This is to certify that the

dissertation entitled

STRESS TESTS OF ANALOG CMOS ICs FOR
GATE-OXIDE RELIABILITY ENHANCEMENT

presented by

Mohammad Athar Khalil

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Elec. & Comp. Engineering

Major professor

Date Aug. 17, 01

STRESS TESTS OF ANALOG CMOS ICs FOR GATE-OXIDE RELIABILITY ENHANCEMENT

By

Mohammad Athar Khalil

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSPHY

Department of Electrical and Computer Engineering

2001

ABSTRACT

STRESS TESTS OF ANALOG CMOS ICs FOR
GATE-OXIDE RELIABILITY ENHANCEMENT

By

Mohammad Athar Khalil

Yield and reliability are two factors affecting the profitability of semiconductor manufacturing. In the manufacturing process of modern VLSI semiconductor devices, gate-oxide defects have been found as one of the major causes for the reliability problems of CMOS integrated circuits (ICs). Stress testing is a technique used to weed-out early life failures by applying higher than usual levels of stress to speed up the deterioration of electronic devices. The industry standard methods for stress testing have been high-temperature burn-in and high-voltage stress. Burn-in decreases failure rate of a product during the early field life, but overall cost and turn around time are of concerns. The added manufacturing cost may range from 5% to 40% of the total product cost, depending on the burn-in time, qualities of ICs, and product complexity. Extreme-voltage stress test aims at enhancing both quality and reliability without performing the high-cost burn-in test process. Extreme-voltage screening has been successfully implemented to enhance gate-oxide reliability of digital CMOS ICs. However, the success has not yet been extended to its analog counterparts. Today, almost all IC manufacturers employ the digital circuits extreme-voltage screening process for the analog modules in mixed-signal CMOS ICs. This study initiates the research on extreme-voltage stress of analog CMOS ICs with the goal to improve their quality and reliability cost effectively.

This study presents an efficient yet effective extreme-voltage stress test process for analog CMOS circuits analytically. It develops the framework of an automatic stress test system for analog circuits, that integrates three major components -- *Stress Vector Generator*, *Stressability Analyzer*, and *Stressability Design Methodology*.

Stress vector generator and stressability analyzer are analysis tools used by the stress test process to determine the stressability of a given circuit. Stress vector generator can provide optimal set of stress vectors and stressability analyzer determines the stressability of the circuit based on the selected stress vectors. Stressability design methodology of the stress test process is employed during design phase to ensure desired stressability of the circuit. This design methodology utilizes the two analysis tools and stressability enhancement process to meet stressability requirements during the design phase.

This work defines the fundamental concepts required for proper stressability of analog circuits. Based on these concepts it develops stress vector generation procedure for extreme-voltage stress test and burn-in of analog CMOS ICs. It also presents stressability analysis process that can locate portions of a circuit having poor stressability. It then describes an effective stressability enhancement process to improve the stressability of problem areas identified by the stressability analysis. The development is readily applicable to digital and mixed signal CMOS ICs. For large circuits, it presents a circuit decomposition model so that the developed processes can be applied in a hierarchical manner. This study also analyzes the stressability of various gate-oxide reliability enhancement techniques currently employed in the industry. It compares the existing methods and identifies the trade-offs that will determine the selection of proper stress test.

To my mother and father.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Chin-Long Wey, for all his assistance with my research and the development of this dissertation. His guidance, care, patience and constant encouragement not only kept me motivated but also were crucial for keeping the research objectives in perspective. Working under his supervision, I have learned a lot during the course of this study. Thanks are also due to the members of my Ph.D. guidance committee, Dr. James A. Resh, Dr. Gregory M. Wierzba, and Dr. Jonathan I. Hall, for all their help, support, and valuable suggestions.

I am very grateful to my family for encouraging me to pursue higher studies, especially to my parents for the encouragement and support they have given me every day of my life. I would like to express my sincere gratitude to my brother and sisters for all the assistance provided by them. I am also very thankful to my wife for her patience and cooperation. I also wish to thank all my friends for their moral support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The world market for electronic systems will reach $1 Trillion within a year with further exponential growth over the next five years [1-3]. The functionality and value of electronics-based systems are increasingly being embedded in integrated circuits (ICs) which are the core elements of these products, most noticeably in the area of communications, network, and multimedia. The functionality of ICs will grow to system-level proportions in which several complex sub-components, including logic, memory, numerical and signal processing, and RF circuits will reside on a single chip or chip set. Creating single chip solutions to system challenges is already the target of intense research and development effort around the world, and system-on-a-chip (SoC) designs have become a reality [4].

A computerized manufacturing procedure is shown in Figure 1.1 [5]. Each project should be verified according to the market's needs, attainable manufacturing techniques, and potential profits. Once a project is approved after careful examination by the engineering, marketing, and financial departments, computer designs can then be commenced followed by the computer aided process planning. Prototypes will then be manufactured for life-testing. If the quality is acceptable, production will then proceed, otherwise, re-design

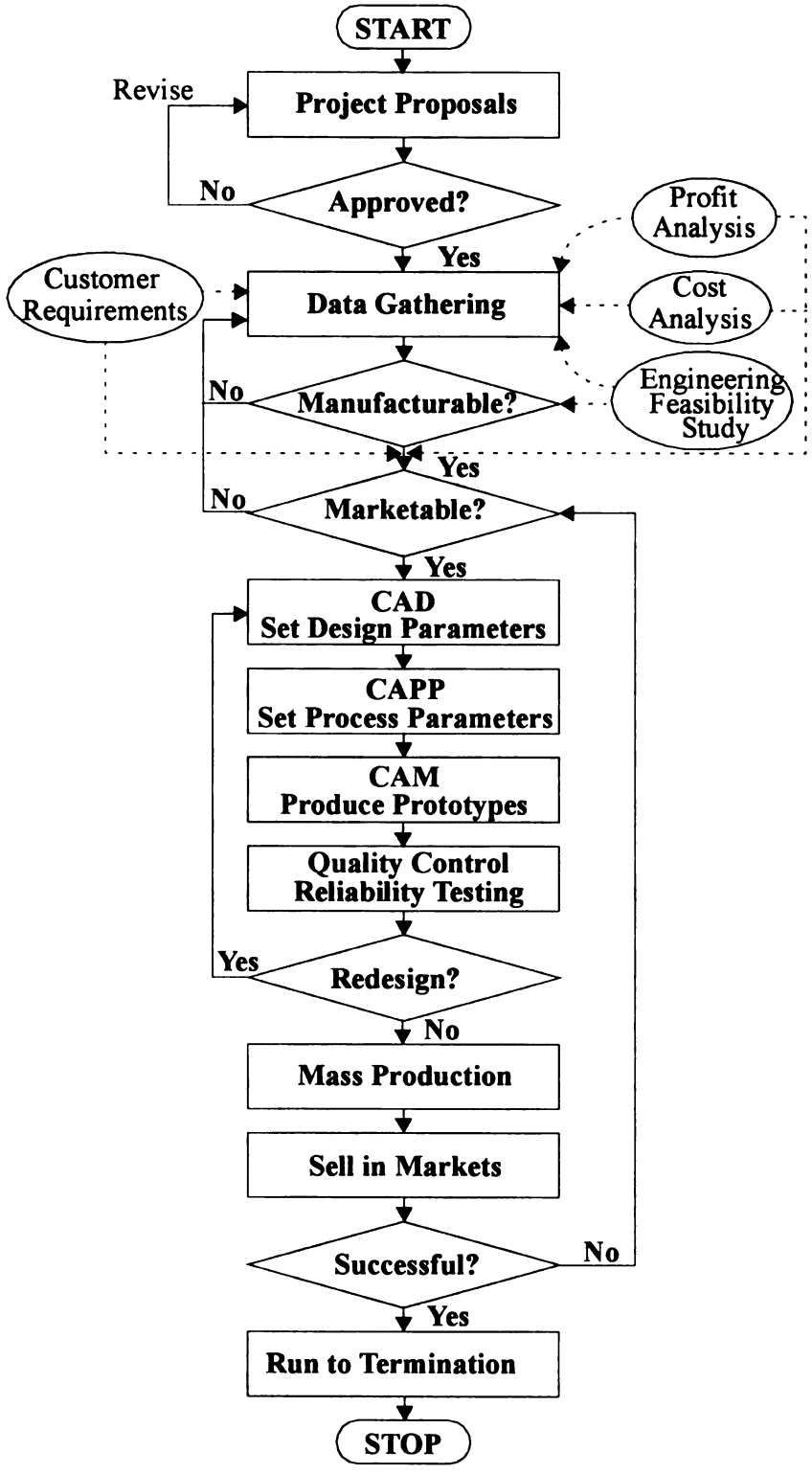


Figure 1.1: A Manufacturing Sequence [5].

might be needed. The popularity of a product is justification for possible re-modeling or modification of selling strategies. If it is found that the product is not competitive, development profiles as well as field reports will be put into a database for future reference. After being in the market for some time, a non-profitable product may be terminated and the production equipment may be either sold or used for making other items.

Today, the microelectronics industry has evolved into the world's largest manufacturing business [6]. The semiconductor manufacturing industry has enjoyed an average compound annual growth rate of 15% per year and is committed to staying on this growth path. The three major types of materials making up circuits are metals, insulators, and semiconductors, among which the basic difference is the magnitude of the energy gap [7]. Metallic materials form the conducting element. The insulator is the dielectric which provides isolation between transistors or between metals; it also serves as a part of transistor structure. Table 1.1depicts the trends of IC devices and technologies. It is evident that CMOS is currently the dominant technology and is expected to continue to dominate in the future.

Table 1.1: Trend of IC Devices and Technologies

Device Type	1987	1992	2000
CMOS	39%	73%	82%
BiCMOS	0%	2%	6%
NMOS	24%	4%	< 1%
PMOS	< 1%	0%	0%
Bipolar (analog)	20%	14%	9%
TTL	12%	4%	1%
ECL	4%	2%	< 1%
GaAs	< 1%	< 1%	1%

The manufacturing process for microcircuits starts with the growth of the crystal and proceeds to packaging. Yield and reliability are the driving forces for the success of any manufacturing scheme. Yield must be maximized for each processing step while maintaining desired reliability [8]. The principal IC manufacturing processes are summarized in Figure 1.2. The front-end fabrication processing sequences proceed from the crystal growth up to the final metallization and the back-end fabrication from the wafer probe to the final test. Typical manufacturing processes start with the conception of new circuits and end with the final electrical tests.

Every user of ICs wants the highest reliability possible, as well as the highest performance, quickest time to market, and lowest cost. However, trade-offs must be made between cost, performance, time to market, reliability and other customer requirements [9]. In designing, manufacturing and testing ICs the manufacturers make technical and economic choices and trade-offs that lead to a best combination of cost, performance, time to market and reliability for the market segment(s) they have targeted.

The sales life of most microelectronic based products is continuously decreasing, since they get superseded by improved or more innovative products. Therefore, any delay in availability involves the risk of losing a considerable share of the market; a six month unforeseen delay in production and sales has been estimated to cost at least 30% loss in total sales of a given product design [10]. Figure 1.3 shows the usual representation of a microelectronic product life cycle and profit revenue; the general effect of being late to market will leave the right hand part of this picture relatively unchanged, but the ramp up to maturity will be depressed thus making total sales and hence total profits smaller [11,12].

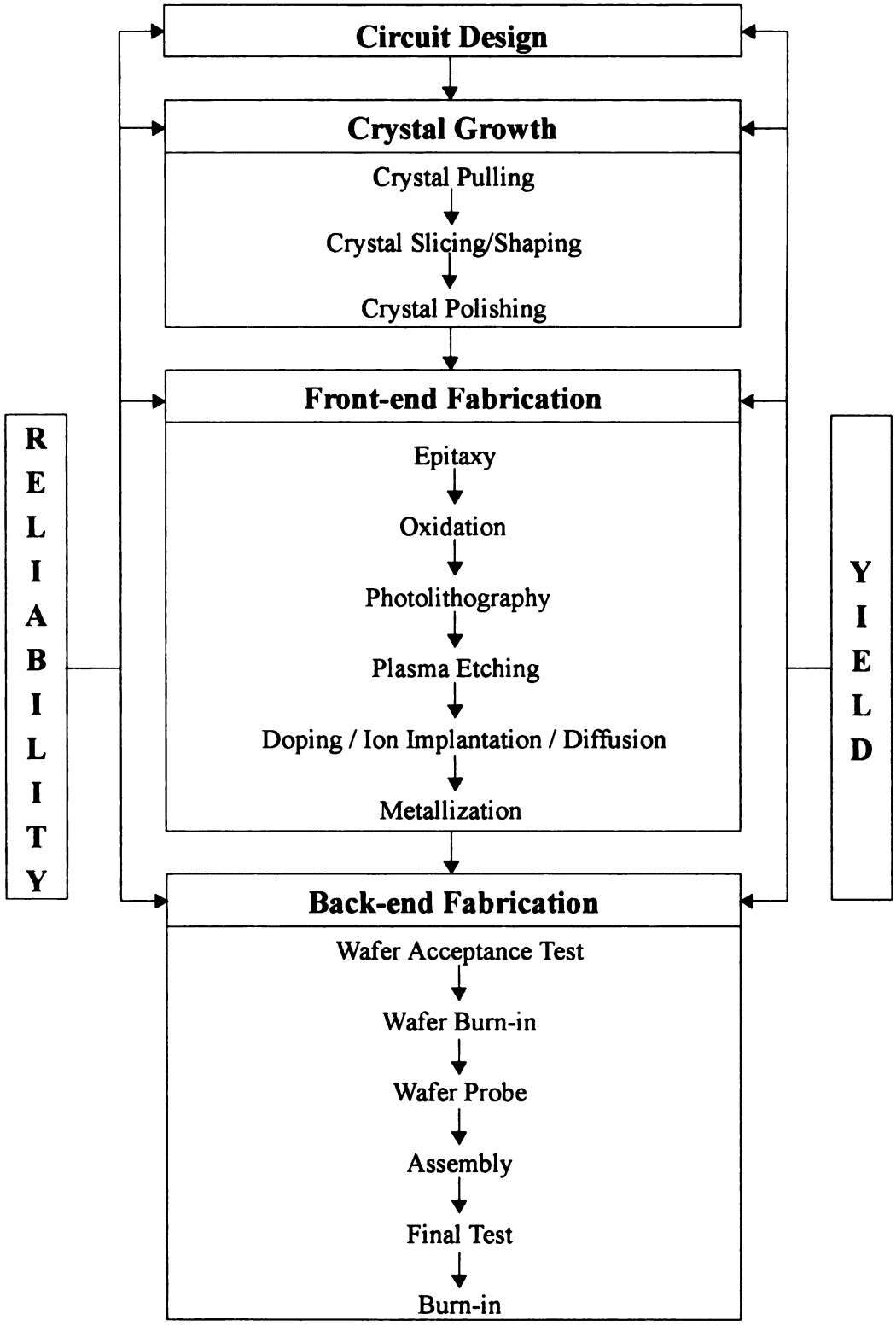


Figure 1.2: Reliability's Influence on IC Manufacturing Process and Yield [5].

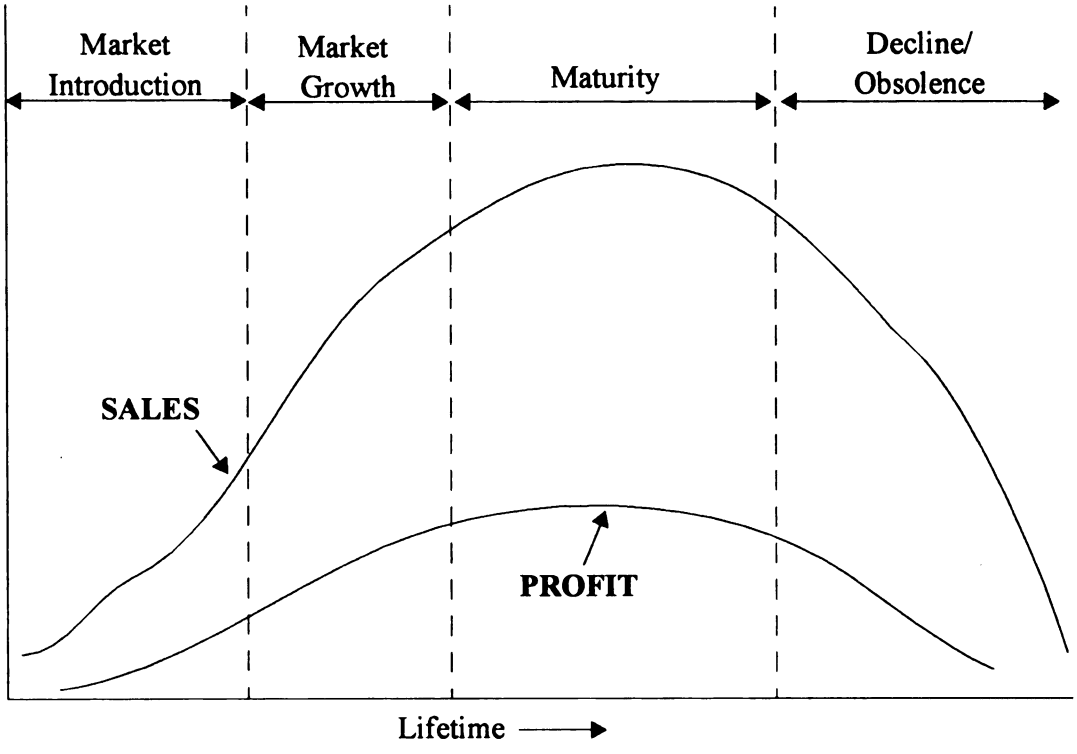


Figure 1.3: Typical Product Life Cycle and Profit Revenue.

Verification and analysis are serious bottlenecks for the timely design of systems and ICs. Verification can take up to one half of the total design process time. The time and added cost associated with life testing and screening is a major issue in product qualification. The dilemma is to determine how much time can be spent on life testing or screening; as broadly indicated in Figure 1.4. If too little qualification is done, then the product will be unreliable and this would eventually result in loss of customer. On the other hand, too much qualification bears the risk of not being competitive in the market at all. Hence, very high accelerated reliability testing techniques need to be developed that can screen products and that can be related to actual field reliability. Electrical or other screening that could eliminate burn-in is highly desirable, so that time to market can be reduced.

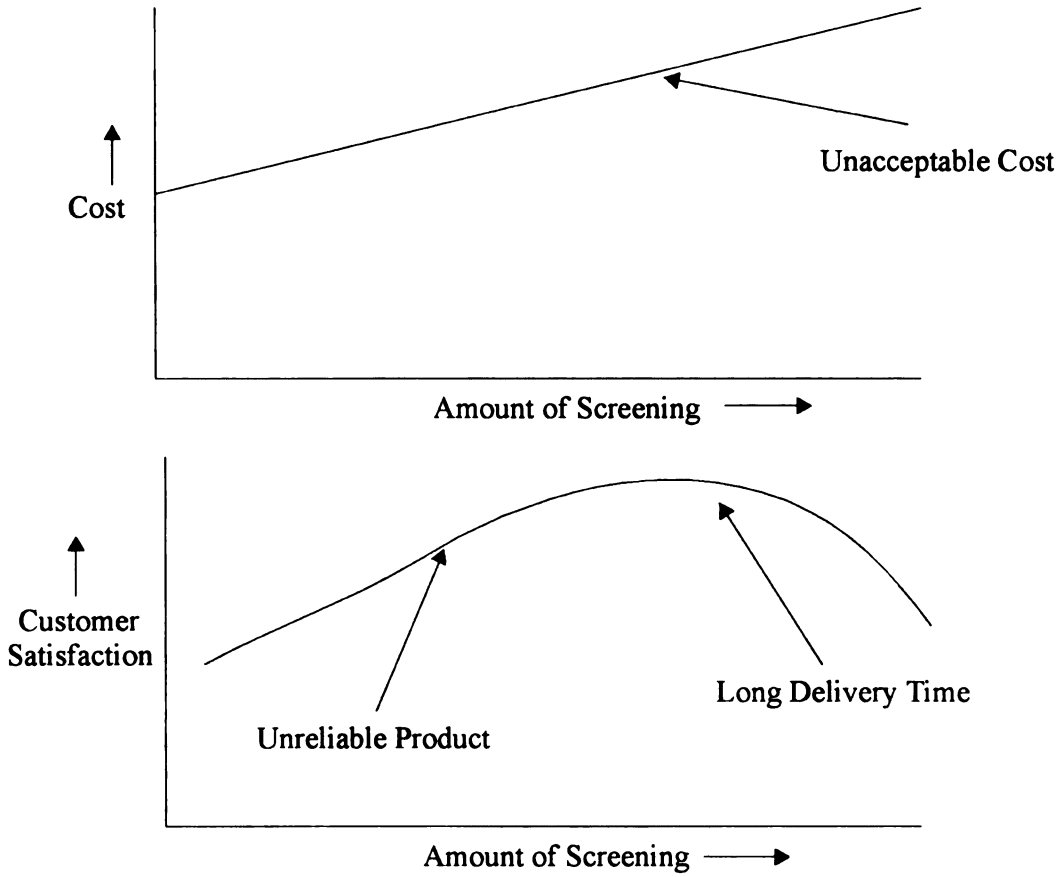


Figure 1.4: Effects of Amount of Screening on (a) Cost; (b) Customer Satisfaction.

There are four major issues in system integration: marketability, manufacturability, testability, and reliability concerns. Yield and reliability are two important factors affecting the profitability of semiconductor manufacturing. In this study, the emphasis is on the quality and reliability enhancement of CMOS analog/mixed-signal ICs at a low cost

Oxide defects have been found as one of the major causes for the reliability problems of CMOS integrated circuits [7,13-16]. Particulate contamination, crystalline defects in the substrate, spot defects, localized thin regions, or surface roughness can cause localized weak spots in an oxide [17]. In the manufacturing process of modern VLSI semiconductor devices, a plasma is often used to deposit or remove material on wafers [18,19]. The

plasmas may cause destructive charges to be built on the wafers. If the charge buildup is large enough, and the charge has no other leakage path to substrate, a current flows through the connected transistor gate-oxide, causing degradation of the gate-oxide. The damaged gate-oxide may result in performance degradation of the affected transistor and cause reliability failures. Moreover, the quality and lifetime of a gate-oxide strongly depend on its thickness [16,20]. Oxide thinning occurs when the oxide thickness of a transistor is physically or effectively thinner than expected. Oxide thinning can be due to localized thin spots, traps in the oxide, surface asperity, or locally reduced tunneling barrier height [21]. It can shorten the lifetime of a gate-oxide, increase oxide leakage current, or cause time-dependent dielectric breakdown.

The term reliability as applied to an IC refers to the time that the circuit continues to work to specification after it has passed its initial tests and been designated as a yielding part [16]. Most electronic devices exhibit a decreasing failure rate in their early life; which results from the weak individuals that have shorter lives than the normal (stronger) ones. If the weak devices are released to customers or are used to assemble modules or systems, many of these defects will cause failure in their early lives; from experience it is known that, quite a few failures can be observed in the first year [5]. These high number of early stage failures are called early life failures, as shown in Figure 1.5. After early life, devices move into the steady-state hazard rate period with an almost constant failure rate. This corresponds to the normal operation of an electronic device and extends well beyond the expected life of most devices. Wear-out failures occur when the strong population reaches its end of life. Due to rapidly evolving manufacturing technology, the wear-out process may

not be detected during product operational life [22]. For most semiconductors, wear-out does not occur until the product becomes obsolete or reaches its end of shelf life in case of military systems (usually 10 years).

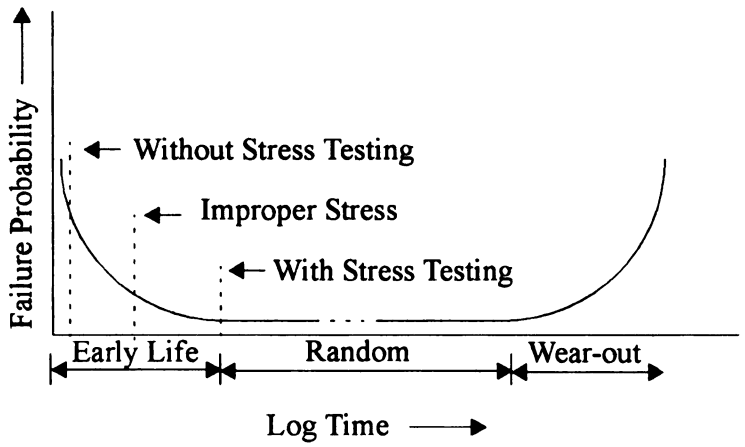


Figure 1.5: Bathtub Curve.

Stress testing is a technique used to weed-out early life failures by applying higher than usual levels of stress to speed up the deterioration of electronic devices. The concept of this screening process is to accelerate the lifetime of devices such that they begin operation with a failure rate beyond the early life region. The industry standard methods for screening have been high-temperature burn-in [23-25] and high-voltage screening [26,27]. Burn-in is effective in varying degrees for almost all circuits and assembly causes of premature failure. Burn-in screening decreases failure rate of a product during the early field life, where overall cost and turn around time are of concern. The added manufacturing cost may range from 5% to 40% of the total product cost, depending on burn-in time, qualities

of ICs, and product complexity. For example, Figure 1.6 shows a typical multichip module (MCM) production flow where burn-in is performed at two places [28], and Figure 1.7 illustrates high-reliability screening flow, for plastic encapsulated packages, used by Maxim, where burn-in is performed for 160 hours [29]. During high-voltage screening, voltages exceeding normal operating conditions are applied to ICs. For example, a 3.3V technology may be stressed at 6V whereas 8V may be used to stress a 5V technology.

1.1 Motivation

This study was motivated from investigating the yield loss of a safety-critical mixed-signal IC, namely Product X. From the manufacturing test data, the yield of Product X was approximately 80%, where 60% of the 20% yield loss, i.e., 12% in total, was caused by its analog modules. The IC was tested by a high-voltage stress test followed by a burn-in test to eliminate early life failures. From these manufacturing data [30], it was concluded that, if the analog modules were stressed properly by high voltage, almost all of them would be identified and screened out during the high-voltage stress test process. As a result, there would be much less or no need of the burn-in test for defective analog modules. As illustrated in Figure 1.5, high failure rate may still occur if the circuit is not properly stressed. This was the major reason causing the lost yield. The question that naturally arises is how do we properly stress analog modules? This led to this study for characterizing the parameters that affect the stressability of analog circuits and developing an analytical approach for conducting the high-voltage stress test.

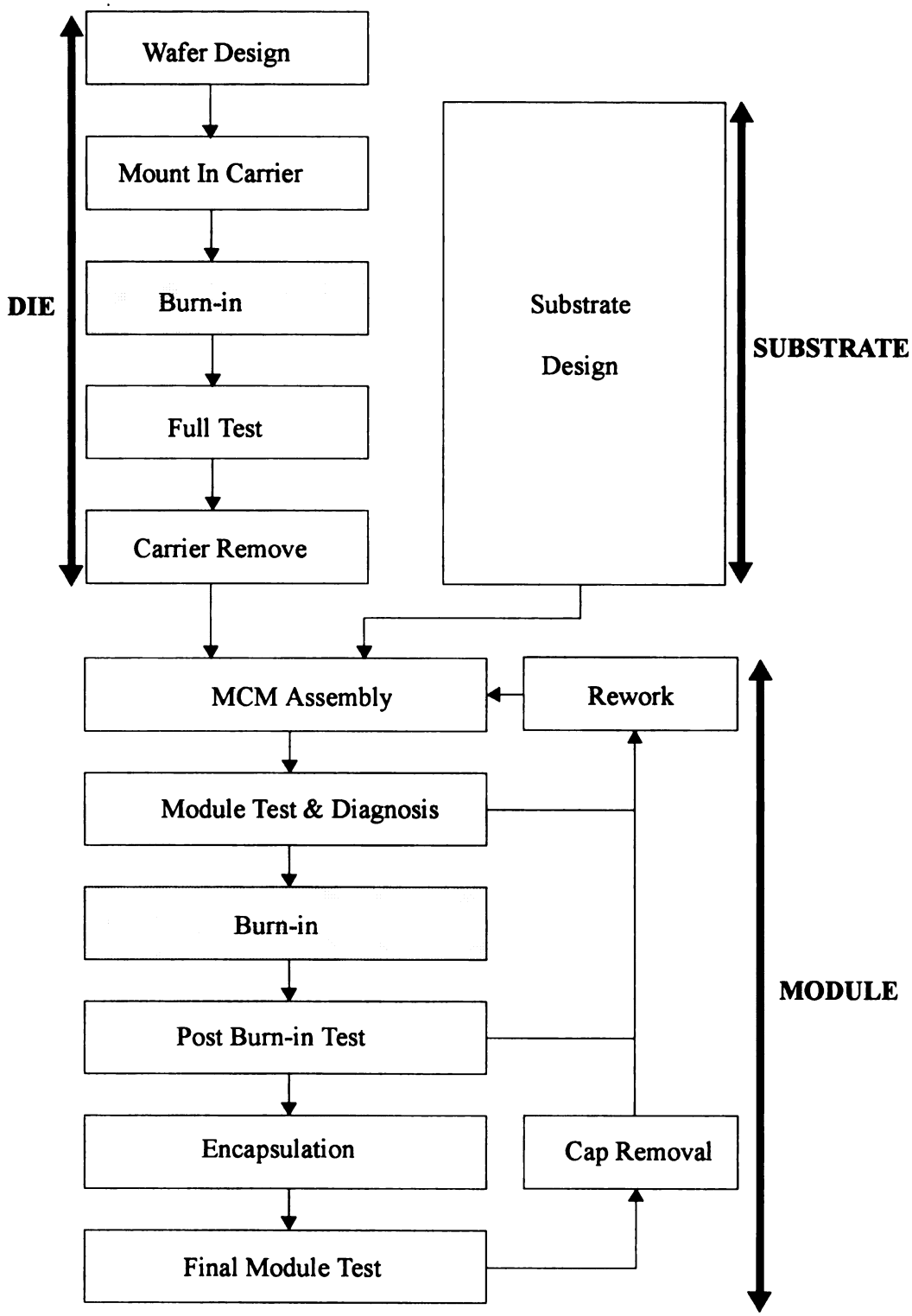


Figure 1.6: MCM Production Flow.

