the energy consumption of all operations decreases due to the increase in transistor threshold voltage.

5.4.3. Radiation Effects

Figure 5.26 shows the mean retention time and its standard deviation when the circuit suffers from radiation only. There are no data in this plot after $t = 6 \cdot 10^7\,\mathrm{s}$ because the circuit fails reading the cells when the total dose is more than $D \cdot t = 1.26\,\mathrm{mrad/s} \cdot 6 \cdot 10^7\,\mathrm{s} = 75.6\,\mathrm{krad}$. Also it again becomes visible that a thicker layer of shielding reduces the effects of space radiation on the DRAM circuit significantly. Table 5.23 presents

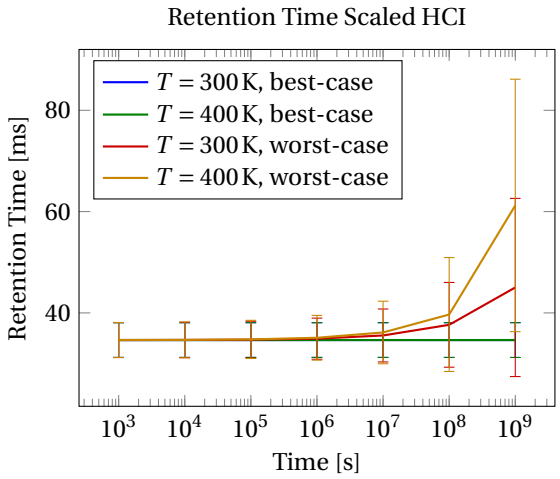


Figure 5.25: Scaled HCI

Table 5.22: Scaling HCI Worst-case workload, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	108.790	21.522	-0.901
1R1	184.499	31.653	-0.883
0W0	133.887	2.176	-0.902
0W1	-58.991	-0.153	-1.606
1W0	-46.663	-0.123	-1.455
1W1	—	—	-0.897

the delay and energy consumption metrics. Note that the samples in this table are taken at $t = 6 \cdot 10^7$ s as the circuit fails after longer aging times. From the table it becomes clear that the sensing delay for a *1R1* operation decreases slightly. Next to that, almost no metrics have changed, which could be attributed to the shorter aging time.

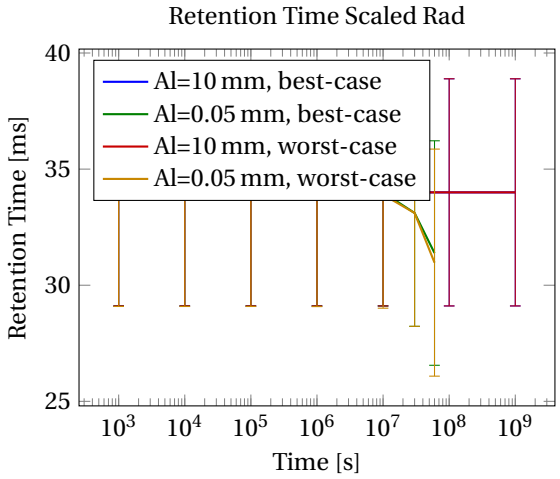


Figure 5.26: Scaled Rad

Table 5.23: Scaling Rad Worst-case workload, $D = 1.26$ mrad/s, $t = 6 \cdot 10^7$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	0.442	0.088	-0.045
1R1	-1.668	-0.286	-0.063
0W0	-1.984	-0.032	-0.045
0W1	-0.480	-0.001	-0.048
1W0	-0.176	-0.000	-0.058
1W1	—	—	-0.063

5.4.4. Bias Temperature Instability and Hot Carrier Injection

The combination of BTI and HCI applied to the downscaled circuit results in the plots in Figure 5.27 and the data in Table 5.24. The retention time increases significantly under the worst-case workload and the sensing delays and *0W0* writing delay do as well. All four metrics vary slightly more from their original value under this combined stress than under HCI stress only, thus suggesting only a small impact of BTI. The energy consumption also decreases more than was the case with either BTI or HCI.

5.4.5. Bias Temperature Instability, Hot Carrier Injection and Radiation Effects

Finally, the effects of all wear-out mechanisms on the downscaled circuit are presented in Figure 5.28 and Table 5.25. From the figure it becomes clear that BTI and HCI increase the retention time and that radiation effects start to reduce it after $t = 1 \cdot 10^7$ s. Note that the circuit fails to operate correctly when the TID becomes more than 126 krad and thus no data can be presented. The table shows the effect where the *0R0* sensing delay increases less than the *1R1* sensing delay which is caused by the combination of reference cell aging and sense amplifier aging. From the table it also becomes clear that the *0W0* writing delay increases, but not

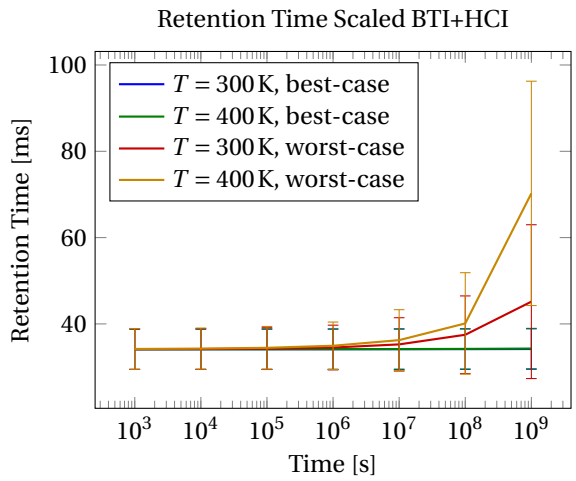


Figure 5.27: Scaled BTI+HCI

Table 5.24: Scaling BTI+HCI Worst-case workload, $t = 1 \cdot 10^8$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	112.048	22.167	-1.076
1R1	191.762	32.899	-1.070
0W0	152.705	2.482	-1.078
0W1	-40.349	-0.105	-2.071
1W0	-65.626	-0.173	-1.782
1W1	—	—	-1.092

as much as it would without radiation effects. Next to that, the overall energy consumption is reduced.

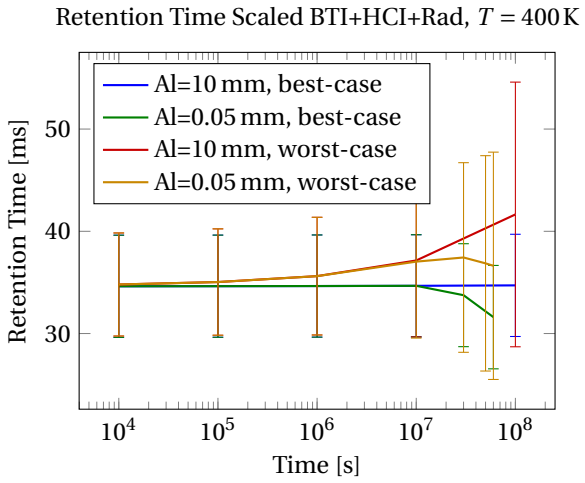


Figure 5.28: Scaled BTI+HCI+Rad

Table 5.25: Scaling BTI+HCI+Rad Worst-case workload, $D = 1.26\,\text{mrad/s}$, $t = 6 \cdot 10^7$ s

Operation	Delay [ps]	Delay [%]	Energy [%]
0R0	80.944	16.013	-0.996
1R1	143.937	24.694	-0.961
0W0	106.806	1.736	-0.995
0W1	-33.290	-0.087	-1.765
1W0	-51.314	-0.135	-1.520
1W1	—	—	-0.979

5.5. Results Summary

This section summarizes the findings of the previous three sections.

It is found that threshold voltage shifts, caused by BTI and HCI, slightly decrease the retention time of a memory cell. This is caused by the increased minimum cell voltage that is written to a cell. In contrast, aging of the reference cells increases the retention time, as the reference voltage of the circuit increases. The voltage at which the sense amplifiers will read a logical '0' is increased, hence increasing the retention time. BTI and HCI affect the sense amplifier as well. An increase of threshold voltage will slow down the sensing operation. Furthermore, the statistical properties of the aging effects increase the mismatch between the transistors in the sense amplifier and therefore increase the standard deviation of the retention time. This mismatch then increases the retention time of the circuit slightly. The impact of BTI and HCI depends not only on the transistor type that is affected (NMOS or PMOS) but also on the duty factor of the transistor. Equally typed transistors with a low duty factor suffer more from BTI wear-out than from HCI, while transistors with a higher duty factor suffer more from HCI. These effects manifests themselves in the DRAM as well. The memory cells have a lower duty factor than column devices like the sense amplifier and the reference cells, and suffer more from BTI. The column devices, however suffer more from HCI effects.

Radiation trapping effects cause the retention time to drop when significant TID is accumulated. This

is caused by the increased leakage current that is generated by the trapping of these charges in the memory cells. When the TID increases even further, the leakage in the other circuit components breaks down the circuit. It becomes then impossible to read out a cell. From the results it follows that a TID of 126 krad starts to degrade the performance of the unscaled circuit and a TID of 37.8 krad degrades the performance of the downscaled circuit. The total amount of charge that is accumulated in the device depends strongly on the thickness of the shielding layer protecting it, the duration of ionization and the radiation environment.

The downscaled version of the circuit suffers from the same problems as the unscaled one. The observed wear-out is more severe, though. This is caused by the fact that the aging effects are more severe in smaller devices. The radiation trapping effects on other subcircuits than the memory cell render the circuit broken at even lower TID.

5.6. Towards a Prediction Model

Now that the circuit performance for all experiments, stressed subcomponents and failure mechanisms is known, a reliability prediction can be made. In Section 3.1 the basics of reliability engineering were introduced including the definition of the reliability function, $R(t) = 1 - F(t)$. In order to define the reliability of the circuit it must therefore first be decided when the circuit fails. In this research, the failure of different subcomponents is defined based on a shift in the metrics which were defined in Section 4.2. The maximum allowed shift in a metric is given by six times its standard deviation at $t = 0\,\text{s}$, or in short $\Delta X < 6\sigma_{t=0\,\text{s}}$, where X is the shifted parameter, e.g. the retention time. Note that this includes a shift of the metric in two directions, the metrics may increase and decrease. This might seem odd as for example a decrease in energy consumption generally is a favorable property for a circuit. However from the definition of reliability it becomes clear why this was chosen: *"Reliability is defined as the ability of a circuit to conform to its specifications over a specified period of time under specified conditions."* [1]. In order to be reliable, the circuit should conform to its specifications. Hence, a shift in either direction should result in failure of the circuit.

The reliability of the circuit can be predicted by multiple methods, as was described in Section 3.1. More extensive reliability models, which for example include redundancy of circuit parts or inter-component correlations, are not useful for this research as the DRAM circuit used in this work lacks among other an address decoder or control logic. It was therefore chosen to evaluate the reliability of the circuit using the method specified by the military handbook [4]. From the handbook it follows that the overall circuit reliability is defined by a weighted multiplication of all subcomponent reliabilities which might also be constructed from such a multiplication. A description of the system reliability in this fashion is presented in Equation 5.1 in which W_k denotes the weighting factor for every subcomponents' $k \in$ *subcomponents* reliability function R_k.

$$R_{sys}(t) = \prod_k W_k \cdot R_k(t) \tag{5.1}$$

The reliability of a subcomponent again depends on a multiplication of the reliability functions of different metrics as can be seen for the memory cell as an example in Equation 5.2.

