On Ageing Effects in Analogue Integrated Circuits

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Informatik und Mathematik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Felix Salfelder
aus Lörrach

Frankfurt (2016)
(D 30)

ABSTRACT

The behaviour of electronic circuits is influenced by *ageing effects*. Modelling the behaviour of circuits is a standard approach for the design of faster, smaller, more reliable and more robust systems. In this thesis, we propose a formalization of *robustness* that is derived from a failure model, which is based purely on the behavioural specification of a system. For a given specification, simulation can reveal if a system does not comply with a specification, and thus provide a failure model. Ageing usually works against the specified properties, and ageing models can be incorporated to quantify the impact on specification violations, failures and robustness. We study ageing effects in the context of analogue circuits. Here, models must factor in infinitely many *circuit states*. Ageing effects have a cause and an impact that require models. On both these ends, the circuit state is highly relevant, an must be factored in. For example, static empirical models for ageing effects are not valid in many cases, because the assumed operating states do not agree with the circuit simulation results. This thesis identifies essential properties of ageing effects and we argue that they need to be taken into account for modelling the interrelation of cause and impact. These properties include frequency dependence, monotonicity, memory and relaxation mechanisms as well as control by arbitrary shaped *stress levels*. Starting from decay processes, we define a class of ageing models that fits these requirements well while remaining arithmetically accessible by means of a simple structure.

Modeling ageing effects in semiconductor circuits becomes more relevant with higher integration and smaller structure sizes. With respect to miniaturization, *digital* systems are ahead of *analogue* systems, and similarly ageing models predominantly focus on digital applications. In the digital domain, the signal levels are either *on* or *off* or switching in between. Given an ageing model as a physical effect bound to signal levels, ageing models for components and whole systems can be inferred by means of average operation modes and cycle counts. Functional and faithful ageing effect models for analogue components often require a more fine-grained characterization for physical *processes*. Here, signal levels can take arbitrary values, to begin with. Such fine-grained, physically inspired ageing models do not scale for larger applications and are hard to simulate in reasonable time. To close the gap between physical processes and system level ageing simulation, we propose a data based modelling strategy, according to which measurement data is turned into ageing models for analogue applications. Ageing data is a set of pairs of stress patterns and the corresponding parameter deviations. Assuming additional properties, such as monotonicity or frequency independence, a learning algorithm can find a complete model that is consistent with the data set.

These ageing effect models decompose into a controlling stress level, an ageing process, and a parameter that depends on the state of this process. Using this representation, we are able to embed a wide range of ageing effects into *behavioural models* for circuit components. Based on the developed modelling techniques, we introduce a novel model for the *BTI effect*, an ageing effect that permits relaxation. In the following, a transistor level ageing model for BTI that targets analogue circuits is proposed. Similarly, we demonstrate how ageing data from analogue transistor level circuit models lift to purely behavioural block models. With this, we are the first to present a data based hierarchical ageing modeling scheme.

An ageing simulator for circuits or system level models computes long term transients, solutions of a differential equation. Long term transients are often close to quasi-periodic, in some sense repetitive. If the evaluation of ageing models under quasi-periodic conditions can be done efficiently, long term simulation becomes practical. We describe an adaptive two-time simulation algorithm that basically skips periods during simulation, advancing faster on a second time axis. The bottleneck of two-time simulation is the extrapolation through skipped frames. This involves both the evaluation of the ageing models and the consistency of the boundary conditions. We propose a simulator that computes long term transients exploiting the structure of the proposed ageing models. These models permit extrapolation of the ageing state by means of a *locally equivalent stress*, a sort of average stress level. This level can be computed efficiently and also gives rise to a dynamic step control mechanism.

Ageing simulation has a wide range of applications. This thesis vastly improves the applicability of ageing simulation for analogue circuits in terms of modelling and efficiency. An ageing effect model that is a part of a circuit component model accounts for parametric drift that is directly related to the operation mode. For example asymmetric load on a comparator or power-stage may lead to offset drift, which is not an empiric effect. Monitor circuits can report such effects during operation, when they become significant. Simulating the behaviour of these monitors is important during their development. Ageing effects can be compensated using redundant parts, and annealing can revert broken components to functional. We show that such mechanisms can be simulated in place using our models and algorithms.

The aim of automatized circuit synthesis is to create a circuit that implements a specification for a certain use case. Ageing simulation can identify candidates that are more reliable. Efficient ageing simulation allows to factor in various operation modes and helps refining the selection. Using long term ageing simulation, we have analysed the fitness of a set of synthesized operational amplifiers with similar properties concerning various use cases. This procedure enables the selection of the most ageing resilient implementation automatically.

ZUSAMMENFASSUNG (GERMAN ABSTRACT)

Viele alltägliche Geräte in Haushalt, Verkehr, zur Kommunikation oder Medizintechnik werden seit Mitte des zwanzigsten Jahrhunderts von elektronischen Schaltungen beherrscht. Solche Schaltungen messen, steuern und regeln wichtige Eigenschaften, Funktionen und Prozesse und ermöglichen immer weiter ausgefeilte Anwendungen, höhere Effizienz und neue Einsatzmöglichkeiten. Grundsätzlich besteht ein Interesse an der einwandfreien Funktionalität solcher Geräte. Mit elektronischen Halbleiterschaltungen werden zunehmend kleinere Bauteile in diese Geräte verbaut. Einerseits führt dies zu einer wachsenden Systemkomplexität bei mitunter gleichbleibenden Abmessungen, andererseits werden im Zuge der Miniaturisierung physikalische Grenzen erreicht und bisweilen überwunden. Die Einzelteile eines Systems sind nach Fertigung nicht zugänglich und können nicht ausgetauscht werden. Als Konsequenz erwartet man einen Ausfall, sobald ein Einzelteil nicht mehr funktioniert. Um solchen Ausfällen vorzubeugen, möchte man gerne vor der Serienproduktion in Erfahrung bringen, wie sie zustande kommen. Hierzu bedarf es geeigneten Modellen und Analysemethoden. Die Entwicklung solcher im Falle von Analogschaltungen ist der Schwerpunkt dieser Arbeit. In dieser Arbeit

- schlagen wir einen Begriff zur Formalisierung von *Robustheit* vor und
- führen ein tragfähiges Konzept zur Modellierung von *Alterungseffekten* ein.
- Wir untersuchen die *Eigenschaften* verschiedener Alterungseffekte und Alterungsmodelle
- und finden *effiziente Modelle*, mit denen sich Alterungseffekte in Transistoren und Transistorschaltungen darstellen lassen.
- Mit einem datenbasierten *Lernalgorithmus* können wir solche Modelle berechnen.
- Wir entwickeln einen *Schaltungssimulator* zur Berechnung von Langzeittransienten.
- Wir demonstrieren die Vereinbarkeit unserer Alterungsmodelle mit gängigen *Verhaltensbeschreibungssprachen*
- und *simulieren die Alterung* einiger analoger integrierter Schaltungen mitsamt Monitoren, redundaten Strukturen und Ausheilvorgängen.