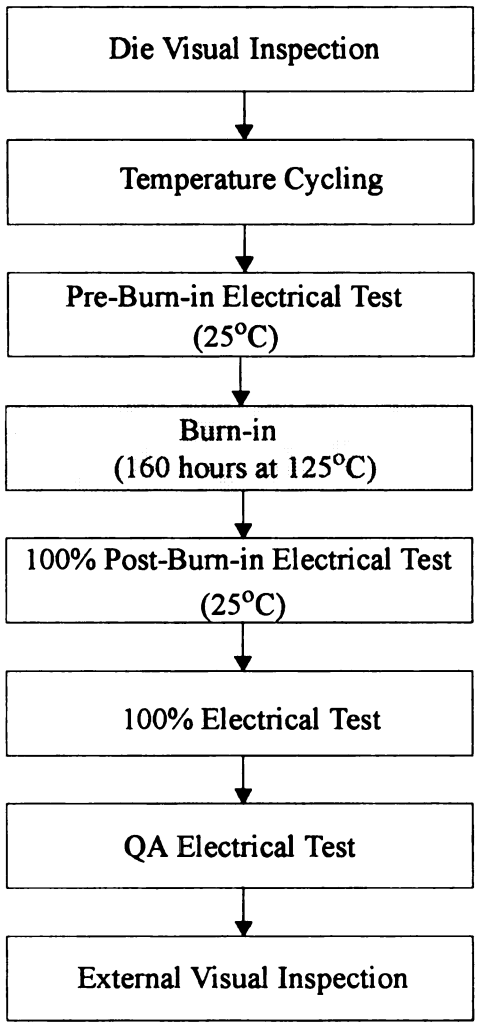


Figure 1.7: Maxim High Reliability Screening Flow.

1.2 Objectives

The research goal is to improve the quality and reliability of CMOS ICs without performing the high-cost burn-in process. Extreme-voltage stress testing can be used to

eliminate early life failures caused by defective gate-oxides in CMOS ICs. The voltage level to be applied is selected to be the highest possible under the constraint that there is no introduction of failure mechanisms which are not encountered during operating life. This voltage level, hereafter, will be referred to as the extreme-voltage. Extreme-voltage screening has been successfully implemented to enhance gate-oxide reliability of digital CMOS ICs. However, the success has not yet been extended to its analog counterparts. Today, almost all IC manufacturers employ the digital circuit screening process for the analog modules in mixed-signal CMOS ICs. In this study, an attempt has been made to analytically develop an efficient yet effective extreme-voltage stress test process for analog CMOS circuits. The research objective is to develop the framework of an automatic stress test system for analog circuits, as depicted in Figure 1.8, that integrates three major components -- stress vector generator, stressability analyzer, and stressability design methodology.

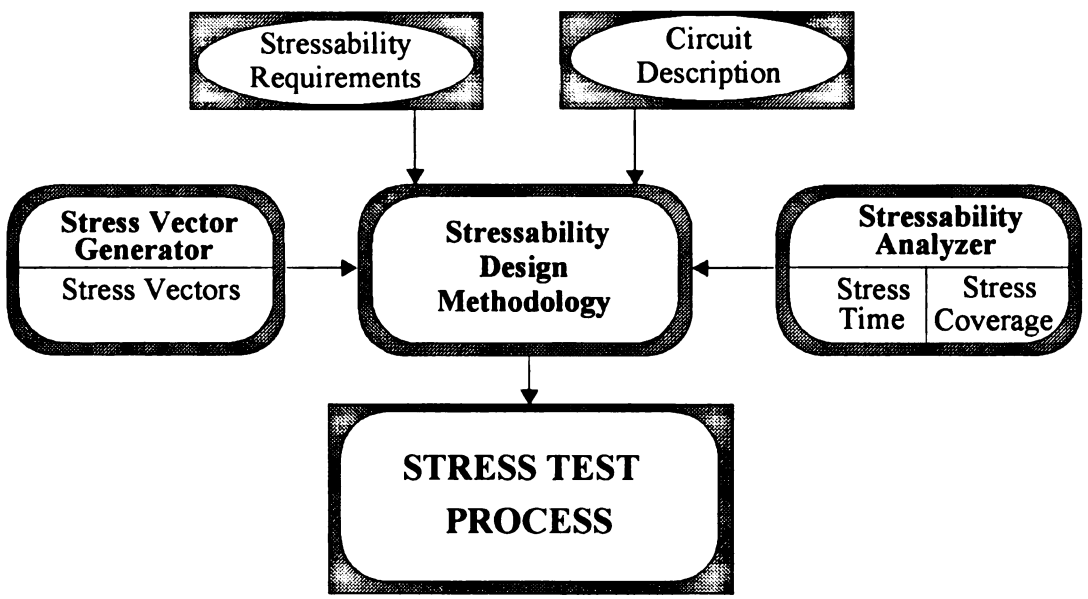


Figure 1.8: Framework of Stress Test System.

Stress vector generator and stressability analyzer are analysis tools used by the stress test process to determine the stressability of a given circuit. Stress vector generator can provide optimal stress vectors for the given circuit and stressability analyzer determines the stressability of the circuit based on the selected stress vectors. Stressability design methodology of the stress test process is employed during design phase to ensure desired stressability of the circuit. This design methodology utilizes the two analysis tools and stressability enhancement strategy to meet stressability requirements during the design phase.

1.3 Organization

The dissertation is organized as follows. Chapter 2 presents information related to the reliability of ICs. It also briefly reviews dominant failure mechanisms and currently used reliability stress tests for the CMOS technology.

Chapter 3 presents the basic concepts and definitions required for proper stressability of analog CMOS ICs. It also describes the development of a stress vector generation procedure and a stressability analysis process. This chapter also presents algorithms to optimize the stress vectors based on the results of stressability analysis of a circuit, under different stressability constraints.

In order to effectively apply the extreme-voltage stress to large circuits, Chapter 4 describes hierarchical stress vector generation based on the topological structure. It also presents a circuit decomposition model to reduce the computational complexity of the problem. This chapter then deals with stressability enhancement of portions of the circuit

having poor stressability, that are identified by stressability analysis, and presents a stressability enhancement strategy using additional hardware. It also describes the developed stressability design methodology with the goal to achieve desired stressability during design phase.

Chapter 5 first introduces the recent IC reliability trends in industry. It then analyzes the existing burn-in and suggests ways to improve it. This chapter compares the developed extreme-voltage stress test with extreme-temperature stress. It also discusses practical issues related to the implementation of extreme-voltage stress test.

Finally, Chapter 6 summarizes the dissertation, outlines major contributions and proposes some directions for future research.

Chapter 2

BACKGROUND

Customers always expect high reliability, low cost, and versatile goods. According to the Sematech Reliability Roadmap, since 1994 reliability levels are no longer continuously improving and there is a real risk of degradation of reliability in the rapid, discontinuous evolution of the IC industry. Recently, CMOS has become the dominant IC technology because of its low power dissipation which is of extreme importance for full portability of electronic systems. However, degradation of CMOS IC reliability is caused by the failure mechanisms such as oxide breakdown, hot carrier aging, electromigration, and etc. Among these failures, gate-oxide shorts have been found to be the most dominant failure mechanism in CMOS ICs.

This chapter first reviews the information related to the IC reliability, describes the principal failure mechanisms that affect the CMOS IC reliability, and finally discusses the existing methods that enhance CMOS gate-oxide reliability.

2.1 IC Reliability

Reliability is defined as the probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered. Hence,

the probability that a system successfully performs as designed is called system reliability. In general, reliability is the ability or capability of the product to operate properly without failure, which is defined as the cumulative probability function at time t for a given time under the operating conditions. True reliability is never exactly known, which means the exact numerical value of the probability of adequate performance is not known, but numerical estimates quite close to this value can be obtained by the use of statistical methods.

2.1.1 Reliability Requirements

There are three reliability requirements. *Initial quality* is determined by the devices that fail incoming inspection and is measured in ppm (parts per million). *Early failures* are the failures during the early life also measured in ppm. *Long term reliability*, measured in fits, is determined by the device failures after the early life period. A FIT is the failure unit and is defined as one failure in 10^9 hours of operation.

2.1.2 Bathtub Curve

Systems and materials start to wear-out when they are used, and they can fail due to various failure mechanisms. A study of many systems during their normal life expectancy has led to the conclusion that failure rates follow a certain basic pattern. It has been found that systems exhibit a high decreasing failure rate during their initial period of operation, called the *early life* period. The operating period that follows the early life has a smaller failure rate and is called the *useful life* period, which tends to remain constant until the beginning of the next phase, called the *aging period* or *wear-out* period. This typical hazard

rate behavior of devices and systems is known as the *bathtub curve*.

The bathtub curve, as shown in Figure 1.5, is basically an addition of three curves. As mentioned, most electronic devices exhibit a decreasing failure rate in their early life, which is caused by the weak individuals in the population. If these weak devices, having a significantly shorter life than the normal ones, are released to customers or are used to assemble modules or systems, many of these will cause failure in their early lives; from experience it is known that, quite a few failures can be observed in the first year. These early-stage high number of failures are called early life failures. Note that early life is viewed for the whole lot instead of for a single device. This failure curve is shown in Figure 2.1(a).

After early life, devices move into the steady-state hazard rate period with an almost constant failure rate. The constant failure rate is defined as the percentage of product failures that occurs during the time period when failures are generally attributed to statistical chance rather than early life or product wear-out. This corresponds to the normal operation of an electronic device and extends well beyond the useful life of most devices. This failure curve is illustrated in Figure 2.1(b). Wear-out failures occur when the strong population reaches its end of life. It has been reported that for most semiconductor devices wear-out does not occur until product obsoletes. The wear-out failure curve is depicted in Figure 2.1(c).

2.2 CMOS Reliability

A failure mechanism is defined as the chemical, electrical, physical, mechanical, or thermal process leading to failure. Failure mechanisms in the microelectronic devices are

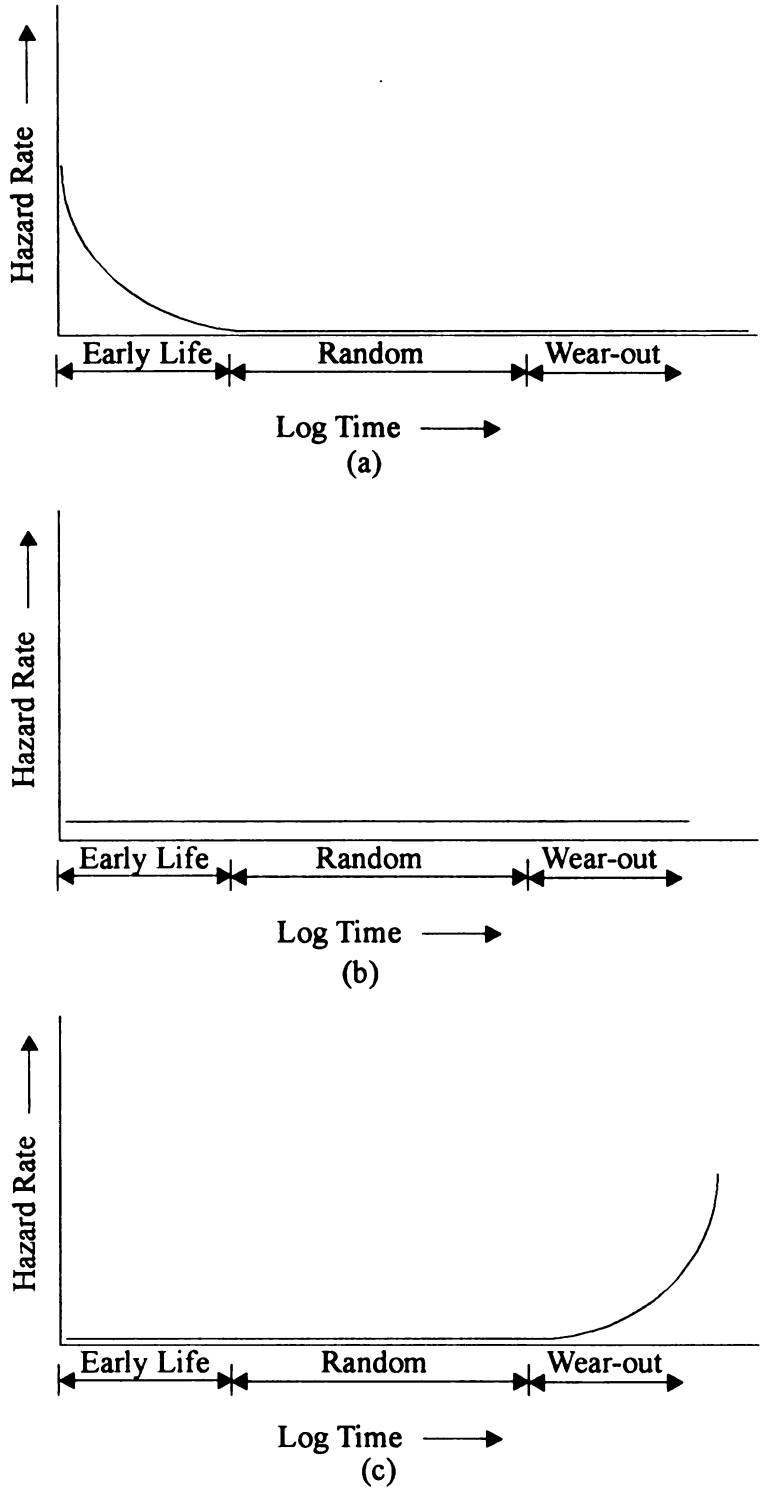


Figure 2.1: Failure Curves: (a) Early Life; (b) Random; and (c) Wear-out.

either categorized according to the type, or mode, of the failure precipitated, or according to the root cause, or mechanism. Failure modes include shorts, opens, or parametric drifts.

2.2.1 Failure Mechanisms

The principal device failure mechanisms in CMOS devices are oxide breakdown, hot carrier aging, and electromigration [31].

Oxide breakdown is a failure mechanism which occurs in thin dielectric films, such as is used for the gate in a MOS device. The dielectric strength of an oxide layer is often expressed in terms of the electric field at which it loses its insulating properties and the insulator is irreversibly damaged. With a sufficiently high applied electric field, and after continued exposure to a constant applied voltage, dielectric films will breakdown. This phenomenon is called Time-Dependent Dielectric Breakdown.

Hot carrier aging is defined as the degradation of intrinsic MOS device characteristics due to the trapping of charge in the gate dielectric. Hot carriers are those carriers (electrons or holes) that are not in thermal equilibrium with the rest of the semiconductor crystal lattice, i.e., the carriers whose energy does not correspond to that of the conduction and valence band edges. These are the carriers in the channel and pinchoff regions of a transistor which have gained much energy from the lateral electric field produced by the drain to source voltage. This situation arises when the electric field seen by the carriers causes carrier acceleration between the collisions with the lattice, i.e., within a mean free path, to be sufficiently high to change their average energy. High energy hot carriers can cause a number of effects within a MOS device, including a change in device threshold voltage and transconductance, eventually resulting in circuit failure.

Electromigration is a failure mechanism which occurs in the interconnect metallization tracks. It is the motion of ions in conductors, such as aluminum, due to the passage of electric current through it. Thermally activated ions of the conductor, which normally self-diffuse in all directions, are given a direction of net motion due to momentum transfer from the conducting electrons. Hence, the ions move downstream with the electrons. A divergence of the ion flux ultimately results in circuit failure. A positive divergence leads to an accumulation of vacancies to form a void in the metal and ultimately an open circuit. A negative divergence leads to a buildup of metal, called hillock, which can eventually lead to a short with adjacent or overlying metal.

2.2.2 CMOS Gate-Oxide Reliability

Gate-oxide shorts have been found to be the most dominant failure mechanism in high reliability CMOS integrated circuits [7,14,15,32]. Gate-oxide defect is a transistor defect that causes a relatively low impedance path between a MOSFET gate and the underlying silicon (substrate, p- or n-well, source, drain). A gate-oxide defect in one or more transistors in a circuit may or may not effect the functionality of the entire circuit at normal operating voltage, i.e., the circuit may continue to meet the design specifications. But, such defects can subsequently change and effect the functionality of circuits. This would cause early-life failures of the circuits that contain transistors with gate-oxide defects. Thus, the devices will not meet the reliability requirements.

Gate-oxide short in a MOS transistor results when the oxide between the gate and the underlying silicon breaks down. If the short is close to the source, a gate to source path is

established. Similarly, a gate to drain path is established if the short is close to the drain. An example of a gate-oxide short in an NMOS transistor is shown in Figure 2.2. Structural non-uniformities in the gate-oxide or in the silicon substrate beneath the gate-oxide or in the silicon substrate can occur during the fabrication process. Metal ions, particulates, and crystalline defects reduce the dielectric strength of the gate-oxide.

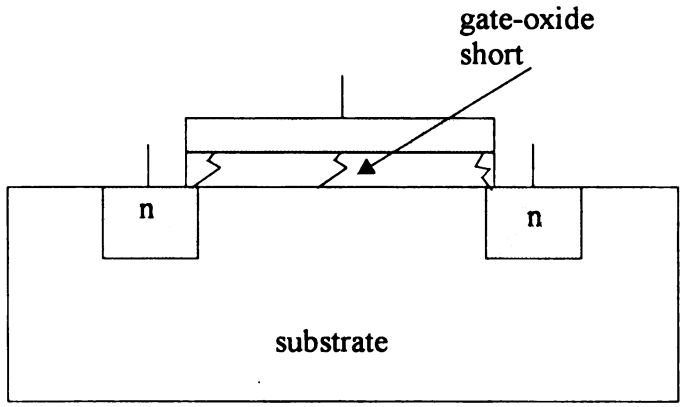
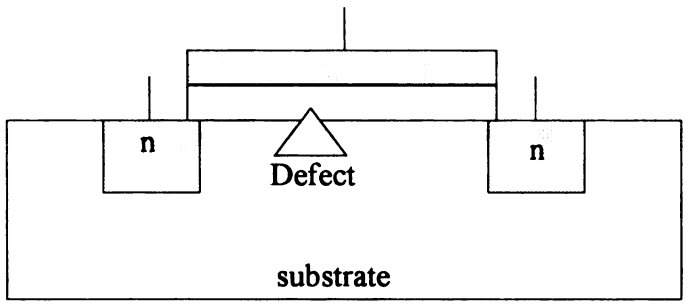
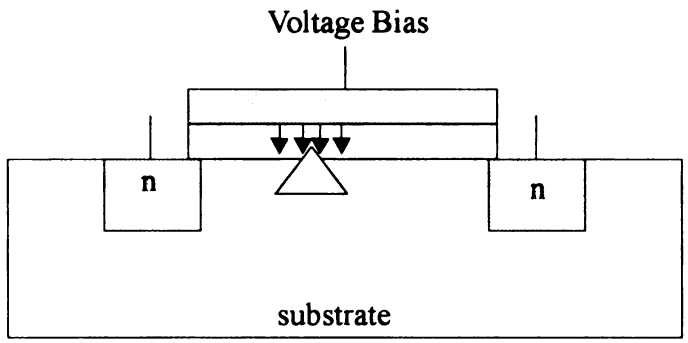


Figure 2.2: Gate-Oxide Shorts.

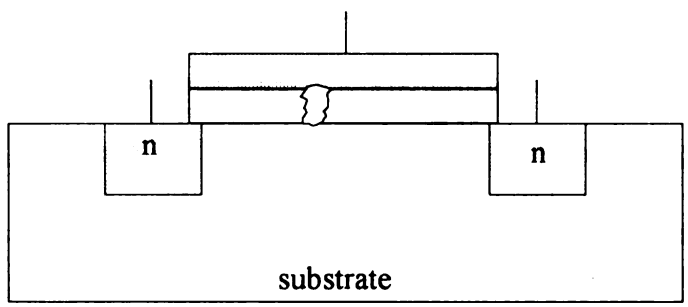
Time dependent shorting process of gate-oxide is shown in Figure 2.3 [33]. Defects, such as in Figure 2.3(a), alter Si crystal or SiO_2 geometry such that an abnormally high local electric field is present when a voltage is applied to the transistor gate as in Figure 2.3(b). This higher electric field causes tunneling of electrons into the gate-oxide and a subsequent acceleration of charge through the gate-oxide. This time dependent buildup of local gate-oxide electric field leads eventually to a rupture as shown in Figure 2.3(c). The time required for rupture to occur depends on defect geometry, duration of applied gate voltage, gate-oxide thickness and magnitude of voltage and temperature applied to the transistor.



(a)



(b)



(c)

Figure 2.3: Conversion of a Gate-Oxide Defect to a Gate-Oxide Short:
(a) a MOSFET having a Physical Defect at Gate-Oxide to Well Interface;
(b) Application of High Electric Field to Gate-Oxide from Gate Bias;
and (c) Conversion of the Gate-Oxide Defect to Gate-Oxide Short.

During the random period of bathtub curve, semiconductor devices normally have low failure rates and long working lives. Failure rate can be down to 0.005% per thousands hours or lower. It is necessary to take actions to force the devices, that lie in the early life region, to fail within convenient time scales. Increase in failure rates is achieved by application of controlled stress. Stress tests or screens are defined as tests that subject 100% of the parts to the test conditions. Product quality and reliability are dependent upon the screens used to detect defects that are introduced during the manufacturing process. Stress tests subject devices to higher than usual levels of stress (e.g., voltage, temperature, humidity, corrosion, magnetic field, current, pressure, radiation, vibration, salt, and loading). Thus, stress testing is a technique used to screen out early life failures by accelerating the deterioration of electronic devices using higher than usual levels of stress.

2.3 Stress Tests

Figure 2.4 illustrates the concept of the screening process that accelerates the lifetime of devices such that they begin operation with a failure rate beyond the early life region [34-36]. Because of rapidly changing technologies, customer expectations for higher reliability, and more complicated products, reliability stress tests have become very important. Environmental Stress Screening (ESS) is a process of eliminating defective parts from the production batch [37]. In order to be competitive in the market, semiconductor companies are trying to achieve a failure rate as low as possible, referred to as the zero defects approach. The current industry practice is to use *burn-in* and *high voltage* stress testing for CMOS technology.

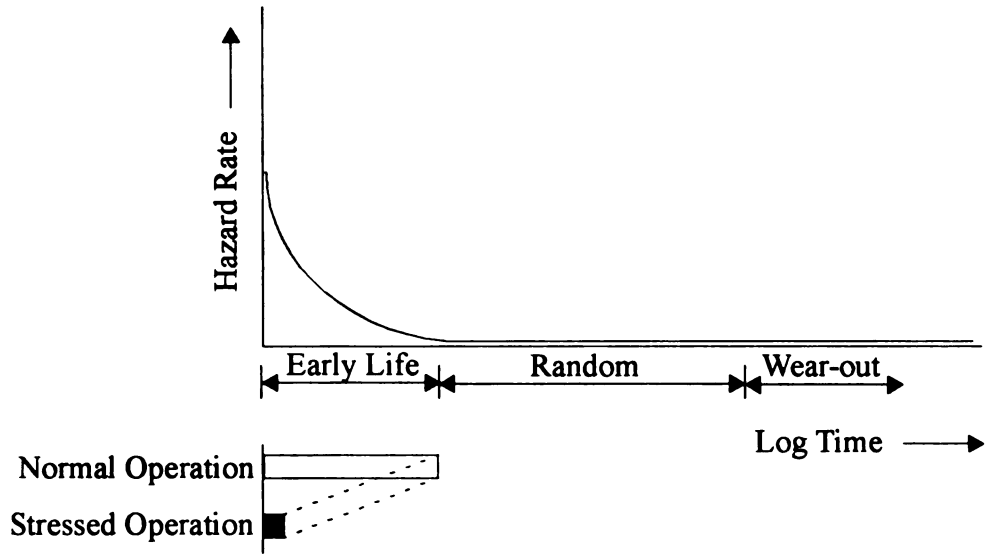


Figure 2.4: Time-to-Failure for Normal and Stressed Operation.

2.3.1 Burn-in

Burn-in can be defined in a number of ways. It is a pre-usage operation of components performed in order to screen out the substandard components, often in a severe environment [38,39]. Burn-in is also described as an effective means for screening out defects contributing to early life failures [40]. It has also been described as running units under design or accelerated conditions for a suitable length of time [41]. However, in some definitions burn-in has been restricted to high stress only [42]. It has also been defined as a technique used to weed-out early life failures by applying higher than usual levels of stress to speed up the deterioration of electronic devices. Burn-in is complete when all the weak devices have failed, thus leaving the devices in a reliable state [5].

Burn-in is an effective screening method used in predicting, achieving, and enhancing field reliability of ICs. It is the most effective screen in detecting die related faults through time, bias, current, and temperature accelerating factors to activate the time-temperature-dependent failure mechanisms to the point of detection in a relatively short period of time. Early life failures are greatly influenced by the burn-in test, and a significant improvement in failure rates have been reported after burn-in. Today, almost all IC manufacturers perform 100% burn-in for various durations to screen defective products. One of the major problems associated with burn-in is the determination of exactly how long the burn-in process should continue, balancing appropriately the needs of reliability and the total costs [38].

Burn-in occurs between two final test stages, which are called the pre- and post-burn-in tests. A typical burn-in chamber consists of a temperature oven and a control system, which is either a PC or a workstation. The temperature oven contains power supplies, compressors, and an interface with the control system. Usually, a driver board goes with each burn-in board (BIB) to send and receive signals to the devices under test. The accuracy, channel number, and precision of the driver boards greatly affect the price of the burn-in chambers. For static and simple dynamic burn-in systems, the driver boards may only send DC bias or simple signals to the devices under test (write operation) and may not be able to get the feedback of the devices (read operation). For these cases, devices are to be tested by independent testers and are to be loaded and unloaded from the burn-in chambers and testers for pre- and post-burn-in tests. This is a time consuming task and liable to damage due to mishandling. Mishandling may cause lead bending, ESD, and lost devices. The

modern test during burn-in (TDBI) system is able to perform write and read operations with short time delays, which makes it capable of detecting any failure during burn-in [43]. The control system is composed of a database to store programs and test results and the software tools for designing test programs, setting biases and temperatures, and writing the test patterns during stressing and functional tests.

Burn-in time can be optimized based on the reliability requirements and trade-off between costs associated with burn-in and the field savings obtained from reduced failures. Two major issues must be checked when considering reduction of burn-in time. First, the stability of the burn-in failure rates must be established. The monitoring duration may be as long as three months to ensure that the trend of the burn-in failure rates can be verified. Second, one or more experiments simulating the effects of reduced burn-in time must be conducted. Another possible way to reduce burn-in time is to develop the wafer level burn-in. However, there are technical difficulties that need to be solved, such as wafer expansion coefficient of temperature, wafer handling, and a feasible method for applying bias and test patterns.

Various burn-in methods are currently being used in the industry [7]. The types of possible defects and the extent to which these defects are activated by the various techniques determine the type of burn-in appropriate for a given device. The type of burn-in is selected based on the trade-offs among customers' requirements, IC performance and characteristics, available equipment and raw materials, technology, time to market, and cost. Some common types of burn-in are *static*, *steady-state*, *dynamic*, *test during burn-in*, *sequential*, and *Markov burn-in*.

2.3.2 Extreme-Voltage Stress

As already discussed, gate-oxide defects are the major cause of quality and reliability problems in CMOS ICs. Time-dependent oxide breakdown occurs at weaknesses in the oxide layer due to poor processing or an uneven oxide growth. Failures of MOS devices due to oxide breakdown during device operational life should be as low as possible, so it is very important that any defective gate-oxides are detected before or at the final testing stage. An extreme-voltage gate stress can be more efficient for screening out defective oxides than burn-in because of the relatively small activation energy of only approximately 0.3 eV for thermally accelerating the oxide failure process [15].

Extreme-voltage stress testing aims at screening early life failures and intermittent failures to improve the reliability and quality level of CMOS ICs at a low cost without performing burn-in. It is useful at wafer sort, as it can screen out weak parts during a wafer-level test and avoid the cost of packaging them. This procedure has been practiced in industry [44,45]. Some data which showed the effectiveness of extreme-voltage stress testing were reported recently [26,27,46].

In general, the breakdown is a result of charge trapping in the oxide due to the excessive field and the current in the SiO_2. The lifetime of an oxide is determined by the time required for the charge to reach a critical value [20]. The time-to-breakdown of the oxide, t_{BD}, can be modeled as follows.

$$t_{BD} = \tau_0 \exp(GT_{eff}/V_{ox}) \qquad (2.1)$$

where V_{ox} is the voltage across the oxide and T_{eff} is the effective oxide thickness, τ_0 is determined by the intrinsic breakdown time under an applied voltage of V_{ox}, and G is the

slope of $\log(t_{BD})$ versus $1/E_{ox}$ (the electrical field across the oxide). The typical values of G=350 MV/cm and τ_0=10 picoseconds were suggested in [20].