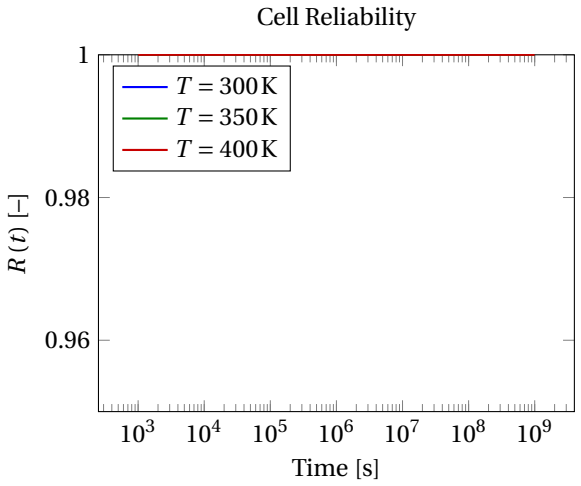
$$R_{cell}(t) = W_{tret} R_{tret}(t) \cdot W_{energy} R_{energy}(t) \cdot W_{delay} R_{delay}(t) \tag{5.2}$$

Again these reliability functions are expressed by multiplication of the reliability of for example the energy consumption per operation for the $R_{energy}(t)$ function as shown in Equation 5.3.

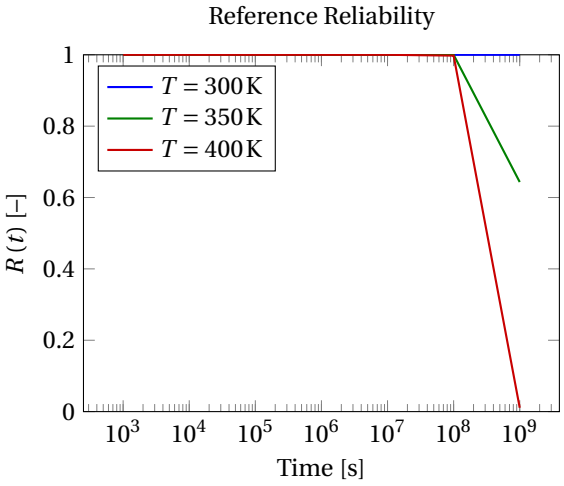
$$R_{energy}(t) = W_{0R0\ energy}(t)\, R_{0R0\ energy}(t) \cdot W_{1W1\ energy}(t)\, R_{1W1\ energy}(t) \cdot W_{0W1\ energy}(t)\, R_{0W1\ energy}(t) \ldots \tag{5.3}$$

This concept is applied to the results from the DRAM aging experiments. The system reliability is given by a equally weighted multiplication of all component reliabilities which are the result of an equally weighted multiplication of the reliabilities for every metric, i.e., $W_{tret} = W_{energy} = W_{delay}$ and $W_{cell} = W_{ref} \ldots$. In Figure 5.29 four reliability functions are plotted for BTI and HCI stress under the worst-case workload for a varying set of aged subcomponents. Figure 5.29d shows the reliability curve for the circuit with all components combined. It becomes clear that the increased delay due to the sense amplifier aging is a larger reliability degrading factor than the increase in retention time caused by the reference cell is. This makes the sense amplifier the most critical part for an aging DRAM. The combination of all subcomponents again results in the highest circuit degradation. From the figures it also becomes clear that temperature has a large influence on the reliability of the circuit; at $T = 300\,\text{K}$ and $t = 1 \cdot 10^8\,\text{s}$ under worst-case workload 84 % of the complete circuit population would still survive while at $T = 400\,\text{K}$ the entire population would have failed already after $t = 1 \cdot 10^7\,\text{s}$.

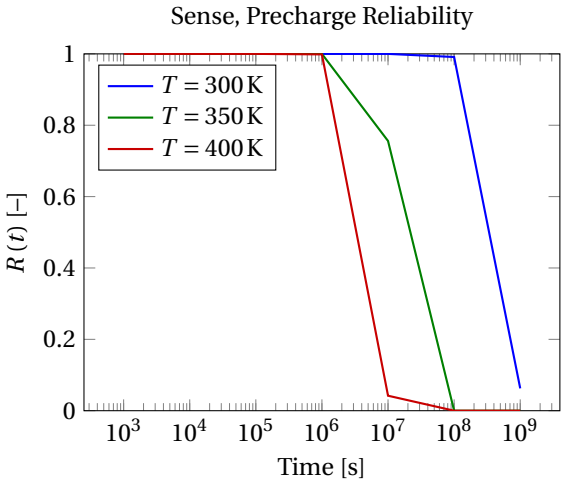
In order to make an accurate reliability prediction for a complete DRAM circuit description, the procedure that was used above can be followed. Based on the relevant metrics, the reliability per subcomponent can be derived which then will be used to calculate the system reliability. It is also important to note that all subcomponents have an equal weighting factor in this prediction. In a complete DRAM circuit the cells might have a higher weighting factor as there are much more of these than column devices like the sense amplifier or reference cells.



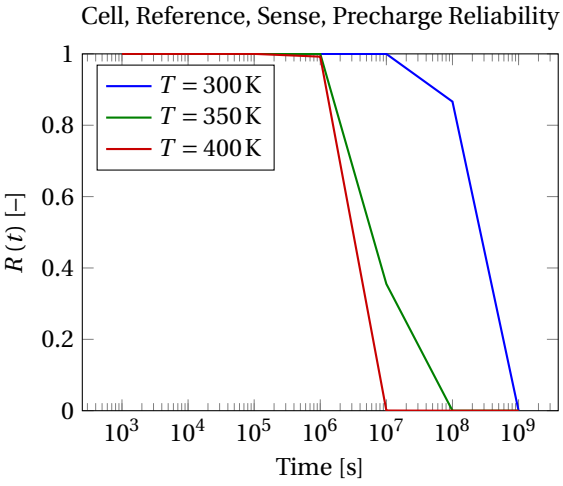
(a) Cell Reliability, Worst-case workload, BTI and HCI



(b) Reference Reliability, Worst-case workload, BTI and HCI



(c) Sense Precharge Reliability, Worst-case workload, BTI and HCI



(d) Cell, Reference, Sense, Precharge Reliability, Worst-case workload, BTI and HCI

Figure 5.29: Reliability functions under BTI and HCI stress under worst-case workload

In Section 3.4.3 and Table 3.6, the ideal reliability prediction method was presented. For ease of readability, Table 3.6 is copied below as Table 5.26. The reliability prediction model presented above, has three important features that other reliability prediction models lack. These properties are: dynamic simulation level, prediction of lifetime and reliability and the ability to evaluate reliability for different metrics per subcomponent.

Dynamic simulation level can be achieved by deriving a reliability function for a subcircuit, metric or a wear-out mechanism. This reliability function can then later be used in the derivation of the system reliability as described in Equation 5.1. The reliability for a subcomponent now only needs to be determined once and can be used in multiple other reliability investigations. This can greatly reduce simulation time for a reliability prediction. The derivation of the reliability functions allows to include different metrics or different weighting functions for the metrics to be used. Hence, the circuit reliability can be determined by the best suited metrics

for every subcomponent. Predicting both reliability and lifetime is possible with this method as the circuit operation is investigated before a reliability function can be made. In this investigation exact lifetimes of the circuit can be found.

Table 5.26: Detailed overview of aging and reliability prediction methods

Name	Netlist	Level	Failure mechanism	Workload aware	Fault dependencies	Prediction Outcome	Failure Metric
BERT [12]	Yes	Transistor	HCI, EM, TDDB	Nodal waveforms; No temperature for HCI	No, only individual failures	Reliability; Lifetime	TDDB: breakdown probability for points in time single devices; EM: time-to-failure for all wires; HCI: degradation of circuit properties
RAMP v1 [5]	No	System	EM, TDDB	Switching activity partially taken into account. Temperature included	Weighted sum; failures independent	Reliability	
RAMP v2 [13]	No	System	EM, TDDB, NBTI	Switching activity partially taken into account. Temperature included	Weighted sum; failures independent	Reliability	MTTF for components in system and complete system
MaCRO [14]	Yes	Transistor	HCI, TDDB, NBTI	Lifetime estimation: average voltages and temperature included; Functionality: waveforms and aged models	Weighted sum; failures independent	Reliability	Degradation of circuit properties resulting in loss of circuit functionality
AgeGate [15]	Yes	Gate	HCI, NBTI	Duty cycle and switching activity; Temperature included	Weighted sum; failures independent	Lifetime	Critical path delay
RSM [16]	Yes	Transistor	HCI, TDDB, BTI	Waveforms and Temperature included	Faults combined in one transistor model; failures independent	Reliability	Circuit parameter like gain, output current, etc.
FaRBS [17]	Yes	Transistor	HCI, TDDB, NBTI, EM	Samples over waveforms. Temperature included.	Weighted sum; failures independent	Lifetime	MTTF for the complete system
SyRA [18]	Yes	Transistor, Gate	HCI, NBTI	Digital: duty cycle; AMS: only generalized AC/DC waveforms; No temperature	Combination unclear; Failures independent	Lifetime	Digital: Path delay; AMS: degradation of circuit properties
Ideal	Yes	Dynamic	All	Actual voltages and currents; temperature included	Yes, dependent failures	Reliability & Lifetime	Different metrics per subcomponent, e.g. Cell: retention time, capacitor voltage; Sense amplifier: detection limit, sensing delay;

6 Conclusion

This chapter rounds off this thesis. First a summary of this thesis will be presented in Section 6.1. A general conclusion will be drawn in Section 6.2 which is followed by a discussion of the limitations of the work in Section 6.3. This chapter finishes with some future work recommendations in Section 6.4.

6.1. Summary

This section will summarize the contents of all previous chapters.

6.1.1. Chapter 2 – DRAM Design and Technology

This chapter introduces computer memory and more specifically DRAM, to the reader. It starts with an introduction of computer memory and its hierarchy. After this, DRAM is introduced in more detail. The differences between the classifications and the resulting differences in architecture are presented. The chapter continues with a description and circuit representation of the subcomponents of DRAM: the memory cell, the precharge circuit, the sense amplifier, reference cells and the memory array. Then, three operations on the DRAM, reading, writing and refreshing, are presented and the corresponding timing diagrams are presented as well. The chapter is finalized with an overview of the technology involved in making a DRAM starting with an introduction of general IC fabrication technology and finishing with the manufacturing technology needed to make the DRAM capacitor as this is an extraordinary structure in comparison with regular IC design.

6.1.2. Chapter 3 – DRAM Reliability

IC and DRAM reliability modeling is introduced in this chapter. Next to that, also the circuit description and the aging models are presented. The reader is presented with the terminology and mathematics related to reliability engineering first. Terms like reliability, failure or hazard rate and mean-time-to-failure are introduced. The reader is also presented with the reliability bathtub curve and the three periods defining it. After this, CMOS wear-out mechanisms are introduced starting with a general introduction of the MOSFET. The background knowledge of MOSFET operation is used to describe the wear-out mechanisms bias temperature instability (BTI), hot carrier injection (HCI), radiation, time dependent dielectric breakdown (TDDB), and electromigration (EM). Then DRAM specific failures, soft and hard faults, are introduced. Now that the background knowledge has become clear to the reader, the model description used in this research is presented. The investigated circuit is a commercial DRAM manufactured in the 1990s. The memory cell, reference cells, sense amplifiers and precharge devices that constitute this circuit are presented in detail. Next, two workloads are defined based on a reference chip. The first workload is the best-case in which the cells are stressed only due to refreshing of the cell contents. The second workload is the worst-case workload which is based on the highest number of L2-cache misses in a benchmark analysis. A comparison of different aging methods and models from literature is then presented. Based on this comparison, the wear-out models used in this research are developed. Transistor variations are modeled using Pelgrom's model. BTI is modeled with the RD-model and Hot carrier injection is implemented as the lucky electron model. Statistical spread was introduced to both BTI and HCI and the parameters for this were experimentally verified. The calibration of both was done after measurements from literature. Both BTI, HCI and the transistor mismatch properties are included as threshold voltage shifts in the transistor model. The radiation effects on the circuit are im-

plemented as a gate-induced-drain leakage current source. This current was derived from irradiated DRAM retention time measurements. The model shows an increase in leakage current for higher accumulated ionizing doses. Next to this leakage based model, a model based on charge trapping was implemented as well. Due to the lack of measurement data, this model was not used in this research. The space radiation environment of the Sentinel-1 mission was presented and is used as the source for the radiation experiments.