Alterungseffekte spielen beim Entwurf von Halbleiterelektronik aufgrund der abnehmenden Strukturgrößen eine zunehmende Rolle. In integrierten Schaltungen sind kleinere Strukturen anfälliger für Alterung. Ist ein alterungsbedingter Ausfall eines Systems unvermeidbar, oder gar

erwünscht, so ist natürlich der erwartete Ausfallzeitpunkt von Interesse. Darüber hinaus ist es vorteilhaft, Ausfallzeiten vorhersagen zu können oder über Maßnahmen zur Einflussnahme auf die Ausfallzeit zu verfügen. So wird es gegebenenfalls möglich, mit weniger Ressourcen langlebigere Bauteile zu entwerfen, oder unter Einsparung von Ressourcen Bauteile mit aufeinander abgestimmten Ausfallzeiten zu einem langlebigeren Gesamtsystem zu kombinieren.

Die Robustheit eines Systems steht im Zusammenhang mit seinem Ausfallverhalten. Die Erhöhung von Robustheit erscheint oft als Zielsetzung beim Entwurf von neuen Systemen. Auf welche Weise sich Robustheit beziffern und gegebenenfalls erhöhen lässt ist jedoch unklar. Für einige Aspekte der Qualität eines Produkts haben verschiedene Maße eine gewisse Verbreitung erzielt. Beispielsweise versteht man unter der *Zuverlässigkeit* eines Systems den Kehrwert seiner erwarteten Ausfallzeit. Diese lässt sich jedoch nicht mit *Robustheit* im Sinne des Sprachgebrauchs vereinbaren. Wir werden dies durch die Beschreibung eines Systems mit sehr hoher Zuverlässigkeit illustrieren, welches sich subjektiv nicht robust „anfühlt“. Hinzu kommt, dass Robustheit implizit mit einem Verwendungszweck einhergeht. Wir werden die Verwendungszwecke durch ein *Mission Profile* modellieren und einen hierzu relativen Robustheitsbegriff auf Basis einer gewöhnlichen Spezifikation vorstellen. Dieser Begriff hält in wichtigen Fällen den Erwartungen stand. Beispielsweise wirkt sich der Einbau von Redundanz positiv auf unsere Robustheit aus. Zudem kann ein burn-in Test diese Robustheit verzehnfachen.

Spezifikationsverletzungen in analogen Schaltungen, speziell solche die auf Alterungseffekte zurückgehen, sind schwer zugänglich. Ob und wann eine Schaltung ihren Dienst quittieren wird, kann unter Umständen durch Simulationen vorhergesagt werden. Simulative Verfahren werden ohnehin eingesetzt, um die Eigenschaften von Schaltungen zu berechnen. Liegt ein *nominales* Modell einer Schaltung in einem bestimmten Alter vor, so kann a priori festgestellt werden, ob zum gegebenen Zeitpunkt erwartungsgemäß noch alles in Ordnung sein wird. Oft hängt dieses gealterte Modell jedoch auf nichttriviale Weise vom Modus ab, in dem seine Bauteile bis dahin betrieben wurden. Man benötigt also vielmehr für jedes Bauteil ein Modell eines Alterungsprozesses einschließlich der Abhängigkeit vom Betriebsmodus. Anstelle gealterter Schaltungen kann so also auch das Altern von Schaltungen simuliert werden. Dieses erstreckt sich typischerweise über grössere Zeiträume und die Simulation ist aufgrund der ineffizienten Laufzeiten ohne weiteres nicht praktikabel.

Die Möglichkeiten für eine praktikablere Analyse des Alterungsverhaltens sind vielfältig. Etwa kommt die Einschränkung auf empirische Daten, oder eine post-mortem Betrachtung in Frage. Die Entwicklung funktionsfähiger Alterungsmodelle, sowie die Modellierung von Alterungseffekten auf hohen Abstraktionsebenen und die Entwicklung eines effizienten Simulationsverfahrens stehen in dieser Arbeit im Vordergrund. Hierzu wird zunächst geklärt, welche Effekte mit welchen Methoden modelliert werden können und welche Modelle sich zur effizienten Simulation eignen.

Alterungsmodelle bringen den Betrieb eines Systems in Zusammenhang mit der betriebsbedingten parametrischen Veränderung des Systems. Digitale Schaltungen, die in der Miniaturisierung den analogen Schaltungen traditionell weit voraus sind, lassen sich relativ leicht spezifizieren und modellieren. Der Einfluß von Alterungseffekten auf digitale Schaltungen ist somit zwar größer, jedoch auch leichter zugänglich. Ein Transistor ist zu jeder Zeit entweder an- oder ausgeschaltet. Für die Alterung ist in guter Näherung das Verhältnis zwischen An- und Auszeit und die Anzahl der Schaltvorgänge relevant. So nimmt die Geschwindigkeit der Schaltung graduell dem Alterungszustand entsprechend ab, und eine Verletzung einer Spezifikation kann

identifiziert werden. Die Einschränkung des Betriebszustands eines Transistors auf einige wenige Konfigurationen ermöglicht eine einfache Modellierung sowohl des Alterungszustandes, als auch der parametrischen Verschiebung des Verhaltens. Beispielsweise ergeben sich im Digitalbetrieb stets dieselben Ströme beim Umschalten und stets dieselben elektrischen Felder, die ursächlich für die Alterungseffekte verantwortlich sind. Kennt man diese Einflüsse und ihre Auswirkungen, ist ein brauchbares Alterungsmodell absehbar.