Magnitude and duration of the applied voltage can be used to control the time-to-breakdown of an oxide. Increasing the magnitude of applied voltage reduces the time to rupture of an oxide exponentially as shown in Figure 2.5. To stress defective oxides effectively while still avoiding damage to flawless oxides, the electrical field E_{ox} must be controlled so that no excess tunneling currents are caused across the oxide. Fowler-Nordheim tunneling currents may occur across flawless oxides if E_{ox} is larger than a critical value. The excess tunneling currents flowing through gate-oxides can damage the oxides resulting in increased oxide leakage currents even after the supply voltage is reduced to the normal operating voltage. Experimental results show that the critical value of E_{ox} is approximately 6 MV/cm [47-51]. Therefore, the stress voltage V_{stress} can be defined as the voltage that will result in application of the critical E_{ox} across the oxide, i.e.,

$$V_{stress} = T_{ox} * 6 \text{ MV/cm} = T_{ox} * 0.6 \text{ V/nm} \qquad (2.2)$$

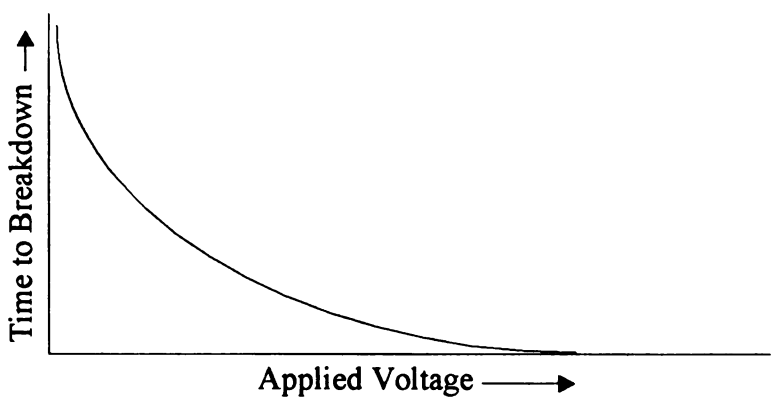


Figure 2.5: Relationship between Time to Breakdown and Applied Voltage of an Oxide.

Consider the HP AMOS14TB 0.5μm process technology [52], where the thickness of the gate-oxide is 9.7nm. Table 2.1 lists the lifetime of gate-oxide for various thicknesses, using Eqn. 2.1, at the normal operating voltage of 3.3V [53].

Table 2.1: Lifetime of Gate-Oxide for Various Thicknesses at 3.3V

Thickness (nm)	Lifetime	Thickness (nm)	Lifetime
9.70 (100%)	1.52×10^{26} years	4.85 (50%)	6.93×10^{3} years
8.73 (90%)	5.16×10^{21} years	3.88 (40%)	86.17 days
7.76 (80%)	1.76×10^{17} years	2.91 (30%)	4.22 minutes
6.79 (70%)	5.98×10^{12} years	1.94 (20%)	8.63 ms
5.82 (60%)	2.04×10^{8} years	0.97 (10%)	294 ns

Since, for this technology the gate-oxide thickness T_{ox}=9.7nm, so the stress voltage using Eqn. 2.2 is 5.82V. The lifetime of gate-oxide for various thicknesses at the stress voltage of 5.82V is listed in Table 2.2.

Table 2.2: Lifetime of Gate-Oxide for Various Thicknesses at 5.82V

Thickness (nm)	Lifetime	Thickness (nm)	Lifetime
9.70 (100%)	6.84×10^{6} years	4.85 (50%)	46.44 seconds
8.73 (90%)	20028.7 years	3.88 (40%)	136 ms
7.76 (80%)	58.65 years	2.91 (30%)	398 μs
6.79 (70%)	62.69 days	1.94 (20%)	1.17 μs
5.82 (60%)	4.41 hours	0.97 (10%)	3.41 ns

For an oxide to have a lifetime of one year at 3.3V, by Eqn. 2.1, the thickness of oxide must exceed 4.016nm. To effectively stress the oxide at 5.82V, by Eqn. 2.1, the stress

duration is approximately 0.31 seconds. It should be mentioned here that a flawless oxide, i.e., with a thickness of 9.7nm, has a life of 1.52×10^{26} years at 3.3V, while it has a life of 6.84×10^{6} years at 5.82V.

Consider the bathtub curve in Figure 1.5. Semiconductor devices normally have low failure rates and long working lives during the Random Period. Failure rate can be down to 0.005% per thousand hours or lower for a qualified process. For CMOS circuits, the early life regime is approximately six months. For simplicity of discussion and having a margin of safety, we shall assume the time period for early life to be one year.

Figure 2.6 plots a general thickness distribution of fabricated devices for a qualified process which meets the reliability mode, i.e., bathtub curve. Here, Point A is the thickness of the oxides corresponding to the end of the early life period. For example, if the time period for early life is one year, the oxide thickness is 4.016nm at 3.3V. In this process, the thickness of the flawless oxide is the nominal 9.7nm. In practice, it is reasonable to assume that the process variation of a qualified process is within 10%. Thus, Point B is the oxide thickness corresponding to the minimum thickness of non-defective oxide. In this case, Point B is 8.73nm, i.e., 90% of the nominal 9.7nm. According to the bathtub curve, the devices distributed between Points A and B are relatively low and the failure rate is acceptable when the time period of early life was defined. As mentioned previously, the stress time for the thickness of 4.016nm is 0.31 seconds at 5.82V. On the other hand, the lifetime for 8.73nm is 20028.7 years at 5.82V. The non-defective oxides can be stressed for a reasonably long time without overstressing the devices [53].

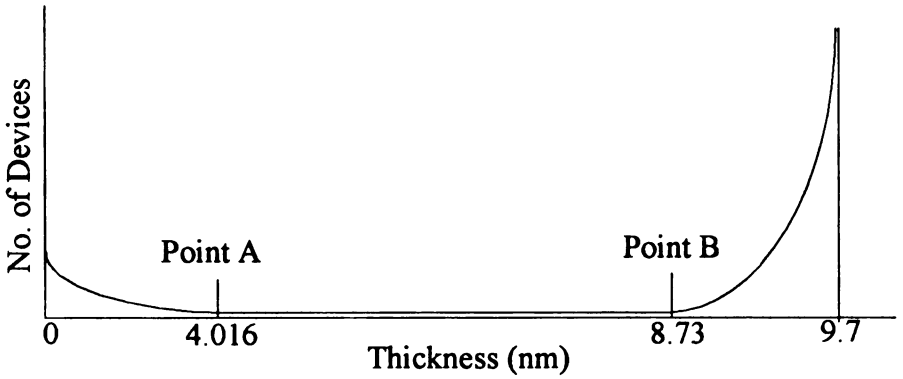


Figure 2.6: General Thickness Distribution of Fabricated Devices.

Recently, an efficient stress test program, namely, SHOVE (Short Voltage Evaluation), was developed [26,27] and it aims at screening early life failures and improving the quality level of digital CMOS ICs. More specifically, a transistor in a CMOS IC must be stressed for enough time during this testing. To effectively stress an NMOS transistor, the gate of the transistor should be held at the stress voltage and the drain and source of the transistor at 0V. Similarly, to effectively stress a PMOS transistor, the gate of the transistor should be held at 0V and the drain and source of the transistor at the stress voltage. The stress voltage is derived from Eqn. 2.2, and the stress time is defined from Eqn. 2.1 with the application of the stress voltage. Each stress vector must be held for at least the stress time for a transistor to make sure all transistors in a CMOS IC are stressed for enough time. Because not all transistors are stressed by each stress vector, some transistors may be stressed longer than others. During SHOVE, test vectors are run at higher-than-normal supply voltage for a short period.

As mentioned, IDDQ test sets that target inter-gate bridging faults are not suitable for SHOVE testing. To stress all the transistors evenly and reduce the stress time for fully complementary CMOS logic gates, all-one and all-zero vectors can perform better than stuck-at test sets as the stress vectors. The algorithm for line justification used in existing ATPG programs can be modified to generate all-zero and all-one vectors. However, justification is more complicated than existing ATPG programs and may be impossible due to the structure of a circuit under test.

Chapter 3

STRESS TEST OF ANALOG CIRCUITS

As mentioned, analog circuits have been stressed improperly by applying the stress vectors generated for digital circuits. This chapter first introduces the fundamental concepts required for stressability of analog circuits, defines the proper stressability conditions, and then presents a stress vector generation procedure. It also describes a stressability analysis process that measures the stressability of a circuit with the generated stress vectors. Stressability analysis allows estimation of a circuit's stressability before the extreme-voltage stress test is attempted by evaluating the stress coverage of the circuit. Based on the trade-offs between the stress coverage and the total stress time of a circuit, this chapter also presents algorithms to optimize stress vectors generated for extreme-voltage stress testing, and finally discusses the stress requirements for CMOS capacitors and resistors.

3.1 Stressability of Analog Circuits

Unless indicated otherwise, hereafter, we define the nominal operating voltage and the stress voltage to be V_{normal} =3.3V and V_{stress}=5.82V; the analog circuits under test are simulated with the HP AMOS14TB 0.5μm process technology and their nominal oxide thickness is 9.7nm. In addition, the period of the early life is assumed to be 1 year for the

sake of discussion.

As discussed in Chapter 2, for digital CMOS circuits, a transistor is fully stressed if its three terminal voltages $(V_D,V_G,V_S)=(V_{ss},V_{stress},V_{ss})$ for an NMOS transistor and $(V_D,V_G,V_S)=(V_{stress},V_{ss},V_{stress})$ for a PMOS transistor, as illustrated in Figure 3.1. Consider a 2-input NOR gate under extreme-voltage stress conditions, as shown in Figure 3.2. When both inputs A and B are equal to V_{ss}, transistors PA and PB are ON while transistors NA and NB are OFF. So, the voltage at nodes X and Z is equal to the V_{stress} and transistors PA and PB are then fully stressed. Similarly, when the inputs A and B are equal to V_{stress}, transistors NA and NB are ON and transistors PA and PB are OFF. As a result, the voltage at node Z is equal to V_{ss} and transistors NA and NB are fully stressed. Hence, the stress vectors for the 2-input NOR gate are $(A,B)=(V_{ss},V_{ss})$ and (V_{stress},V_{stress}).

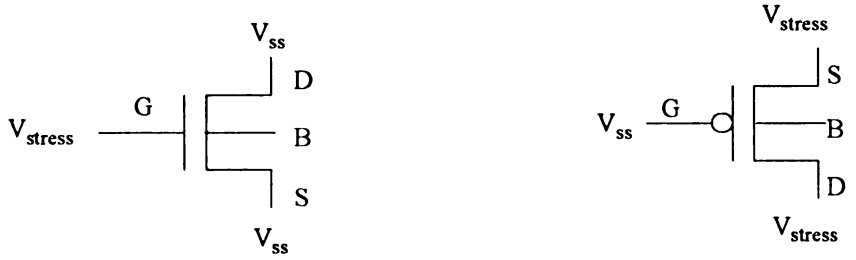


Figure 3.1: Digital Stressability Definition (a) NMOS; (b) PMOS.

Each transistor in a CMOS IC must be stressed long enough to make sure the defective oxide deteriorates significantly so that either oxide breakdown or stress-induced oxide leakage occurs in the defective oxide. To optimize the stress effect of each stress vector

and thus reduce the total stress time, each signal should be held at its full-swing signal level for enough time.

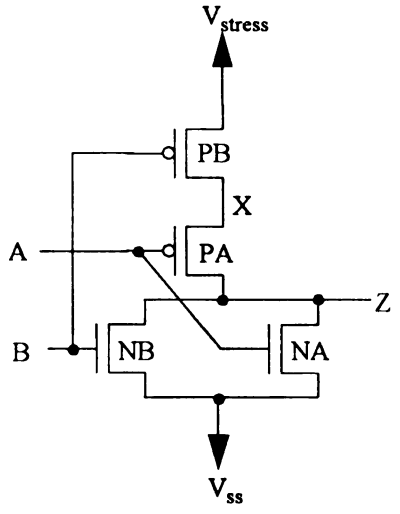


Figure 3.2: A 2-Input NOR gate.

In order to evaluate the feasibility of the digital definition for analog circuits, consider the simple bias voltage generator circuit in Figure 3.3. It generates a bias voltage of 0.045V at 3.3V and a 0.111V at 5.82V. Simulation results of the circuit show that both transistors, M10 and M11, are not fully stressed under the definitions described in Figure 3.1. The question that naturally arises is how to properly stress analog circuits?

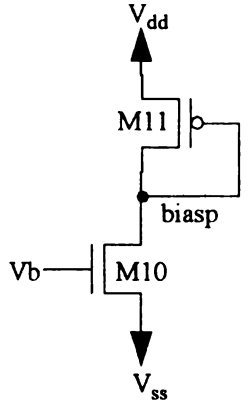


Figure 3.3: A Bias Voltage Generator Circuit.

Two basic definitions for properly stressing analog circuits are introduced in this sub-section: *Full Stressability* and *Stress Time*.

3.1.1 Full Stressability

Gate-oxide distributed in a transistor can be divided into three regions: GD (between Gate and Drain), GB (Gate and Substrate), and GS (Gate and Source). Breakdown in the region GD causes a gate-oxide short between gate and drain, so do the other regions. Let V_{GD}, V_{GB}, and V_{GS}, be the voltages across the three regions, respectively. When a transistor in digital circuits is stressed, the voltages V_{GD}, V_{GB}, and V_{GS} are the same, i.e., the stress voltage. In fact, the stress voltages can be easily applied to the terminals with the carefully selected stress vectors, where the stress time is determined by the stress voltage. However, for analog circuits, the voltages V_{GD}, V_{GB}, and V_{GS} may be different due to the work function difference and/or the circuit topology. So, for analog circuits, the regions GD, GB, and GS need to be considered independently while evaluating the stress conditions [53].

3.1.2 Stress Time

Another problem associated with analog circuits is that it may not be possible to achieve the desired stress voltage across a region due to the operating conditions and circuit topology. For example, the voltage across region GS of transistor M11 in Figure 3.3 will depend on the driving transistor M10. This issue can be resolved by using the fact that magnitude and duration of applied voltage can be used to control the stressing of a region. When

desired stress voltage cannot be obtained across a region, the required amount of stress can be achieved by applying the available lower voltage for a longer time [53]. Let *t2* < *t1* and V_{dd} < V2 < V1, then the following property holds.

Property 3.1

A region that requires *t2* seconds to be fully stressed at V1 can also be fully stressed by the application of V2 for *t1* seconds.

The *stress time* of a region is determined by the maximum voltage appearing across it during normal operation. It should be mentioned here that the time to breakdown of flaw-less oxides is much higher than the lifetime of the circuits, so almost any stress time can be selected without overstressing or damaging the circuit. Let V_{yn} be the maximum voltage across the region y (GD, GB, or GS) of a transistor at V_{normal}. It can either be given by the circuit designer, or obtained from Monte Carlo simulations with DC analysis during the design phase. Let V_{yst} be the voltage across the region at V_{stress}.

Property 3.2

The stress time for the region y for a lifetime of Z seconds can be expressed as

$$t_{stress} = \tau_0(Z/\tau_0)^{(Vyn/Vyst)} \qquad (3.1)$$

Proof: By Eqn. 2.1, the thickness of the gate-oxide for t_{BD} = Z seconds is

$$T_{eff} = (V_{yn}/G)\ln(Z/\tau_0)$$

Therefore, by Eqn. 2.1 with $t_{BD} = t_{stress}$ and $V_{ox} = V_{yst}$, we have

$$t_{stress} = \tau_0 \exp[(G/V_{yst})(V_{yn}/G) \ln(Z/\tau_0)] = \tau_0 \exp[(V_{yn}/V_{yst}) \ln(Z/\tau_0)]$$

This concludes that $t_{stress} = \tau_0(Z/\tau_0)^{(Vyn/Vyst)}$

Consider the stress time of the circuit in Figure 3.3. The maximum voltages across the three regions of the NMOS transistor at 3.3V are V_{GDn}=3.255V and $V_{GSn}=V_{GBn}$=3.3V. On the other hand, the voltages across the regions at 5.82V are V_{GDst}=5.709V and $V_{GSst}=V_{GBst}$=5.82V. By Eqn. 3.1, the stress times are $t_{stress\text{-}GD}$=0.35 seconds and $t_{stress\text{-}GS}=t_{stress\text{-}GB}$=0.31 seconds. Similarly, for the PMOS transistors, $V_{SGn}=V_{BGn}$=3.255V, V_{DGn}=0V, $V_{SGst}=V_{BGst}$=5.709V and V_{DGst}=0V, by Eqn. 3.1, the stress times are $t_{stress\text{-}SG}=t_{stress\text{-}BG}$=0.35 seconds. Note that the region DG with 0V does not need to be stressed, and can be excluded from the calculation process, or $t_{stress\text{-}DG}$=0 seconds. Therefore, the circuit in Figure 3.3 can be fully stressed at 5.82V in 0.35 seconds, or the circuit has a full stressability at 5.82V in 0.35 seconds.

The above stressability definition can also be applied to the digital circuits. Consider the 2-input NOR gate of Figure 3.2. The maximum voltages across the three regions of the NMOS transistors, NA and NB, at 3.3V are $V_{GDn}=V_{GSn}=V_{GBn}$=3.3V. On the other hand, the voltages across the regions at 5.82V are $V_{GDst}=V_{GSst}=V_{GBst}$=5.82V. By Eqn. 3.1, the stress times are $t_{stress\text{-}GD}=t_{stress\text{-}GS}=t_{stress\text{-}GB}$=0.31 seconds. Similarly, for the PMOS transistors, PA and PB, $V_{SGn}=V_{BGn}= V_{DGn}$=3.3V, $V_{SGst}=V_{BGst}=V_{DGst}$=5.82V, by Eqn. 3.1, the stress times are $t_{stress\text{-}SG}=t_{stress\text{-}BG}=t_{stress\text{-}DG}$=0.31 seconds.

The results obtained using this procedure are identical to the one obtained from digital extreme-voltage stress test process. Hence, it can be concluded that extreme-voltage stress test of digital circuits is a special case of extreme-voltage stress test of analog circuits.

3.2 Stress Vector Generation

The stress vectors are selected from the vectors for all possible combinations of the input voltages of V_{ss} and V_{dd} of the analog circuit under test. The stress vector generation process is comprised of two steps: *primary stress vector generation* and *stress time calculation*.

3.2.1 Primary Stress Vectors

For simplicity, the level converter circuit [54] in Figure 3.4 is used to describe the developed stress vector generation procedure. The circuit is comprised of 6 transistors and has two inputs, in+ and in-. Thus, there are four possible candidate stress vectors, i.e., {(in+,in-)=(0,0), (0,3.3), (3.3,0), (3.3,3.3)}. The circuit is simulated with each candidate stress vector, and the resultant voltages across the regions of all transistors are listed in Table 3.1, where the row "SV max" records the maximum voltage of the region corresponding to a column and the row "MC max" is the resultant maximum voltages with the Monte Carlo simulations or given by the circuit designer. The table shows that both rows "SV max" and "MC max" are identical. As discussed in [55], a MOS transistor has a nonlinear resistance. The resistance of a MOS transistor increases monotonically as the supply voltage decreases. This has been matched to our empirical results, the "SV max" is the same as

"MC max" for all example circuits, and endorsed the effectiveness of using the digital-like input combinations as the candidate stress vectors for the circuit simulations.

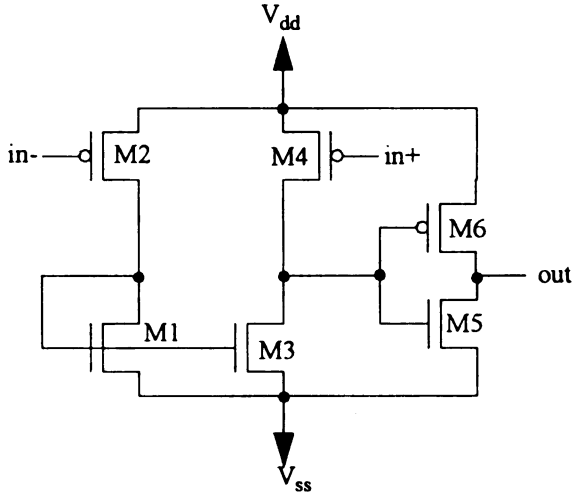


Figure 3.4: A Level Converter Circuit.

Table 3.1: Level Converter Circuit Simulation Results

in+	in-	M1 GS	M1 GD	M1 GB	M2 SG	M2 DG	M2 BG	M3 GS	M3 GD	M3 GB	M4 SG	M4 DG	M4 BG	M5 GS	M5 GD	M5 GB	M6 SG	M6 DG	M6 BG
0.0	0.0	2.73	0	2.73	3.3	2.73	3.3	2.73	0	2.73	3.3	2.73	3.3	2.73	2.73	2.73	0.57	-2 7	0.57
0.0	3.3	0.26	0	0.26	0	-3.04	0	0.26	-3.04	0.26	3.3	3.3	3.3	3.3	3.3	3.3	0	-3.3	0
3.3	0.0	2.73	0	2.73	3.3	2.73	3.3	2.73	2.73	2.73	0	-3.3	0	0	-3.3	0	3.3	3.3	3.3
3.3	3.3	0.26	0	0.26	0	-3.04	0	0.26	0	0.26	0	-3.04	0	0.26	-3.04	0.26	3.04	3.04	3.04
SV max		2.73	0	2.73	3.3	2.73	3.3	2.73	2.73	2.73	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
MC max		2.73	0	2.73	3.3	2.73	3.3	2.73	2.73	2.73	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3

The next step is to build up a stress vector selection table. The intersection of a row and a column in Table 3.2 is assigned a 1 if the voltage of the corresponding entry in Table 3.1 is the maximum voltage, i.e., if it is equal to the entry in the row "SV max" of the same

column. Otherwise, a 0 is assigned. In other words, the entry with a 1 means that the region can be fully stressed when the candidate stress vector is applied. Thus, the stress vector generation problem is to select a minimal set of stress vectors from the candidate stress vectors such that all regions are fully stressed. In effect, it is a minimal covering problem, i.e., a minimal set of rows is chosen to cover all columns. A heuristic algorithm for minimum covering can be employed to find the solution. In Table 3.2, the stress vectors (in+,in-)= (0.0,3.3) and (3.3,0.0) are selected. In this study, these vectors are referred to as *primary stress vectors*.

Table 3.2: Stress Vector Selection Table

in+	in-	M1 GS	M1 GD	M1 GB	M2 SG	M2 DG	M2 BG	M3 GS	M3 GD	M3 GB	M4 SG	M4 DG	M4 BG	M5 GS	M5 GD	M5 GB	M6 SG	M6 DG	M6 BG
0.0	0.0	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	0
0.0	3.3	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
3.3	0.0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1
3.3	3.3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.2.2 Stress Time Calculation

The stress time of the circuit is estimated based on the selected stress vectors. Table 3.3(a) lists the selected stress vectors for the level converter circuit of Figure 3.4, where only 1-entries are shown. The circuit is again simulated with the selected stress vectors, where the stress voltage 5.82V replaces the normal operating voltage 3.3V in both supply voltage and stress vectors. The resultant voltages across those regions which can be fully stressed, i.e., 1-entries in Table 3.3(a), are recorded in Table 3.3(b). Finally, the bottom row lists the "SV max" voltages shown in Table 3.1.

Table 3.3: Level Converter Circuit Simulation Results with the Stress Voltage

(a)

in+	in-	M1 GS	M1 GD	M1 GB	M2 SG	M2 DG	M2 BG	M3 GS	M3 GD	M3 GB	M4 SG	M4 DG	M4 BG	M5 GS	M5 GD	M5 GB	M6 SG	M6 DG	M6 BG
0.0	3.3		1								1	1	1	1	1	1			
3.3	0.0	1	1	1	1	1	1	1	1	1							1	1	1

(b)

in+	in-	M1 GS	M1 GD	M1 GB	M2 SG	M2 DG	M2 BG	M3 GS	M3 GD	M3 GB	M4 SG	M4 DG	M4 BG	M5 GS	M5 GD	M5 GB	M6 SG	M6 DG	M6 BG
0.0	5.82		0								5.82	5.82	5.82	5.82	5.82	5.82			
5.82	0.0	4.73	0	4.73	5.82	4.73	5.82	4.73	4.73	4.73							5.82	5.82	5.82
SV max		2.73	0	2.73	3.3	2.73	3.3	2.73	2.73	2.73	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3

In this implementation, a reduced table, as shown in Table 3.4, is used to calculate the stress times. The reduced table is constructed by grouping the equivalent regions in Table 3.3, where two regions are equivalent if they have the same column data.

Table 3.4: Level Converter Reduced Table for Stress Time Calculation

in+	in-	C1	C2	C3	C4
0.0	5.82		0		5.82
5.82	0.0	4.73	0	5.82	
SV max		2.73	0	3.3	3.3

C1: M1_GS, M1_GB, M2_DG, M3_GS, M3_GD, M3_GB;
C2: M1_GD; C3: M2_SG, M2_BG, M6_SG, M6_DG, M6_BG;
C4: M4_SG, M4_DG, M4_BG, M5_GS, M5_GD, M5_GB.

As shown in Eqn. 3.1, the stress time is a function of V_{yn}/V_{yst}. There are two distinct pairs of V_{yn}/V_{yst} in Table 3.4, i.e., (2.73/4.73) and (3.3/5.82). Thus, the stress times are $t_{stress}(2.73/4.73)=0.47$ seconds and $t_{stress}(3.3/5.82)=0.31$ seconds. As mentioned, the

region with V_{yn}=0 does not need to be stressed, and can be excluded from the stress time calculation process. The stress times for these pairs, TG_i, i=0,1,2, are listed in Table 3.5.

Table 3.5: Level Converter Calculated Stress Times

	V_{yn}/V_{yst}	Stress Time
TG0	0	0 seconds
TG1	3.3/5.82	0.31 seconds
TG2	2.73/4.73	0.47 seconds

3.2.3 Generation Procedure

Algorithm I summarizes the stress vector generation process presented in the previous sub-section. Here, the fully differential folded-cascode operational transconductance amplifier (OTA) [56], as shown in Figure 3.5, is used to describe the stepwise procedure of the algorithm.

The circuit has 22 transistors and 6 inputs (VCM, Vbias2, Vbias3, Vbias4, V+,V-). By Step 1.1, there exist 64 input combinations of 0.0V and 3.3V, i.e., there exist 64 candidate stress vectors. In Step 1.2, the circuit is simulated with each candidate stress vector, and the resultant voltages across the regions of all transistors are listed in a table similar to Table 3.1. The table contains 72 columns (i.e., 6 columns for the inputs and 66 columns for the transistor regions) and 65 rows (i.e., 64 rows for the stress vectors and one row for recording "SV max." In Step 1.3, vector selection table is built where the intersection of a row and a column is assigned a 1 if the voltage of the corresponding entry in the table generated in Step 1.2 equals the entry in the row "SV max" of the same column. By Step 1.4,

Algorithm I

Step 1: Primary Stress Vector Generation

1.1: Generate candidate stress vectors.

1.2: Simulate the circuit with various candidate vectors and tabulate the simulated results.

1.3: Generate a stress vector selection table.

1.4: Find a minimal set of stress vectors.

Step 2: Stress Time Calculation

2.1: Simulate the circuit with each primary stress vector at V_{stress}, and tabulate the simulated results.

2.2: Construct a reduced table.

2.3: Find distinct pairs V_{yn}/V_{yst} and calculate their stress times.

among the 64 candidate stress vectors, eight primary stress vectors in Table 3.6 are selected, where the vector (0,0,1,1,1,1) means the input voltages are assigned as VCM=Vbias2= 0.0V, and Vbias3=Vbias4=V+=V-=3.3V.

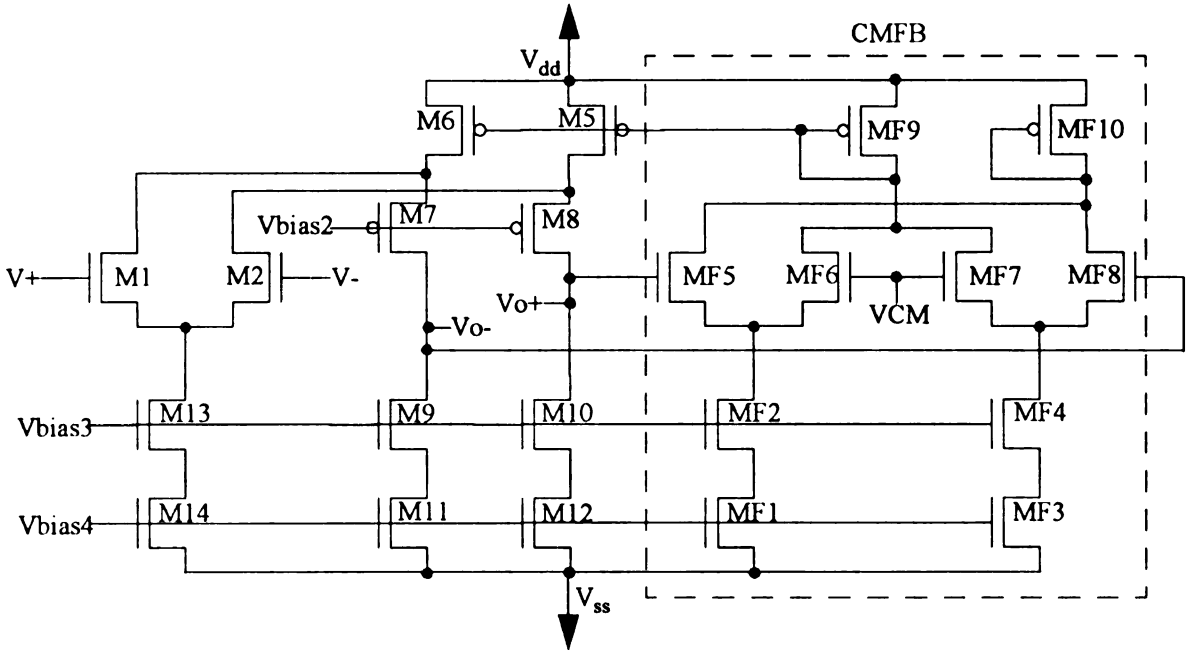


Figure 3.5: Fully Differential Folded-Cascode OTA.