6.1.3. Chapter 4 – Framework

The simulation framework is described in this chapter. First, the software that is used in this research is presented. Then, the five metrics that are used to determine the DRAM degradation are presented next. These are: retention time, internal cell voltage, sensing and writing delay and energy consumption of the circuit. Graphs illustrate the measurement descriptions where necessary. Next, the operating conditions of the DRAM are defined. The workload is set to the best-case and the worst-case. The circuit is stressed on three different temperatures, 300 K, 350 K and 400 K, and its performance is verified on one temperature: 300 K. Three experiments are described after this. First, the impact of the wear-out mechanisms individually is examined, then the combination of them. In the third experiment, a downscaled version of the circuit is subjected to individual and combined stresses. The chapter concludes with an overview of the initial circuit performance for the previously presented metrics.

6.1.4. Chapter 5 – Experimental Results and Reliability Prediction Model

The results from the experiments defined in the previous chapter are presented and discussed in this chapter. First the individual impact of wear-out mechanisms is presented. It is found that the wear-out factor that has the most influence on the circuit degradation is determined by the duty factor of the transistor. The memory cells that have a relative low workload, suffer more from BTI effects while the column devices, which have a higher workload then the memory cells, suffer more from HCI. The increase in threshold voltage caused by BTI and HCI reduces the retention time when applied to cells only because the minimal voltage written to the cell increases due to the increase of the threshold voltage. Aging increases the reference voltage. This increases the threshold voltage where a logical ‘1’ is read, therefore increasing the retention time. Degradation of the sense amplifier and the precharge devices results in a slight increase in retention time as the mismatch between the transistors in the sense amplifier is increased. Next to that, the sensing delays increase significantly due to their threshold voltage increase. Application of radiation wear-out to the circuit, results in a decrease of retention time for higher total ionizing dose (TID). The leakage currents caused by these high doses become so large that storing data in the memory cells becomes impossible. If the TID becomes larger than 126 krad, the circuit is unable to perform a successful read operation and thus fails completely.

Combination of the wear-out mechanisms results in a combination of the individual effects. Combination of BTI and HCI effects results in more degradation of the retention time when only the memory cell is stressed, but is mainly caused by the BTI degradation. The reference voltage also increases slightly more with the combination of both of them, resulting in increased retention times. The inclusion of radiation again decreases retention time when high doses are accumulated. Combining BTI, HCI and radiation effects results in an increase of retention time for low ages while it decreases at higher ages. The radiation shielding thickness and the workload of the DRAM, are important factors in determining the overall degradation observed.

The effects seen in the previous two experiments manifest themselves again in the downscaled circuit. Downscaling of the circuit results in a degradation of overall retention time due to the increased internal leakage currents. The reference voltage increases due to aging, increasing the retention time with it. The degradation of the sense amplifier also increases the sensing delays as was seen in the unscaled circuit. Also the effects of radiation seen on the circuit are equal, retention time drops with higher TID accumulation.

The chapter concludes with a first-order reliability prediction model. The system reliability is described as the multiplication of the component reliability which again is a multiplication of the reliability per experimental metric. It becomes clear that the increase in sensing delay is a more important factor when evaluating the reliability of the DRAM than the increase of reference voltage is, which thus points to the sense amplifier as the most critical part in DRAM reliability. Furthermore, a comparison with state-of-the-art reliability prediction methods and this work was made. The ability to predict reliability on a dynamic level as well as the prediction based on varying metrics is new and proves useful in the reliability analysis of DRAM.

6.2. Conclusion

This thesis presented a framework to analyze DRAM aging and presented a reliability prediction model. It was shown that BTI and HCI wear-out can increase the overall retention time of a DRAM chip due to the increase of the reference voltage. Next to the retention time, the sensing delays increase as well which lead to timing violations if they increase too much and thus reduce reliability. Radiation trapping effects proved to be a severe degrading factor for DRAMs because of the cell leakage current increase it generates. Overall degradation of the circuit is worse under increasing aging temperatures and with higher workloads of the chip. Based on the first-order reliability prediction model, it can be concluded that the sense amplifier is the most vulnerable component of this DRAM as it has the highest impact on the system reliability degradation.

6.3. Limitations

The results presented in this research are limited by the following aspects.

- **Circuit Model**
  The circuit used in this research was designed and produced in the 1990s. More modern circuits might be using different techniques to reduce for example leakage or speed up overall operation. It is important to remember that the results in this research are relevant to this old design but not necessarily to newer ones. The methodology however, remains valid for newer and older circuits.

  Downscaling of the circuit was done manually as no other sources were available. The involved scaling was performed using standard CMOS scaling techniques and assumptions. Since DRAM transistors are designed to have a low leakage characteristic this scaling method might not be completely suited for this circuit.

- **Aging Models**
  Since a fitting function was used to include duty factor into the BTI model, an error may arise from this for low duty factors, which are present in the memory cells. For these low duty factors it might be beneficial to include either measurements of these low duty factors into the curve fitting or use a different model to determine the impact of these low duty factors. That being said, from the results it follows that both BTI and HCI have an impact on the retention time of the memory but are negligible when also other subcomponents are experiencing aging as well. This would suggest that the small error that might exist from the duty factor fitting is small and can be neglected.

  The HCI model was calibrated with averaged peak current measurements on the unaged circuit. These peak currents might result in an overestimation of the actual HCI degradation. The model should be calibrated based on actual transistor measurements to guarantee a good prediction.

  The model used to simulate the impact of radiation on the circuit was based on secondary DRAM retention time measurements and not verified by actual data. Therefore, important characteristics for different kinds and sizes of transistors could not be derived properly, which might present a source of error.

- **Operating Conditions** The workload of the DRAM and hence the duty factors of the subcomponents that arise from this, are subjected to some assumptions as well. The best-case workload, in which only refresh operations form the DRAM workload, of the chip might be too optimistic in real-world applications. Next to that, the worst-case scenario assumes that a block with a size of 32 kB is replaced on a cache miss. The total size of this block strongly impacts the worst-case workload experienced by the circuit and might overestimate the impact.

- **Reliability Prediction Model**
  The circuit reliability is predicted by a first order prediction model which does not include correlation effects of the different subcomponents. Next to that, due to the lack of a complete DRAM circuit, reliability prediction of a complete chip could not be made.

6.4. Future Work

The work presented in this thesis allows for the analysis of DRAM wear-out due to aging. The research could be improved and extended as described below.

- **Circuit Model**
  The methods presented in this research could be applied to a more modern DRAM circuit and transistor library to find out how it would perform under aging. Also the addition of missing components, e.g. address decoders or control logic, to the DRAM circuit would improve the accuracy of the final reliability prediction.

- **Aging Models**
  When this research is applied to modern devices, the impact of BTI for low duty factors as well as the impact of radiation in the models should be verified and set to the according technology constraints. It could also prove useful to measure the trap forming in transistors caused by various wear-out mechanisms and add all these effects in a trap based model as proposed in [85].

- **Reliability Prediction Model**
  The reliability prediction model in this work could be used as a basis for a more advanced prediction model. The model should be extended with reliability functions of the missing circuit components and appropriate weighting functions. Further improvement could be achieved by including correlations of the components and wear-out mechanisms.

- **Extension**
  This research could also be extended in the depth of its analysis. Future work should include variations and wear-out of the cell capacitor like cell capacitance mismatch, imperfections that increase leakage current or decrease charging current and dielectric breakdown. TDDB should also be included in the transistor aging model to estimate the impact of its increased leakage current. Furthermore, if a complete design becomes available, an analysis of EM degradation should be included as well.

A

Radiation GIDL Verilog-A Model

Below the two Verilog-A modules for NMOS and PMOS transistors respectively are presented. The modules are written in Verilog-A and were used by the Virtuoso Spectre circuit simulator, version 14.1.0 64bit [93].

A.1. Radiation GIDL Verilog-A Model - NMOS

```
1  // Implementation of additional GIDL caused by radiation trapping
2
3  `include "disciplines.h"
4  `include "constants.h"
5
6  module radiation_gidl_nmos(d, g, s, b);
7
8          inout       d, g, s, b;
9          electrical d, g, s, b;
10
11         // Alterable parameters
12         parameter real W = 600e-9; // m
13         parameter real L = 450e-9; // m
14         parameter real Tox = 8e-9; // m
15         parameter real NF = 1; // number of fingers (finFet)
16
17         parameter real krad = 0; // Amount of trapped radiation in krad
18
19         parameter real AGIDL = 0.0002;
20         parameter real BGIDL = 2.1e9;
21         parameter real CGIDL = 0.0002;
22         parameter real EGIDL = 0.8;
23
24         // Add an multiplier based on DF to deal with bias conditions during
               irradiation
25         parameter real DF = 1;
26         parameter real ON_RAD = 250;
27         parameter real OFF_RAD = 100;
28
29         real multiplier = ((ON_RAD-OFF_RAD)*DF+OFF_RAD) / ON_RAD;
30
31         // Transistor parameters
32         real TOXE = Tox;
33         real WL = 0;
34         real WLC = WL;
```

```
35          real WINT = 0;
36          real DWC = WINT;
37          real DWJ = DWC;
38          real XW = 0;
39          real Wdrawn = W;
40          real WLN = 1;
41          real WWN = 1;
42          real WWL = 0;
43          real WWLC = WWL;
44          real WWC = WWL;
45          real WeffCJ = Wdrawn/NF+XW -2*(DWJ+WLC/(L**WLN)+WWC/(W**WWN)+WWLC/(L**WLN *
                W**WWN));
46
47          real rad_factor;
48          real Igidl;
49
50          analog begin
51
52                  @(initial_step)begin
53                          rad_factor = multiplier * 0.050 * $exp(krad * 0.1315); //
                                derived from fitting script
54                  end
55
56                  if((V(g,s) < 1.0) && (V(d,s) > (V(g,s)+EGIDL))) begin
57                          Igidl = rad_factor * AGIDL * WeffCJ * NF * (V(d,s)-V(g,s)-
                                EGIDL)/(3*TOXE) * $exp(-(3*TOXE*BGIDL)/(V(d,s)-V(g,s)-
                                EGIDL)) * (V(d,b)**3)/(CGIDL+V(d,b)**3) + 1e-20; //Added
                                1e-20 to prevent too small currents
58                          I(d,b) <+ -abs(Igidl);
59                          I(b,s) <+ 1e-20;
60                          end
61                  else if((V(g,d) < 1.0) && (V(s,d) > (V(g,d)+EGIDL))) begin
62                          Igidl = rad_factor * AGIDL * WeffCJ * NF * (-V(d,s)-V(g,d)-
                                EGIDL)/(3*TOXE) * $exp(-(3*TOXE*BGIDL)/(-V(d,s)-V(g,d)-
                                EGIDL)) * (V(b,s)**3)/(CGIDL+V(b,s)**3) + 1e-20; //Added
                                1e-20 to prevent too small currents
63                          I(s,b) <+ -abs(Igidl);
64                          I(d,b) <+ 1e-20;
65                          end
66                  else begin
67                          Igidl = 1e-20;
68                          I(d,b) <+ -Igidl;
69                          I(s,b) <+ -Igidl;
70                  end
71          end
72  endmodule
```