In analogen Schaltungen hingegen treten alle erdenklichen Beschaltungen von Transistoren auf, die Betriebszustände jenseits der im Digitalbetrieb möglichen nach sich ziehen – beispielsweise kann sich ein Transistor im Sättigungsbereich befinden. Auch solche Modi wirken sich auf den Alterungszustand individuell aus und müssen berücksichtigt werden, insbesondere, wenn man sie nicht ausschliessen kann. Sind die durch analoge Beschaltung erreichbaren Alterungszustände modelliert, gilt es noch die Verschiebung der Bauteilparameter von diesen abzuleiten. Auf diese Weise werden wir Alterungsmodelle für Transistoren erstellen.

In analogen Schaltungen gibt es typischerweise keine Entsprechung zu An- und Auszeiten oder zu Schaltvorgängen in digitalen Schaltungen. Es Bedarf eines direkteren Zugangs zu Ursachen und Auswirkungen von Alterungseffekten. Diese stehen im Zusammenhang mit dem Betriebsmodus einer Schaltung. Zur Beschreibung der Alterungsursachen in einer Schaltung definieren wir den Begriff eines *Stress-Levels*. Der Zustand kann sich sich gemäß eines durch den Stress-Level gesteuerten *Prozess* über die Zeit verändern. Von einem solchen Prozess können dann wiederum *Alterungsparameter* abhängen. Die Modellierung gelingt etwa durch physikalische Betrachtungen von Hand, ist aber auch automatisiert mit Lernalgorithmen möglich. Auch Spezifikationsverletzungen sind im Betrieb einer Analogschaltung schwerer auszumachen als in einer digitalen Schaltung. In dieser Arbeit wird die Simulation des Verhaltens und der durch Alterungseffekte induzierten Veränderung des Verhaltens eines nominalen Schaltungsmodells behandelt. Mit Hilfe solcher Simulationen können wir Spezifikationsverletzungen prinzipiell identifizieren.

Modelle für Alterungseffekte stellen die Veränderung von bestimmten Parametern mit dem Alter eines Bauteils nach. Solche Parameter stehen hierbei in direkter Verbindung mit dem Verhalten des Bauteils. Modelle für Alterungseffekte liegen in verschiedenen Varianten vor. So gibt es Modelle, die einen Parameter auf das Alter einer Schaltung zurückführen, manche auf einen durch Belastungseinflüsse herbeigeführten Verschleiß. Wiederum andere Modelle bilden physikalische Prozesse nach und leiten Parameter von deren Zustand ab. Neuerdings spielen auch Prozesse mit Ausheil- oder Relaxationseigenschaften eine Rolle. Auch gibt es rein empirische, phänomenologische Modelle, die zum Beispiel Verteilungen von Ausfallzeiten beschreiben.

Für all diese Modelle stellt sich grundlegend die Frage, wie Alterungs- oder Ausfallmodelle von Einzelteilen auf das Gesamtsystem übertragen werden können. In dieser Arbeit wird eine Methode zur Modellierung von Alterung in analogen Bauteilen, die aus mehreren Transistoren bestehen, vorgestellt und untersucht. Hierzu wird eine Modellierungsmethode entwickelt, die auf Basis von bekannten Daten alterungsbedingte Parameterverschiebungen nachempfindet. Ein Grundbaustein des hier entwickelten allgemeinsten Modells ist ein geregelter Zerfallsprozess. Ein Zerfallsprozess ist ein stochastischer Prozess, der einen Übergang von einem Zustand in einen anderen beschreibt. Durch Steuerung der Zeitkonstante für den Übergang durch eine externe Größe erhält man einen geregelten Prozess. Solche Prozesse können dann gegenläufig gekoppelt werden, um auch Ausheilung zu modellieren. Die Parametrisierung solcher Prozesse zur Nachbildung konkreter Messdaten stellt sich als durchaus machbar heraus. Ein zweistufiges Verfahren zur Suche

von prozessbasierten Modellen hat sich dafür als effektiv erwiesen. Es gelingt zuerst, die Güte eines Modells als Optimum eines konvexen Optimierungsproblems zu formulieren, das effizient berechnet werden kann. Dieses Optimum bewertet die Vereinbarkeit von konkreten Parameterdaten mit Zuständen von gewählten Prozessen. Ein Zustand eines Prozesses lässt sich wiederum explizit aus dem steuernden Stress-Level berechnen. Auf zweiter Stufe sucht ein Lernalgorithmus dann durch Rekombination und Mutation der Prozesse ein Modell mit großer Güte.

Mit Transistormodellen, die Alterungseingenschaften abbilden, ist es prinzipiell möglich, die Parameterverschiebungen ganzer Bauteile zu simulieren. Solche Verschiebungen können anschliessend auch in Verhaltensmodelle der entsprechenden Bauteile aufgenommen werden. Diese Bauteilmodelle werden üblicherweise in Verhaltensbeschreibungssprachen formuliert, jedoch sehen gängige Sprachen keine Alterungseffekte vor. Daher ist zwar ein Verhaltensmodell eines gealterten Bauteils leicht zu erstellen, während alternde Bauteile relativ schwer zu modellieren sind. Solche Modelle sind zur differenzierten Simulation jedoch unverzichtbar. Üblicherweise liegt der Verhaltensmodellierung analoger Bauteile das Paradigma des konservativen Systems zugrunde. Hierbei werden Netze modelliert auf denen sich Potentiale in Entsprechung eingespeister Flüsse einstellen. Die physikalische Interpretation macht daraus etwa ein elektrisches Netzwerk mit Spannungen und Strömen. In dieser Semantik lassen sich gesteuerte Alterungsprozesse ankoppeln, indem der Stress-Level als Fluß und der Alterungszustand als Potential aufgefasst wird. Trotz dieser Kopplung gibt es keine praktikable Möglichkeit, Alterungsprozesse als konventionelles Verhaltensmodell darzustellen, oder gar effizient auszuwerten. Ein gewöhnlicher Simulator interpretiert solche Modelle schliesslich als konservatives System. Praktikabel erweist sich die Auslagerung der Prozessmodelle in Unterbauteile, die – in einer geeigneten Sprache geschrieben – lediglich instanziert werden und mit Netzen des beschriebenen Typs angeschlossen werden. Stehen geeignete Alterungsprozesse zur Verfügung, so obliegt dem Verhaltensmodell des alternden Bauteils nur noch die Modellierung entsprechender Stress-Level und die Interpretation des Alterungszustands als Parameterverschiebung.