Table 3.6: OTA with CMFB Stress Vectors

VCM	Vbias2	Vbias3	Vbias4	V+	V-	
0	0	1	1	1	1	15
1	0	0	0	1	0	34
1	0	1	0	0	1	41
1	0	1	0	1	0	42
1	0	1	1	0	0	44
1	0	1	1	0	1	45
1	0	1	1	1	0	46
1	1	1	1	0	0	60

By Step 2.1, the circuit is simulated with each selected stress vector, and the simulation results are tabulated. In Step 2.2, a reduced table, i.e., Table 3.7, is constructed. By Step 2.3, seven distinct pairs of (V_{yn}/V_{yst}) are identified, i.e., (3.29/5.81), (3.3/5.82), (2.65/4.37), (2.19/3.38), (1.44/2.17), (1.40/1.92), and (0.47/0.49). The stress times for these pairs, TG_i, i=0,1,2,...,7, are listed in Table 3.8.

Table 3.7: OTA with CMFB Reduced Table for Stress Time Calculation

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
15			5.82		5.82	5.82	5.82	5.82	5.82	5.82	5.82									0
34		5.81	5.82								5.82						5.81		5.82	0
41			5.82	5.81		5.82		5.82					0.49						5.82	0
42		5.81	5.82					5.82			5.82					0.49			5.82	0
44			5.82				5.82	5.82		5.82								1.92	5.82	0
45			5.82			5.82		5.82		5.82		2.17							5.82	0
46			5.82					5.82		5.82	5.82				2.17				5.82	0
60	3.38						5.82	5.82	5.82	5.82				4.37					5.82	0
*	2.19	3.29	3.3	3.29	3.3	3.3	3.3	3.3	3.3	3.3	3.3	1.44	0.47	2.65	1.44	0.47	3.29	1.40	3.3	0

*: SV max; Input vectors: (VCM,Vbias2,Vbias3,Vbias4,V+,V-)
C1: M5_SG,M5_DG,M5_BG,M6_SG,M6_DG,M6_BG,MF9_GS,MF9_GB,MF6_GD,MF7_GD;
C2: M8_SG,M8_DG; C3: M8_BG,M7_BG; C4: M7_SG,M7_DG,MF8_GB; C6: M2_GB;
C5: M2_GS,M2_GD,M1_GS,M1_GD,MF3_GD,MF4_GS,MF4_GD,MF1_GD,MF2_GS,MF2_GD;
C7: M13_GS,M13_GD,M14_GD; C8: M13_GB,M10_GB,MF4_GB,MF2_GB,M9_GB;
C9:M9_SG, M9_DG, M10_GS, M10_GD, M12_GD, M11_GD; C11: M1_GB; C12:MF8_GS;
C10:M12_GS,M12_GB,M11_GS,M11_GB,M14_GS,M14_GB,MF13_GS,MF3_GB,MF1_GS,MF1_GB;
C13: MF8_GD; C14: MF6_GS,MF7_GS; C15: MF5_GS; C16: MF5_GD; C17: MF5_GB;
C18: MF10_SG,MF10_BG; C19: MF6_GB,MF7_GB; C20: MF10_DG,MF9_DG.

Table 3.8: OTA with CMFB Stress Times

	V_{yn}/V_{yst}	Stress Time		V_{yn}/V_{yst}	Stress Time
TG0	0	0 seconds	TG4	2.19/3.38	9.68 seconds
TG1	3.29/5.81	0.30 seconds	TG5	1.44/2.17	18.87 seconds
TG2	3.3/5.82	0.31 seconds	TG6	1.40/1.92	5.14 minutes
TG3	2.65/4.37	1.65 seconds	TG7	0.47/0.49	64.16 days

3.3 Stressability Analysis

A circuit is fully stressed, or has a full stressability, if all regions of all the transistors in that circuit can be stressed completely within a feasible stress time period. Table 3.9 rearranges the columns in Table 3.4 for the level converter by sorting their stress times in an ascending order. The number of regions included in each column and the stress time of each column are also listed.

Table 3.9: Level Converter Stressed Regions and Stress Times

in+	in-	C2	C3	C4	C1
0.0	3.3	1		1	
3.3	0.0	1	1		1
0.0	5.82	0		5.82	
5.82	0.0	0	5.82		4.73
Number of regions		1	5	6	6
Column stress time		0 seconds	0.31 seconds		0.47 seconds

Based on the information given in Table 3.9, for various stress times TG_i's, the number of regions are plotted in Figure 3.6(a). When a circuit is stressed in a time period of TG_i, the regions with the stress time TG_j's, where $j \leq i$, will all be stressed. Thus, the number of regions can be stressed in a time period of TG_i is the total number of the regions for TG_j's, $j \leq i$. Figure 3.6(b) plots the accumulated number of regions for each stress time TG_i. The total number of regions in this circuit is 18. The stress coverage for a stress time TG_i is defined as the ratio of the accumulated number of regions for TG_i over the total number of regions. Figure 3.6(b) also lists the stress coverages for various stress times, and it is referred to as a *coverage chart*. A 100% stressability can be achieved for this circuit in a

total stress time of 0.78 seconds by applying stress vectors (0,5.82) and (5.82,0) for 0.31 seconds and 0.47 seconds, respectively.

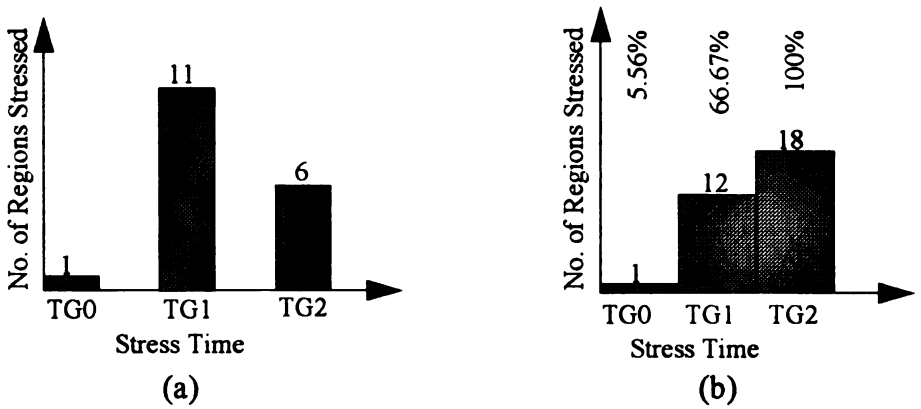


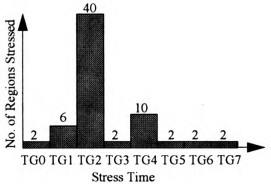
Figure 3.6: Level Converter Coverage Chart: (a) Stress Regions; and (b) Stress Coverage.

Similarly, for the OTA with CMFB circuit shown in Figure 3.5, Table 3.10 rearranges the columns in Table 3.6, where the stress times are sorted in an ascending order and both the number of regions included in each column and the stress time of each column are tabulated. Based on the information given in Table 3.10, for various stress times TG_i's, the corresponding coverage chart is given in Figure 3.7, where the total number of regions in this circuit is 66.

Table 3.10: OTA with CMFB Stressed Regions and Stress Times

							C20	C2	C4	C17	C3	C5	C6	C7	C8	C9	C10	C11	C19	C14	C1	C12	C15	C18	C13	C16
15	0	0	1	1	1	1	1				1	1	1	1	1	1	1	1								
34	1	0	0	0	1	0	1	1		1	1							1	1							
41	1	0	1	0	0	1	1		1		1		1		1				1						1	
42	1	0	1	0	1	0	1	1			1				1			1	1							1
44	1	0	1	1	0	0	1				1			1	1		1		1					1		
45	1	0	1	1	0	1	1				1		1		1		1		1			1				
46	1	0	1	1	1	0	1				1				1		1	1	1				1			
60	1	1	1	1	0	0	1							1	1	1	1		1	1	1					
Number of regions							2	2	3	1	2	10	1	3	5	6	10	1	2	2	10	1	1	2	1	1
Column stress time							0	0.30			0.31									1.65	9.68	18.87		5.14	64.16	

seconds minutes days



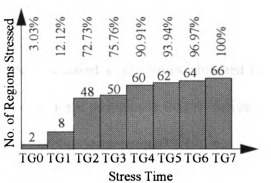


Figure 3.7: The Coverage Chart for OTA: (a) Stress Regions; and (b) Stress Coverage.

3.4 Stress Vectors Optimization

There exists a trade-off between stress coverage and the total stress time for a given set of stress vectors [57]. The stress coverage may be increased at the cost of increasing the total stress time. The stress vector optimization problem can be formulated as follows.

Problem I. (Stress Coverage Constraint)

To generate a set of stress vectors that meet stress coverage requirements and result in the minimal total stress time.

Problem II. (Total Stress Time Constraint)

To generate a set of stress vectors that meet total stress time constraints and result in the maximum stress coverage.

This sub-section presents two algorithms for the above two problems.

3.4.1 Stress Coverage Constraint

Based on the primary stress vectors and the calculated stress times obtained in Algorithm I for a given circuit, the stress coverage constraint problem is first to derive a primary solution. The primary solution is then refined to obtain the optimal solution.

We first consider the derivation of a primary solution. Given a required stress coverage q, the least stress time, TG_i, is defined from the generated coverage chart, where TG_i is the minimum stress time in the coverage chart that the corresponding stress coverage is greater than or equal to the required stress coverage q. The stress coverage of the regions in groups C_j's, whose stress times TG_j's exceed TG_i, are discarded. In other words, those stress vectors which were used to stress these regions can be removed from the selection table. Therefore, the selection table is reduced by removing the columns whose stress times are greater than TG_i, and removing the redundant rows, if they exist, after removing columns. The remaining stress vectors in the reduced table must be linearly independent, and

their stress times are calculated. The set of stress vectors is referred to as the *primary solution*.

For example, assume a stress coverage of 92% is required for the OTA circuit in Figure 3.5. From the coverage chart in Figure 3.7(b), the least stress time is TG5, or 18.87 seconds, for a stress coverage of 93.94%. As listed in Table 3.10, the columns with the stress time of 18.87 seconds are "C12" and "C15" and stress times in that table are in an ascending order. Therefore, the columns "C18", "C13", and "C16" in Table 3.10 are removed. After removing these columns, the vectors "42" and "44" become redundant and are also removed. The primary solution {15,34,41,45,46,60} is resulted from Table 3.11. The next step is to calculate the stress times for these vectors.

Table 3.11: Reduced Table for Stress Vector Selection

							C20	C2	C4	C17	C3	C5	C6	C7	C8	C9	C10	C11	C19	C14	C1	C12	C15
15	0	0	1	1	1	1	1				1	1	1	1	1	1	1	1					
34	1	0	0	0	1	0	1	1		1	1							1	1				
41	1	0	1	0	0	1	1		1		1		1		1				1				
45	1	0	1	1	0	1	1				1		1		1		1		1			1	
46	1	0	1	1	1	0	1				1				1		1	1	1				1
60	1	1	1	1	0	0	1							1	1	1	1		1	1	1		

Let $t_{stress}(Ci)$ denote the stress time for the column Ci. The stress time of a stress vector is the maximum stress time of the columns covered by the vector. For example, in Table 3.11, the columns with the maximum stress time in the vector "15" is "C11". Thus, the stress time of the vector "15" is $TS_{15}=t_{stress}(C11)=0.31$ seconds. Similarly, $TS_{34}=t_{stress}(C19)=0.31$ seconds, $TS_{41}=t_{stress}(C19)=0.31$ seconds, $TS_{45}=t_{stress}(C12)=9.87$

seconds, $TS_{46}=t_{stress}(C15)=18.87$ seconds, and $TS_{60}=t_{stress}(C1)=18.87$ seconds. Thus, the total stress time is $TS_{15}+TS_{34}+TS_{41}+TS_{45}+TS_{46}+TS_{60}=48.35$ seconds. This concludes that application of stress vectors {15,34,41,45,46,60} in a total stress time of 48.35 seconds will achieve a stress coverage of 93.94% which meets the coverage requirement of 92%.

Once the primary solution is derived, the next attempt is to refine the primary solution for the optimal solution, where the total stress time is decreased while keeping the stress coverage within limit. Note that the total stress time may be reduced by decreasing the stress coverage.

Recall that the primary solution was derived by selecting the least stress time TG_i whose corresponding coverage meets the requirement. For simplicity of discussion, let r be the number of columns having the stress time TG_i, R_{TGi} be the total number of regions included in the columns having the stress time TG_i, and R_{diff} be the minimum number of regions needed, together with all regions with the stress times less than TG_i, to meet the coverage requirement. For example, in Table 3.10, r=2, i.e., there are two columns, "C12" and "C15", having the stress time TG5=18.87 seconds. In Figure 3.7(a), $R_{TG5}=2$. Finally, $R_{diff}=61-60=1$ because the number of regions needed for the 92% coverage is 61 and the accumulated number of regions for TG4 is 60.

Note that if r=1, i.e., there is only one column with TG_i, or if $r > 1$ and the regions included in any combinations of any (r-1) columns is less than R_{diff}, the coverage will not meet the requirement when a column is removed. Thus, there will be no improvement and the process will terminate. This is summarized in the following property.

Property 3.3

There will be no improvement if r=1; or r > 1 and the number of regions included by any combinations of any (r-1) columns is less than R_{diff}.

If r > 1, a number of tables are constructed for generating the stress vectors and their stress times and coverages. A combination of z columns, where $1 \leq z < r$, including the number of regions larger than R_{diff} is referred to as a candidate combination. A table is constructed by including each candidate combination with all columns whose $TG_j < TG_i$. Its reduced table results in a set of linearly independent stress vectors. Based on the vectors, the total stress time TS_x and stress coverage SC_x are calculated. The solution is updated if this solution meets the coverage requirement and has lesser stress time. Otherwise, the next candidate combination is selected. The above process is repeated until there are no more candidates to be chosen.

In this example, r=2. Two possible combinations are concluded: either including "C12" or "C15". In both cases, the same stress coverage of 61/66=92.4% is resulted. We first consider the first combination. The column "C12" combines with the columns with stress times less than TG5 to form a table. In fact, this table is equivalent to removing the column "C15" from Table 3.11. The corresponding reduced table is given in Table 3.12, where the linearly independent stress vectors are {13,34,41,45,60}. Therefore, the total stress time is $TS_{15}+TS_{34}+TS_{41}+TS_{45}+TS_{60}$=29.48 seconds. Similarly, with "C15", the linearly independent stress vectors {13,34,41,46,60} are obtained from the corresponding reduced table. The total stress time is also 29.48 seconds.

Table 3.12: Reduced Table after Excluding C15 from Table 3.11

							C20	C2	C4	C17	C3	C5	C6	C7	C8	C9	C10	C11	C19	C14	C1	C12
15	0	0	1	1	1	1	1				1	1	1	1	1	1	1	1				
34	1	0	0	0	1	0	1	1		1	1							1	1			
41	1	0	1	0	0	1	1		1		1		1		1				1			
45	1	0	1	1	0	1	1				1		1		1		1		1			1
60	1	1	1	1	0	0	1							1	1	1	1		1	1	1	

This concludes that the total stress time is improved from 48.35 seconds to 29.48 seconds at the cost of decreasing the coverage from 93.4% to 92.4%. But, it still meets the requirement.

The above stress vector generation is summarized in Algorithm II.

3.4.2 Total Stress Time Constraint

Given a total stress time requirement, T_{req}, the coverage chart is used to define an initial stress time and the associated stress coverage to work with. Let TG_i be the maximum stress time that is less than or equal to T_{req}. Thus, a table is constructed by including all columns with $TG_j \leq TG_i$. A set of linearly independent stress vectors is generated from the corresponding reduced table. The total stress time for these stress vectors are calculated. If the total stress time is greater than T_{req}, then the next smaller stress time, i.e., TG_{i-1}, is selected. The above process is repeated until the resultant total stress time is less than or equal to T_{req}. Without loss of generality, say, TG_j is the selected stress time, and its corresponding coverage is SC_j. The resultant set of stress vectors is referred to as the primary solution, and its coverage is SC_j.

Algorithm II

Step 1: Generate the Primary Stress Vectors and calculate Stress Times (Algorithm I).

Step 2: Derive the Stress Coverage Chart.

Step 3: Generate the Primary Solution.

3.1: Let TG_i be the least stress time with a coverage exceeding the requirement.

3.2: Construct reduced table by removing the columns with stress times $TG_j > TG_i$.

3.3: Generate linearly independent vectors from reduced table.

3.4: Calculate the total stress time TS_x for the stress vectors.

Step 4: Improvement

4.1: Find r and calculate R_{diff}.

4.2: IF Property 3.3 holds, STOP.

4.3: Derive candidate combinations.

4.4: Construct a table containing a candidate and the columns with $TG_j < TG_i$.

4.5: Calculate the stress time TS_x for the stress vectors from the reduced table.

4.6: The solution is updated if the stress time is smaller.

4.7: Repeat 4.4 until no candidates exist.

Suppose that the total stress time for the OTA with CMFB is allowed to be within 30 seconds, i.e., T_{req}=30 seconds. From the coverage chart in Figure 3.7(b), TG5=18.87 seconds $\leq T_{req}$ is the maximum one. A table that excludes columns "C18", "C13", and "C16" from Table 3.10 is constructed. The corresponding reduced table is exactly the same as Table 3.11. As discussed previously, the total stress time for the resultant stress vectors is 48.35 seconds. The time is greater than T_{req}=30. Thus, the stress time TG4=8.97 seconds is selected. A table that removes columns "C12" and "C15" from Table 3.11 is constructed. Table 3.13 shows the corresponding reduced table, where the set of linearly independent stress vectors {15,34,41,60} is the primary solution with the total stress time of 19.80 seconds < T_{req}. By Figure 3.7(b), its stress coverage is SC4=90.91%.

Table 3.13: Reduced Table after Excluding C12 and C15 from Table 3.11

							C20	C2	C4	C17	C3	C5	C6	C7	C8	C9	C10	C11	C19	C14	C1
15	0	0	1	1	1	1	1				1	1	1	1	1	1	1	1			
34	1	0	0	0	1	0	1	1		1	1							1	1		
41	1	0	1	0	0	1	1		1		1		1		1				1		
60	1	1	1	1	0	0	1							1	1	1	1		1	1	1

Based on the primary solution with a coverage SC_j, the next step is to increase the stress coverage while keeping the total stress time within the limit. The stress coverage may be increased at the cost of increasing the total stress time. In fact, the maximum stress coverage for the circuit lies between SC_j and SC_{j+1}.

Consider the table consisting of all columns which have the stress time less than or equal to TG_{j+1}. Let r be the number of columns having the stress time TG_{j+1} and T_{diff} is the

difference between the required stress time and the total stress time for the primary solution. Similar to Property 3.3, the following property results.

Property 3.4

There will be no improvement if r=1, or r > 1 and $T_{diff} < TG_{j+1}$.

If r > 1 and $T_{diff} \geq TG_{j+1}$, then a number of tables are constructed for generating the stress vectors and their stress times and coverages. A combination of z columns, where $1 \leq z < r$, is referred to as a candidate combination. A table is constructed by including each candidate combination with all columns whose $TG_i \leq TG_j$. Its reduced table results in a set of linearly independent stress vectors. Based on the vectors, the total stress time TS_x and stress coverage SC_x are calculated. The solution is updated if $TS_x < T_{req}$ and $SC_x > SC_j$. Otherwise, the next candidate combination is selected. The above process is repeated until no more candidates are left to be chosen.

In this example, r=2 and T_{diff}=30-19.80=10.20 > 9.68 seconds. Two possible candidate combinations are resulted, i.e., "C12" or "C15". We first consider the table consisted of "C12" and the columns with stress times $TG_i \leq$ TG4. This table is exactly the same as Table 3.12, where the linearly independent stress vectors {13,34,41,45,60} are resulted and its total stress time is $TS_{15}+ TS_{34}+TS_{41}+TS_{45}+TS_{60}$ =29.48 seconds and the stress coverage is 92.4%. Comparing with the primary solution, this solution meets the requirement and improves the coverage from 90.91% to 92.4%. Similarly, the candidate combination "C15" is processed, it results in a set of stress vectors {13,34,41,46,60} with a total stress time of 29.48 seconds and a coverage of 92.4%.

The above stress vector generation process is summarized in Algorithm III.

3.5 Stress Vectors for CMOS Capacitors and Resistors

MOS capacitors are usually formed as parallel plate devices with gate-oxide as an insulator. In mixed-signal CMOS ICs, the digital circuitry usually requires only a single poly-silicon (poly) layer for the gates. However, for linear operation, the analog stages generally require high linearity capacitors which are normally realized between two poly layers, adding a second poly layer for this purpose. This increases the cost of fabrication significantly. Using poly-metal, metal-metal, or special poly-metal-metal sandwich construction, the second poly layer can be avoided, but the oxide used in these structures requires a large chip area per unit capacitance and yields a large parasitic capacitance to the substrate. Recently, there have been several attempts to use the gate to channel capacitance of MOSFETs as the capacitors needed in analog CMOS circuits [58-63], because of the several advantages gained. Hence, the stress test paradigms can readily be applied to such analog capacitors and the developed process can be used to generate the stress vectors.

A MOSFET can be biased to provide a non-linear resistor. Such a resistor provides much greater values than diffused ones while occupying a much smaller area. When the gate is shorted to the drain in a MOSFET, a quadratic relation between current and voltage exists and the device conducts current only when the voltage exceeds the threshold voltage. Under these circumstances, the current flowing in this resistor (i.e., the MOSFET drain current) depends on the ratio of channel width-to-length [64]. Hence, to increase the resistor value, the aspect ratio of the MOSFET should be reduced to give longer channel length and

Algorithm III

Step 1: Generate the Primary Stress Vectors and Calculate Stress Times (Algorithm I).

Step 2: Derive the Stress Coverage Chart.

Step 3: Generate the Primary Solution.

3.1: Let T_{req} be the required stress time.

Let TG_i be the maximum stress time in Stress Coverage Chart with $TG_i \leq T_{req}$.

3.2: Construct a table consisted of the columns with stress time $\leq TG_i$.

3.3: Generate linearly independent vectors and calculate total stress time TS_x.

3.4: IF $TS_x > T_{req}$, i=i-1; GOTO 3.2.

Otherwise, the set of vector is the primary solution.

(TS_x, SC_j, and TG_j be the total stress time, coverage, and selected stress time.)

Step 4: Improvement

4.1: Find r and calculate $T_{diff}=TG_j-T_{req}$

4.2: IF Property 3.4 holds, STOP.

4.3: Derive the candidate combinations.

4.4: Construct table containing a candidate and the columns with stress time $< TG_j$.

4.5: Calculate TS_x and CV_x for the stress vectors from the reduced table.

4.6: The solution is updated if $TS_x < T_{req}$ and CV_x increases.

4.7: Repeat 4.4 until no candidates exist.

narrower channel width. Transistor M11 of voltage bias generator circuit, shown in Figure 3.3, provides a resistor of this kind and its stress requirements have already been discussed in Section 3.2. Previously in most standard MOS processes monolithic resistors were implemented as polysilicon or n-well or n+/p+ diffusion strips acting as passive devices but current trend is to use MOSFETs for resistor implementation [56]. Resistors implemented using polysilicon or n-well or n+/p+ diffusion do not involve any gate-oxide and hence, do not need to be stressed.

3.6 Discussion

This chapter showed that applying the extreme-voltage stress process for digital circuits is inadequate for analog circuits, i.e., the analog circuits can not be properly stressed using the digital procedure. Basic concepts required for proper stressability of analog circuits have been presented. It has been concluded that extreme-voltage stress test of digital circuits is a special case of analog circuits. Based on the fundamental concepts, stress vectors generation procedure for extreme-voltage stress of analog circuits has been described.

It has been found that stress time is a major factor for determining the stressability of analog circuits. So, the stressability of an analog circuit should be defined as the percentage of regions in a circuit that can be fully stressed in a given stress time. Full or 100% stressability can be achieved if there is no limit on stress time. However, in practice the stress time for a circuit will be selected by the manufacturer based on the trade-offs among customers' reliability requirements, available equipment, technology, time to market and cost. Hence, in order to evaluate the stressability of a circuit for the selected stress vectors,

a stressability analysis procedure has been developed.

Stressability analysis determines the relationship between stress coverage and stress time of the circuit based on the selected stress vectors. In order to improve the efficiency of extreme-voltage stress test, algorithms have been developed for two stress vectors optimization problems. Stress coverage constraint problem is to generate a set of stress vectors that meet the desired stress coverage requirements and result in the minimal stress time. Total stress time constraint problem is to generate a set of stress vectors that meet the required stress time constraints and result in the maximum stress coverage.

The stressability analysis process allows estimation of a circuit's stressability during the design phase. Once areas having poor stressability are found, the circuit may be redesigned or modified with extra circuitry in order to improve the final stressability. The goal will be to achieve the required stress coverage within the desired feasible time. Thus, it is necessary to develop a design methodology that provides some stressability enhancement strategy. Development of such a strategy is presented in the next chapter.

Chapter 4

STRESSABILITY DESIGN METHODOLOGY

This chapter presents the stressability design methodology, a major component of the framework of an automatic stress test system for analog/mixed-signal circuits, as illustrated in Figure 1.8. The stressability design methodology integrates the developed procedures of *stress vector generation*, *stressability analysis*, and *stressability enhancement process* in order to facilitate the design of analog circuits that can be stressed optimally. The methodology can be utilized in the design stage to evaluate the stressability of the circuit before the designs are finalized. Based on the stress vectors generated, stressability analysis will be employed to determine the feasibility of extreme-voltage stress testing under the desired stressability constraints. If it is not feasible to achieve the required stressability, modifications will be introduced in the design using stressability enhancement process such that the circuit meets the desired requirements. Stressability design methodology will be employed throughout the design cycle of the product to ensure the feasibility of extreme-voltage stress test of final design.

Generalized overall flow of the developed stressability design methodology is shown in Figure 4.1. Given an analog circuit, and stressability requirements, in terms of either stress coverage or stress time or both, the first step is to generate the stress vectors.

Stressability analysis is then performed using the selected stress vectors. In case the desired stressability requirements are met, the process terminates. If it is found that the selected vectors do not adhere to the specified stressability, optimization of stress vectors is attempted under the given stressability constraint. The process stops in case the required stressability is achieved with the optimized stress vectors. However, if these vectors still fail to meet the desired criterion, stressability enhancement strategy is employed to obtain the required level of stressability. In case of modifications to the circuit for stressability enhancement, the entire process is repeated starting from the stress vector generation of the redesigned circuit.

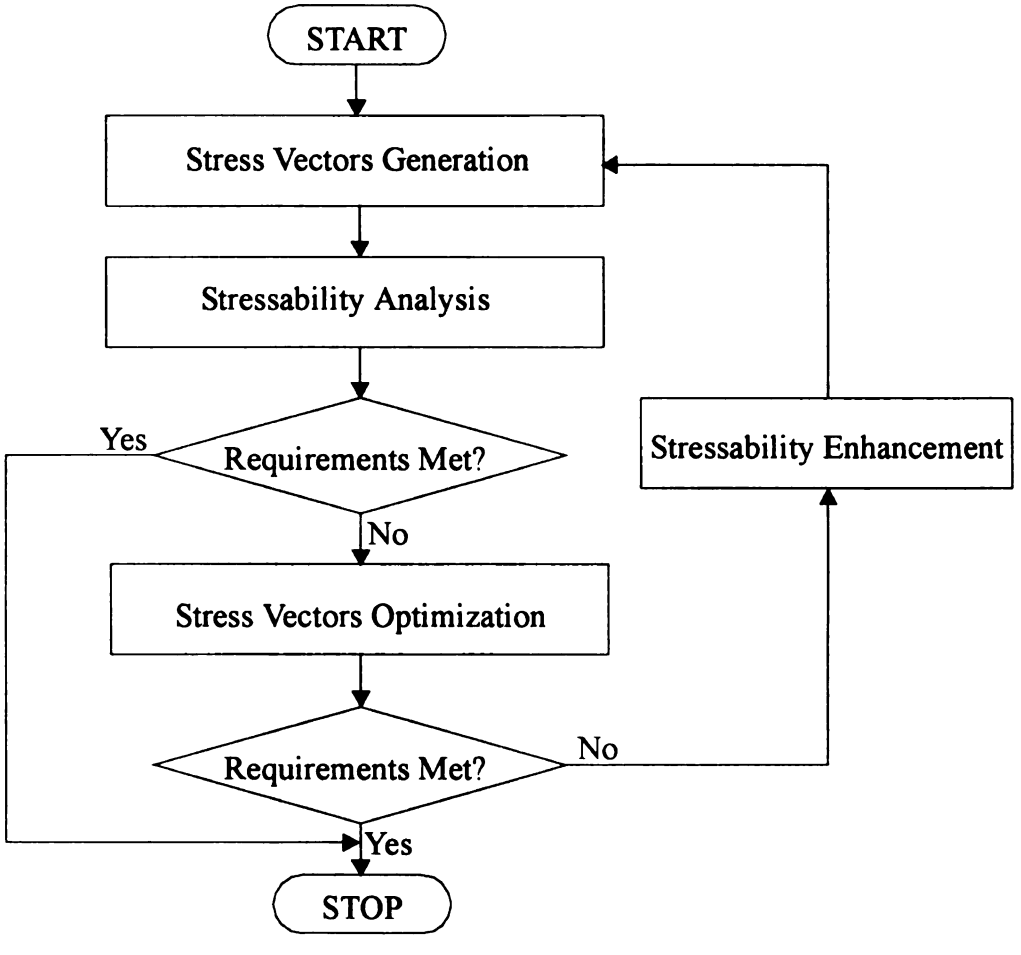


Figure 4.1: Stressability Design Methodology Flow Chart.