A.2. Radiation GIDL Verilog-A Model - PMOS

```
1   // Implementation of additional GISL caused by radiation trapping
2
3   `include "disciplines.h"
4   `include "constants.h"
5
6   module radiation_gidl_pmos(d, g, s, b);
7
8           inout       d, g, s, b;
```

```
9           electrical d, g, s, b;
10
11          // Alterable parameters
12          parameter real W = 600e-9; // m
13          parameter real L = 450e-9; // m
14          parameter real Tox = 8e-9; // m
15          parameter real NF = 1; // number of fingers (finFet)
16
17          parameter real krad = 0; // Amount of trapped radiation in krad
18
19          parameter real AGIDL = 0.0002;
20          parameter real BGIDL = 2.1e9;
21          parameter real CGIDL = 0.0002;
22          parameter real EGIDL = 0.8;
23
24          // Add an multiplier based on DF to deal with bias conditions during
                irradiation
25          parameter real DF = 1;
26          parameter real ON_RAD = 400;
27          parameter real OFF_RAD = 350;
28
29          real multiplier = ((ON_RAD-OFF_RAD)*DF+OFF_RAD) / ON_RAD;
30
31          // Transistor parameters
32          real TOXE = Tox;
33          real WL = 0;
34          real WLC = WL;
35          real WINT = 0;
36          real DWC = WINT;
37          real DWJ = DWC;
38          real XW = 0;
39          real Wdrawn = W;
40          real WLN = 1;
41          real WWN = 1;
42          real WWL = 0;
43          real WWLC = WWL;
44          real WWC = WWL;
45          real WeffCJ = Wdrawn/NF+XW -2*(DWJ+WLC/(L**WLN)+WWC/(W**WWN)+WWLC/(L**WLN *
                W**WWN));
46
47          real rad_factor;
48          real Igidl;
49
50          analog begin
51
52                  @(initial_step)begin
53                          rad_factor = multiplier * 0.050e7 * $exp(krad * 0.1315); //
                                derived from fitting script
54                  end
55
56
57                  if((V(s,g) < 1.0) && (V(s,d) > (V(s,g)+EGIDL))) begin
58                          Igidl = abs(rad_factor * AGIDL * WeffCJ * NF * (V(s,d)-V(s,g
                                )-EGIDL)/(3*TOXE) * $exp(-(3*TOXE*BGIDL)/(V(s,d)-V(s,g)-
                                EGIDL)) * (V(b,d)**3)/(CGIDL+V(b,d)**3) + 1e-20); //Added
                                 1e-20 to prevent too small currents
```

```
59                          I(d,b) <+ -Igidl;
60                          I(b,s) <+ -1e-20;
61                          end
62                  else if((V(d,g) < 1.0) && (V(d,s) > (V(d,g)+EGIDL))) begin
63                          Igidl = abs(rad_factor * AGIDL * WeffCJ * NF * (-V(s,d)-V(d,
                                g)-EGIDL)/(3*TOXE) * $exp(-(3*TOXE*BGIDL)/(-V(s,d)-V(d,g)
                                -EGIDL)) * (V(s,b)**3)/(CGIDL+V(s,b)**3) + 1e-20); //
                                Added 1e-20 to prevent too small currents
64                          I(b,s) <+ Igidl;
65                          I(d,b) <+ -1e-20;
66                          end
67                  else begin
68                          Igidl = -1e-20;
69                          I(d,b) <+ Igidl;
70                          I(b,s) <+ Igidl;
71                  end
72          end
73  endmodule
```

B

Ionizing Radiation Modeling in DRAM Transistors

This Appendix contains the paper submitted to the 2018 IEEE Latin-American Test Symposium (LATS2018): M. Fieback, M. Taouil, S. Hamdioui and M. Rovatti, "Ionizing Radiation Modeling in DRAM", in *2018 IEEE Latin-American Test Symposium*

Ionizing Radiation Modeling in DRAM Transistors

Moritz Fieback Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Mekelweg 4, 2628CD Delft, The Netherlands

Marco Rovatti
European Space Agency (ESA)
European Space Technology Center ESTEC
Noordwijk, The Netherlands

***Abstract*—Electronics in space suffer from increased wear-out due to the accumulation of high concentrations of ionizing dose. The costs of a space mission in combination with the harsh space environment force space agencies to demand electronic components with extreme high reliability to guarantee mission success. One of the main reliability concerns for DRAM is the retention time degradation due to radiation, as radiation increases the Gate Induced Drain Leakage (GIDL). In this work we present a methodology to develop a Spice-based radiation model that could be used to simulate this retention time degradation. The model estimates the GIDL based on existing silicon measurements of the retention time and gives designers the opportunity to measure the impact of radiation during the design stage. Simulation results show a strong retention time degradation for small Total Ionizing Dose (TID) while this stabilizes with larger TID. The application of the model with space radiation environment data shows that the damage that spacecrafts suffer depends strongly on altitude and aging time.**

***Index Terms*—Retention Time, Radiation, DRAM, GIDL, Leakage**

I. Introduction

Radiation is a prime wear-out mechanism for IC's in space [1]. Ionized particles may get trapped in the transistors which leads to increased leakage currents or even oxide breakdown [1], [2]. The breakdown or leakage currents of one of these transistors may result in a complete system failure, rendering the spacecraft useless. This is especially true for DRAMs, where low leakage is vital for functional cells. The failure probability in this harsh space environment forces space agencies to demand an extremely high reliability for their components [3]. Increase in Total Ionizing Dose (TID) increases the off-current of a transistor and causes a threshold voltage shift [2], [4]. With the gate oxide becoming thinner, the impact of TID effects becomes less apparent in the transistor itself but is still an important parameter as the isolating structures surrounding it form parasitic channels that degrade the transistor performance [5], [6]. During the design phase, the designer may increase the reliability of a system. For example, a clever circuit design could result in hardened circuits to radiation effects [1], [7]. Another solution is mitigating charge trapping by having thick layers of shielding material. This is however an unwanted property due to its weight. To verify the design with respect to radiation impact, appropriate aging models are required. Lack of these models requires expensive and tedious test chips and structures that need to be measured under irradiation. A faster and cheaper characterization method during the design phase is therefore desired.

Several researchers have investigated the effects of radiation on transistor performance and develped models for it. Barnaby *et al.* presented a model that simulates the parasitic effects through calculations of the surface potential [7]. This physics based surface modeling approach was later extended to include the effects of traps generated by Bias Temperature Instability (BTI) [8]. The aforementioned physics based approach has the benefit that it can include the effects of charge trapping on the transistor performance very accurately. However, a disadvantage is that this high accuracy level requires very accurate and extensive measurements of transistors, making it a tedious task to develop these models. In addition, the extensive accuracy can result in long computation times. Li *et al.* proposed a simulation model that relies on body biasing to incorporate radiation effects in transistors [9]. This simpler model achieves reasonable accuracy when the transistor is on but is not accurate enough in the sub-threshold region. Therefore, this model is not suitable enough for low leakage designs like DRAMs. Some specific research is performed on the effects of radiation on DRAMs. Most of these works focus on Single Event Upsets (SEU) caused by radiation [10], [11] rather than on actual circuit damage caused by the radiation effects. In [12], the authors presented the effects of ionizing radiation on the retention time of a DRAM. They found that the ionizing radiation causes traps which increases the transistor's Gate Induced Drain Leakage (GIDL) current. The above clearly shows that there is a need for a model that allows for fast simulation of radiation impact on the retention time that is valid for sub-threshold currents.

This work proposes a novel model that allows to measure the impact of radiation on the retention time by creating a Spice model of GIDL. The model is based on indirect measurements of the retention time of the cells presented in [12]. From these secondary measurements, the GIDL current is estimated and included as a current source in the transistor model. The benefits of the proposed method is that is simple (i.e., a low computation complexity as it is not based on physical modeling) and that is does not require an extensive characterization based on test chips. The contributions of this paper are as follows.

- A novel radiation model based on existing silicon measurements.
- Verification of the model using Spice simulations.
- Case studies that show the impact of the altitude and

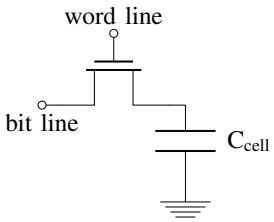


Fig. 1. DRAM Memory Cell

aging time on the retention time.

The remainder of this paper is organized as follows. Section II provides background information on the DRAM memory cell, GIDL and radiation trapping. Section III presents the derivation and implementation of the radiation trapping transistor model, as well as the experiments that are performed in this paper. Section IV shows the results of these experiments. Finally Section VI concludes this paper.

II. Background

In this section, background information will be provided on the DRAM memory cell, gate induced drain leakage and trap formation due to radiation will be given.

A. DRAM Memory Cell

A good description of DRAM architectures and their components is given in [13]. A short summary on the memory cell is given below. In a DRAM, data is stored in a memory array consisting of memory cells. The memory cells store the data as charge on capacitors. Figure 1 shows a typical DRAM memory cell. The capacitor can be accessed via a pass transistor that is connected to the word line and the bit line. The word lines are connected to the row address decoder and the bit lines to the sense amplifiers and precharge circuit. The data flows to and from the cell via the bit line.

The charge on the capacitor slowly leaks away. Hence, after a certain time the data written to the cell is lost. This time is called the retention time and should be as high as possible to prevent data loss in DRAMs. Therefore, cells have to be refreshed regularly to compensate for the charge loss due to leakage. When the leakage current becomes so high that the cell is unable to hold data for its designed retention time, the cell and thus the memory will fail.

B. Gate Induced Drain Leakage

Gate Induced Drain Leakage occurs in the overlapping area between the gate and the drain of a transistor [14]. If the gate of a NMOS device is biased at a low voltage, while the gate is at a high voltage, a deep-depletion region is formed in the overlapping region. For a PMOS device, this occurs when the gate is biased at high voltage while the drain is biased at a low voltage. Carriers can tunnel through the drain-to-bulk band gap and are then swept away [14]. This process is schematically presented in Figure 2 [14]. GIDL is an important leakage mechanism in DRAMs. Since the cells are accessed

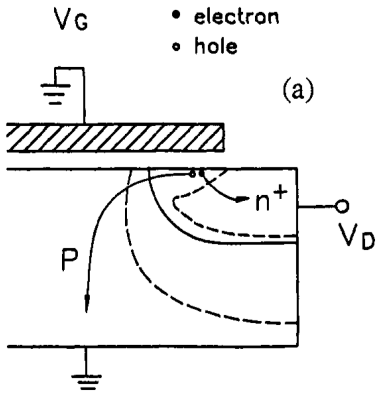


Fig. 2. Deep-depletion in the gate drain overlapping region causes GIDL [14]

infrequently, the charge stored on the capacitor creates a strong electric field between in the gate and drain; this gives rise to GIDL currents. Hence GIDL decreases the retention time of a memory cell.

When there are traps present in the gate drain overlapping region, GIDL current increases. The traps form energy states in the band gap that can be occupied by the carriers. From these states, the carriers may either continue to cross the band gap or fall back in their previous energy band. Since these energy states are within the band gap a carrier needs less energy to cross the complete band gap [15].

C. Radiation Trap Formation

Highly energized particles generate electron-hole pairs in a semiconductor. These holes will travel through the oxide and once they arrive at the oxide-substrate interface they are swept away in the substrate [2], [15]. However, some of the holes can occupy trapping sites in the oxide-silicon interface. This shifts the threshold voltage and may give rise to additional leakage current as explained in the previous section [2], [14], [15].