Wir zeigen anhand von Beispielen, wie auf diese Weise simulationsfähige Bauteilmodelle mit Alterungseffekten in einer gängigen Modellierungssprache dargestellt werden können. Etwa können betriebsbedingte Parameterverschiebungen in Feldeffekttransistoren und Variationen von Verhaltensparametern einer Inverterschaltung unter Berücksichtigung beliebig allgemeiner Betriebsbedingungen modelliert werden. Dieses Verfahren ist ohne weiteres auf beliebige Verhaltensmodelle übertragbar.

Die Simulation von Alterungseffekten auf Schaltungs- und Systemebene kann mit Modellen, die die Auswirkung eines Betriebs auf die parametrischen Eigenschaften der Komponenten darstellen, durchgeführt werden. Physikalische Modelle können sehr genau sein und sind mit großem Rechenaufwand verbunden. Empirische Modelle erlauben die Entkopplung der Simulation von Alterungseffekten von der Simulation des betroffenen Systems. Diese Modelle ähneln Tabellen und sind effizient auszuwerten, können jedoch sehr ungenaue Ergebnisse liefern, insbesondere im Zusammenhang mit Ausheilung. Zur vernünftigen Simulation von alternden Schaltungen auf jeglicher Ebene bedarf es also zunächst eines Konzepts zur Modellierung von durch Alterungseffekte bedingten parametrischen Verschiebungen. Wir entwickeln ein Alterungseffektmodell, das sowohl den Betriebsbedingungen entsprechend Parameterverschiebungen darstellt, als auch sehr effizient auszuwerten ist. Auf dieser Basis können wir einen Algorithmus zur Langzeitsimulation auf Schaltungsebene formulieren. Dieser Algorithmus macht Gebrauch davon, dass die Alterungspa-

rametermodelle auf gesteuerten Prozessen basieren. Die steuernden Stress-Level lassen eine Art Mittelung zu, die uns eine Extrapolation des Zustands des Effektmechanismus erlaubt. Mit einer Implementierung auf Basis eines Simulators für analoge Schaltungen belegen wir die Durchführbarkeit entsprechender Simulationen. Diese ist stark von der Wahl der Alterungsmodelle abhängig.

Effiziente Simulation von Schaltungen samt Alterungseffekten ermöglicht verschiedene Anwendungen, die auf Simulation aufbauen. Etwa kann der Entwurf von Monitorschaltungen simulativ unterstützt werden. Obliegt die Auswertung der Alterungsmodelle dem Simulator, kann man sich dem Design der Schaltung selbst zuwenden. Eine Monitorschaltung kann dann zum Beispiel ein System überwachen, so dass Systemfehler im laufenden Betrieb gemeldet werden oder gleich Gegenmaßnahmen ausgelöst werden können. Es ist uns gelungen, solche Monitore als Teil eines alternden und wieder ausheilenden Gesamtsystems korrekt zu simulieren. Grundsätzlich verbessert effiziente Simulation auch alle Verfahren zur Ausbeute- oder Parameteroptimierung. Wir zeigen, dass Alterungseffekte während der Schaltungssynthese berücksichtigt werden sollten. Bei der Schaltungssynthese geht es darum, die Performanz verschiedener Schaltungsimplementierungen für ein und dieselbe Spezifikation zu vergleichen. Ist auch der Einsatzzweck einer Schaltung zum Zeitpunkt der Synthese bekannt, so kann beispielsweise eingeplant werden, wie lange eine Implementierung dem Betrieb standhält, oder wie stark Alterungseffekte den Betrieb individuell beeinflussen können. Empirische Alterungsmodelle scheiden in diesem Zusammenhang sicher aus, da während der Synthese beliebig viele Beschaltungen auftreten, in denen Alterungseffekte unterschiedliche Auswirkungen haben können. Wir haben im Zusammenhang mit der Synthese von Operationsverstärkern eine Machbarkeitsstudie zur anwendungsfallbezogenen Auswahl durchgeführt: Mithilfe von Simulationen stellt sich heraus, dass je nach Einsatzzweck unterschiedliche Implementierungen vorteilhaft sein können. Auf diese Weise können wir somit effizient einem Verwendungszweck eine geeignete Implementierung zuordnen.

LIST OF SYMBOLS

σ-robustness (R_σ) the σ-robustness

a norm ($|x|$) a norm or the absolute value of $x \in \mathbb{R}^n$
ageing parameter (p) a parameter subject to ageing
ageing process (Q) a stresswave dependent ageing state
ageing state space ($\mathbb{A}$) set containing all ageing related states of a system, circuit, component
append ($\odot$) append stresswaves (infix operator)

bias temperature instability (BTI) a transistor ageing effect
Boltzmann constant (k_B) fundamental physical constant $[\mathrm{m}^2\,\mathrm{kg}\,\mathrm{s}^{-2}\,\mathrm{K}^{-1}]$
branch current (I_{xy}) current from `x` to `y` (through a component)
branch voltage (V_{xy}) voltage at a node `x` relative to `y`

Cartesian product ($\times$) the cartesian product (infix operator)
composition ($\circ$) concatenate maps (infix operator)

domain (dom) the domain (of a function)

ERCD (ERCD) RCD with exponential control
expectation (E) the expected value (of a random variable)
exponential function (exp) The map $x \mapsto \mathrm{e}^x$

field effect transistors (FET, n-FET, p-FET) semiconductor device, n- and p-channel variants

Heaviside function (Θ) the Heaviside function $\mathbb{R} \to \{0, 1\}$
hot carrier injection (HCI) a transistor ageing effect

Landau symbol (O) Landau big O
level (L) a level function $\mathbb{T} \to \mathbb{R}$

mathematical constants (e, π) Euler's number, half perimeter of unit circle (unitless)
mean time to failure (MTTF) the expected failure time
mission (M) a specified fixed operation of a system.

natural logarithm (ln) The inverse of exp
number sets ($\mathbb{N}, \mathbb{Z}, \mathbb{R}$) (nonnegative) integers and reals

probability (P) the probability (of an event)
probability space (Ω) a set of samples together with a set of events

rate (λ) a (transition) rate [Hz]
RCD (RCD) reversely coupled decay process
reaction-diffusion (RD) process involving transformation and spatial distribution of particles