In the stress vector generation process presented in the previous chapter, exhaustive simulation was used to generate the primary stress vectors. Use of exhaustive simulations is a major concern for large circuits as simulation times may become infeasible for very large circuits with large number of primary inputs. Thus, it is necessary to develop efficient, yet effective procedures for stress vector generation of large circuits. Section 4.1 presents a partitioning scheme based on the control flow model and a hierarchical stress vector generation based on the topological structure of the circuit.

The stressability analysis process locates areas of a circuit having poor stressability, which usually cannot be discovered by the designers beforehand. The information obtained from stressability analysis allows estimation of a circuit's stressability during the design phase. Hence, potential problem areas can be located earlier. Once those areas are found, the circuit may be redesigned or modified with extra circuitry in order to improve the final stressability. A strategy for stressability enhancement is presented in Section 4.2.

4.1 Stress Vector Generation

Exhaustive simulation was used to generate the primary stress vectors in Chapter 3. More specifically, the stress vectors are selected from the vectors for all possible combinations of the input voltages of 0V and 3.3V of the analog circuit under test. The simulation results for all possible combinations are used to construct the stress vector selection table, such as Table 3.2. In case the number of inputs of a circuit is n, then it is simulated with 2^n candidate stress vectors. Simulation of large circuits or circuits with large number of inputs may take a significantly long time.

In order to reduce the computational complexity, this section presents a hierarchical stress vector generation process and a hierarchical stressability analysis process. The stress vectors are generated based on the topological structure of the circuits/sub-circuits without performing circuit simulations. Both stress vector generation process and stressability analysis process can be performed in a hierarchical fashion. In order to allow both processes to be performed hierarchically, a control-flow model (CFM) of CMOS circuits is introduced.

4.1.1 Control-Flow Model

A MOS network can be described by a flow graph and a control graph. Since the drain current and gate voltage are two major parameters in the DC analysis of a MOS transistor, the *flow graph* describes the current flow, while the *control graph* represents the gate control voltage. More specifically, the flow graph is constructed by converting each three-terminal MOS transistor in a MOS circuit network into a two-terminal device, where the gate connection of each transistor is removed.

A ***valid path*** is defined as a path that travels from V_{dd} to V_{ss}. Both V_{dd} and V_{ss} are referred to as *terminal nodes* and the others are *primary nodes*. A simple path partitioning scheme was developed to partition a flow graph using the following simple rules:

Rule 1:

Any valid paths which share at least a primary node will be in the same partition.

Rule 2:

A valid path forms a partition if it does not share a primary node with others.

Consider an operational transconductance amplifier (OTA) [65], as shown in Figure 4.2(a). The circuit comprises of four valid paths, as shown in Figure 4.2(b), where its flow graph can be divided into three partitions, namely P1, P2, and P3.

A primary node of a partition is called a *controlling node* if it connects to at least one gate of a transistor in the other partition, e.g., node n9 in P1, nodes n15 and n22 in P2. The flow "A-->B" in the flow graph, as shown in Figure 4.2(c), means a controlling node in Partition A connects to the gate of a transistor in Partition B. A *primary partition* is the one which is not controlled by any other partitions, e.g., P2. Figure 4.2(d) shows the circuit partitions resulted from the original circuit in Figure 4.2(a).

4.1.2 Generation Process

For simplicity of discussion, we first consider a partition with a single valid path, as shown in Figure 4.3(a). The partition contains r cascaded transistor TR_i, i=1,2,...,r, where each transistor has an input IN_j, j=1,2,...,r. The transistor can be either NMOS or PMOS. A primary node PN_j is located between two transistors TR_j and TR_{j+1}. Without loss of generality, the primary nodes are lined up and labeled in an ascending order, where let $PN_0=V_{ss}$ and $PN_r=V_{dd}$. The transistor TR_j is located between nodes PN_{j-1} and PN_j. This circuit is equivalent to a number of NMOS or PMOS switches connected in series. Therefore, the node voltage to be V_{ss} or V_{dd} can be easily obtained by turning ON and OFF the switches as described in the following property.

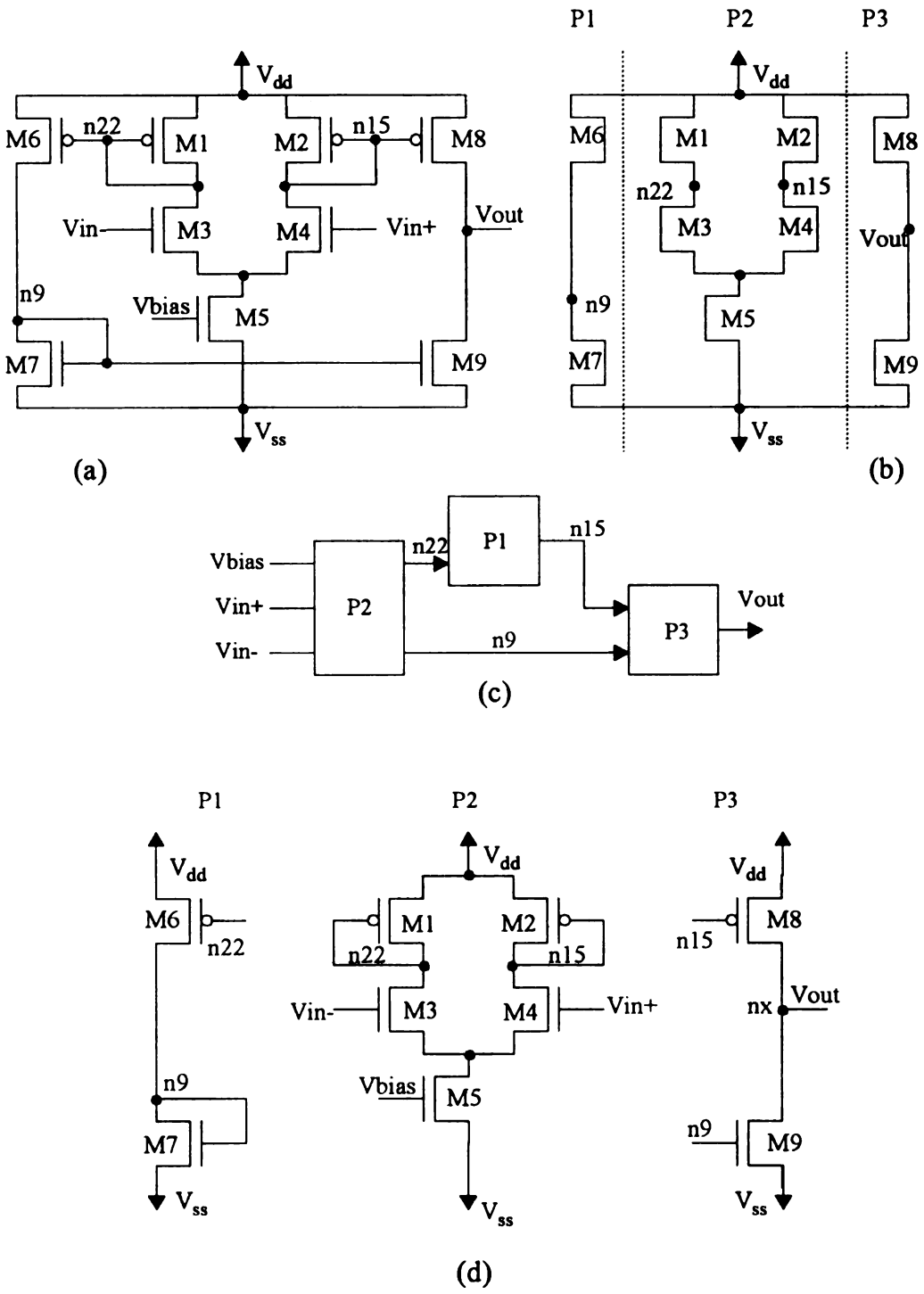


Figure 4.2: OTA Circuit: (a) Schematic; (b) Flow Graph;
(c) Control Graph; and (d) Circuit Partitions.

Property 4.1

(a) The voltage at node PN_j is V_{ss} if the transistors TR_i, i=1,2,...,j, are all ON and at least one TR_i, i=j+1,...,r, is OFF.

(b) The voltage at node PN_j is V_{dd} if the transistors TR_i, i=j+1,...,r, are all ON and at least one TR_i, i=1,2,...,j, is OFF.

Note that an NMOS transistor is turned ON and OFF by applying V_{dd} and V_{ss} to the gate terminal, respectively. Similarly, a PMOS transistor is turned ON and OFF by applying V_{ss} and V_{dd} to the gate terminal, respectively. Diode-connected transistors are very often used as a resistor in analog circuits, where a diode-connected transistor connects its gate to its drain or source terminal. Property 4.1 also holds when the partition contains some diode-connected transistors.

Consider a partition with single valid path which contains only one diode-connected transistor, say, TR_{r-2}. Assume that the diode-connected transistor does not have an input to its gate terminal. To properly stress the GS region of TR_{r-2}, its gate voltage at node PN_{r-2} must be V_{dd} and its source voltage at node PN_{r-3} should be V_{ss}. By Property 4.1(a), the voltage at node PN_{r-3} is V_{ss} if all TR_i, i=1,2,...,r-3, are ON, and either TR_{r-1} or TR_r is OFF. By Property 4.1(b), the voltage at node PN_{r-2} is V_{dd} if both TR_{r-1} and TR_r are ON, and at least one TR_i, i=1,2,...,r-3, is OFF. However, both conditions are conflicting to each other. This implies that it is virtually impossible to make $PN_{r-2}=V_{dd}$ and $PN_{r-3}=V_{ss}$ due to the diode-connected structure. However, when all transistors TR_i, i=1,2,...,r-3,r-1,r,

are ON, the maximum current of that circuit occurs. Therefore, the voltage drop from the gate (or drain) to the source will be the maximum among the voltage in the GS region of that transistor. The following property results.

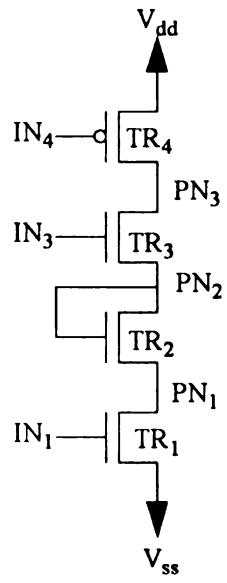
Property 4.2

Consider the diode-connected transistors in a partition containing a single valid path, the regions GS (or SG) and GD (or DG) are fully stressed if all transistors in the valid path are turned ON.

Both Properties 4.1 and 4.2 are used to generate the candidate stress vectors for the partitions with single valid path.

Consider a partition with a single valid path shown in Figure 4.3(a), where the circuit contains 4 transistors, 3 primary nodes, and 3 inputs. TR2 is a diode-connected NMOS transistor. By Property 4.1(a), the voltage at node PN_1 is V_{ss} if TR_1 is ON, i.e., $V_{IN1}=V_{dd}$, and either TR_3 or TR_4 is OFF, i.e., $V_{IN3}=V_{ss}$ or $V_{IN4}=V_{dd}$. Similarly, the voltage at node PN_3 is V_{dd}, by Property 4.1(b), if TR_4 is ON, i.e., $V_{IN4}=V_{ss}$, and either TR_3 or TR_1 is OFF, i.e., $V_{IN3}=V_{ss}$ or $V_{IN1}=V_{ss}$.

The first step of the stress vector generation process is to tabulate the required conditions to fully stress the regions of all transistors in that circuit. The first column in Figure 4.3(b) lists all regions of the transistors in Figure 4.3(a). The next step is to define the stress conditions for each region. The stress conditions contains two parts, one for inputs and the other for primary nodes. To fully stress an NMOS transistor, the node voltages are



(a)

	Stress Conditions						Derived Vectors		
regions	IN1	IN3	IN4	PN1	PN2	PN3	IN1	IN3	IN4
$GS(TR_1)$	1						1		
$GB(TR_1)$	1						1		
$GD(TR_1)$	1			0			1	0	0
								1	1
								0	1
$GS(TR_2)$							1	1	0
$GB(TR_2)$					1		0	1	0
$GD(TR_2)$									
$GS(TR_3)$		1			0		1	0	0
								1	1
								0	1
$GB(TR_3)$		1							
$GD(TR_3)$		1				0	1	1	1
$SG(TR_4)$			0						
$BG(TR_4)$			0						
$DG(TR_4)$			0			1	1	0	0
							0	1	
							0	0	

(b)

	Valid Vectors		
regions	IN1	IN3	IN4
$GS(TR_1)$	1	x	x
$GB(TR_1)$	1	x	x
$GD(TR_1)$	1	0	x
	1	x	1
$GS(TR_2)$	1	1	0
$GB(TR_2)$	0	1	0
$GD(TR_2)$	x	x	x
$GS(TR_3)$	1	1	1
$GB(TR_3)$	x	1	x
$GD(TR_3)$	1	1	1
$SG(TR_4)$	x	x	0
$BG(TR_4)$	x	x	0
$DG(TR_4)$	x	0	0
	0	x	0

(c)

Figure 4.3: Stress Vector Generation: (a) Example Circuit; (b) Stress Conditions & Derived Stress Vectors; and (c) Valid Stress Vectors.

$(V_D,V_G,V_S)=(0,1,0)$. For a PMOS transistor, $(V_D,V_G,V_S)=(1,0,1)$, where "1" represents V_{dd} and "0" represents V_{ss}. For example, to fully stress the region GD of the NMOS transistor TR_1, the required condition is $V_G=V_{IN1}=1$ and $V_D=V_{PN1}=0$. Similarly, for the regions GS of TR_3, the conditions are $V_G=V_{IN3}=1$ and $V_S=V_{PN2}=0$. Note that no stress conditions are considered for both GS (or SG) and GD (or DG) regions for the diode-connected transistors. For example, for the diode-connected transistor TR_2, no stress conditions are given for GD(TR_2) and GS(TR_2).

Once all stress conditions are defined, the next step is to derive the required stress vector(s) from the stress conditions for the primary nodes using Properties 4.1 and 4.2. For fully stressing a region, all its stress conditions must be met. In other words, a stress vector derived from the stress conditions defined by the primary nodes is invalid if it conflicts to the stress conditions defined by the input part.

For example, for the region GD of TR_1, the stress condition is $V_G=V_{IN1}=1$ and $V_D=V_{PN1}=0$. By Property 4.1(a), we can obtain "$V_D=V_{PN1}=0$" by setting TR_1 to be ON, and TR_3 or TR_4 to be OFF, i.e., $V_{IN1}=1$ and $(V_{IN3},V_{IN4})=(0,0)$, (1,1), or (0,1), or $(V_{IN3},V_{IN4})=(0,x)$ or (x,1), where x means "don't care" and can be either 0 or 1. Similarly, the stress condition for the region GS of TR_3 is $V_G=V_{IN3}=1$ and $V_S=V_{PN2}=0$. The condition $V_S=V_{PN2}=0$ results that $V_{IN1}=1$, and $(V_{IN3},V_{IN4})=(0,0)$, (1,1), or (0,1). However, the derived vectors $(V_{IN1},V_{IN3},V_{IN4})=(1,0,0)$ and (1,0,1) do not agree with the condition $V_{IN3}=1$. Thus, only the stress vector $(V_{IN1},V_{IN3},V_{IN4})=(1,1,1)$ is valid. Finally, for the region GS(TR_2), by Property 4.2, the valid stress vector is $(V_{IN1},V_{IN3},V_{IN4})=(1,1,0)$. The table in Figure 4.3(c) lists the valid stress vectors for all regions of the transistors in Figure

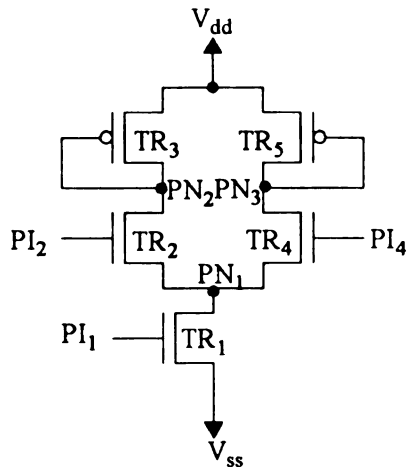
4.2(a). The stress vectors are the candidate stress vectors.

The primary nodes in the partition can be the controlling nodes and connect to the other partitions. As shown in Figure 4.3(b), we can obtain the primary node PN_3 as V_{dd} if we apply the stress vector $(V_{IN1},V_{IN3},V_{IN4})=(1,0,0)$, and obtain the maximum voltage, but not V_{dd}, at node PN_2 if the vector $(V_{IN1},V_{IN3},V_{IN4})=(0,1,0)$ is applied. In this implementation, a logical "1" is used to represent the case that the node voltage reaches its maximum, and a logical "0" represents the case that the minimum voltage is reached.

The above results can be generalized for a partition with more than one valid paths. Basically, we will activate only one valid path which contains the node-under consideration and disable the remaining valid paths. Thus, the problem is equivalent to the case discussed above.

Consider a partition, as shown in Figure 4.4(a), containing two valid paths, namely, PATH23 and PATH45, where path PATH23 contains TR_3, TR_2, and TR_1, and path PATH45 includes TR_5, TR_4, and TR_1. Figure 4.3(b) lists the stress conditions, derived stress vectors, and the candidate stress vectors for fully stressing all regions of the transistors in that circuit. The stress condition for the regions GD of TR_2 is $V_{IN2}=1$ & $V_{PN2}=0$. Since node PN2 is located in path PATH23, we disable path PATH45 by setting $V_{IN4}=0$ and consider the valid path PATH23. Similar to the previous cases, the condition $V_{PN2}=0$ is obtained by switching ON both TR_1 and TR_2, i.e., setting $V_{PI1}=V_{PI2}=1$. As a result, the stress vector $(V_{PI1},V_{PI2},V_{PI4})=(1,1,0)$ is generated.

Once the candidate stress vectors for all regions of the transistors in a circuit are derived properly, the next step is to minimize the number of stress vectors. Since some



(a)

Primary Vectors		
IN1	IN2	IN4
1	0	0
1	1	0
1	0	1

(c)

	Stress Conditions						Derived Vectors			**Candidate Vectors**		
regions	IN1	IN2	IN4	PN1	PN2	PN3	IN1	IN2	IN4	IN1	IN2	IN4
$GS(TR_1)$	1									1	x	x
$GB(TR_1)$	1									1	x	x
$GD(TR_1)$	1			0			1	0	0	1	0	0
$GS(TR_2)$		1		0			1		0	1	1	0
$GB(TR_2)$		1								x	1	x
$GD(TR_2)$		1			0		1	1	0	1	1	0
$SG(TR_3)$					0		1	1	0	1	1	0
$BG(TR_3)$					0		1	1	0	1	1	0
$DG(TR_3)$										x	x	x
$GS(TR_4)$			1	0			1	0		1	0	1
$GB(TR_4)$			1							x	x	1
$GD(TR_4)$			1			0	1	0	1	1	0	1
$SG(TR_5)$						0	1	0	1	1	0	1
$BG(TR_5)$						0	1	0	1	1	0	1
$DG(TR_5)$										x	x	x

(b)

Figure 4.4: Stress Vector Generation: (a) A Partition with Two Valid Paths; (b) Derivation of Candidate Vectors; and (c) Primary Stress Vectors.

candidate stress vectors contain "don't care" terms which can be either 0 or 1, the number of candidate stress vectors can be minimized by using the simple rules in the minimal covering problem. Figure 4.4(c) lists the minimum set of stress vectors derived from the candidate stress vectors in Figure 4.4(b). This concludes that the three stress vectors listed in Figure 4.4(c) can fully stress all regions of the transistors in Figure 4.4(a).

4.1.3 Hierarchical Generation Process

As described in the sub-section 4.1.1, an analog circuit can be decomposed into a number of partitions using the control-flow model. The stress vector generation process presented in the previous sub-section can be applied to each partition for generating stress vectors. The stress vectors of the original circuit can be obtained by combining those for all decomposed partitions.

For example, the circuit in Figure 4.2(a) can be decomposed into three partitions P1, P2, and P3 shown in Figure 4.2(d). The partition in Figure 4.4(a) is exactly the same as the partition P2, and thus the resultant stress vectors in Figure 4.4(c) are the stress vectors for P2. Figure 4.5 shows the stress vectors generated for both partitions P1 and P3. The next step is to derive the input-output relationships of all partitions based on their topological structures. Figure 4.6(a) shows the input-output relationship of P2, where "x" means "don't care" and "-" means the value determined by the transistor dimensions. Figures 4.6(b) and 4.6(c) illustrate those for P1 and P3, respectively.

Finally, the stress vectors and output results of the original circuit are developed by combining the stress vectors which are the primary inputs of that circuit. The combination

	Stress Conditions		Derived Vectors	Candidate Vectors
regions	n22	n9	n22	n22
GS(M7)		1	0	0
GB(M7)		1	0	0
GD(M7)				x
SG(M6)	0			0
BG(M6)	0			0
DG(M6)	0	1	0	0

(a)

Primary Vectors
n22
0

(b)

	Stress Conditions			Derived Vectors		Candidate Vectors	
regions	n9	n15	nx	n9	n15	n9	n15
GS(M9)	1					1	x
GB(M9)	1					1	x
GD(M9)	1		0		1	1	1
SG(M8)		0				x	0
BG(M8)		0				x	0
DG(M8)		0	1	0		0	0

(c)

Primary Vectors	
n9	n15
1	1
0	0

(d)

Figure 4.5: Stress Vector Generation for Partitions P1 & P3 in Figure 4.2(d):
For P1: (a) Derivation of Candidate Vectors; (b) Primary Vectors; and
For P3: (c) Derivation of Candidate Vectors; (d) Primary Vectors.

Stress Vectors			Outputs(P2)	
Vbias	Vin-	Vin+	n15	n22
1	0	0	1	1
1	1	0	0	1
1	0	1	1	0
0	0	x	-	1
0	x	0	1	-

(a)

I(P1)	O(P1)
n22	n9
0	1
1	0

(b)

I(P3)		O(P3)
n9	n15	nx
1	1	0
0	0	1
0	1	-
1	0	-

(c)

Stress Vectors			Outputs(P2)		O(P1)	O(P3)
Vbias	Vin-	Vin+	n15	n22	n9	nx
1	0	0	1	1	0	-
1	1	0	0	1	0	1
1	0	1	1	0	1	0

(d)

Figure 4.6: Input-output Relationship for Partitions in Figure 4.2(d);
(a) for P2; (b) for P1; (c) for P3; (d) for Entire Circuit.

of those stress vectors are virtually applied to the original circuit following the control graph in the control-flow mode. A combination of the stress vectors in all partition is valid if it does not violate the input-output relationships of the partitions it applies. More specifically, since all primary input vectors are applied to P2, the stress vectors are Vbias, Vin+, and Vin-. Following the control graph in Figure 4.2(c) and the input-output relationship in Figure 4.6(a), P2 produces the outputs n15 and n22. Similarly, P1 takes n22 as its input and produces an output n9, as shown in the column O(P1) of Figure 4.6(d), while P3 takes inputs n9 and n15 and has an output nx, as shown in the column O(P3) of Figure 4.6(d).

In summary, given a practical, large analog circuit, its stress vectors can be generated by first decomposing the circuit into a number of smaller partitions, deriving the stress vectors and input-output relation for all partitions, and combining the generated stress vectors to result the valid stress vectors. In this approach, generating stress vectors for smaller partitioned sub-circuits can reduce the computational complexity significantly.

The stress time can also be calculated using the partitions based on the control flow model. For example, the partition P2 is simulated with the three stress vectors and with the normal and stress voltages, respectively, where the stress times for the columns of regions were computed. Results show that the maximum stress time of P2 is 18.1 seconds. The values obtained for nodes n15 and n22 are also recorded for the stress vectors. These values are used as input values while simulating the partitions controlled by P2. P1 is then simulated with the values of n22 obtained for stress vector (1,0,1) under normal and stress voltages. The results show that the maximum stress time for P1 is 84.7 seconds. Similarly, P3 is simulated with appropriate values of inputs for its stress vectors to calculate the stress time of various regions. The maximum stress times for P3 is 18.1 seconds.

Simulating the partitioned sub-circuits generally require much less computation time. This partitioning may also result in the reduction of number of simulations depending on the manner in which primary inputs of the circuit are distributed in the partitions. For example, let h be the total number of primary inputs in the circuit partitioned into g partitions, denoted by P1, P2,..., Pg. Each partition Pi may take α_i primary inputs, say u_i, where $1 \leq \alpha_i \leq h$. If $u_i \cap u_j = \varnothing$, for any i and j, then $h=\alpha_1+\alpha_2+...+\alpha_g$. Thus even the exhaustive number of simulations required is reduced from 2^h to $(2^{\alpha_1}+2^{\alpha_2}+...+2^{\alpha_g})$.

4.1.4 Hierarchical Stressability Analysis

Similar to the hierarchical stress vector generation, the stressability analysis of a practically large analog circuit can also be obtained from the analysis results of its all partitions. The original circuit is first decomposed into a number of partitioned sub-circuits. The stressability analysis discussed previously can be applied to each partition to evaluate its stressability. Figure 4.7(a) shows the stressability analysis results for P2. Similarly, stressability analysis results for partitions P1 and P3 are listed in Figures 4.7(b) and 4.7(c), respectively.

Figure 4.7(d) shows the stressability analysis results of the original OTA circuit in Figure 4.2(a), obtained by merging the results in Figures 4.6(a)-(c). The results show that stress time of the OTA circuit is 84.7 seconds.

Most analog designers use wherever possible predesigned commercially available circuit elements (amplifiers, filter networks, comparators, etc.) for the system design. For such cases, each circuit element can be considered a module and their stress vectors and

Vbias	Vin+	Vin-	C1	C2	C3	C4	C5	C6	C7	C9	C10
1	0	0	1	1	1						
1	0	1	1	1		1		1		1	
1	1	0	1	1			1		1		1
Number of regions			2	2	1	1	1	1	1	3	3
Column stress time			0	0.31				0.32		18.1	

C1: M1_DG,M2_DG; C2: M5_GS,M5_GB; C3: M5_GD;
C4: M3_GB; C5: M4_GB; C6: M3_GS; C7: M4_GS;
C9: M1_SG,M1_BG,M3_GD; C10: M2_SG,M2_BG,M4_GD.

(a)

Vbias	Vin+	Vin-	C1	C8	C9	C11
1	0	0	1			
1	0	1	1	1	1	1
1	1	0	1			
Number of regions			1	2	2	1
Column stress time			0	0.37	18.1	84.7

C1: M7_GD; C8: M7_GS,M7_GB;
C9: M6_SG,M6_BG; C11: M6_DG.

(b)

Vbias	Vin+	Vin-	C8	C10
1	0	0		
1	0	1	1	
1	1	0		1
Number of regions			3	3
Column stress time			0.37	18.1

C8: M9_GS,M9_GD,M9_GB;
C10: M8_SG,M8_DG,M8_BG.

(c)

Vbias	Vin+	Vin-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	0	0	1	1	1								
1	0	1	1	1		1		1		1	1		1
1	1	0	1	1			1		1			1	
Number of regions			3	2	1	1	1	1	1	5	5	6	1
Column stress time			0	0.31				0.32		0.37	18.1		84.7

C1: M1_DG,M2_DG,M7_GD; C2: M5_GS,M5_GB; C3: M5_GD;
C4: M3_GB; C5: M4_GB; C6: M3_GS; C7: M4_GS;
C8: M7_GS,M7_GB,M9_GS,M9_GD,M9_GB; C9: M1_SG,M1_BG,M3_GD,M6_SG,M6_BG;
C10: M2_SG,M2_BG,M4_GD,M8_SG,M8_DG,M8_BG; C11: M6_DG.

(d)

Figure 4.7: Stressability Analysis for the Partitions in Figure 4.2(d):
(a) for P2; (b) for P1; (c) for P3; and (d) for Entire Circuit.

stressability analysis results can be determined and stored into a design database. Both stress vector generation process and stressability analysis process can then be applied effectively and efficiently in a hierarchical fashion for systems containing these modules.

4.2 Stressability Enhancement

The stressability analysis provides a measure of stress coverage for a given set of stress vectors. As discussed previously, all circuits can be fully stressed, but many of them may take a huge amount of stress time to accomplish the task. Stressing a circuit with such a long time may not be acceptable to the manufacturers and is then referred to as infeasible. A stress time is *feasible* if within a pre-specified limit. Hence, design strategies are required that can enable us to achieve desired stressability of the circuits in a feasible time.