An illustration of the space radiation environment is presented in Figure 3 [16]. It can be seen that the total amount of charge that will get trapped strongly depends on the altitude of the spacecraft which thus makes it mission dependent. The total trapped charge is measured in krad for a specific material, in this case silicon.

III. Framework

This section describes first the radiation GIDL model subsequently followed by the experimental set-up.

A. Radiation Gate Induced Drain Leakage Model

In [12], the authors developed a model that relates the measurements of retention time under irradiation to GIDL leakage in transistors. The authors link the degradation of retention time to the generation of interface traps in the oxide of the transistor, which increases the gate induced drain leakage current. In our model we try to fit the degraded parameters to physical ones that can be used on transistor

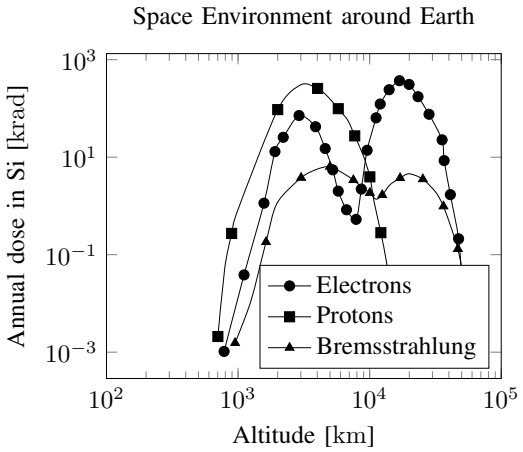


Fig. 3. Space Radiation Environment around earth in a 4 mm diameter aluminium sphere [16]

TABLE I
GIDL CURRENT PARAMETERS

Parameter	Description
I_{GIDL}	Gate Induced Drain Leakage Current
W_{effCJ}	Effective Source/Drain Diffusion Width
t_{ox}	Transistor Oxide Thickness
V_{db}	Drain-Bulk Voltage
V_{ds}	Drain-Source Voltage
V_{gs}	Gate-Source Voltage
AGIDL	Pre-exponential Coefficient
BGIDL	Exponential Coefficient
CGIDL	Parameter for Body-Bias Effect
EGIDL	Fitting Parameter for Band Bending

level Spice simulations. It is important to mention that the data presented in [12] are all measurements from a single commercial DRAM chip. Therefore, the accuracy of the model could be improved if more chips are taken into account.

In order to model the effect of the increased leakage current on the retention time, first a general GIDL model was implemented based on the one used in the BSIM4 physics based transistor model [17]. Equation 1 shows this GIDL current. The parameters of this equation are described in Table I.

$$I_{GIDL,\ normal} = \text{AGIDL} \cdot W_{eff} \cdot \frac{V_{ds} - V_{gs} - \text{EGIDL}}{3 \cdot t_{ox}} \cdot \exp\left(-\frac{3 \cdot t_{ox} \cdot \text{BGIDL}}{V_{ds} - V_{gs} - \text{EGIDL}}\right) \cdot \frac{V_{db}^3}{\text{CGIDL} + V_{db}^3} \quad (1)$$

To allow modeling of trapped charge due to radiation, we add an extra multiplier to the GIDL current equation. Its value depends solely on the total amount of radiation trapped in the oxide, T_D measured in krad, and the duty factor, DF, or activity factor of the transistor. The resulting GIDL current, $I_{Rad,\ GIDL}$, is presented in Equation 2.

$$I_{Rad\ GIDL} = I_{GIDL} \cdot f\left(T_D, DF\right) \quad (2)$$

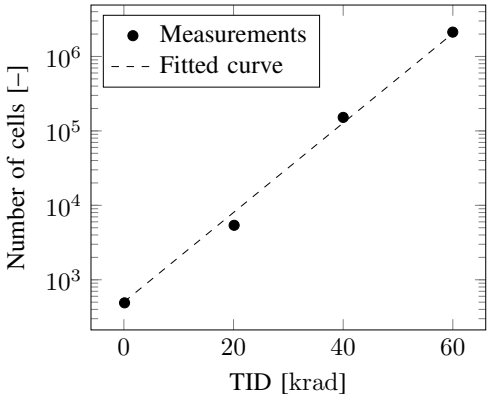


Fig. 4. Cells with $t_{ret} < 300$ ms and fitted curve versus TID [12]

Figure 4 shows the amount of cells with a retention time less than 300 ms based on the data from [12], as well as a fitting curve through these points. The data suggests that the degradation of the circuit is exponentially dependent on TID.

The multiplier function $f\left(T_D, DF\right)$ is further described in Equation 3. In this function, $f\left(DF\right)$ is a function that describes the impact of the duty factor on total radiation and X_1 is a dimensionless fitting parameter. E_a presents the activation energy in eV, k is the Boltzmann constant and T the temperature in K. The exponential term links the TID to the shift in retention time under radiation. The term -1 ensures that $f\left(T_D, DF\right) = 0$ when $T_D = 0$ krad. Below these terms will be explained in more detail.

$$f\left(T_D, DF\right) = f\left(DF\right) \cdot X_1 \left(\exp\left(\frac{E_a \cdot T_D}{kT}\right) - 1\right) \quad (3)$$

In [6], [18], the authors show that the bias conditions during radiation have an impact on the total amount of charges that get trapped in a transistor. This suggests that the duty factor of the transistor needs to be included in the model as well via $f\left(DF\right)$. However [6], [18] provide no specific details regarding this function, except for the on-state ($DF = 1$) and off-state ($DF = 0$). They mention that more charges get trapped during irradiation when the device is on. Since the device under test in [12] was unbiased during irradiation, this model sets $f\left(DF = 0\right) = 1$, effectively excluding the duty factor effects in this work. Nevertheless, the impact of this duty factor can be added in case the same experiments of [12] are repeated for different duty factors.

X_1 and E_a can be found by analyzing the discharge current graphs of a memory cell for different TID. In this work, we calcualte the discharge current from the retention time measurements in [12]. Figure 5 shows the change in the retention time density after a total dose of 60 krad is accumulated. The mean retention time shifts down from $\mu_{0\,\text{krad}} \approx 2800$ ms to $\mu_{60\,\text{krad}} \approx 1300$ ms, which is a factor $\frac{\mu_{0\,\text{krad}}}{\mu_{60\,\text{krad}}} \approx 2.15$.

The retention time of a cell is described by Equation 4 [19].

$$t_{ret} = \frac{(V_1 - V_2) \cdot C_{cell}}{I_{leak,\ average}} \quad (4)$$

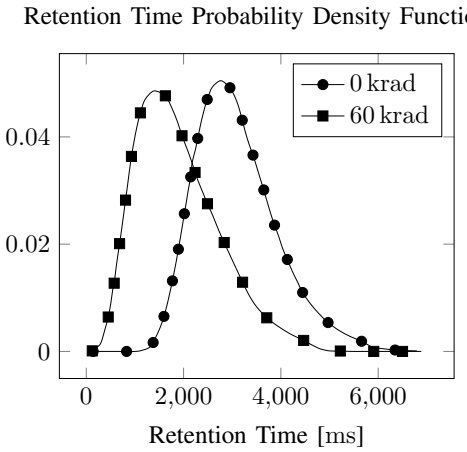


Fig. 5. Change in retention time density function due to radiation [12]

In this Equation, V_1 denotes the high voltage (i.e., the voltage on the capacitor right after writing to it), V_2 the voltage where the logic cell value flips, C_{cell} is the cell capacitance and t_{ret} is the retention time of the cell. The values of V_1, V_2 and C_{cell} are design dependent. Finally, $I_{leak,\ average}$ is the average current that flows when the capacitor is discharging and consists of two components, $I_{leak,\ average} = I_{Rad\ GIDL,\ average} + I_{Other,\ average}$. $I_{Other,\ average}$ is the leakage current in the cell caused by sources different from radiation trapping effects, and $I_{Rad\ GIDL,\ average}$ is the current that is caused by radiation trapping effects. When no charges are trapped (TID $= 0\,\mathrm{krad}$), there is no additional leakage current and hence $I_{Rad\ GIDL,\ average} = 0\,\mathrm{A}$. The retention time of the cell is affected only by $I_{Other,\ average}$. As was seen in the previous paragraph, the mean retention time decreases 2.15 times once $60\,\mathrm{krad}$ of charge is accumulated in the transistor. This means that the average leakage current ($I_{leak,\ average}$) must have increased 2.15 times. Since $I_{Other,\ average}$ does not change (as we assume that this is unaffected by radiation), $I_{Rad\ GIDL,\ average}$ must have increased. Setting $t_{ret,1} = 2.15 \cdot t_{ret,2}$ and solving for $I_{Rad\ GIDL,\ average}$ yields $I_{Rad\ GIDL,\ average} = 1.15 \cdot I_{Other,\ average}$. Therefore, the exponential term in Equation 3 should have a slope of $\frac{E_a}{kT} = 0.0128$ in order to satisfy this relation. Calculating the average current is nearly impossible in an actual circuit because the current strongly depends on the bias voltages of the pass transistor. These voltages change over time as the charge leaks away from the capacitor and the voltage on the bit line may change depending on the performed operations. We therefore chose to fit X_1 based on retention time simulations.

Figure 5 clearly shows not only a degradation of the mean retention time but also an increase in standard deviation. These statistical properties are not taken into account in this model due to a lack of detailed measurement data. It is for example impossible to know how a single cell degrades as only complete chip distributions are presented. This makes it harder to derive how radiation will affect the tail distribution, i.e., the cells with the lowest retention times, as they have the most impact on the complete DRAM reliability [20].

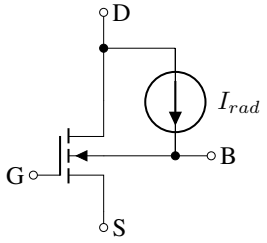


Fig. 6. NMOS trapped radiation GIDL model

The extra GIDL current generated due to trapped charges is implemented as a Verilog-A controlled current source. This simplifies the simulation of the radiated circuit. The current source discharges either from drain to bulk for NMOS transistors or from bulk to source for PMOS transistors. Figure 6 shows the NMOS radiation circuit.

B. Experimental Set-up

The experiments with the presented model are described next. We perform two types of experiments with respect to the presented radiation model; they are described below.

1) GIDL Current: The induced GIDL current is measured for varying TID and varying drain voltage (V_{DS}). This verifies the correct operation of the model.

2) Retention Time: The radiation model is applied to a commercial DRAM netlist. The retention time of the circuit is measured to see how it changes under irradiation.

IV. Results

A. GIDL Current

Figure 7 and 8 show the leakage currents $I_{Rad\ GIDL}$ due to radiation (as described in Equation 2) and the normal GIDL current $I_{GIDL,\ normal}$ without radiation (as described in Equation 1) for a single NMOS and PMOS transistor, respectively. The horizontal axis shows the drain voltage relative to V_{DD}. All other nodes are grounded for the NMOS transistor and connected to V_{DD} for the PMOS transistor. From the figures it becomes clear that the total current depends strongly on the total amount of trapped charges, as was expected from Equation 3 and the results presented in [6]. For the NMOS transistor, $I_{Rad\ GIDL}$ increases fast with increasing drain voltage. This can be explained by the increasing voltage difference between gate and drain. In the PMOS transistor, $I_{Rad\ GIDL}$ increases with decreasing drain voltage. Note that GIDL in a PMOS occurs when the gate is biased high while the drain is biased low. From both figures it also becomes clear that the GIDL current becomes negligible when the electric field between the gate and the drain weakens.