SI-units (A, V, Ω, m, s...) various SI-units

SRC (SRC) switch capacitor process
state space ($\mathbb{X}$) set containing states of a system, circuit, component
stress level function (L) a stress valued function on $\mathbb{T}$
stress wave (L, w) a stress level on $[0, T] \subset \mathbb{T}$

temperature (ϑ) a temperature [K]
threshold voltage (V_{th}) threshold voltage [V]
time (t, T, τ) a time [s]
times ($\mathbb{T}$) the set of times [s] starting from zero
total differential (D) the total differential (of a function)
transpose ($\cdot^{\mathsf{t}}$) the transpose (of a vector or matrix)

variance ($\mathrm{Var} = \sigma^2$) the variance (of a random variable)

year (yr) a year ($\approx \pi \times 10^7$s)

CONTENTS

CHAPTER

ONE

INTRODUCTION

1.1 Integrated Circuit Design

Moore's law [Moo65] and its variants have been predicting a decrease in structure size of electronic systems with a surprising accuracy over decades. No matter, if the prognosis will remain this accurate or not, there will always be pressure to miniaturise parts for the sake of higher integration of more complex systems. Because less material will always be cheaper, particularly material of the expensive kind. Clearly, smaller parts are more susceptible to process fluctuations and to random physical displacement effects [GR10]. Hence, with increasing integration it becomes harder to maintain designs that are functional. Aging effects aggravate the situation, demanding designs that also maintain functionality during operation, e. g. a resilience with respect to aging.

Process fluctuations and randomness are usually addressed by simulation. So called *yield analysis* performs nominal simulations of not-yet manufactured component models with unknowns assigned randomly (Fig. 1.1). This gives an estimate of the percentage of components that will eventually work after manufacture. *Yield optimization* is devoted to increase this percentage changing nominal parameters of the original design.

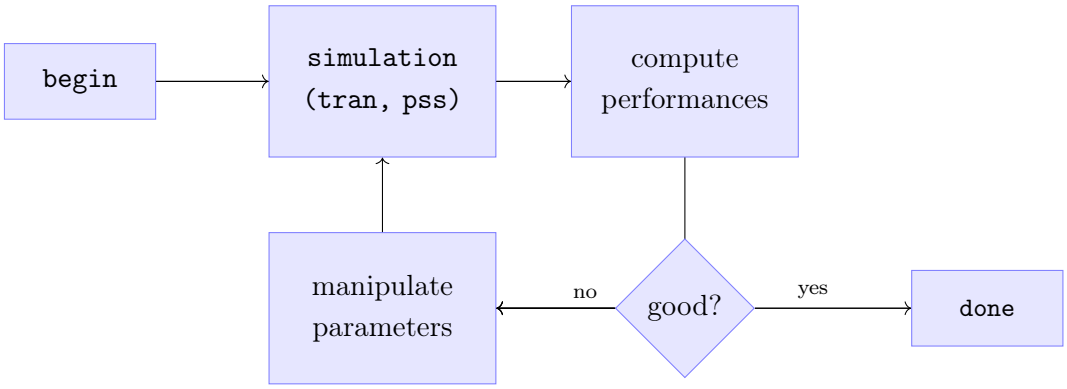


Figure 1.1: Yield optimization.

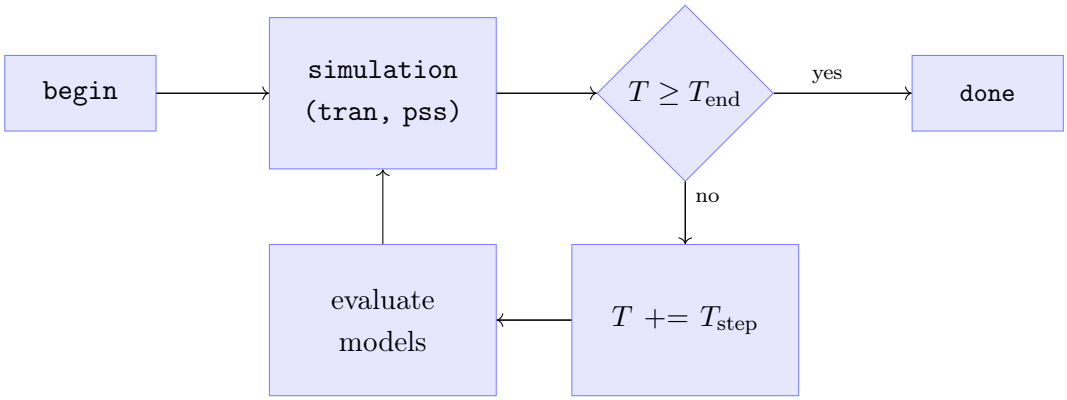


Figure 1.2: Reliability simulation: simulating circuits with increasing age.

In contrast, aging effects change the device properties and must be taken into account for sensible yield analysis and effective optimization. Parametric shift induced by aging is hardly correlated with process variation, as the mechanisms are independent. Reliability simulation (Fig. 1.2) refers to simulating circuits with parameters affected by ageing effects. Parametric drift can be attributed to the operation or task a system is performing. In general, it is challenging to take into account the operation environment of a component before manufacture. Various physical or empirical models exist that predict nominal parametric shifts under some conditions [Ber+06]. Often simulation is required to complement empirical data. Particularly optimization highly depends on nominal data that traces back to input and environmental parameters.

1.2 Contribution

This work introduces a new notion of *robustness* that is based on previous research. We define robustness of a system as a property of a failure model with respect to a specification of the system. With this, robustness is directly coupled to wearout and ageing effects.

An abstacted sustainable view on ageing effects and ageing effect models reveals *properties* of ageing effects that are relevant during modelling. Ageing effect parameter models can be described in terms of stress levels and ageing processes. We introduce a class of aging process models with advantageous arithmetic properties. This class permits fast evaluation and is provably well suited for data based modelling of various ageing effects. It covers stress level dependent ageing effects involving relaxation. We introduce a *data based learning* algorithm in order to find model parameters. Using suitable optimization routines we sucessfully learn ageing data sets for stress level controlled parametric effects. With these techniques we create the first model for the *BTI effect* that works in efficient *analogue circuit simulation*.