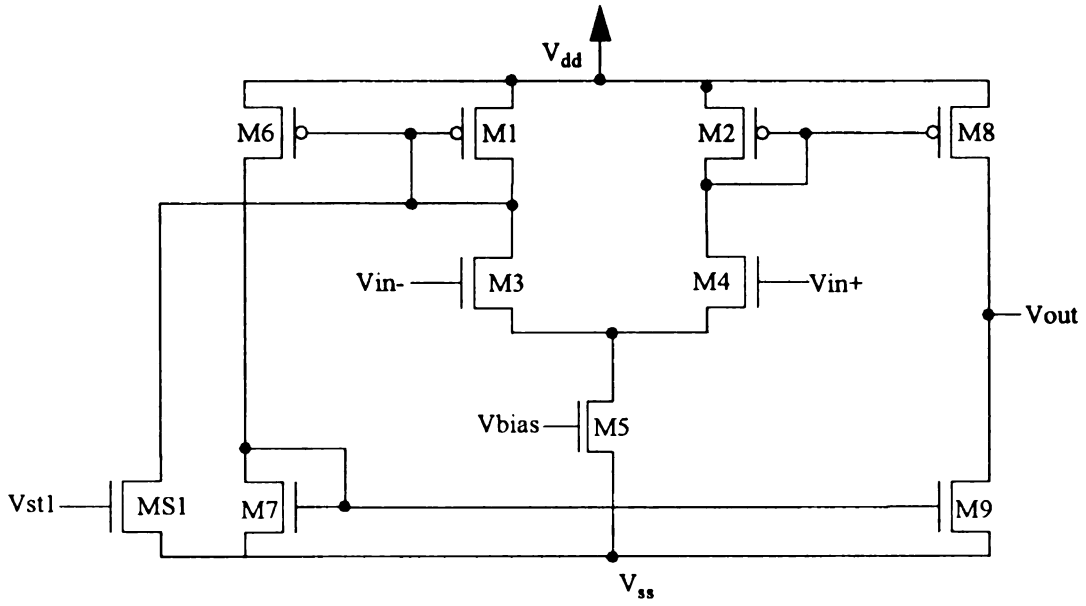
This section presents a stressability enhancement strategy using additional hardware. This strategy uses a two step approach. In the first step, referred to as Stress Time Reduction, stress time of the portion having poor stressability can be improved by one or more switches along with inputs to control these switches. Reducing stress times of such portions of the circuit will result in stressability enhancement. The control inputs of the switches used are referred to as the Pin Overhead of this implementation. Since, pin overhead is costly, so it is desirable to minimize the number of external control inputs. Once the desired stressability is achieved using stress time reduction, the next step is Pin Overhead Reduction, in which reduction of control inputs used for switches is attempted. The purpose of this step is to reduce the overall cost of stressability enhancement, and hence resulting in cost effective designs.

4.2.1 Stress Time Reduction

A portion of the circuit having poor stressability generally means that the stress time for that portion is not feasible. By Eqn. 3.1, the stress time of a region is determined by its voltage ratio V_{yn}/V_{yst}. The smaller the voltage ratio for a region, the lesser is the time required to stress it. A large voltage ratio V_{yn}/V_{yst} occurs when the gate node has a weak-1 (0) condition to turn ON an NMOS (PMOS) transistor. The weak-1 condition means that the gate node may reach its maximum voltage but far beyond the supply voltage and stress voltage during the normal and stress test modes, respectively. The node with a weak-1 condition can be improved by connecting it to V_{dd} via a switch, where the switch is realized by a PMOS transistor. Similarly, a weak-0 condition can be enhanced by connecting to V_{ss}, where the switch is realized by an NMOS transistor It is important to note that the added switch transistor must also be fully stressable.

For example, consider the stressability analysis, as shown in Figure 4.7(d), for the OTA circuit in Figure 4.2(a). The longest stress time, 84.7 seconds, occurs at the region DG of M6, where the voltage ratio V_{yn}/V_{yst} is 1.81/2.59. Figure 4.8 shows a solution to improve the stress time, where an NMOS transistor MS1 is added as the switch to connect the gate terminal of M6 to V_{ss}, and the switch is controlled by an input Vst1. Based on the stress vectors shown in Figure 4.7(b), the stressability analysis results are tabulated in Figure 4.8(c), where the stress time of the region DG of M6 is improved from 84.7 seconds to 0.6 seconds. Note that the extra switch transistor MS1 is also fully stressed.

It should be mentioned that the stress time can be improved by using extra switch, and the improvement in stress time will be determined by the aspect ratio (W/L) of the



(a)

Vst1	Vbias	Vin+	Vin-
1	1	0	0
1	1	0	1
0	1	1	0

(b)

Vst1	Vbias	Vin+	Vin-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	1	0	0	1	1	1	1				1			
1	1	0	1	1	1	1			1		1	1	1	
0	1	1	0	1		1		1		1				1
Number of regions				3	2	2	1	1	2	1	5	6	1	6
Column stress time				0	0.31				0.32		0.39	0.43	0.6	18.1

C1: M1_DG,M2_DG,M7_GD; C2: MS1_GS,MS1_GB; C3: M5_GS,M5_GB;
C4: M5_GD; C5: M4_GB; C6: M3_GS,M3_GB; C7: M4_GS;
C8: M7_GS,M7_GB,M9_GS,M9_GD,M9_GB; C10: M6_DG;
C9: MS1_GD,M1_SG,M1_BG,M3_GD,M6_SG,M6_BG;
C11: M2_SG,M2_BG,M4_GD,M8_SG,M8_DG,M8_BG.

(c)

Figure 4.8: Stressability Enhancement for the OTA Circuit in Figure 4.2(a):
(a) Adding a Switch; (b) Stress Vectors; and (c) Stressability Analysis.

transistor used in switch. Table 4.1 lists the stress time of the region DG of M6 for various aspect ratios of the extra transistor MS1. Results show that the stress time decreases as the aspect ratio increases. Note that the aspect ratio (W/L)=300 was selected for MS1 in Figure 4.7(a) for a stress time of 0.6 seconds. Therefore, depending upon the required stress time in the design specifications, one can improve the stress time by adding extra switches with appropriate transistor size.

Table 4.1: M6_DG Stress Times for Various MS1 Aspect Ratios

W/L(MS1)	Stress Time
1	66.53 seconds
5	31.95 seconds
9	19.60 seconds
10	17.92 seconds
50	2.25 seconds
100	1.17 seconds
200	0.67 seconds
300	0.60 seconds
400	0.53 seconds
500	0.49 seconds
1000	0.36 seconds

Consider the same OTA circuit in Figure 4.2(a) that has a stress time of 84.7 seconds. For sake of discussion, the following cases are discussed:

Case 1: with a feasible stress time of 20 seconds, and

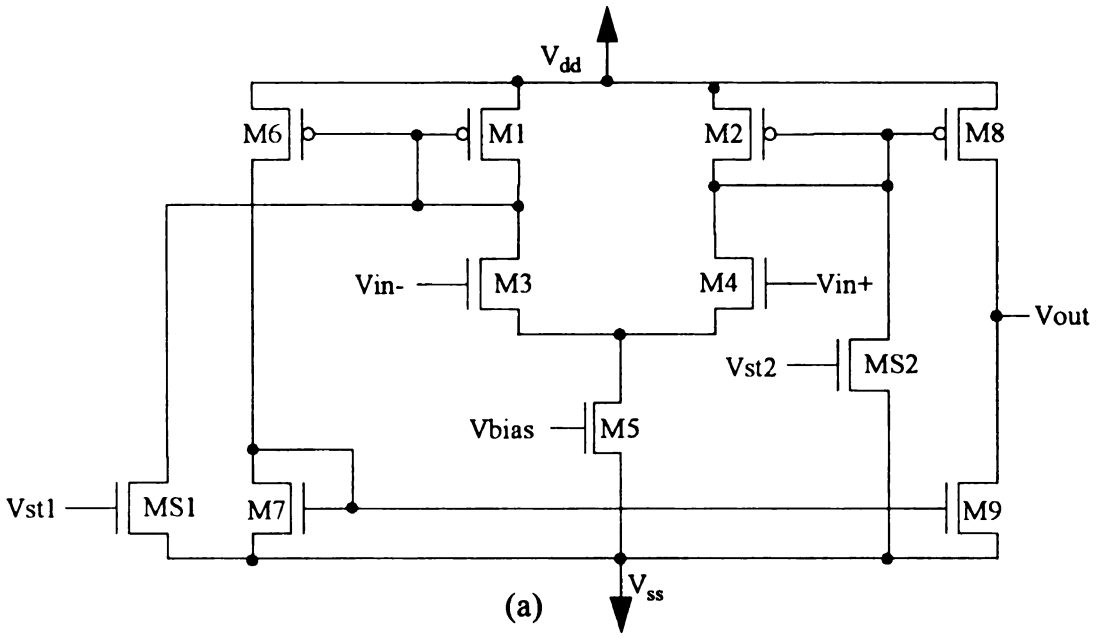
Case 2: with a feasible stress time of 0.6 seconds

Our goal is to provide optimal design solutions for enhancing the stress time. According to the stressability analysis results shown in Figure 4.7(d), the longest stress time is 84.7 seconds, and the next stress time is 18.1 seconds. For Case 1, one can add an extra switch to reduce the stress time for the regions in the column "C11", while one needs two or more extra switches to reduce the stress times for the regions in both columns "C10" and "C11".

For Case 1, an extra switch can be added, as illustrated in Figure 4.8, to improve the stress time. According to Table 4.1, the stress time of 19.6 seconds can be achieved if we choose an aspect ratio (W/L)=9 for the transistor MS1. The stress time then meets the design requirements, that is, the maximum stress time for any region of the circuit is less than 20 seconds.

For Case 2, adding an extra switch MS1 with an aspect ratio of 300 can improve the stress time of C11 from 84.7 seconds to 0.6 seconds which meets the desired requirement. However, the stress time of 18.1 seconds for C10 still exceeds the requirement of 0.6 seconds. Therefore, an additional switch MS2 with the aspect ratio 300, as shown in Figure 4.9(a), is added in order to meet the requirement. Based on the generated stress vectors shown in Figure 4.9(b), the stressability analysis is given in Figure 4.9(c). Results show that the modified circuit fulfil the design requirements, that is, the stress time of none of the regions in the circuit exceed 0.6 seconds. Note that the solution requires two external inputs, Vst1 and Vst2, to control both switches MS1 and MS2. Here, additional switches are used to improve the stressability of the given circuit. The use of extra chip area for adding switches is tolerable in today's chip design.

The above stress time reduction procedure is summarized in Algorithm IV.



(a)

Vst1	Vst2	Vbias	Vin+	Vin-
1	1	1	0	0
1	1	1	0	1
0	0	1	0	1
1	1	1	1	0

(b)

Vst1	Vst2	Vbias	Vin+	Vin-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	1	1	0	0	1	1	1	1					1				
1	1	1	0	1	1	1	1		1		1		1	1		1	
0	0	1	0	1	1		1		1			1					
1	1	1	1	0	1	1	1			1			1		1		1
Number of regions					3	4	2	1	1	2	1	1	4	6	6	1	1
Column stress time					0	0.31						0.37	0.39	0.43		0.60	

C1: M1_DG,M2_DG,M7_GD; C2: MS1_GS,MS1_GB,MS2_GS,MS2_GB;
C3: M5_GS,M5_GB; C4: M5_GD; C5: M3_GB; C6: M4_GS,M4_GB;
C7: M3_GS; C8: M9_GD; C9: M7_GS,M7_GB,M9_GS,M9_GB;
C10: MS1_GD,M1_SG,M1_BG,M3_GD,M6_SG,M6_BG;
C11: MS2_GD,M2_SG,M2_BG,M4_GD,M8_SG,M8_BG;
C12: M6_DG; C13: M8_DG.

(c)

Figure 4.9: Stressability Enhancement for the OTA Circuit in Figure 4.2(a) with Two Additional Switches: (a) Schematic; (b) Stress Vectors; and (c) Stressability Analysis.

Algorithm IV

Stress Time Reduction

1: Let T_{max} be the maximum allowable stress time for a region.

Let TG_{max} be the stress time with a coverage exceeding the requirement.

2: Identify the node responsible for TG_{max}.

3: Enhance the controllability of the node responsible for TG_{max}.

4: Generate the Stress Vectors for the modified circuit.

5: Perform Stressability Analysis and update TG_{max}.

6: IF $TG_{max} > T_{max}$; GOTO 2.

Otherwise, the desired stressability has been achieved.

4.2.2 Pin Overhead Reduction

In this approach, the stressability of analog circuits has been improved using additional transistors and input pins. However, pin overhead is generally very expensive. Therefore, how to reduce the number of external inputs for such applications becomes a very important and practical issue. In order to accomplish this, the modified circuit, that meets the desired stressability requirements, is examined again to determine whether it is possible to reduce the number of external inputs. Let q be the number of switches added in the stressability improved circuit. Since, additional circuitry has been used to enhance the stressability of the circuit which needs to be disabled during normal operation. So, at least one input pin is required that controls the normal and stressed operation and the following property results.

Property 4.3

The minimum value of q = 1, for a stressability enhanced circuit.

Input optimization is based on the selected stress vectors and the conditions required for normal operation of the stressability enhanced circuit. First step is to build an input minimization table, which lists the conditions of each additional input required for the selected stress vectors. The number of columns in this table is q. If the table does not contain a row with values required for normal operation of the circuit, another row is added containing these values. For example, the input minimization table for the OTA circuit of Figure 4.9(a) is shown in Table 4.2. The normal operation of the circuit requires

that Vst1 = 0.0 and Vst2 = 0.0. Since, first row of the table already contains these values, no additional row is required. There are two columns in this table, so q = 2.

Table 4.2: Input Minimization Table for OTA

Vst1	Vst2
0	0
1	1
1	1
1	1

If two columns of an input minimization table are identical, it means that the switches belonging to these columns can be controlled by the same external input. So, a reduced input table is constructed by merging the duplicated columns in the above table. Let s be the number of columns in the reduced table. Hence, the circuit will require s additional input pins and the following property results.

Property 4.4

If s < q, then the number of additional inputs required is s.

For the OTA circuit, since q is greater than one, a reduced input minimization table is built as shown in Table 4.3. The table contains only one column, indicating that s=1, and no further improvement is possible according to Property 4.3. Thus both inputs can be shared, i.e., both inputs can be merged as a single control input, as shown in Figure 4.10.

Table 4.3: Reduced Input Minimization Table for OTA

Vst1
0
1
1
1

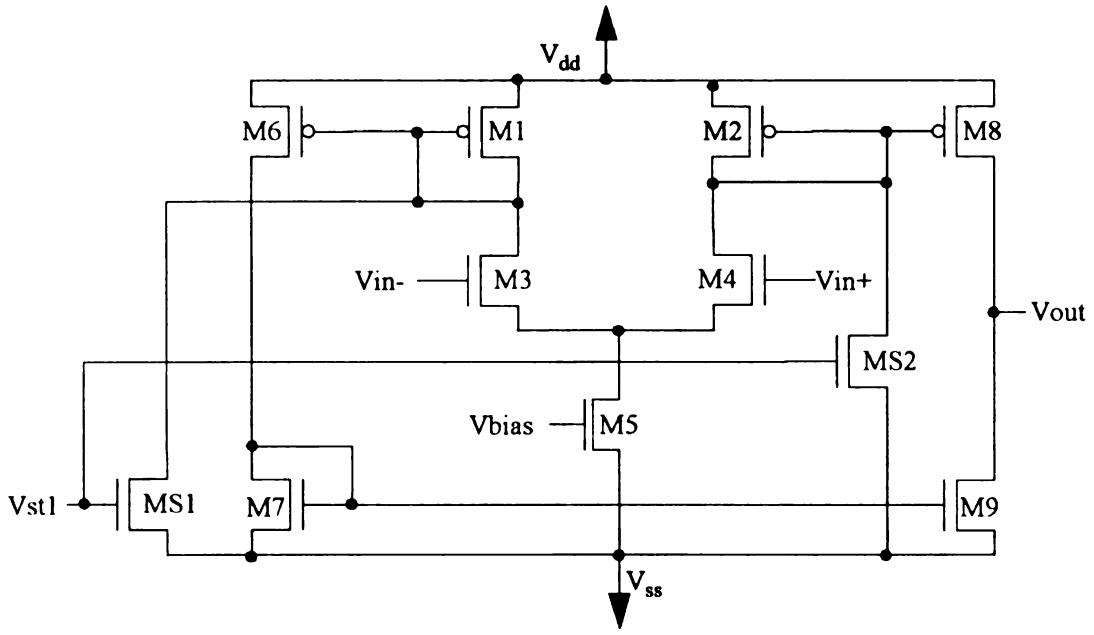


Figure 4.10: Modified OTA Circuit of Figure 4.2(a) with Two Additional Switches and One Input.

As, an inverted output of an additional pin can be generated using an inverter. So, input minimization table is further reduced by eliminating the columns that are complement of other columns. This will remove those columns from the table that belong to the switches which can be controlled by true and complement values of the same external input pin. Let m and n be the number of rows and columns, respectively, in the updated reduced table.

The number of additional inputs required will then be determined based on the values of m and n. If $n \leq \lceil \log_2 m \rceil$, then the number of inputs required is n. However, if $n > \lceil \log_2 m \rceil$, the number of inputs can be reduced to $\lceil \log_2 m \rceil$ by designing a combinational circuit having $\lceil \log_2 m \rceil$ inputs and n outputs. This is summarized in the following property.

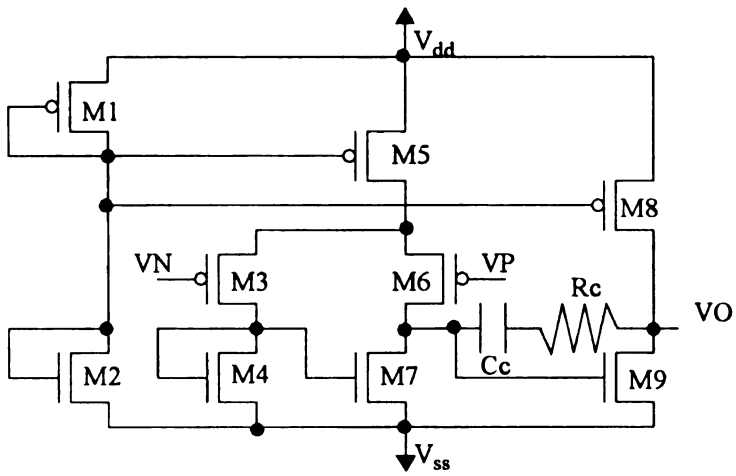
Property 4.5

The number of additional inputs required is n if $n \leq \lceil \log_2 m \rceil$, otherwise, can be reduced to $\lceil \log_2 m \rceil$ by adding extra hardware.

The drawback associated with reduction of inputs using additional combinational circuit is that it is to be ensured that the added circuit can also be stressed with the already selected stress vectors, otherwise, stress vectors for the added combinational circuit need to be added in the set of stress vectors for the circuit.

Consider the benchmark opamp circuit shown in Figure 4.11(a) [66]. The circuit is comprised of transistors, capacitor, and resistor. In this implementation, capacitor is realized using two poly layers and the resistor is implemented using n-well. The circuit has two inputs (VN and VP) and produces an output VO. Based on the selected stress vectors, Figure 4.11(b) shows the stressability analysis results, where the stress time is 202.4 seconds.

Suppose that a feasible stress time of 0.60 seconds is required for the benchmark opamp circuit. Based on the stressability analysis results in Figure 4.11(b), two switches are



(a)

VN	VP	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	1	1	1	1			1	1	1		1
1	0	1	1		1	1			1	1	
Number of regions		3	2	1	5	1	1	6	6	1	1
Column stress time		0	0.02	0.31		0.34	24.3	65.6	202.4		

C1: M1_DG,M2_GD,M4_GD; C2: M2_GS,M2_GB; C3: M3_BG;
C4: M6_SG,M6_DG,M6_BG,M9_GS,M9_GB; C5: M9_GD; C6: M3_SG;
C7: M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB; C9: M5_DG;
C8: M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG; C10: M8_DG.

(b)

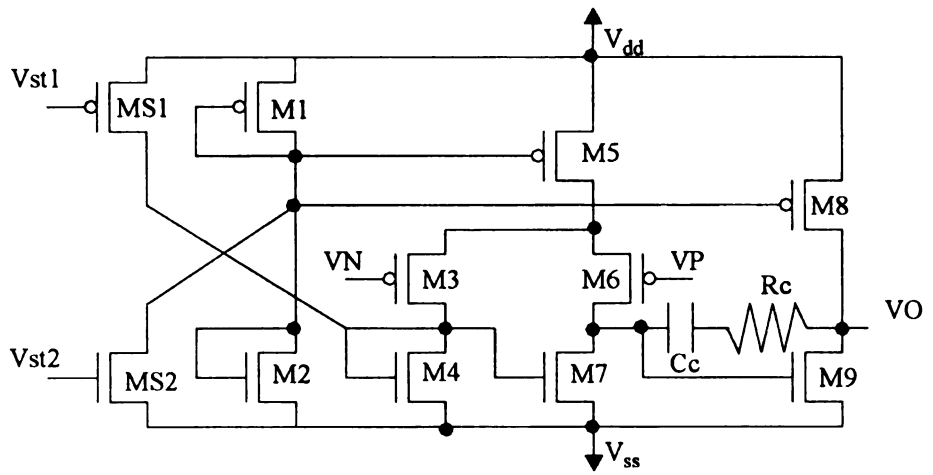
Figure 4.11: Benchmark Opamp Circuit: (a) Schematic; and (b) Stressability Analysis.

added as the modified circuit shown in Figure 4.12(a), one is realized by the PMOS transistor MS1 and the other by the NMOS transistor MS2, where both transistors are respectively controlled by the inputs Vst1 and Vst2. With the aspect ratios 20 for MS1 and 4 for MS2, the stressability analysis results are shown in Figure 4.12(b). The results show that the maximum stress time is 0.57 seconds which meets the requirement, and both the added transistors are fully stressed. Table 4.4 is the input minimization table for this circuit. For normal operation of the circuit, both MS1 and MS2 should be off which is accomplished by applying Vst1=3.3V and Vst2=0.0V. Since, second row of the table meets such conditions, no additional row is needed for normal operation. As the input minimization table contains two columns, so q = 2.

Table 4.4: Input Minimization Table for Opamp

Vst1	Vst2
0	1
1	0
0	1

Since, q is greater than one, a reduced input minimization table is constructed as shown in Table 4.5. There are no identical columns in this table, therefore, s=2. However, the two columns of this table are complements of each other, so further reduction is possible. The updated reduced table for the opamp is shown in Table 4.6. Since, there is only one column in Table 4.6, therefore n=1, and, no further improvement is possible according to



(a)

Vst1	Vst2	VN	VP	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	1	0	1	1		1		1		1	1		1
1	0	1	0	1	1		1		1				
0	1	1	1	1		1					1	1	
Number of regions				3	2	4	5	1	1	1	8	1	7
Column stress time				0	0.02	0.31			0.34	0.41	0.54		0.57

C1: M1_DG,M2_GD,M4_GD; C2: M2_GS,M2_GB; C3: MS2_GS,MS2_GB,MS1_SG,MS1_BG;
C4: M6_SG,M6_DG,M6_BG,M9_GS,M9_GB; C5: M3_BG; C6: M9_GD;
C7: M3_SG; C8: MS2_GD,M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_DG,M8_BG;
C9: M5_DG; C10: MS1_DG,M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB.

(b)

Figure 4.12: Modified Benchmark Opamp Circuit: (a) Schematic with (W/L)(MS2)=4 and (W/L)(MS1)=20; and (b) Stressability Analysis.

Property 4.3. This implies that the patterns of both Vst1 and Vst2 are complemented to each other and the input signal of Vst2 can be generated by inverting the signal Vst1. Therefore, as shown in Figure 4.13(a), a modified circuit is given with a simple CMOS inverter. The stressability analysis results in Figure 4.13(b) show that the regions of all transistors in Figure 4.13(a) are fully stressed including the additional circuitry.

Table 4.5: Reduced Input Minimization Table for Opamp

Vst1	Vst2
0	1
1	0
0	1

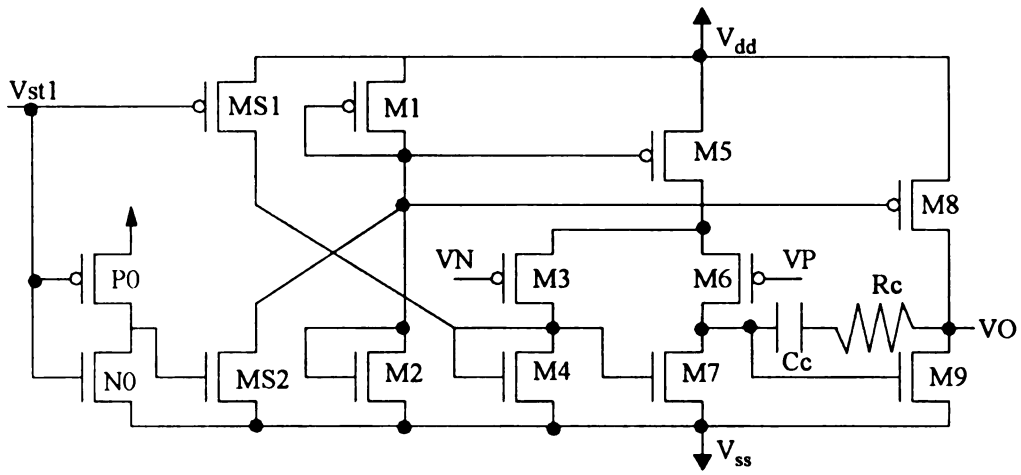
Table 4.6: Updated Reduced Input Minimization Table for Opamp

Vst1
0
1
0

The above pin overhead reduction process can be summarized as in Algorithm V.

4.3 Discussion

This chapter presented a stressability design methodology for circuit designs that can not meet the desired stressability requirements. The goal is to enhance the stressability



(a)

Vst1	VN	VP	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	0	1	1		1		1			1		1
1	1	0	1	1		1		1	1			
0	1	1	1		1					1	1	
Number of regions			3	2	5	8	3	1	1	8	1	7
Column stress time			0	0.02	0.31			0.34	0.41		0.54	0.57

C1: M1_DG,M2_GD,M4_GD; C2: M2_GS,M2_GB; C5 M3_BG; C6: M9_GD;
C3: MS2_GS,MS2_GB,MS1_SG,MS1_BG,P0_SG,P0_DG,P0_BG;
C4: M6_SG,M6_DG,M6_BG,M9_GS,M9_GB,N0_GS,N0_GD,N0_GB;
C7: M3_SG; C8: MS2_GD,M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_DG,M8_BG;
C9: M5_DG; C10: MS1_DG,M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB.

(b)

Figure 4.13: Modified Benchmark Opamp Circuit with Only One Input: (a) Schematic; and (b) Stressability Analysis.

Algorithm V

Pin Overhead Reduction

1: Generate the input minimization table and find q.

Let EI_{min} be the minimum number of extra inputs required.

2: IF Property 4.3 holds, STOP.

3: Construct the reduced input minimization table;

Update the reduced input minimization table to find m and n.

4: IF $n \leq \lceil \log_2 m \rceil$; $EI_{min} = n$; STOP.

5: $EI_{min} = \lceil \log_2 m \rceil$;

Design combinational circuit with m inputs and n outputs.

6: Check added combinational circuit stressability;

Add stress vectors, if required.

of a given circuit, if required. Stressability requirements become a part of the design specifications and will be determined based on the trade-offs among desired reliability, time to market, and cost. Stressability design methodology utilizes stressability enhancement supplemented by stress vector generation and stressability analysis to provide the means to design circuits that can be stressed under predefined constraints. It should be pointed out here that although the design methodology has been developed for analog circuits, yet it is readily applicable to digital and mixed-signal CMOS circuits.

Computational complexity is a major concern for large circuits, but can be handled using hierarchical approach. A circuit partitioning scheme based on the control flow model has been proposed to reduce the computational complexity of stress vector generation and stress time calculation. A hierarchical stress vector generation process based on the topological structure of the circuit has also been developed, in order to improve the efficiency of primary stress vector generation for large circuits. The effectiveness of using the hierarchical approach for large circuits has also been discussed.

It can be concluded that it is feasible to enhance the stressability of a circuit by using extra hardware. A greedy approach is being employed to improve the stressability of a circuit. Stressability analysis of the circuit is performed and the regions with maximum stress time are identified. Stress time is then reduced by enhancing the controllability of the node responsible for the maximum time. The rationale behind using this greedy approach is that in any case the stress time of the region having the maximum stress time has to be reduced for stressability enhancement. Also, the reduction of stress time for such a region generally results in improvement of stress time of some other regions having poor stress time. After enhancing the controllability of the region having the maximum stress time, stressability

analysis is performed again to check whether the requirements are met or not. In case further improvement is required, again the stress time of the region with current maximum stress time is reduced and the entire process is repeated until the desired requirements are met. The area overhead of the developed stressability enhancement technique is limited, and the performance overhead is negligible. The additional transistors consume DC power during stress testing but they are turned off during normal operation. The amount of improvement in stress time using switches can be controlled by varying the area of the switch used. There exists a trade-off between stress time and area of the switch implementation. In case there are no constraints on the area of extra hardware, minimum stress time can be achieved. It has also been shown that the number of additional inputs can be reduced at the expense of supplemental hardware.

Chapter 5

EXTREME-VOLTAGE STRESS VS. EXTREME-TEMPERATURE STRESS

The previous chapters have presented the development of various components of extreme-voltage stress test of analog CMOS ICs, such as stress vector generator, stressability analyzer, and stressability design methodology. This chapter evaluates the performance of the developed extreme-voltage stress test based on the recent IC reliability trends in industry. Section 5.1 introduces the contemporary reliability calculations being used for ICs. Section 5.2 analyzes the existing burn-in process being followed by majority of the IC manufacturers and also suggests ways to improve it. Section 5.3 compares the developed extreme-voltage stress test with extreme-temperature stress and discusses practical issues related to the implementation of extreme-voltage stress test.

5.1 Infant Mortality Period for IC Reliability

As mentioned in Chapter 1, the IC reliability refers to the time that the circuit continues to work to specification after it has passed its initial tests and been designated as a yielding part. Most electronic devices exhibit a decreasing failure rate in their early life, which results from the weak individuals in the total population. These weak devices have

significantly shorter lives compared to the normal (stronger) ones. Such devices are responsible for reliability problems in case these are released to customers or are used to assemble modules or systems. Majority of the failures caused by these devices lie in the early life period of the bathtub curve shown in Figure 1.5. With the advancement in technology and process stabilization, two distinct failure rates have been observed in the early life period. A very high decreasing failure rate initially, referred to as the infant mortality, followed by another decreasing failure rate which is substantially less than the one observed in infant mortality. After early life, devices move into the random regime with an almost constant failure rate. Random regime is followed by the wear-out period. Presently, wear-out does not occur until the product becomes obsolete or reaches its end of shelf life in case of military systems, for most semiconductor devices.