B. Retention Time

Figure 9 presents the simulated retention time for varying TID. Here, the radiation model is applied to the pass transistor

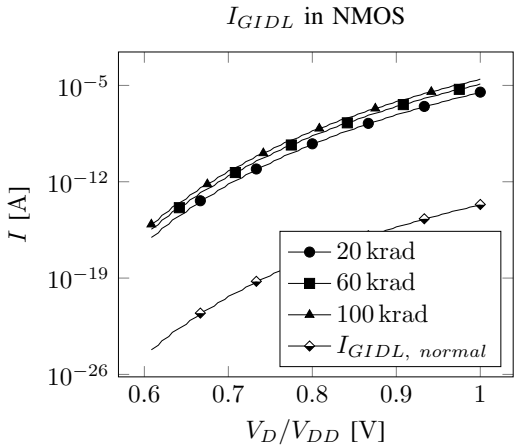


Fig. 7. GIDL current due to charge trapping in a NMOS transistor

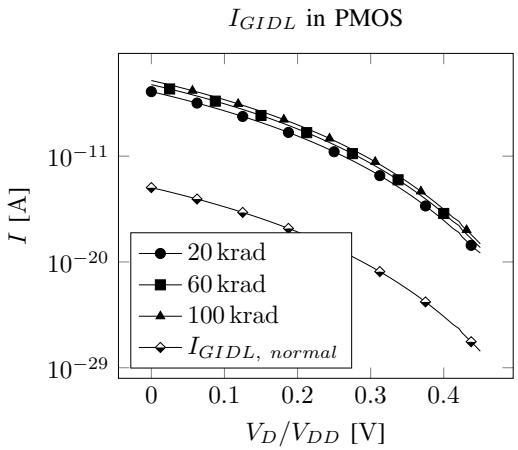


Fig. 8. GIDL current due to charge trapping in a PMOS transistor

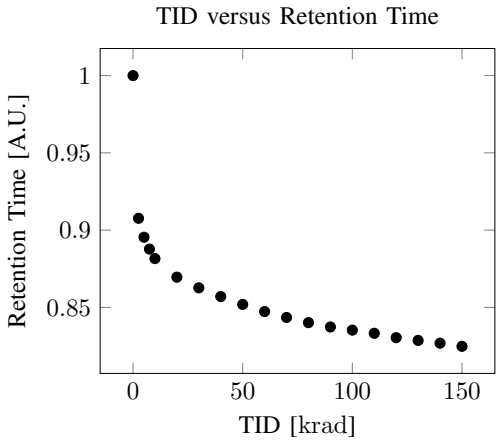


Fig. 9. Retention time degradation in a DRAM due to radiation trapping

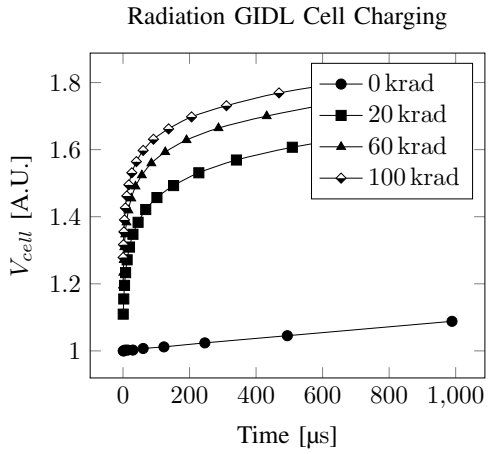


Fig. 10. Increase of memory cell voltage directly after a write operation for varying TID

of a commercial cell design. The figure shows a strong dependency on low TID. When TID becomes larger than $10\,\mathrm{krad}$ the retention time starts to decrease slower. This degradation can be explained by Figure 10. The figure shows the increase of the cell voltage directly after a low voltage was written. The DRAM netlist uses PMOS transistors as pass transistors in the cells and therefore the leakage increases towards V_{DD}. The leakage currents increase with increasing TID. This results in faster discharging of the capacitor. The impact of $I_{Rad\ GIDL}$ reduces when V_D approaches GND for NMOS transistors and V_{DD} for PMOS transistors. The capacitor discharges then via other leakage mechanisms. This effects explains why the irradiated cells in Figure 10 discharge very fast for a short period of time only, before other leakage mechanisms take over with lower leakage currents.

Next, the retention time degradation of the DRAM netlist was investigated for different altitudes and aging times. The radiation data from [16] is used to evaluate the retention time after 1 month, 1 year and 3 years of operation. Four different altitudes were investigated, $900\,\mathrm{km}$, $2000\,\mathrm{km}$ and $35\,786\,\mathrm{km}$. The first two altitudes are in the lower earth orbit, while the thirth altitude is at the geostationary orbit. Figure 11 presents the results of this analysis. It becomes clear that the altitude has a significant impact on the retention time, as expected from the TID in Figure 3 [16]. If the circuit is irradiated for a longer period, the retention time will drop further.

V. Discussion

This study presented a radiation model which includes the leakage effects of ionizing radiation in DRAM transistors. The effects are modeled based on retention time measurements of an irradiated DRAM. This secondary metric allows for a faster characterization of the leakage currents that are caused by radiation trapping. The model can be used during the design phase of a circuit to estimate the effects of radiation on it.

The model simulates the effects of radiation trapped charges as a current source. This current is a function of the nodal voltages and TID. It is important to note though that this

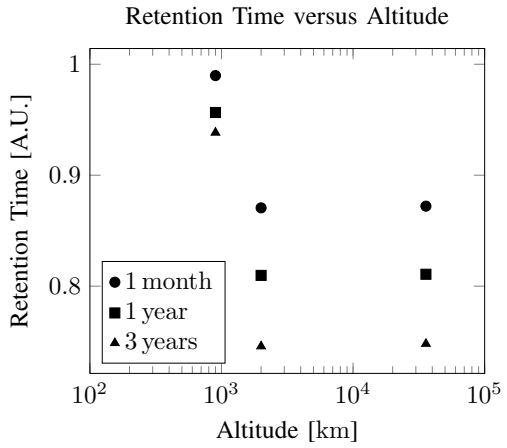


Fig. 11. Retention time versus altitude for varying time

does not include all effects in an irradiated transistor. For example, the trapped charged may shift the threshold voltage or damage the structure of the transistor which can lead to oxide breakdown. In addition, the stochastic nature of radiation trapping effects is not included in the model due to a lack of data. Nevertheless, the model can be extended by incorporating the statistics of the retention time measurements. To do this, more measurements on DRAMs need to be performed to measure the stochastic nature of individual cells. Similarly, the impact of the duty factor could be investigated as well.

Experiments of the model show a strong increase in leakage current with respect to TID when the transistor is in its off-state. Application of the model to a commercial DRAM netlist shows the importance of analyzing radiation effects. It is found that the retention time drops steeply when small doses get trapped. With further increasing doses, the retention time degradation weakens. Note that thicker layers of shielding can reduce the TID significantly [2].

Although the radiation leakage model may be less accurate than a physics based one, the methodology presented in this paper can be used to easily incorporate radiation effects in transistor models. The model aids the designer when designing a circuit that is hardened to radiation. This is a favored option over shielding the circuit with thicker, heavier layers of shielding material, which increases the costs of a space mission. Furthermore, since the model does not simulate the physical phenomena in detail, the simulation time is very low.

VI. Conclusion

This work presented a simple radiation model that estimates the effect of gate induced drain leakage current on the retention time for space applications, by evaluating the impact of total ionizing dose. This model can be used during the design phase for circuits that will endure high doses of radiation. This allows designers to verify correct operation prior to manufacturing.

References

[1] S. S. Rathod, A. K. Saxena, and S. Dasgupta, "Radiation Effects in MOS-based Devices and Circuits: A Review," *IETE Technical Review*, vol. 28, no. 6, pp. 451–469, 2011.

[2] K. Galloway and R. Schrimpf, "MOS device degradation due to total dose ionizing radiation in the natural space environment: A review," *Microelectronics Journal*, vol. 21, no. 2, pp. 67 – 81, 1990.

[3] R. DiBari, "What reliability engineers should know about space radiation effects," in *2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS)*, 1 2013, pp. 1–2.

[4] B. Sigfridsson, "Application of a new MOS-transistor model to total dose radiation effects in CMOS-devices," in *RADECS 91 First European Conference on Radiation and its Effects on Devices and Systems*, 9 1991, pp. 538–543.

[5] J. Jiang, W. Shu, K. S. Chong, T. Lin, N. K. Z. Lwin, J. S. Chang, and J. Liu, "Total Ionizing Dose (TID) effects on finger transistors in a 65nm CMOS process," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 5 2016, pp. 5–8.

[6] S. C. Witczak, R. C. Lacoe, J. V. Osborn, J. M. Hutson, and S. C. Moss, "Dose-rate sensitivity of modern nMOSFETs," *IEEE Transactions on Nuclear Science*, vol. 52, no. 6, pp. 2602–2608, 12 2005.

[7] H. J. Barnaby, M. L. McLain, I. S. Esqueda, and X. J. Chen, "Modeling Ionizing Radiation Effects in Solid State Materials and CMOS Devices," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 8, pp. 1870–1883, 8 2009.

[8] I. S. Esqueda and H. J. Barnaby, "A defect-based compact modeling approach for the reliability of CMOS devices and integrated circuits," *Solid-State Electronics*, vol. 91, no. Supplement C, pp. 81 – 86, 2014.

[9] T. Li, Y. Yang, J. Liu, and X. Huang, "Forward body bias for characterizing TID effect in CMOS integrated circuits," in *2017 IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 7 2017, pp. 1–5.

[10] A. Bougerol, F. Miller, N. Guibbaud, R. Leveugle, T. Carriere, and N. Buard, "Experimental Demonstration of Pattern Influence on DRAM SEU and SEFI Radiation Sensitivities," *IEEE Transactions on Nuclear Science*, vol. 58, no. 3, pp. 1032–1039, 6 2011.

[11] C. Miller, R. Owen, M. Rose, P. M. Rutt, J. Schaefer, and I. A. Troxel, "Trends in radiation susceptibility of commercial DRAMs for space systems," in *2009 IEEE Aerospace conference*, 3 2009, pp. 1–12.

[12] A. Bacchini, G. Furano, M. Rovatti, and M. Ottavi, "Total Ionizing Dose Effects on DRAM Data Retention Time," *IEEE Transactions on Nuclear Science*, vol. 61, no. 6, pp. 3690–3693, 12 2014.

[13] B. Jacob, S. W. Ng, and D. T. Wang, *Memory systems: Cache, DRAM, Disk*. Burlington, MA: Morgan Kaufmann Publishers, 2008.

[14] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling," in *1987 International Electron Devices Meeting*, vol. 33, 12 1987, pp. 718–721.

[15] D. Neamen, *Semiconductor Physics and Devices: Basic Principles*. New York, NY: McGraw-Hill, 2012.

[16] E. Daly, G. Drolshagen, A. Hilgers, and H. Evans, "Space environment analysis: Experience and trends," in *Environment Modeling for Space-Based Applications*, vol. 392, 1996, p. 15.

[17] N. Paydavosi, T. H. Morshed, D. D. Lu, W. M. Yang, M. V. Dunga, X. J. Xi, J. He, W. Liu, Kanyu, X. J. M. Cao, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, *BSIM4v4.8.0 MOSFET Model - User's Manual*, Department of Electrical Engineering and Computer Sciences University of California, Berkeley, 2013.

[18] M. Gaillardin, V. Goiffon, S. Girard, M. Martinez, P. Magnan, and P. Paillet, "Enhanced Radiation-Induced Narrow Channel Effects in Commercial 0.18 µm Bulk Technology," *IEEE Transactions on Nuclear Science*, vol. 58, no. 6, pp. 2807–2815, 12 2011.