Our abstraction plays well with the semantics of todays *hardware description* languages. The concept addresses the gap between physical level ageing effects and behavioural level circuit simulation. Our approach enables the integration of ageing effects into functional behavioural block models. Our simulator demonstrates the feasibility of long term ageing analysis of such

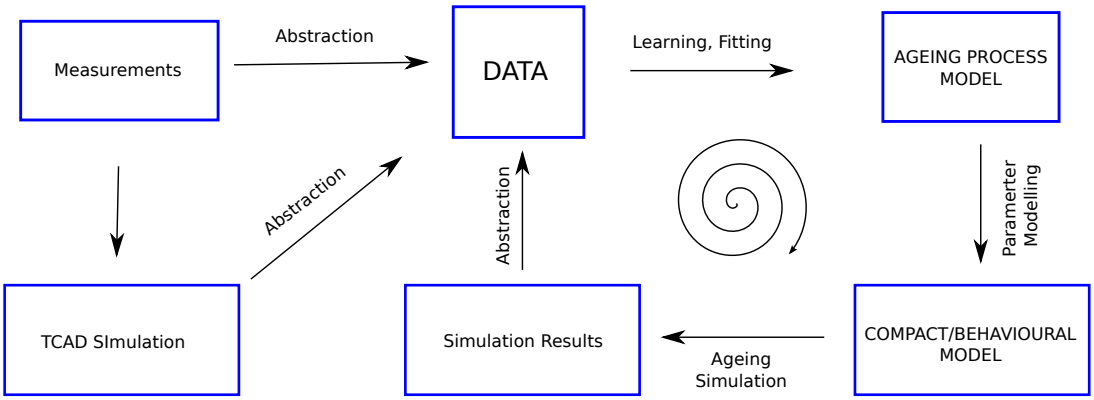


Figure 1.3: Hierarchial Ageing Models from Data.

models. It provides nominal analogue circuit and analogue ageing effect evaluation, computes long term transients efficiently, exploiting the arithmetic structure of the ageing effect models we introduce. This simulator can be used to simulate analogue circuits with ageing effects under heterogeneus conditions such as prescribed by a *mission profile*. We demonstrate various applications including the analysis of on-line ageing monitors and redundant structures and ageing aware circuit synthesis.

We identify and solve problems with existing ad-hoc approaches on ageing effect modelling and simulation. Our solutions address the essentials of an automatic hierarchical data based ageing modelling and analysis workflow (Fig. 1.3) that works for analogue and mixed-signal components. The proof of concept implementations are based on existing free (as in freedom) software and are publically available for the most part.

CHAPTER

TWO

STATE OF THE ART

Dependability, lifetime, reliability or robustness of electronic circuits are a major concern during the design of integrated circuits. Yet these properties are hard to measure or to even define. The *mean time to failure* (MTTF) is a standardized term [08] that is deemed to characterize the reliability of a system. Similarly, the *mean time between failures* (MTBF) is meant to account for the reliability of systems with (manually) replaceable parts [IEC14]. These numbers are useful for identical systems that run under constrained and approximately the same conditions and with a common failure time distribution. Recent work [ROB09] is targeting a measure for robustness, which seeks a complementing characterization. It is meant to incorporate a relation to the task or *mission* that is to be carried out [Nir+14] and to take into account subjective connotation of the word *robustness* [Rad+10; Bar+12]. See Chapter 3 for an examination, further details and related material.

2.1 Ageing Effect Mechanics

Behavioural parameters in electronic circuit components are not constant throughout lifetime. This traces back to physical properties of the used materials. In most cases, a specific functional part of a device wears out, gradually reducing its fitness for the mission it has been assigned to. Electronic circuits are electrical networks that consist of conductor, semiconductor and insulator parts. When they are small, the parts are exposed to extremely high electrical fields and dense currents. The corresponding forces can affect the geometric configuration as well as electrical properties. Several *ageing effects* have been observed, studied and modelled in various ways. Roughly, ageing effects can be categorized as either continuous or discontinuous and some involve annealing, see Table 2.1 for an overview.

2.1.1 Electromigration (EM)

An electron that collides with a metal atom can displace the metal atom. A current flowing through a conducting path on a silicon chip has densities much higher than in usual electric

Table 2.1: Typical ageing effects

effect/level	continuous	impact
EM drift	✓	(none)
EM tear/shortcut	✗	permanent, destructive
HCI	✓	parametric ΔV_{th} drift
TDDB	✗	temporary, permanent
BTI noise	✗	single events
BTI degradation	(✓)	temporary, ΔV_{th} shift

devices [Bla69]. Hence, the probability of a displacement during operation is relatively high. In practice, the conducting material moves with the electric current. Depending on the situation, a conductor may for example tear or build up a short circuit to a neighboring conductor, breaking functionality in either case.

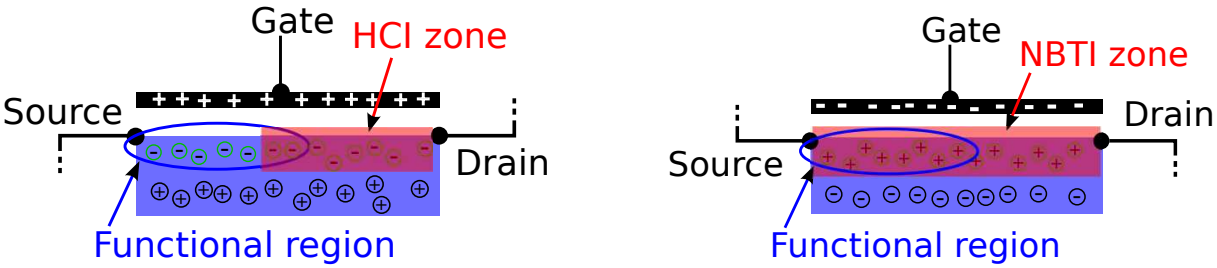
Whilst electromigration, as a macroscopic effect, is relatively easy to understand, to observe and to measure, simulation is complex. Due to the impact on the geometry, a simulation model must factor in the positions of the involved atoms so it can predict how and when a conductor will tear or have moved into a neighboring one. Thus, these breakdown events are usually modelled statistically [YN00] and avoided by increasing the structure size. Fortunately, in current use cases, migration effects in conductors that precede a breakdown do not significantly affect the circuit. Changes in position or diameter are irrelevant, as long as the Ohmic resistance is not affected much. After a significant move, the effect will accelerate, thus lead to a quick breakdown and will be covered by the statistics.