Recent trend in IC reliability is to calculate reliability at the three periods illustrated in Figure 5.1: Infant mortality reliability, early life reliability, and long term reliability. For safety critical and highly reliable circuits, the devices lying in the early life period are screened out. However, for standard commercial products, elimination of devices that fall in infant mortality period meets the desired reliability requirements in most cases. The long term reliability is also referred to as the reliability of the expected life, a part of the product useful life period. For example, Intel Corporation calculates the reliabilities of infant mortality, early life, and long term using the periods 0-50 hours, 0-1 year, and 1-10 years, respectively [67].

For standard commercial products, early life is generally one year and the infant mortality period varies from fifty hours to about a few months depending on the process

stability. Reliability requirements of the product will determine the infant mortality period to be selected for stress testing. Selecting infant mortality period of 1 year will ensure that the devices start their operation with a failure rate of random regime. However, if a higher failure rate is acceptable, then infant mortality period defined for stress testing can be reduced accordingly. The higher the acceptable failure rate, the lower the infant mortality period will be.

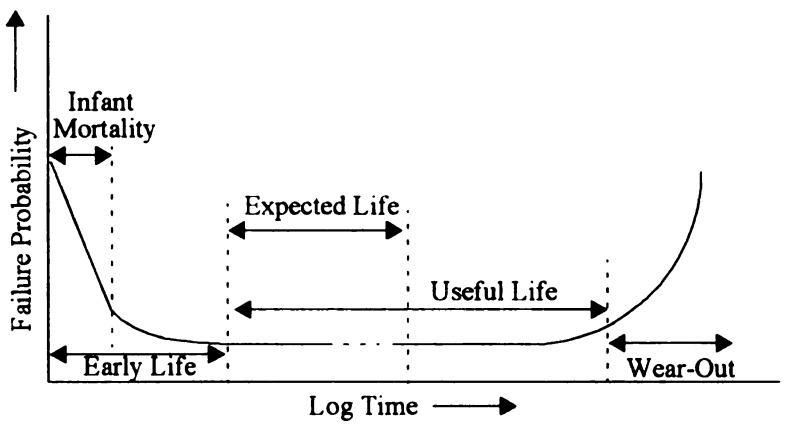


Figure 5.1: Bathtub Curve for Semiconductor Devices.

The stress time required to screen out weak devices using extreme-voltage stress strongly depends on the time period defined for infant mortality. Table 5.1 lists the stress times for different infant mortality periods ranging from the standard commercial products with 50 hours to the safety-critical with 1 year under extreme-voltage stress, with the stress voltage of 5.82V. Results show that, for the standard commercial products, the infant mortality is assumed to be 50 hours and thus the stress time is 0.02 seconds. The stress time increase as the period defined for infant mortality is increased, and, hence for safety-critical products, where the infant mortality is considered to be 1 year, the required stress time is 0.31 seconds.

Table 5.1: Extreme-Voltage Stress Times for Various Infant Mortality Periods

Infant Mortality Period	Stress Time
1 year	310 ms
6 months	210 ms
3 months	140 ms
1 month	70 ms
15 days	50 ms
10 days	40 ms
7 days	30 ms
50 hours	20 ms

5.2 Burn-In Stress

Burn-in is an effective screening method used in predicting, achieving, and enhancing field reliability of ICs. Today, almost all IC manufacturers perform 100% burn-in for various durations to screen defective products. The major problems associated with burn-in are the determination of exactly how long the burn-in process should continue, balancing appropriately the needs of reliability and the total costs, and what stress vectors should be applied? This section describes the temperature-dependent oxide thinning model [68,69]. The model can be used to calculate the stress time required for high temperature and/or high voltage stress. This section also presents stress vector generation and stressability analysis for burn-in along with a comparison of extreme-voltage and extreme-temperature stresses.

5.2.1 Temperature-Dependent Oxide Thinning Model

Figure 5.2 illustrates the concept of oxide thinning. Let Y and L be the gate oxide thickness and length, respectively. The defect size x refers to the amount of oxide thinning

at the localized defective spot. The effective oxide thickness due to defects is defined as $Y_{eff} = (Y - x)$. The oxide breakdown model is given by

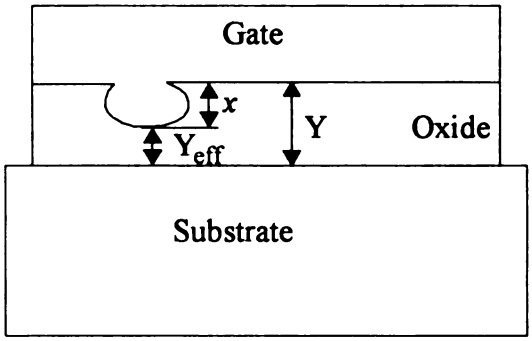


Figure 5.2: Oxide Thinning Concept.

$$t_{BD} = \tau_0(T) \exp[G(T)Y_{eff}/V_{ox})] \qquad (5.1)$$

where

t_{BD} time to breakdown (s);

Vox voltage across the oxide (V);

$\tau_0(T)$ temperature-dependent constant

$$\tau_0(T) = \{\exp[(-E_a/k)((1/T)-(1/300))]\}x10^{-11} \qquad (5.2)$$

k Boltzmann's constant ($8.617x10^{-5}$ eV/K)

E_a Activation energy;

T Absolute Temperature (^oK);

G(T) Temperature-dependent constant (MV/cm)

$$G(T) = 350[1+(\delta/k)\{(1/T)-(1/300)\}] \quad (5.3)$$

δ characteristic constant for G(T).

Note that the parameters $\delta = 0.0167$ and $E_a = 0.28$ eV are best fitted in the temperature range between 25°C and 125°C. For T=300°K, i.e., the room temperature 27°C, G(T)=350 MV/cm and $\tau_0(T)$=10 picoseconds, Eqn. 5.1 is exactly the same as Eqn. 2.1. In other words, the oxide breakdown mode in Eqn. 2.1 assumed a fixed temperature of 27°C. The thickness of oxide Y_{eff} that will be broken when a voltage V_{ox} is applied across it at a given temperature T for time period Z in seconds can be calculated as follows.

$$Y_{eff} = [V_{ox}/G(T)] \ln[Z/\tau_0(T)] \quad (5.4)$$

Based on Eqn. 5.4, Figure 5.3 plots the effective thickness of oxide broken at voltages of 3.3V, 4.5V, 5V, and 5.82V, and at the temperatures of 27°C, 70°C, and 125°C. Tables 5.2-5.4 summarizes the sample points of the plot in Figure 5.3.

Table 5.2: Oxide Thickness(nm) Broken at 27°C for Various Voltages

27°C	3.3V	4.5V	5V	5.82V
0.1 second	2.171	2.960	3.289	3.829
1 second	2.388	3.257	3.618	4.212
1 minute	2.774	3.783	4.203	4.893
1 hour	3.16	4.309	4.788	5.573
6 hours	3.329	4.540	5.044	5.871
8 hours	3.356	4.577	5.085	5.919
24 hours	3.46	4.718	5.242	6.102

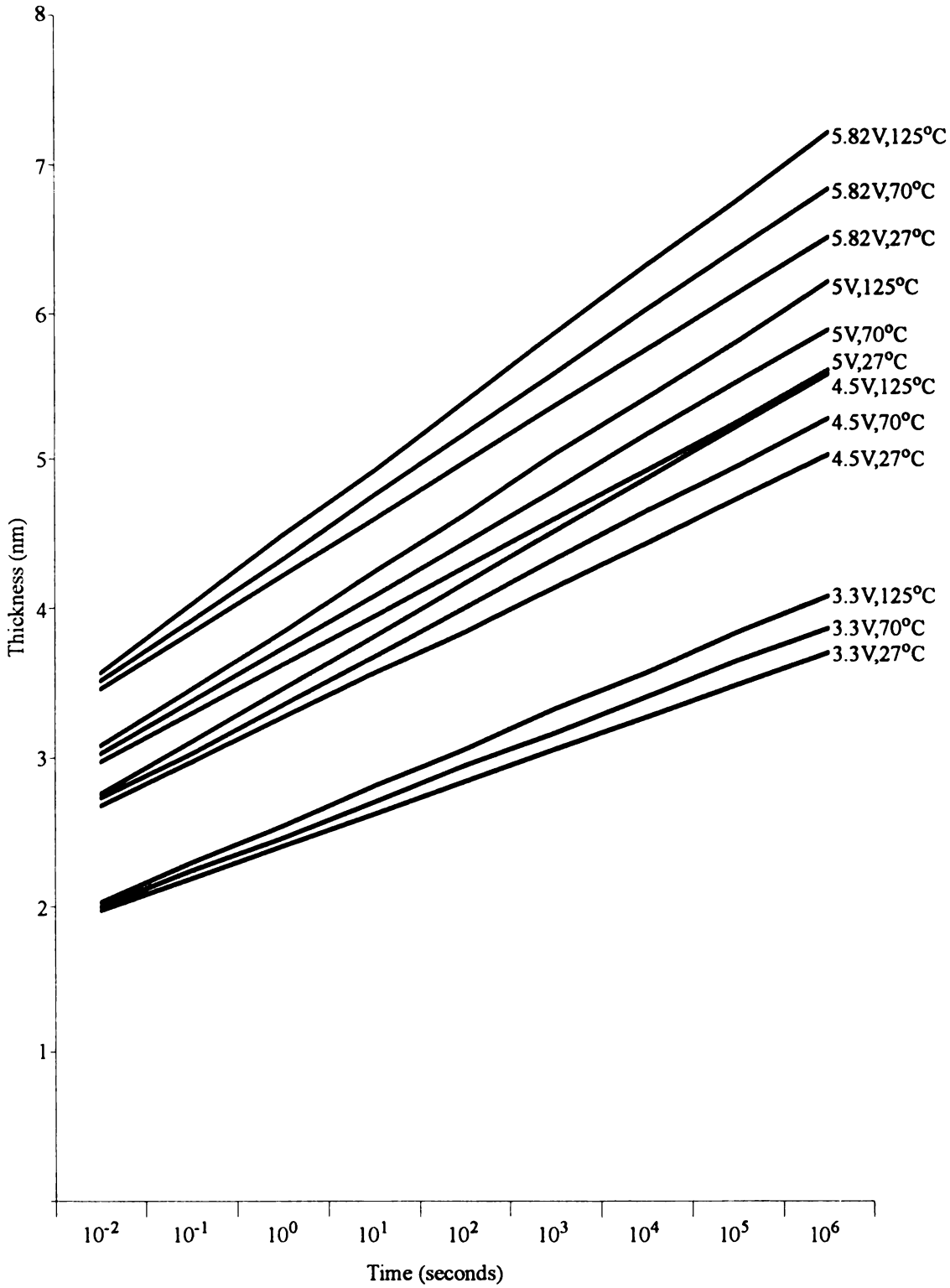


Figure 5.3: Thickness of Oxide Broken at Various Voltages and Temperatures.

Table 5.3: Oxide Thickness(nm) Broken at 70°C for Various Voltages

70°C	3.3V	4.5V	5V	5.82V
0.1 second	2.223	3.031	3.368	3.921
1 second	2.459	3.354	3.726	4.337
1 minute	2.879	3.926	4.363	5.078
1 hour	3.299	4.499	4.999	5.819
6 hours	3.483	4.750	5.278	6.143
8 hours	3.513	4.790	5.322	6.195
24 hours	3.625	4.944	5.493	6.394

Table 5.4: Oxide Thickness(nm) broken at 125°C for Various Voltages

125°C	3.3V	4.5V	5V	5.82V
0.1 second	2.283	3.113	3.459	4.026
1 second	2.541	3.465	3.85	4.481
1 minute	3.0	4.091	4.545	5.291
1 hour	3.459	4.717	5.241	6.1
6 hours	3.66	4.991	5.545	6.455
8 hours	3.692	5.035	5.594	6.511
24 hours	3.815	5.203	5.781	6.729

Based on both Eqn. 5.1 and Eqn. 5.4, the stress time calculation, stress vector generation, and stressability analysis can be developed for burn-in stress process in a manner similar to the one used for the development of these processes in case of extreme-voltage stress test.

5.2.2 Stress Time Calculation

Let V_{yn} be the maximum voltage across the region y (GD, GB, or GS) of a transistor at V_{normal} and room temperature T_{yn}. Let V_{yst} be the voltage across the region at V_{stress} at temperature T_{yst}.

Property 5.1

The stress time for the region y for a lifetime of Z in seconds can be expressed as

$$t_{stress} = \tau_0(T_{yst})[(Z/\tau_0(T_{yn})]^{[G(Tyst)/G(Tyn)(Vyn/Vyst)]} \qquad (5.5)$$

Proof: By Eqn. 5.1 with $t_{BD} = t_{stress}$, $T=T_{yst}$, $V_{ox} = V_{yst}$, and Y_{eff} in Eqn. 5.4, we have

$$t_{stress} = \tau_0(T_{yst}) \exp\{(G(T_{yst})/V_{yst})[V_{yn}/G(T_{yn})] \ln[Z/\tau_0(T_{yn})]\}$$

$$= \tau_0(T_{yst}) \exp[(G(T_{yst})/G(T_{yn}))(V_{yn}/V_{yst}) \ln(Z/\tau_0(T_{yn}))]$$

This concludes that$t_{stress} = \tau_0(T_{yst})[(Z/\tau_0(T_{yn})]^{[G(Tyst)/G(Tyn)(Vyn/Vyst)]}$

For conventional burn-in, we only consider the high temperature stress. The value of voltage appearing across various regions of transistors will depend on the stress vectors applied. If the circuit is subjected only to high temperature stress and vectors are selected in such a way that all the regions get their maximum voltage (V_{yn}), then the stress time will be the same for all regions. In such a case, both voltages V_{yn} and V_{yst} will be the same. Substituting $V_{yn}/V_{yst}=1$ into Eqn. 5.5, the stress time for high temperature stress with proper stress vectors can be calculated as in the following corollary.

Corollary 5.1

The stress time for the region y for a lifetime of Z seconds, stressed only by temperature using vectors satisfying $V_{yn}/V_{yst}=1$, can be expressed as

$$t_{stress} = \tau_0(T_{yst})[(Z/\tau_0(T_{yn})]^{[G(Tyst)/G(Tyn)]} \qquad (5.6)$$

It is evident from Eqn. 5.6 that the stress time required for high temperature stress also depends on the time period defined for infant mortality. The stress times required for various infant mortality periods ranging from 50 hours to 1 year under high temperature stress with the stress temperature of 125°C are listed in Table 5.5. The stress times in Table 5.5 are calculated under the assumption that proper stress vectors are applied, i.e., $V_{yst} = V_{yn}$.

Table 5.5: High-Temperature Stress Times for Various Infant Mortality Periods

Infant Mortality Period	Stress Time
1 year	6 days
6 months	3.3 days
3 months	1.85 days
1 month	17.6 hours
15 days	9.83 hours
12 days	8.15 hours
10 days	6.99 hours
7 days	5.18 hours
50 hours	1.87 hours

Eqn. 5.6 also shows that the stress time for high temperature stress depends on the temperature used. Currently, high temperature stress is applied in the industry using temperatures ranging from 70°C to 125°C. Table 5.6 lists the required stress time for an infant mortality period of 50 hours when the devices are stressed at different temperatures with proper stress vectors.

Table 5.6: High Temperature Stress Times for 50 hours Infant Mortality

Temperature	Stress Time
70°C	9.38 hours
100°C	3.67 hours
120°C	2.12 hours
125°C	1.87 hours

5.2.3 Stress Vector Generation and Stressability Analysis

The modern TDBI (Test During Burn-In) system is capable of applying test vectors during burn-in and is used for efficient high temperature stress in the industry. For digital high-voltage stress, all multi-input logic gates in an IC must experience in-phase toggling to ensure complete stress coverage of the gate oxide, i.e., in case of two inputs A and B, the stress vectors will be (A,B)=(0,0) and (1,1).

For example, consider the benchmark opamp circuit in Figure 4.11(a). Suppose that the circuit is stressed at 125°C with the stress vectors (VN,VP)=(0.0,0.0) and (3.3,3.3). The stressability analysis results are shown in Figure 5.4(a), where the infant

(VN,VP) = (0,0) and (1,1)

Column	Regions	No. of Regions	Stress Time (hours)
C1	M1_DG,M2_GD,M4_GD	3	0
C2	M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG	6	1.87
C3	M5_DG,M8_DG	2	1.87
C4	M2_GS,M2_GB,M3_BG	2	1.87
C5	M3_BG,M6_BG	2	1.87
C6	M3_SG	1	16.63
C7	M3_DG,M4_GS,M4_GB,M7_GS,M7_GB	5	*
C8	M6_SG	1	*
C9	M6_DG,M9_GS,M9_GB	3	*
C10	M9_GD	1	*
C11	M7_GD	1	*

*: Stress time greater than 30 hours.

(a)

(VN,VP) = (0,1) and (1,0)

Column	Regions	No. of Regions	Stress Time (hours)
C1	M1_DG,M2_GD,M4_GD	3	0
C2	M2_GS,M2_GB	2	1.87
C3	M3_BG	1	1.87
C4	M6_SG,M6_DG,M6_BG,M9_GS,M9_GB	5	1.87
C5	M9_GD	1	1.87
C6	M3_SG	1	1.87
C7	M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB	6	1.87
C8	M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG	6	1.87
C9	M5_DG	1	1.87
C10	M8_DG	1	1.87

(b)

Figure 5.4: Stressability Analysis with 50 hours Infant Mortality and Stress Vectors (VN,VP): (a) (0,0) and (1,1); and (b) (0,1) and (1,0).

mortality is assumed to be 50 hours and stress times are calculated using Eqn. 5.5. Since an infant mortality period of 50 hours is considered and stress temperature is 125°C, so using Table 5.6 stress time of 1.87 hours is selected. However, the stressability analysis of Figure 5.4(a) shows that with stress time of 1.87 hours, only 15 regions out of the total 27 regions can be fully stressed, that is, the stress coverage is 15/27 = 56%. The reason for such a low stress coverage is that the selected vectors, (VN,VP)=(0.0,0.0) and (3.3,3.3), can not provide the desired condition of $V_{yst} = V_{yn}$. In other words, when these vectors are applied the voltage appearing across regions of columns C6 to C11 of Figure 5.4(a) is less than V_{yn}.

Now consider C6 which contains region M3_SG, the stress time for this region calculated using Eqn. 5.5 is 16.63 hours. Hence, if a stress time of 16.63 hours is acceptable, then the stress coverage can be increased to 59%. Similarly, the stress time required for column C7, using Eqn. 5.5, is 46.73 hours. Since, the maximum time used for burn-in to eliminate infant mortality is usually less than 30 hours, therefore, if a region requires a stress time greater than 30 hours it is considered that it can not be stressed. Hence, the regions belonging to column C7, i.e., M3_DG, M4_GS, M4_GB, M7_GS, and M7_GB, which require a stress time of 46.73 hours are not stressed. Likewise, all the regions belonging to columns C8, C9, C10, and C11 of Figure 5.4(a), have stress times greater than that of column C7, and, can not be stressed with the selected stress vectors.

From Eqn. 5.6, it can be implied that if the circuit is subjected only to high temperature stress and vectors are selected in such a way that all the regions get their maximum voltage (V_{yn}), then the stress time will be the same for all regions. As discussed in Figure

4.11(b), with the application of the stress vectors (VN,VP)=(0.0,3.3) and (3.3,0.0), all the regions will get their maximum voltage. Now, under the same stress conditions, i.e., the benchmark opamp circuit is stressed at 125°C for an infant mortality period of 50 hours, but with the stress vectors (VN,VP)=(0.0,3.3) and (3.3,0.0). Stressability analysis results are shown in Figure 5.4(b) for this case, where 100% stress coverage can be achieved with the stress time of 1.87 hours.

The above results have demonstrated that the conventional burn-in process is not properly stressing analog circuits because it takes the digital stress vector generation concept for generating stress vectors for analog circuits. Based on the stress vector generation process developed in this study, the stress coverage for the conventional burn-in stress process with normal supply voltage can be improved significantly. The importance of proper vector selection for stressing is obvious when the results are compared for the above two cases. 100% stressability is achieved when the vectors (0.0,3.3) and (3.3,0.0), generated by stress vector generation process, are used whereas the maximum stressability achieved in case of (0.0,0.0) and (3.3,3.3) is only 56% for high temperature stress.

Since for the benchmark opamp circuit, there are two inputs and two vectors are being used to stress it, there can be a total of six combinations of the input vectors that can be used for stressing the circuit. Two of the combinations (0.0,3.3), (3.3,0.0) and (0.0,0.0), (3.3,3.3) have already been discussed. Rest of the combinations were also analyzed to have an exhaustive solution and results are shown in Table 5.7. The exhaustive results also endorse the importance of proper stress vector selection.

Table 5.7: Stressability for Various Stress Vectors

Vectors	Stressability
(0.0,0.0) and (3.3,0.0)	70%
(3.3,0.0) and (3.3,3.3)	70%
(0.0,0.0) and (0.0,3.3)	78%
(0.0,3.3) and (3.3,3.3)	78%

Hence, the stress coverage for conventional burn-in stress process depends on the stress vectors used. The stress time can be improved further if the circuit is also stressed with high voltage in addition to high temperature.

5.2.4 Extreme-Voltage and Extreme-Temperature Stress

The plots presented in Figure 5.3 show that the stress time decreases considerably as both stress temperature and voltage increase. Thus, the stress-time for conventional burn-in process can be reduced significantly by applying extreme-voltage stress simultaneously. For example, consider the same benchmark opamp circuit which is stressed at extreme-temperature i.e. 125°C with the same infant mortality period of 50 hours, but this time with the extreme stress voltage of 5.82V, i.e., using the stress vectors (VN,VP)=(0.0,5.82) and (5.82,0.0). The stressability analysis results, as shown in Figure 5.5, show that the maximum stress time is 979 milliseconds. The improvement is significant compared to the case when only extreme-temperature was used to stress the circuit.

VN	VP	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	1	1	1	1			1	1	1		1
1	0	1	1		1	1			1	1	
Number of regions		3	2	1	5	1	1	6	6	1	1
Column stress time (ms)		0	1	8		9	205	426	979		

C1: M1_DG,M2_GD,M4_GD; C2: M2_GS,M2_GB; C3: M3_BG;
C4: M6_SG,M6_DG,M6_BG,M9_GS,M9_GB; C5: M9_GD; C6: M3_SG;
C7: M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB; C9: M5_DG;
C8: M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG; C10: M8_DG.

Figure 5.5: Stressability Analysis of Benchmark Opamp for 50 hours Infant Mortality under Extreme-Temperature and Extreme-Voltage Stress.

5.2.5 Stress Test for Safety Critical Products

For safety critical products stress vector generation and stressability analysis are similar to the one discussed for standard commercial products with the only difference that the period of infant mortality in this case is 1 year. For example, again consider the benchmark opamp circuit and suppose that the circuit is stressed using conventional burn-in at 125°C with the stress vectors (VN,VP)=(0.0,0.0) and (3.3,3.3). Assuming an infant mortality of 1 year, the stressability analysis results are shown in Figure 5.6(a), where the stress times are calculated using Eqn. 5.5. Table 5.5 shows a stress time of 6 days for an infant mortality of one year at 125°C. Similar to the case of 50 hours with the stress time of 6 days a stress coverage of only 56% can be achieved and can be increased to 59% if a stress time of 72 days is acceptable. Generally, the maximum time allowed for burn-in of safety critical products is 3000 hours, so, it is considered that a region can not be stressed if it requires a stress time greater than 3000 hours. There are eleven regions with stress

(VN,VP) = (0,0) and (1,1)

Column	Regions	No. of Regions	Stress Time (days)
C1	M1_DG,M2_GD,M4_GD	3	0
C2	M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG	6	6
C3	M5_DG,M8_DG	2	6
C4	M2_GS,M2_GB,M3_BG	2	6
C5	M3_BG,M6_BG	2	6
C6	M3_SG	1	72.12
C7	M3_DG,M4_GS,M4_GB,M7_GS,M7_GB	5	*
C8	M6_SG	1	*
C9	M6_DG,M9_GS,M9_GB	3	*
C10	M9_GD	1	*
C11	M7_GD	1	*

*: Stress time greater than 3000 hours.

(a)

(VN,VP) = (0,1) and (1,0)

Column	Regions	No. of Regions	Stress Time (days)
C1	M1_DG,M2_GD,M4_GD	3	0
C2	M2_GS,M2_GB	2	6
C3	M3_BG	1	6
C4	M6_SG,M6_DG,M6_BG,M9_GS,M9_GB	5	6
C5	M9_GD	1	6
C6	M3_SG	1	6
C7	M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB	6	6
C8	M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG	6	6
C9	M5_DG	1	6
C10	M8_DG	1	6

(b)

Figure 5.6: Stressability Analysis with 1 year Infant Mortality and Stress Vectors (VN,VP): (a) (0,0) and (1,1); and (b) (0,1) and (1,0).

time greater than 3000 hours as shown in Figure 5.6(a). However, when the same circuit is stressed using proper stress vectors i.e., it is stressed at 125°C for an infant mortality period of 1 year, but with the stress vectors (VN,VP)=(0.0,3.3) and (3.3,0.0), the stressability analysis results are shown in Figure 5.6(b). Using proper stress vectors 100% stress coverage can be achieved with the stress time of 6 days.

Hence, proper stress vector selection is also necessary for safety critical products. As expected from the results of standard commercial products, the stress time of safety critical products can be improved if the circuit is stressed using high voltage and high temperature simultaneously. For example, consider the same benchmark opamp circuit which is now stressed at extreme-temperature i.e. 125°C with the infant mortality period of 1 year, and also with the extreme-voltage of 5.82V, i.e., applying the stress vectors (VN,VP)=(0.0,5.82) and (5.82,0.0). The stressability analysis results are shown in Figure 5.7. The results show considerable improvement in the maximum stress time from 6 days of extreme-temperature stress to 22.28 seconds.

VN	VP	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0	1	1	1	1			1	1	1		1
1	0	1	1		1	1			1	1	
Number of regions		3	2	1	5	1	1	6	6	1	1
Column stress time (sec)		0	0.01	0.1			3.75	8.64	22.28		

C1: M1_DG,M2_GD,M4_GD; C2: M2_GS,M2_GB; C3: M3_BG;
C4: M6_SG,M6_DG,M6_BG,M9_GS,M9_GB; C5: M9_GD; C6: M3_SG;
C7: M3_DG,M4_GS,M4_GB,M7_GS,M7_GD,M7_GB; C9: M5_DG;
C8: M1_SG,M1_BG,M5_SG,M5_BG,M8_SG,M8_BG; C10: M8_DG.

Figure 5.7: Stressability Analysis of Benchmark Opamp for 1 year Infant Mortality under Extreme-Temperature and Extreme-Voltage Stress.

5.3 Cost Effective Stress Approach

The section compares the stress times for various types of stress employed and also discusses practical issues related to the cost effective implementation of stress testing.

5.3.1 Comparison of Various Stresses

In the following discussion, three types of stress discussed in the previous section, i.e., extreme-temperature (3.3V,125°C), extreme-temperature-voltage (5.82V,125°C), and extreme-voltage (5.82V, 27°C), are compared. It has been assumed that the stress vectors are generated using the developed procedure described in Chapter 3. Table 5.8 lists the maximum stress time for different circuits for the three types of stress. For example, the mixed signal benchmark opamp of Figure 4.11(a) requires stress time of 1.87 hours, 979 milliseconds, and 4.93 seconds for extreme-temperature, extreme-temperature-voltage, and extreme-voltage stress, respectively.

Table 5.8: Maximum Stress Time for Various Stresses with 50 hours Infant Mortality

Circuit	Maximum Stress Time		
	ET(125°C)	ET,EV(125°C,5.82V)	EV(5.82V)
Figure 4.11(a)	1.87 hours	979 ms	4.93 seconds
Figure 4.13(a)	1.87 hours	13 ms	28 ms
Figure 4.2(a)	1.87 hours	514 ms	2.29 seconds
Figure 4.8(a)	1.87 hours	164 ms	590 ms
Figure 4.10	1.87 hours	13 ms	30 ms
Figure 3.4	1.87 hours	11 ms	24 ms
Figure 3.3	1.87 hours	9 ms	18 ms
Figure 3.2	1.87 hours	8 ms	16 ms

As already discussed in Chapter 3, the total stress time for a circuit depends on the number of stress vectors and the maximum stress time for each vector. For example, let the maximum stress time for vectors (0.0,5.82) and (5.82,0.0) of benchmark opamp be denoted by TS_1 and TS_2, respectively. In case of extreme-temperature stress, since all the regions have identical stress time of 1.87 hours, so $TS_1=TS_2=1.87$ hours. Thus, the total stress time is $TS_1+TS_2=3.74$ hours. For extreme-temperature-voltage stress, using Figure 5.5, $TS_1=TS_2=979$ milliseconds and hence, total stress time is 1.96 seconds. Similarly, for the extreme-voltage stress of this opamp, $TS_1=TS_2=4.93$ seconds and the total stress time is 9.86 seconds. The total stress times for the circuits listed in Table 5.8, with an infant mortality period of 50 hours, are tabulated in Table 5.9.