[19] Z. Al-Ars, "DRAM Fault Analysis and Test Generation," Ph.D. dissertation, Delft University of Technology, Delft, Netherlands, 6 2005.

[20] M. White, J. Qin, and J. B. Bernstein, "A study of scaling effects on DRAM reliability," in *2011 Proceedings - Annual Reliability and Maintainability Symposium*, 1 2011, pp. 1–6.

Bibliography

[1] G. Gielen, P. D. Wit, E. Maricau, J. Loeckx, J. Martin-Martinez, B. Kaczer, G. Groeseneken, R. Rodriguez, and M. Nafria, "Emerging yield and reliability challenges in nanometer CMOS technologies", in *2008 Design, Automation and Test in Europe*, Mar. 2008, pp. 1322–1327.

[2] SPACECAST. (). Satellite anomalies, [Online]. Available: `http://fp7-spacecast.eu/help/bg_sa.pdf` (visited on 11/07/2017).

[3] M. Bushnell and V. Agrawal, *Essentials of electronic testing for digital, memory and mixed-signal VLSI circuits*. New York: Kluwer Academic, 2000, ISBN: 0-7923-799-1-8.

[4] U.S. Department of Defense, *Military Handbook – Reliability Prediction of Electronic Equipment*, MIL-HDBK-217F, Dec. 2, 1991.

[5] J. Srinivasan, S. V. Adve, P. Bose, J. Rivers, and C.-K. Hu, "Ramp: A model for reliability aware microprocessor design", *IBM Research Report*, 2003.

[6] Statista Inc. (2017). Quarterly DRAM revenues worldwide from 2011 to 2017, by manufacturer (in million U.S. dollars), [Online]. Available: `https://www.statista.com/statistics/271725/global-dram-revenues-of-semiconductor-companies-since-2010/` (visited on 10/26/2017).

[7] C. Kale. (May 31, 2016). Semiconductor Sales over last 25 years [1988 – 2015], [Online]. Available: `http://rvantage.net/tracking-semiconductor-sales/` (visited on 10/26/2017).

[8] R. DiBari, "What reliability engineers should know about space radiation effects", in *2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS)*, Jan. 2013, pp. 1–2.

[9] S. S. Rathod, A. K. Saxena, and S. Dasgupta, "Radiation Effects in MOS-based Devices and Circuits: A Review", *IETE Technical Review*, vol. 28, no. 6, pp. 451–469, 2011.

[10] K. Galloway and R. Schrimpf, "MOS device degradation due to total dose ionizing radiation in the natural space environment: A review", *Microelectronics Journal*, vol. 21, no. 2, pp. 67–81, 1990. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/002626929090027Z`.

[11] P. Lee. (Jun. 27, 2017). New and old tech required to develop space apps, [Online]. Available: `http://www.newelectronics.co.uk/electronics-technology/space-applications-require-new-and-old-technologies/156842/` (visited on 11/21/2017).

[12] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami, K. Quader, P. K. Ko, and C. Hu, "Berkeley reliability tools - BERT", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, no. 10, pp. 1524–1534, 1993.

[13] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers, "Exploiting structural duplication for lifetime reliability enhancement", in *32nd International Symposium on Computer Architecture (ISCA'05)*, Jun. 2005, pp. 520–531.

[14] X. Li, J. Qin, B. Huang, X. Zhang, and J. B. Bernstein, "A new SPICE reliability simulation method for deep submicrometer CMOS VLSI circuits", *IEEE Transactions on Device and Materials Reliability*, vol. 6, no. 2, pp. 247–257, Jun. 2006.

[15] D. Lorenz, M. Barke, and U. Schlichtmann, "Aging analysis at gate and macro cell level", in *2010 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov. 2010, pp. 77–84.

[16] E. Maricau and G. Gielen, "Efficient Variability-Aware NBTI and Hot Carrier Circuit Reliability Analysis", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 12, pp. 1884–1893, Dec. 2010.

[17] J. Qin, H. Avshalom, and J. B. Bernstein, "FaRBS: A new PoF based VLSI reliability prediction method", in *2011 Proceedings - Annual Reliability and Maintainability Symposium*, Jan. 2011, pp. 1–6.

[18] Y. Cao, J. Velamala, K. Sutaria, M. S. W. Chen, J. Ahlbin, I. S. Esqueda, M. Bajura, and M. Fritze, "Cross-Layer Modeling and Simulation of Circuit Reliability", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 1, pp. 8–23, Jan. 2014.

[19] D. Patterson, *Computer organization and design: the hardware/software interface*. Oxford Waltham, MA, USA: Morgan Kaufmann, 2014, ISBN: 978-0124077263.

[20] J. C. McCallum. (Jul. 22, 2017). Price and Performance Changes of Computer Technology with Time: Memory, Disk Drives, Flash Memory ans SSDs, [Online]. Available: `http://jcmit.net/` (visited on 10/04/2017).

[21] B. Jacob, S. W. Ng, and D. T. Wang, *Memory systems: Cache, DRAM, Disk*. Burlington, MA: Morgan Kaufmann Publishers, 2008, ISBN: 978-0-12-379751-3.

[22] M. Bellis. (Aug. 25, 2016). Who Invented the Intel 1103 DRAM Chip – The World's First Available DRAM Chip, [Online]. Available: `https://www.thoughtco.com/who-invented-the-intel-1103-dram-chip-4078677` (visited on 10/31/2017).

[23] Transcend Information Inc. (). What is the difference between SDRAM, DDR1, DDR2, DDR3 and DDR4?, [Online]. Available: `https://www.transcend-info.com/Support/FAQ-296` (visited on 11/10/2017).

[24] Z. Al-Ars, "DRAM Fault Analysis and Test Generation", PhD thesis, Delft University of Technology, Delft, Netherlands, Jun. 2005.

[25] J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, ser. Prentice Hall electronics and VLSI series. Pearson Education, 2003, ISBN: 0130909963.

[26] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging", in *2007 IEEE International Electron Devices Meeting*, Dec. 2007, pp. 247–250.

[27] E. Keyes and D. Scansen. (Nov. 12, 2006). DRAM makes move to copper, [Online]. Available: `https://www.eetimes.com/document.asp?doc_id=1281149` (visited on 11/09/2017).

[28] E. Atanassova and A. Paskaleva, "Challenges of ta2o5 as high-k dielectric for nanoscale {drams}", *Microelectronics Reliability*, vol. 47, no. 6, pp. 913–923, 2007, Modelling the Negative Bias Temperature Instability. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0026271406001466`.

[29] E. Wyrwas, L. Condra, and A. Hava, "Accurate quantitative physics-of-failure approach to integrated circuit reliability", *IPC APEX Expo, Las Vegas, NV*, vol. 13, 2011.

[30] S. R. Hofstein and F. P. Heiman, "The silicon insulated-gate field-effect transistor", *Proceedings of the IEEE*, vol. 51, no. 9, pp. 1190–1202, Sep. 1963.

[31] R. Jaeger, *Introduction to microelectronic fabrication*. Upper Saddle River, N.J: Prentice Hall, 2002, ISBN: 978-0201444940.

[32] D. A. Neamen, *Semiconductor Physics and Devices - Basic Principles*. 2 Penn Plaza, New York, NY 10121: McGraw-Hill Education, 2012, ISBN: 9780071089029.

[33] M. Gaillardin, V. Goiffon, S. Girard, M. Martinez, P. Magnan, and P. Paillet, "Enhanced Radiation-Induced Narrow Channel Effects in Commercial 0.18 µm Bulk Technology", *IEEE Transactions on Nuclear Science*, vol. 58, no. 6, pp. 2807–2815, Dec. 2011.

[34] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling", in *1987 International Electron Devices Meeting*, vol. 33, 1987, pp. 718–721.

[35] S. Mahapatra and M. A. Alam, "Defect Generation in p-MOSFETs Under Negative-Bias Stress: An Experimental Perspective", *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 1, pp. 35–46, Mar. 2008.

[36] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the Physical Mechanism of NBTI in Silicon Oxynitride p-MOSFETs: Can Differences in Insulator Processing Conditions Resolve the Interface Trap Generation versus Hole Trapping Controversy?", in *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, Apr. 2007, pp. 1–9.

[37] R. Fernandez, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodriguez, M. Nafria, and G. Groeseneken, "AC NBTI studied in the 1 Hz – 2 GHz range on dedicated on-chip CMOS circuits", in *2006 International Electron Devices Meeting*, Dec. 2006, pp. 1–4.

[38] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress", in *2010 IEEE International Reliability Physics Symposium*, May 2010, pp. 7–15.

[39] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. J. Wagner, F. Schanovsky, J. Franco, M. T. Luque, and M. Nelhiebel, "The Paradigm Shift in Understanding the Bias Temperature Instability: From Reaction–Diffusion to Switching Oxide Traps", *IEEE Transactions on Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov. 2011.

[40] S. Mahapatra, A. E. Islam, S. Deora, V. D. Maheta, K. Joshi, A. Jain, and M. A. Alam, "A critical re-evaluation of the usefulness of R-D framework in predicting NBTI stress and recovery", in *2011 International Reliability Physics Symposium*, Apr. 2011, 6A.3.1–6A.3.10.

[41] M. Makabe, T. Kubota, and T. Kitano, "Bias-temperature degradation of pMOSFETs: mechanism and suppression", in *2000 IEEE International Reliability Physics Symposium Proceedings. 38th Annual (Cat. No.00CH37059)*, 2000, pp. 205–209.

[42] P. Hehenberger, "Advanced Characterization of the Bias Temperature Instability", PhD thesis, Technische Universität Wien, 2011.

[43] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in digital circuits", in *2009 Asia and South Pacific Design Automation Conference*, Jan. 2009, pp. 284–289.

[44] E. Maricau, P. D. Wit, and G. Gielen, "An analytical model for hot carrier degradation in nanoscale {cmos} suitable for the simulation of degradation in analog {ic} applications", *Microelectronics Reliability*, vol. 48, no. 8–9, pp. 1576–1580, 2008, 19th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF 2008). [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0026271408002060`.

[45] K. N. Quader, C. Li, R. Tu, E. Rosenbaum, P. Ko, and C. Hu, "A new approach for simulation of circuit degradation due to hot-electron damage in NMOSFETs", in *International Electron Devices Meeting 1991 [Technical Digest]*, Dec. 1991, pp. 337–340.

[46] J. Franco and B. Kaczer, "Channel Hot Carriers in SiGe and Ge pMOSFETs", in *Hot Carrier Degradation in Semiconductor Devices*, T. Grasser, Ed. Cham: Springer International Publishing, 2015, pp. 259–285, ISBN: 978-3-319-08994-2. [Online]. Available: `https://doi.org/10.1007/978-3-319-08994-2_9`.

[47] B. Yan, J. Qin, J. Dai, Q. Fan, and J. B. Bernstein, "Reliability Simulation and Circuit-Failure Analysis in Analog and Mixed-Signal Applications", *IEEE Transactions on Device and Materials Reliability*, vol. 9, no. 3, pp. 339–347, Sep. 2009.

[48] X. Wang, J. Keane, T. T. H. Kim, P. Jain, Q. Tang, and C. H. Kim, "Silicon Odometers: Compact In Situ Aging Sensors for Robust System Design", *IEEE Micro*, vol. 34, no. 6, pp. 74–85, Nov. 2014.

[49] K. H. Koo, *The Comparison Study of Future On-chip Interconnects for High Performance VLSI Applications*. Stanford University, 2011.

[50] A. Bacchini, G. Furano, M. Rovatti, and M. Ottavi, "Total Ionizing Dose Effects on DRAM Data Retention Time", *IEEE Transactions on Nuclear Science*, vol. 61, no. 6, pp. 3690–3693, Dec. 2014.