2.1.2 Time Dependent Dielectric Breakdown (TDDB)

Insulators in transistor devices are subject to high electric fields. These fields pull on particles charged by impact ionization. The displaced particles may build up conducting paths, similar to dendrites, between channel and gate. Depending on the circumstances, a dendrite may either feature a permanent short cut or fall victim to a sufficiently high short current. Usually, statistical models predict the TDDB rate from empirical data under constant stress conditions [KGA13]. Hardening against TDDB faults becomes increasingly challenging with newer CMOS technologies. Notably, the sata controller breakdown in Intels Sandy Bridge chipset was reportedly caused by TDDB and cost about $700 million.

2.1.3 Hot Carrier Injection (HCI)

Hot Carrier Injection (HCI), a. k. a. Channel Hot Carrier (CHC), is a semiconductor wearout effect that mainly affects n-FET devices [WWS86]. Electrons that are crossing the channel of an n-FET are experiencing a high acceleration due to the electric field and their high mobility. Similar to electromigration, collisions with high energy electrons lead to permanent damage of



(a) HCI mainly affects the drain region.

(b) BTI may degrade the whole channel.

Figure 2.1: Physical locations of FET ageing effects in forward mode.

the channel material. Electrons are fastest close to the drain region, see Fig. 2.1a. The change in device characteristics concerns the switching capabilities can be measured in terms of a shift in threshold voltage. The first and still state-of-the-art model [WWS86; Tu+93; LMM06] integrates a harm of level L of a leakage current over time. The harm is attributed to the fast electrons and their impact on the oxide. It is usually derived from the currents that are essential part of the n-FET compact model [She+87]. The total damage then relates to the induced parameter shift, which also corresponds to the damage state of the device. This model,

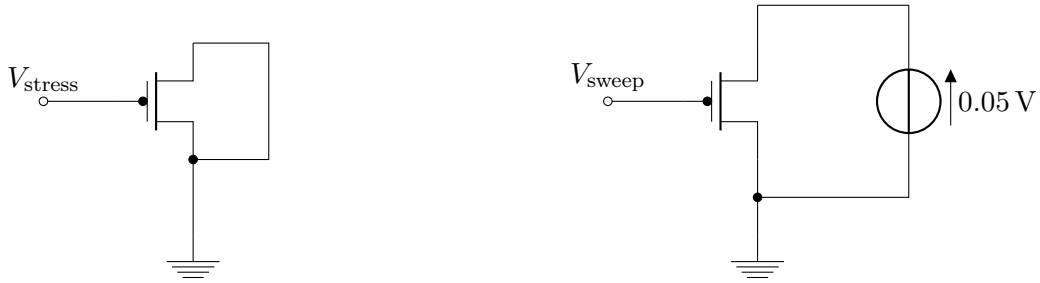
$$\Delta V_{\text{th}}(t) = \Delta V_{\text{th}} \left(\int_0^t L(t')\mathrm{d}t' \right), \tag{2.1}$$

has been extensively fitted to various technology nodes and applies to transistors in digital as well as analogue circuits. Unlike the electromigration model that establishes an expected failure time, HCI models predict a time continuous shift in device parameters.

Notably, the commonly used HCI effect model is asymmetric. Neither does the defect state model account for reverse currents, nor does the parameter shift (Eq. 2.1) even approximate the reverse characteristic in a useful way. This is interesting, because the actual parameter shift can be much bigger [Cho12]. In analogue circuit simulation, the MOSFETs are traditionally symmetric, w. r. t. drain and source. The HCI effect is not (see Fig. 2.1a). In Section 6.1.1 we provide an approach to work around this limitation.

2.1.4 Bias Temperature Instability (BTI)

The most intriguing ageing effect presently known is *Bias Temperature Instability (BTI)*. BTI affects the properties of field effect transistors, primarily shifting their threshold voltages [JS77]. A field effect transistor relies on moving charge carriers in a semiconductor material. These carriers are influenced by an electric field that in turn controls the conductivity through a *channel*. The very same electric field also supports a slow decay of the carriers (Fig. 2.1b), thus the field required to maintain conductivity increases. The absence of the electric field causes the transistor device to recover the charge carriers. This way a dysfunctional component left to itself may regain its use. Also intervals in operation can significantly increase the lifespan.



(a) BTI stress conditions.

(b) BTI measurement.

Figure 2.2: The usual BTI modelling is based on specific stress situation and parameter observation.

Similar to HCI, a basis for modelling is an additive impact ΔV_{th} on the threshold voltage V_{th}, sometimes also an impact on carrier mobility. Common models essentially express a functional dependency

$$\Delta V_{\text{th}}(t) = \Delta V_{\text{th}}(E_t, \vartheta_t) \tag{2.2}$$

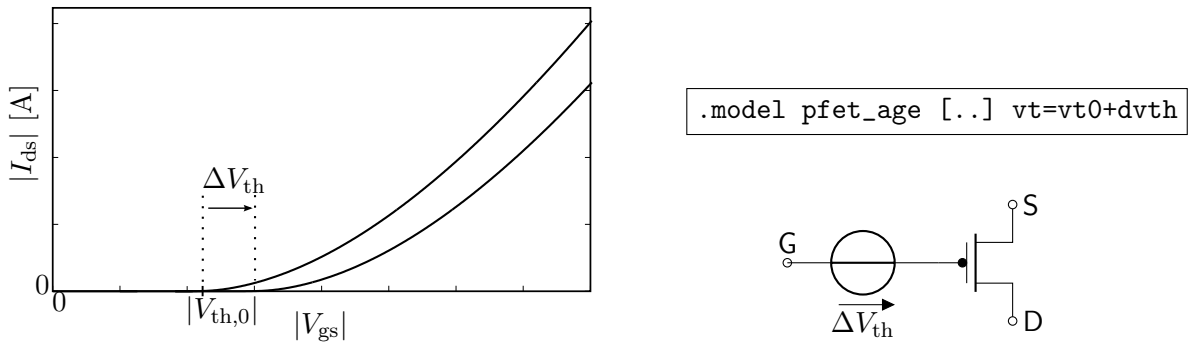
where E_t and ϑ_t are functions for an electric field and a temperature over time, until t, respectively. The stress field model traces back to transistors used in digital circuits. Here, the electric field and degradation impact can be considered homogeneous in space. Also, the defects close to the source are most relevant (see Fig. 2.1b). The derived static threshold voltage shift (Fig. 2.3a) then accounts for the performance of the degraded device. Stated more clearly, the model Eq. 2.2 is meant to describe the following interrelation. A transistor exposed to V_t at a temperature of ϑ_t as in Fig. 2.2a, put in the test bench in Fig. 2.2b, will exhibit a threshold voltage shifted by $\Delta V_{\text{th}}(t)$. Circuit models such as in Fig. 2.3b incorporate the shifted characteristic.