Table 5.9: Total Stress Time for Various Stresses with 50 hours Infant Mortality

Circuit	Total Stress Time		
	ET(125°C)	ET,EV(125°C,5.82V)	EV(5.82V)
Figure 4.11(a)	3.74 hours	1.96 seconds	9.86 seconds
Figure 4.13(a)	5.61 hours	35 ms	76 ms
Figure 4.2(a)	5.61 hours	686 ms	2.9 seconds
Figure 4.8(a)	5.61 hours	185 ms	636 ms
Figure 4.10	7.48 hours	43 ms	95 ms
Figure 3.4	3.74 hours	19 ms	40 ms
Figure 3.3	1.87 hours	9 ms	18 ms
Figure 3.2	3.74 hours	16 ms	32 ms

It is obvious from Table 5.8 and Table 5.9 that extreme-temperature-voltage approach provides the least stress times and extreme-voltage stress provides significant

reduction in stress times compared to the extreme-temperature stress. It should be noted here that both extreme-temperature-voltage and extreme-temperature stress require additional expensive equipment whereas extreme-voltage stress can be handled using the existing equipment. So, the decision to go either for extreme-temperature-voltage or extreme-voltage only will be based on the trade-off between the costs associated with the equipment required for high temperature stress and the costs resulting from longer stress times of extreme-voltage stress. The cost factors will also include customers' reliability requirements and time-to-market while making this decision.

5.3.2 Extreme-Voltage Vs. High-Voltage

The stress time also depends on the stress voltage applied. Note that the extreme voltage is referred to as the stress voltage that is the maximum voltage that can be applied without damaging the device, and the high voltage is a voltage applied to stress the devices ranging between the normal supply voltage and the extreme stress voltage.

For normal operating voltage V_{yn} = 3.3V, the stress times required for infant mortality period of 50 hours when the stress voltage V_{yst} ranges from 5.0V to 5.82V are tabulated in Table 5.10. It is clear that stress time can be reduced by increasing the voltage applied. The goal will be to stress the devices at the highest possible voltage in order to minimize the stress time. However, there can be limitations based on either the technology or the test equipment used. In case the extreme-voltage can not be provided by the existing test equipment, the circuits can be stressed using high-voltage at the expense of longer stress times. Again the decision to upgrade test equipment or to use high-voltage less than

the extreme-voltage will depend on the costs associated with the upgrade and the costs incurred due to longer stress times.

Table 5.10: High-Voltage Stress Times for 50 hours Infant Mortality

Stress Voltage (V)	Stress Time (seconds)
5.0	0.54
5.1	0.33
5.2	0.21
5.3	0.13
5.4	0.09
5.5	0.06
5.6	0.04
5.7	0.03
5.8	0.02
5.82	0.02

5.3.3 Safety Critical Products

The concepts related to the three types of stress and high-voltage stress discussed for standard commercial products with an infant mortality of 50 hours are readily applicable to safety critical products with the only difference that infant mortality period is considered to be one year. The maximum and total stress times for different circuits with extreme-temperature, extreme-temperature-voltage, and extreme-voltage stress using proper stress vectors are listed in Table 5.11 and Table 5.12, respectively. Also, like the previous case stress time depends on the applied stress voltage. For normal operating voltage V_{yn} = 3.3V, the stress times required for infant mortality period of 1 year with values of stress voltage V_{yst} ranging from 5.0V to 5.82V are listed in Table 5.13. It is clear from these

tables that safety critical products follow the same pattern as already discussed for standard commercial products. Hence, the selection of the type of stress and the value of stress voltage to be used will be based on the same cost factors defined for standard commercial products.

Table 5.11: Maximum Stress Time for Various Stresses with 1 year Infant Mortality

Circuit	Maximum Stress Time		
	ET(125°C)	ET,EV(125°C,5.82V)	EV(5.82V)
Figure 4.11(a)	6 days	22.28 seconds	202.4 seconds
Figure 4.13(a)	6 days	0.16 seconds	0.57 seconds
Figure 4.2(a)	6 days	10.71 seconds	84.7 seconds
Figure 4.8(a)	6 days	2.92 seconds	18.1 seconds
Figure 4.10	6 days	168 ms	600 ms
Figure 3.4	6 days	137 ms	470 ms
Figure 3.3	6 days	107 ms	350 ms
Figure 3.2	6 days	95 ms	310 ms

Table 5.12: Total Stress Time for Various Stresses with 1 year Infant Mortality

Circuit	Total Stress Time		
	ET(125°C)	ET,EV(125°C,5.82V)	EV(5.82V)
Figure 4.11(a)	12 days	44.56 seconds	404.8 seconds
Figure 4.13(a)	18 days	0.43 seconds	1.52 seconds
Figure 4.2(a)	18 days	13.73 seconds	103.11 seconds
Figure 4.8(a)	18 days	3.18 seconds	19.01 seconds
Figure 4.10	24 days	0.45 seconds	1.88 seconds
Figure 3.4	12 days	232 ms	780 ms
Figure 3.3	6 days	107 ms	350 ms
Figure 3.2	12 days	190 ms	620 ms

Table 5.13: High-Voltage Stress Times for 1 year Infant Mortality

Stress Voltage (V)	Stress Time (seconds)
5.0	16.19
5.1	9.33
5.2	5.49
5.3	3.30
5.4	2.02
5.5	1.26
5.6	0.80
5.7	0.51
5.8	0.34
5.82	0.31

5.4 Discussion

This chapter was devoted to a comparison between the developed extreme-voltage stress test and the stress tests currently being employed in industry for gate-oxide reliability. Recent trend of calculating IC reliability for three regimes, i.e., infant mortality, early life and long term, has been discussed. Stressability analysis has been performed for the currently used burn-in process by majority of the manufacturers based on the temperature dependent oxide thinning model. A comparison was made between extreme-temperature, extreme-temperature-voltage, and extreme-voltage stresses both for standard commercial and safety critical products.

It has been observed that selection of stress vectors plays a vital role in determining the stressability of a circuit. The results show that the stress vectors currently being used can not achieve 100% stress coverage in case of analog circuits, i.e., they can not properly stress analog circuits, for any type of stress discussed. However, when the stress

vectors generated for extreme-voltage stress are used the three types of stress can provide 100% stress coverage for analog/mixed-signal circuits.

It has also been demonstrated that stress time is dependent on the value of parameter being used for stressing the circuit. Stress time can be reduced by increasing the value under the constraint that the value does not effect the devices adversely. The maximum possible values have been referred to as the extreme value, e.g., extreme-temperature is 125°C and extreme-voltage is 5.82V for the case discussed here. Reduction in stress time has been observed when either the temperature and/or voltage used for stressing is increased. The values lying between normal and extreme are referred to as the high values.

Conventional burn-in is a batch process where up to a thousand assembled units are simultaneously stressed at elevated temperatures in order to accelerate latent reliability defects and processing problems to failure. The key challenge at burn-in is to keep the burn-in time low in order to decrease throughput time and minimize equipment and processing costs. Burn-in time is a function of many variables including the outgoing failure rate, yield, die size, voltage, and junction temperature. The outgoing failure rate is defined by corporate policy while yield and die size are process and product attributes, respectively [70]. Since it has been shown that voltage yields a higher acceleration factor than temperature, so burn-in time can be reduced if the highest possible voltage is used during burn-in.

Hence, extreme-temperature-voltage stress with proper stress vectors results in the least amount of stress time. As discussed earlier, high temperature stress requires expensive burn-in equipment. It was also noticed that extreme-voltage stress provides considerable amount of improvement in stress time compared to the extreme-temperature stress.

The stress time for extreme-voltage stress can be kept within feasible limits using the developed stressability design methodology. The use of only extreme-voltage stress testing has two major advantages, first, no additional infra-structure is required for high temperature stress, and second, it can be incorporated at the wafer level testing resulting in cost savings as the defective devices will be identified earlier in the flow. It was also shown that if for some reason extreme-voltage can not be used then a high voltage can be used at the cost of increased stress time. So, it can be concluded that high voltage stress is generally the most cost effective way to enhance gate-oxide reliability of CMOS circuits. However, cost factors including reliability requirements and time-to-market may influence the decision to select high temperature-voltage stress.

Chapter 6

CONCLUSIONS

In the past two decades, the microelectronics revolution has fundamentally reshaped the global computing and telecommunication industries. The remarkable evolution of microelectronics technology is still continuing, and is now creating economic opportunities in new fields. Reliability and yield are two important factors affecting the profitability of semiconductor manufacturing. Manufacturers are continuously trying to reduce cost and time to market. It is also expected that the zero defects approach will soon become mandatory for the semiconductor companies in order to be competitive in the market.

The development of an extreme-voltage stress test provides reliability enhancement of CMOS ICs cost effectively. It can be used to eliminate early life failures caused by defective gate oxides in CMOS ICs. Extreme-voltage is a voltage level higher than V_{dd}, and is selected based on the required acceleration factor for the target failure mechanism. Extreme-voltage stress test can be used at wafer sort as opposed to burn-in. This reduces assembly costs by identifying defective die at the wafer level so that these devices are not assembled. Defective die identified at wafer level, allows for the elimination of some downstream processes that ultimately result in considerable capital savings and faster time to market.

In this study an efficient yet effective extreme-voltage stress test process for analog CMOS circuits has been developed analytically. A framework of an automatic stress test system for analog circuits, as depicted in Figure 1.8, is developed. Three major components -- stress vector generation, stressability analysis, and stressability design methodology have been integrated into the extreme-voltage stress test system.

6.1 Summary

In the manufacturing process of modern VLSI semiconductor devices, a plasma is often used to deposit or remove material on wafers. The gate oxide may be damaged during plasma etching resulting in reliability failures. Burn-in and high-voltage screening has become the industry standard methods to detect and eliminate early-life failures due to gate-oxide defects. However, the added manufacturing cost of burn-in process may range from 5% to 40% of the total product cost, depending on the burn-in time, qualities of ICs and product complexity. The developed extreme-voltage screening is a practical solution for reliability enhancement and cost reduction.

Extreme-voltage stress test comprises of stress vector generator, stressability analyzer, and stressability design methodology. The basic concepts of stress vector generation and stressability analysis for analog CMOS ICs were presented in Chapter 3. Using the trade-off between the stress coverage and stress time, efficient algorithms were developed to optimize stress vectors under stress coverage and total stress time constraints.

For large circuits, a circuit decomposition model was presented in Chapter 4, that reduces the computational complexity of the problem. Stress vector generation based on the

topological structure was described. Use of hierarchical approach for stress vector generation and stressability analysis was introduced. A stressability design methodology to be incorporated in the design phase was also described. This methodology, in addition to the stress vector generator and stressability analyzer, utilizes the developed stressability enhancement process to improve the stressability of portions of the circuit that do not meet the desired stressability requirements.

Chapter 5 provided a comparison of the extreme-voltage stress test with other tests currently employed in the industry for reliability enhancement of CMOS analog/mixed-signal ICs. It also discussed the practical issues involved in the implementation of extreme-voltage stress test, and proposed techniques to tackle the limitations caused by technology and/or test equipment. It concluded that the stress vectors currently used are inadequate for stressing analog circuits and with proper stress vector selection stressing the devices using only high voltage can be a cost effective solution.

6.2 Major Contributions

This study initiated the research on extreme-voltage stress of CMOS analog circuits to enhance reliability of CMOS ICs without performing the high-cost burn-in. Fundamental concepts required for proper stressability of analog circuits were defined. Based on these concepts stress vector generation procedure for extreme-voltage stress test was developed. An efficient stressability analysis process was developed to evaluate the stressability of a given circuit with the selected stress vectors. This process can also locate portions of the circuit having poor stressability during design phase. An effective stressability enhancement process has also been developed to improve the stressability of problem areas that are

identified by stressability analysis. A stressability design methodology has also been developed that ensures that the design meets the desired stressability requirements utilizing the developed stress vector generation, stressability analysis, and stressability enhancement processes. The above modules laid the foundation for the development of a framework of an automated extreme-voltage stress test system for analog circuits as depicted in Figure 1.8. Although the development was dedicated to analog circuits only, but it is readily applicable to digital and mixed signal CMOS ICs, as well. In order to determine the stress applied to gate-oxides by the existing burn-in, stressability analysis process for burn-in was developed. Various methods for gate-oxide reliability enhancement were compared and the trade-offs that will determine the selection of proper stress test were identified.

6.3 Future Research Area

In the course of this study, a number of interesting yet practical issues have been identified for further study. Analog circuit faults can be classified into *catastrophic (hard)* and *parametric (soft) faults*. The changes to the circuit that cause the circuit to fail catastrophically are catastrophic faults whereas the changes that cause performance degradation are parametric faults [71]. The catastrophic failures are generally detected by DC testing or power supply monitoring [72,73]. In this development extreme-voltage stress is being used to convert parametric faults caused by gate-oxide defects to catastrophic faults which can then be detected easily and cost effectively using DC or functional testing. It was verified from circuit simulation results, so it should be validated physically using test chips. Current testing techniques should be investigated to determine whether they can be used to supplement the detection of such defects.

Efficient and effective partitioning schemes are needed to improve the efficiency of stressability analysis, so that the developed extreme-voltage stress test can be applied effectively to large mixed signal ICs and system-on-chips.

Stress vectors optimization can be improved by developing an algorithm for a minimum weighted covering (MWC) problem. The problem can be stated as follows. Consider a matrix $A=\{a_{ij}\}_{kxm}$, where each entry is either a 0 or a 1, referred to as a 0-entry or a 1-entry, respectively. The matrix describes the relationship among the k row vectors $\{R_1,R_2,..., R_k\}$ and m components $\{g_1,g_2,...,g_m\}$. An entry $a_{ui}=1$ means that there exists a relationship between R_u and g_i, or the component g_i is covered by the row R_u. Each component g_i is assigned a weight w_{gi}. Without loss of generality, the weights are in an ascending order, i.e., $w_{g1} \le w_{g2} \le ... \le w_{gm}$. The row vectors $R_u=<a_{u*}>$ are linearly independent, and the union of all rows R_u's cover all components $\{g_1,g_2,...,g_m\}$. A row weight is defined as $w_{Ru} = \max\{w_{gi} \mid a_{ui}=1, \text{ for } i=1,2,...,m\}$, i.e., the maximum weight of the components g_i's covered by R_u. A MWC-weight is defined as the sum of all row weights, i.e., $w_{MWC}=w_{R1}+w_{R2}+...+w_{Rk}$. Let S_u be an arbitrary subset of R_u with a row weight w_{Su}, by definition, $w_{Su} \le w_{Ru}$. The minimum weighted covering problem is To find a set of S_u's, $S_u \subseteq R_u$, u=1,2,...,k, such that the union of S_u's covers $\{g_1,g_2,...,g_m\}$, and the MWC-weight is minimum.

Stressability enhancement optimization is currently being performed based on the selected primary stress vectors. Improvement in optimization may be achieved if candidate stress vectors are considered instead of primary stress vectors. This issue needs to be investigated and efficient algorithms should be developed to handle the complexity involved in such a case.

The developed extreme-voltage stress test should also be evaluated for other analog CMOS defects. For example, its effectiveness for via defects needs to be verified. The effect of extreme-voltage stress on hot carrier injection also needs to be investigated.

Recently, analog test bus (ATB) (IEEE Standard 1149.4) has been introduced to enhance the testability of analog circuits. The 1149.4 standard requires every chip pin to have an analog boundary module (ABM), which provides the capability of disconnecting the analog circuitry from analog pin and applying the desired voltage on the pin. This is accomplished using the EXTEST instruction. Complex mixed signal ICs and SOCs are expected to use this standard. This will result in the ability to apply developed extreme-voltage stress test more efficiently, due to the enhanced controllability.

REFERENCES

REFERENCES

[1] C.Y. Chang and S.M. Sze, *ULSI Devices*, John Wiley & Sons, Inc., New York, 2000.

[2] *1999 Electronic Market Data Book*, Electronic Industries Association, Washington, DC, 1999.

[3] *1999 Semiconductor Industry Report*, Industrial Technology Research Institute, Hsinchu, 1999.

[4] International Technology Roadmap for Semiconductors, [Online] Available http:// public.itrs.net/Files/1999_SIA_Roadmap/Home.htm, 1999.

[5] W. Kuo, W-T.K. Chien, and T. Kim, *Reliability, Yield, and Stress Burn-in, A Unified Approach for Microelectronics Systems Manufacturing and Software Development,* Kluwer Academic Publishers, Boston, 1998.

[6] D.E. Swanson, "Forty years and looking forward," Semiconductor International, vol. 11, no. 1, January 1988.

[7] E.R. Hnatek, *Integrated Circuit Quality and Reliability*, 2nd Edition, Marcel Dekker, Inc., New York, 1995.

[8] A. Christou, *Integrating Reliability into Microelectronics Manufacturing*, John Wiley & Sons, Chichester, 1994.

[9] Technology Roadmap for Integrated Circuits used in Critical Applications, [Online] Available http://www.sandia.gov/eqrc/critical/critical.html, September 1998.

[10] S.L. Hurst, *VLSI Testing, Digital and Mixed Analogue/Digital Techniques*, IEEE, UK, 1998.

[11] S.L. Hurst, *VLSI Custom Microelectronics: Digital, Analog and Mixed Signal*, Marcel Dekker, Inc., New York, 1998.

[12] W.D. Needham, *Designer's Guide to Testable ASIC Devices*, Van Nostrand Reinhold, 1991.

[13] C.F. Hawkins and J.M. Soden, "Electrical Characteristics And Testing Considerations for Gate Oxide Shorts in CMOS ICs," Proc. International Test Conference, Philadelphia, PA, pp. 544-555, 1985.

[14] C.F. Hawkins and J.M. Soden, "Reliability and Electrical Properties of Gate Oxide Shorts in CMOS ICs," Proc. International Test Conference, Washington, DC, pp. 443-451, 1986.

[15] M.H. Woods, "MOS VLSI Reliability and Yield Trends," Proceedings of IEEE, vol.74, no. 12, pp.1715-1729, December 1986.

[16] T. Kim and W. Kuo, "Modeling Manufacturing Yield and Reliability," IEEE Trans. on Semiconductor Manufacturing, Vol. 12, No. 4, pp. 485-492, November 1999.

[17] M. Syrzycki, "Modelling of Spot Defects in MOS transistors," Proc. International Test Conference, Washington, DC, pp.148-157, 1987.

[18] P. Simon, J.M. Luchies, and W. Maly, "Identification of Plasma-Induced Damage Conditions in VLSI Designs," IEEE Trans. on Semiconductor Manufacturing, vol. 13, no. 2, pp.136-144, May 2000.

[19] T. Brozek, V.R. Rao, A. Sridharan, J.D. Werking, Y.D. Chan, and C.R. Viswanathan, "Charge Injection Using Gate-Induced-Drain-Leakage Current for Characterization of Plasma Edge Damage in CMOS Devices," IEEE Trans. on Semiconductor Manufacturing, vol. 11, no. 2, pp.211-216, May 1998.

[20] J.C. Lee, I. Chen, and C. Hu, "Modelling and Characterization of Gate Oxide Reliability," IEEE Trans. on Electron Devices, vol. 35, no. 12, pp. 2268-2278, December 1988.

[21] K.F. Schuegraf and C. Hu, "Effects of Temperature and Defects on Breakdown Lifetime of Thin SiO2 at Very-Low Voltages," Proc. IEEE International Reliability Physics Symp., San Jose, CA, pp.126-135, 1994.

[22] C.M. Bailey, "Effect of burn-in and temperature cycling on the corrosion resistance of plastic encapsulated integrated circuits," Proc. IEEE International Reliability Physics Symp., Las Vegas, pp.120-124, 1977.

[23] T. Barrette, et. al., "Evaluation of Early Failure Screening Methods," Dig. of Papers, 1996 International Workshop on IDDQ Testing, Washington, DC, pp.14-17, 1996.

[24] R. Kawahara, O. Nakayama, and T. Kurasawa, "The Effectiveness of IDDQ and High Voltage Stress for Burn-in Elimination," Dig. of Papers, 1996 International Workshop on IDDQ Testing, Washington, DC, pp.9-13, 1996.

[25] A.W. Righter, C.F. Hawkins, J.M. Soden, P. Maxwell, "CMOS IC Reliability Indicators and Burn-in Economics," Proc. International Test Conference, Washington, DC, pp.194-203, 1998.

[26] T.Y.J. Chang and E.J. McCluskey, "SHOrt Voltage Elevation (SHOVE) Test for Weak CMOS ICs," Proc. VLSI Test Symposiums, Monterey. pp. 446-451, 1997.

[27] T.Y.J. Chang and E.J. McCluskey, "SHOrt Voltage Elevation (SHOVE) Test," Dig. of Papers, 1996 IEEE International Workshop on IDDQ Testing, Washington, DC, pp. 45-49, 1996.

[28] Y. Zorian, "Tutorial: MCM Testing Strategies," Proc. International Test Conference, Washington, DC, 1995.

[29] Maxim Reliability Report RR-1, [Online] Available http://www.maxim-ic.com/TechSupport/reliability.htm, 2001.

[30] C.L. Wey, "Development of High-Voltage Stressability Analysis and Stressability Design Methodologies for CMOS Analog Circuits," Final Report, Bosch Microelectronics Design Center, Reultingen, Germany, 1999.

[31] M.G. Pecht, R. Radojcic, and G. Rao, *Guidebook for Managing Silicon Chip Reliability,* CRC Press, Boca Raton, Fla., 1999.

[32] R. Degraeve, J.L. Ogier, and R. Bellens, "On the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," Proc. International Reliability Physics Symp., Dallas, pp.44-54, 1996.

[33] J.M. Soden and C.F. Hawkins, "Reliability of CMOS ICs with Gate Oxide Shorts," Semiconductor International, pp. 240-245, May 1987.

[34] W. Kuo and T. Kim , "An Overview of Manufacturing Yield and Reliability Modeling for Semiconductor Products," Proceedings of the IEEE, vol.87, no. 8, pp.1327-1344, August 1999.

[35] H.H. Huston and C.P. Clarke, "Reliability Defect Detection and Screening during Processing-Theory and Implementation," Proc. International Reliability Physics Symp., San Diego, pp.268-275, 1992.

[36] F. Kuper, J. van der Pol, E. Ooms, T. Johnson, R Wijburg, W. Koster, and D. Johnston, "Relation between Yield and Reliability of Integrated Circuits: Experimental Results and Application to Continuous Early Failure Rate Reduction Programs," Proc. International Reliability Physics Symp., Dallas, pp.17-21, 1996.

[37] A.G. Sabnis, *VLSI Electronics Microstructure Science V.22 VLSI Reliability,* Academic Press, Inc., San Diego, CA, 1990.

[38] F. Jenson and N.E. Peterson, *Burn-in An Engineering Approach to Design and Analysis of Burn-in Procedures,* John Wiley & Sons, Inc., New York, 1982.

[39] B. Bergman, "On Reliability Theory and its Applications," Scand. J. Statis. 12, pp. 1-41, 1985.

[40] AT&T, *AT&T Reliability Manual,* Ed. by D.J. Klinger, Y. Nakada, and M.A. Meenedez, Van Nostrand Rheinhold, New York, 1990.

[41] W. Nelson, *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses,* John Wiley & Sons, Inc., New York, 1990.

[42] P.A. Tobias and D.C. Trindade, *Applied Reliability,* Van Nostrand Company, New York, 1995.

[43] E.R. Hnatek, *I* "A Realistic View of VLSI Burn-in," Evaluation Engineering, 28(2), 1989.

[44] S.J. Duey, A.R. Harvath, and P.G. Kowalczyk, "Improved Quality and Reliability Using Operating Extremes Test Methods," Proc. 1993 Customed Integrated Circuit Conference, San Diego, CA, pp.30.5.1-30.5.4, 1993

[45] P. Kowalczyk, "Wafer Level Stress Testing: Containment of Infant Mortality Failures," 1990 WLR Workshop Final Report, pp.169-176, 1990.

[46] R. Roy, K. Roy, and A. Chatterjee, "Stress Testing: A Low Cost Alternative for Burn-in," Ed. by R. Reis, and L. Claesen, *VLSI: Integrated Systems on Silicon,* pp.526-539, Chapman & Hall, 1997.

[47] R. Moazzami and C. Hu, "Stress-Induced Current in Thin Silicon Dioxide Films," Proc. 1992 International Electronics Devices Meeting, San Francisco, CA, pp.139-142, 1992.

[48] D.J. Dumin and J. Maddux, "Correlation of Stress-Induced Leakage Current in Thin Oxides with Trap Generation Inside the Oxides," IEEE Trans. on Electronics Devices, vol. 40, no.5, pp.986-993, May 1993.

[49] D.J. Dumin, J.R. Maddux, R.S. Scott, and R. Subramoniam, "A Model Relating Wearout to Breakdown in Thin Oxides," IEEE Trans. Electronics Devices, vol. 41, no. 9, pp. 1570-1580, September 1994.

[50] H. Watanabe, S. Aritome, G.J. Hemink, T. Maruyama, and R. Shirota, "Scaling of Tunnel Oxide Thickness for Flash EEPROMs Realizing Stress-Induced Leakage Current Reduction," Dig. of Tech. Papers, 1994 Symposium on VLSI Technology, pp. 47-48, 1994.

[51] M. Depas, et al., "Critical Processes for Ultrathin Gate Oxide Integrity," The Physics and Chemistry of SiO_2 and the Si-SiO_2 Interface - 3, vol. 96-1, pp. 352-366, 1996.

[52] MOSIS parametric test results, [Online] Available http://www.mosis.org/cgi-bin/cgiwrap/umosis/swp/params/hp-amos14tb/t07k-params.txt, September 2000.

[53] M.A. Khalil and C.L. Wey, "High-Voltage Stress Test Paradigms of Analog CMOS ICs for Gate-Oxide Reliability Enhancement," Proc. IEEE VLSI Test Symposium, pp. 333-338, Marina del Rey, CA, 2001.

[54] B. Thompson, H-S. Lee, and L.M. DeVito, "A 300-MHz BiCMOS Serial Data Transceiver," IEEE Journal of Solid State Circuits, vol. 29, no. 3, pp. 185-192, March 1994.

[55] T.Y.J. Chang, and E.J. McCluskey, "Quantitative Analysis of Very-Low-Voltage Testing," Proc. IEEE VLSI Test Symp., Princeton, NJ, pp. 332-337, 1996.

[56] R.J. Baker, H.W. Li, and D.E. Boyce, *CMOS Circuit Design, Layout, and Simulation, IEEE Press*, Piscataway, NJ, 1998.

[57] M.A. Khalil and C.L. Wey, "Extreme-Voltage Stress Vector Generation of Analog CMOS ICs for Gate-Oxide Reliability Enhancement," Proc. International Test Conference, Baltimore, MD, 2001.

[58] H. Yoshizawa, Y. Huang, A. Sridharan, P.F. Ferguson, Jr., and G.C.Temes, "MOSFET-Only Switched-Capacitor Circuits in Digital CMOS Technology," IEEE Journal of Solid State Circuits, vol. 34, no. 6, pp.734-747, June 1999.

[59] A.T. Behr, M.C. Schneider, S.N. Filho, and C.G. Montoro, "Harmonic Distortion caused by Capacitors Implemented with MOSFET Gates," IEEE Journal of Solid State Circuits, vol. 27, no.10, pp.1470-1475, October 1992.

[60] J.C.M. Bermudez, M.C. Schneider, and C.G. Montoro, "Linearity of Switched Capacitor Filters employing Nonlinear Capacitors," Proc. IEEE International Symp. on Circuits and Systems, San Diego, pp. 1211-1214, 1992.

[61] S.S. Bazarjani and W.M. Snelgrove, "A 4th Order SC Bandpass Sigma-Delta Modulator designed on a Digital CMOS Process," Proc. 38th Midwest Symp. on Circuits and Systems, Brazil, pp. 1345-1348, 1995.

[62] H. Yoshizawa, G.C.Temes, P.F. Ferguson, Jr., and F. Krummenacher, "Novel Design Techniques for high-lineairty MOSFET-only Switched Capacitor Circuits," Proc. IEEE VLSI Circuit Symp., Honolulu, pp. 152-153, 1996.

[63] H. Yoshizawa and G.C.Temes, "High-Linearity Switched-Capacitor Circuits in Digital CMOS Technology," Proc. IEEE International Symp. on Circuits and Systems, Seattle, pp. 1029-1032, 1995.

[64] W.K. Chen, *The VLSI Handbook*, CRC Press, Boca Raton, Fla., 2000.

[65] B.L. Barranco, A.R. Vazquez, E.S. Sinencio, and J.L. Huertas, "CMOS OTA-C High Frequency Sinusoidal Oscillators," IEEE Journal of Solid State Circuits, vol. 26, no. 2, pp.160-165, February 1991.

[66] Mixed Signal Benchmark Operational Amplifier netlist, [Online] Available http://faculty.washington.edu/manisoma/madtest/benchmarks/OpAmp.htm, April 2001.

[67] Intel Reliability Calculations, [Online] Available http://support.intel.com/support/fittable.htm, April 2000.

[68] J.C. King, W.Y. Chan, and C.Hu, "Efficient gate oxide defect screen for VLSI reliability," Proc. International Electron Devices Meeting, San Francisco, pp. 597-600, 1994.

[69] K.F. Schuegraf and C. Hu, "Reliability of thin SiO2," Semiconductor Science and Technology, vol. 9, no. 5, pp. 989-1004, May 1994.

[70] P. Tadayon, "Thermal Challenges during Microprocessor Testing," Intel Technology Journal, Q3, [Online] Available http://www.intel.com/technology/itj/q32000/articles/art_7.htm, August 2000.

[71] B. Vinnakota, *Analog and Mixed Signal Test*, Prentice Hall Inc., Upper Saddle River, NJ, 1999.

[72] M. Soma, "Fault Coverage of DC Parameteric tests for Embedded Analog Amplifiers," Proc. IEEE International Test Conference, Baltimore, pp. 566-571, 1993.

[73] S.S. Somayajula, E. Sanchez-Sinencio and J.P. deGyvez, "Analog Fault Diagnosis based on Ramping Power Supply Current Signature Clusters," IEEE Trans. on Circuits and Systems Part II, Vol. 43, No 10, pp.703-712, October 1996.