[51] Y. Kim, R. Daly, J. Kim, C. Fallin, J. H. Lee, D. Lee, C. Wilkerson, K. Lai, and O. Mutlu, "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors", *SIGARCH Comput. Archit. News*, vol. 42, no. 3, pp. 361–372, Jun. 2014. [Online]. Available: `http://doi.acm.org/10.1145/2678373.2665726`.

[52] J. Liu, B. Jaiyen, Y. Kim, C. Wilkerson, and O. Mutlu, "An Experimental Study of Data Retention Behavior in Modern DRAM Devices: Implications for Retention Time Profiling Mechanisms", *SIGARCH Comput. Archit. News*, vol. 41, no. 3, pp. 60–71, Jun. 2013. [Online]. Available: `http://doi.acm.org/10.1145/2508148.2485928`.

[53] A. Bacchini, M. Rovatti, G. Furano, and M. Ottavi, "Characterization of data retention faults in DRAM devices", in *2014 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, Oct. 2014, pp. 9–14.

[54] A. Chan, D. Lam, W. Tan, and S. Y. Khim, "Electrical failure analysis in high density DRAMs", in *Proceedings of IEEE International Workshop on Memory Technology, Design, and Test*, Aug. 1994, pp. 26–31.

[55] T. Joseph, K. Varn, N. Arnold, and D. Griffiths, "Vertical transistor deep trench DRAM failure analysis and failure mechanisms", in *Proceedings of the 10th International Symposium on the Physical and Failure Analysis of Integrated Circuits. IPFA 2003*, Jul. 2003, pp. 183–186.

[56] *64MBit Synchronous DRAM – HYB39S64400/800/160AT*, Rev. 10.98, Siemens.

[57] Standard Performance Evaluation Corporation. (). SPEC CPU 2006, [Online]. Available: `https://www.spec.org/cpu2006/` (visited on 09/27/2017).

[58] Intel Corporation. (). Intel® Core™2 Duo Processor E6700, [Online]. Available: `https://ark.intel.com/products/27251/Intel-Core2-Duo-Processor-E6700-4M-Cache-2_66-GHz-1066-MHz-FSB` (visited on 12/07/2017).

[59] A. Kejariwal, A. V. Veidenbaum, A. Nicolau, X. Tian, M. Girkar, H. Saito, and U. Banerjee, "Comparative architectural characterization of SPEC CPU2000 and CPU2006 benchmarks on the Intel Core 2 Duo processor", in *2008 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation*, Jul. 2008, pp. 132–141.

[60] S. Birdj, A. Phansalkarj, L. K. Johnj, A. Mericasa, and R. Indukurua, "Performance Characterization of SPEC CPU Benchmarks on Intel's Core Microarchitecture based processor", University of Texas at Austin and IBM Austin, Tech. Rep.

[61] C. C. Li, K. N. Quader, E. R. Minami, C. Hu, and P. K. Ko, "A new bi-directional PMOSFET hot-carrier degradation model for circuit reliability simulation", in *1992 International Technical Digest on Electron Devices Meeting*, Dec. 1992, pp. 547–550.

[62] K. N. Quader, C. C. Li, R. Tu, E. Rosenbaum, P. K. Ko, and C. Hu, "A bidirectional NMOSFET current reduction model for simulation of hot-carrier-induced circuit degradation", *IEEE Transactions on Electron Devices*, vol. 40, no. 12, pp. 2245–2254, Dec. 1993.

[63] J. B. Bernstein, M. Gurfinkel, X. Li, J. Walters, Y. Shapira, and M. Talmor, "Electronic circuit reliability modeling", *Microelectronics Reliability*, vol. 46, no. 12, pp. 1957–1979, 2006. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0026271406000023`.

[64] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors", *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[65] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits", *IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, Jun. 2005.

[66] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs", in *2010 IEEE International Reliability Physics Symposium*, May 2010, pp. 26–32.

[67] P. Weckx, B. Kaczer, M. Toledano-Luque, T. Grasser, P. J. Roussel, H. Kukner, P. Raghavan, F. Catthoor, and G. Groeseneken, "Defect-based methodology for workload-dependent circuit lifetime projections - Application to SRAM", in *2013 IEEE International Reliability Physics Symposium (IRPS)*, Apr. 2013, 3A.4.1–3A.4.7.

[68] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices", *Journal of Applied Physics*, vol. 48, no. 5, pp. 2004–2014, 1977. [Online]. Available: `http://dx.doi.org/10.1063/1.323909`.

[69] M. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation", *Microelectronics Reliability*, vol. 45, no. 1, pp. 71–81, 2005. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0026271404001751`.

[70] N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, and T. Horiuchi, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling", in *1999 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.99CH36325)*, Jun. 1999, pp. 73–74.

[71] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing", *Journal of Applied Physics*, vol. 94, no. 1, pp. 1–18, 2003. [Online]. Available: `http://dx.doi.org/10.1063/1.1567461`.

[72] Y. Wang, "Aging Assessment and Reliability Aware Computing Platforms", PhD thesis, Delft University of Technology, Delft, Netherlands, Sep. 2013.

[73] K. Kang, S. P. Park, K. Roy, and M. A. Alam, "Estimation of Statistical Variation in Temporal NBTI Degradation and Its Impact on Lifetime Circuit Performance", in *Proceedings of the 2007 IEEE/ACM International Conference on Computer-aided Design*, ser. ICCAD '07, San Jose, California: IEEE Press, 2007, pp. 730–734, ISBN: 1-4244-1382-6. [Online]. Available: http://dl.acm.org/citation.cfm?id=1326073.1326227.

[74] K. B. Sutaria, P. Ren, A. Mohanty, X. Feng, R. Wang, R. Huang, and Y. Cao, "Duty cycle shift under static/-dynamic aging in 28nm HK-MG technology", in *2015 IEEE International Reliability Physics Symposium*, Apr. 2015, CA.7.1–CA.7.5.

[75] S. Mukhopadhyay, N. Goel, and S. Mahapatra, "A Comparative Study of NBTI and PBTI Using Different Experimental Techniques", *IEEE Transactions on Electron Devices*, vol. 63, no. 10, pp. 4038–4045, Oct. 2016.

[76] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection", *IEEE Electron Device Letters*, vol. 4, no. 4, pp. 111–113, Apr. 1983.

[77] M. Cho, P. Roussel, B. Kaczer, R. Degraeve, J. Franco, M. Aoulaiche, T. Chiarella, T. Kauerauf, N. Horiguchi, and G. Groeseneken, "Channel Hot Carrier Degradation Mechanism in Long/Short Channel *n*-FinFETs", *IEEE Transactions on Electron Devices*, vol. 60, no. 12, pp. 4002–4007, Dec. 2013.

[78] A. Strong, E. Wu, R.-P. Vollertsen, J. Suñé, G. LaRosa, T. Sullivan, and S. Rauch, *Reliability wearout mechanisms in advanced CMOS technologies*. Piscataway, NJ Hoboken, NJ: IEEE Press Wiley, 2009, ISBN: 9780471731726.

[79] X. Li, J. Qin, and J. B. Bernstein, "Compact Modeling of MOSFET Wearout Mechanisms for Circuit-Reliability Simulation", *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 1, pp. 98–121, Mar. 2008.

[80] G. L. Rosa, F. Guarin, S. Rauch, A. Acovic, J. Lukaitis, and E. Crabbe, "NBTI-channel hot carrier effects in PMOSFETs in advanced CMOS technologies", in *1997 IEEE International Reliability Physics Symposium Proceedings. 35th Annual*, Apr. 1997, pp. 282–286.

[81] P. Magnone, F. Crupi, N. Wils, R. Jain, H. Tuinhout, P. Andricciola, G. Giusi, and C. Fiegna, "Impact of Hot Carriers on nMOSFET Variability in 45- and 65-nm CMOS Technologies", *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2347–2353, Aug. 2011.

[82] N. Paydavosi, T. H. Morshed, D. D. Lu, W. (Yang, M. V. Dunga, X. (Xi, J. He, W. Liu, Kanyu, X. J. M. Cao, J. J. Ou, M. Chan, A. M. Niknejad, and C. Hu, *BSIM4v4.8.0 MOSFET Model - User's Manual*, Department of Electrical Engineering and Computer Sciences University of California, Berkeley, 2013.

[83] S. C. Witczak, R. C. Lacoe, J. V. Osborn, J. M. Hutson, and S. C. Moss, "Dose-rate sensitivity of modern nMOSFETs", *IEEE Transactions on Nuclear Science*, vol. 52, no. 6, pp. 2602–2608, Dec. 2005.

[84] M. White, J. Qin, and J. B. Bernstein, "A study of scaling effects on DRAM reliability", in *2011 Proceedings - Annual Reliability and Maintainability Symposium*, Jan. 2011, pp. 1–6.

[85] I. S. Esqueda, H. J. Barnaby, and M. P. King, "Compact Modeling of Total Ionizing Dose and Aging Effects in MOS Technologies", *IEEE Transactions on Nuclear Science*, vol. 62, no. 4, pp. 1501–1515, Aug. 2015.

[86] H. J. Barnaby, M. L. McLain, I. S. Esqueda, and X. J. Chen, "Modeling Ionizing Radiation Effects in Solid State Materials and CMOS Devices", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 8, pp. 1870–1883, Aug. 2009.

[87] I. S. Esqueda and H. J. Barnaby, "A defect-based compact modeling approach for the reliability of CMOS devices and integrated circuits", *Solid-State Electronics*, vol. 91, no. Supplement C, pp. 81–86, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038110113003055.

[88] E. Daly, G. Drolshagen, A. Hilgers, and H. Evans, "Space environment analysis: Experience and trends", in *Environment Modeling for Space-Based Applications*, vol. 392, 1996, p. 15.

[89] E. Stassinopoulos and J. P. Raymond, "The space radiation environment for electronics", *Proceedings of the IEEE*, vol. 76, no. 11, pp. 1423–1442, 1988.

[90] Royal Belgian Institute for Space Aeronomy. (). SPENVIS – Space Environment Information System, [Online]. Available: https://www.spenvis.oma.be/ (visited on 09/25/2017).

[91] European Space Agency. (). Sentinel-1 Mission Overview, [Online]. Available: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-1 (visited on 09/25/2017).

[92] European Space Agency, "Copernicus Sentinel-1 Spacecraft Radiation, Meteoroids and Debris Environment", European Space Agency, Tech. Rep., Feb. 2016.

[93] Cadence Design Systems, Inc., "Virtuoso Spectre Circuit Simulator Reference", Cadence Design Systems, Inc., Tech. Rep., Oct. 2014.

[94] Cadence Design Systems, Inc., "Virtuoso Simulator Measurement Description Language User Guide and Reference", Cadence Design Systems, Inc., Tech. Rep., Oct. 2014.

[95] Verilog-AMS Working Group. (2014). Verilog-AMS, [Online]. Available: http://www.accellera.org/downloads/standards/v-ams (visited on 10/05/2017).

[96] MathWorks. (). Matlab R2015b, [Online]. Available: https://nl.mathworks.com/products/new_products/release2015b.html (visited on 10/24/2017).

[97] G. E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics Magazine*, vol. 38, no. 8, Apr. 1965.

[98] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, L. Rideout, E. Bassous, and A. R. Leblanc, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions", *IEEE Journal of Solid-State Circuits*, vol. SC-9, no. 5, pp. 256–268, Oct. 1974.

[99] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design", *IEEE Transactions on Electron Devices*, vol. 31, no. 4, pp. 452–462, Apr. 1984.