The mechanism that underlies BTI has been controversial from the day of its discovery, which likely dates back to 1966 [GHN66]. One of the first physics based models [JS77] describes a chemical reaction at the transistor channel that releases hydrogen that diffuses through the gate oxide (and back).

Free hydrogen leaves a defect at the interface and is responsible for the threshold voltage shift. The exact interrelation is likely more complex [Sch06]. Multiple variants and extensions of this *Reaction-Diffusion* (RD) model have been proposed and considered to explain experimental data. For example, multiple sorts of diffusors [OS95; KA86], various diffusion speeds, gate thickness [Bha+06], or dispersion parameters [Kac+08] have been contemplated.

Recent models [Sch+10; Gra+10; Gra+11] restrict to what happens at the interface, and do without considering diffusing particles. A *defect site* or *charge trap* is the location of a potential defect. An actual defect constitutes a contribution to the threshold voltage shift, depending on where it is located in the channel (Fig. 2.1b). The electrical field controls how the defects arise and heal up, see Section 4.3.4.

It is commonly believed that the temperature dependence of the BTI mechanism is uniform. According to [Gra+09]temperature affects the rates of the *stress* and *relaxation* mechanisms the



(a) Shifting the threshold.

(b) Static BTI impact modelling.

Figure 2.3: Static ΔV_{th} drift expressed as model or as netlist.

same way, in compliance with an Arrhenius law. This means that a model which is valid for one temperature will also work for another after rescaling the time appropriately [Gra+09]. With this, we may drop the temperature dependency completely during modelling. The dependency on frequency is more likely to be complex. While it has been reported that BTI is frequency independent for a wide range up to 10 Mhz [Ala03], some researchers claim otherwise and investigate in physically accurate models that reflect frequency dependence properly [Gra+12]. See Section 4.1.3 for more details and modelling aspects.

2.2 Analogue Ageing Effect Models

On circuit level, ageing effects manifest as parameter deviation. Ageing parameter models consist of two parts. One part models a devices overall ageing state that is meant to express the constitution of the material, see Fig. 2.1. A second part maps this ageing state to behavioural parameters. For example, electromigration on circuit level behaves like a switch that changes its state at the time determined by the physical constitution.

The HCI model Eq. 2.1 exhibits a simple, deterministic and continuous structure. The integral of the harm function L is the device state, and the parameter deviation function maps back to a behavioural parameter.

The BTI case is considerably more complex. The defects area is bigger (Fig. 2.1b), and also relaxation is possible. Corner cases, mostly of the digital kind, are well charted [VWC06; Wan+07]. These BTI models require that the shortened drain and source (Fig. 2.2a) suffiently approximates the stress conditions and ΔV_{th} in Fig. 2.3b correctly models the device properties. This is essentially valid for digital applications and probably for low level MOS models. In analogue applications either of these simplifications is questionable (see Section 6.1.3).

Often, measurements taken for custom applications still provide parameter drifts over time (such as Eq. 2.3) and are then used for reliability simulation purposes. This approach is referred

to as *semi-empirical* modelling, and may well be in tune with a circuits performance in retrospect. If E is restricted to be homogeneous in space and rectangular over time, a stress pattern can be described easily. For example switching back and forth n times between 0 to E_0 for some time t, with an effective on-time βt (see [VWC06; KKS06]). Then a threshold voltage shift model is a function

$$\Delta V_{\text{th}}(t) = \Delta V_{\text{th}}(E_0, \beta, n, \vartheta, t).$$

According to [Ber+06] similar models are based on functions f_1, f_2 to predict a shift that can be expressed as

$$\Delta V_{\text{th}}(t) \sim f_1(t) f_2(E),$$

which is only valid for specific interpretations of E, such as an average over time and space. Perhaps this latter approach targets empirical exploration and is meant to fit known trajectories.

2.3 Modelling and Simulation

Circuit reliability simulation has its roots in the digital domain. State of the art digital semiconductor circuits have always been built from much smaller components than their analogue counterparts. Smaller components are more susceptible to wearout and displacement effects. To no surprise, the modelling of ageing effects in digitally operated circuitry takes the lead.

For most practical considerations, the signal levels in digital circuits are restricted to two levels plus switching in between. This simplifies the mechanics of ageing effects dramatically. Ageing mechanisms that cause a permanent degradation and are susceptible to currents (such as EM and HCI) may have an impact Δp on a parameter $p = p_0 + \Delta p$. This impact can be modelled as a relative shift of the number of switching cycles n, and $\Delta p = \Delta p(n)$. Similarly, parameters that are susceptible to high or to low signal levels such as BTI (without relaxation) can be expressed by means of the total on-time t_{on} with $\Delta p = \Delta p(t_{\text{on}})$. Taking into account relaxation and restricting to periodic conditions, a model $\Delta p(t, t_{on})$ can be used for the parameter shift at a time t [VWC06; Wan+07; Ala03; OS95; KKS06; Wit07]. Ageing effect parameters for a fixed component in a fixed application may collapse to a time dependent parameter shifted by $\Delta(t)$ and only the total age is of relevance.

Such parameters models give rise to reliability simulation [Tu+93; Cad]. A reliability simulator assigns time dependent parameters $p = p(t)$, and simulates a circuit at different values for t Fig. 1.2. These often follow a **power-law**

$$t \mapsto p_0 + A \cdot t^n. \tag{2.3}$$

The coefficients A and n may be user input or as well be deduced from models that process the output of transient simulations. Reliability simulation works well if the stress patterns involved in parameter deviation are of the digital kind, where Eq. 2.3 can be adopted from corresponding lower level simulations, where similar conditions are assumed. In other cases reliability simulation, requires that parameters are known for the particular context, e. g. from measurements or if they are computed by concurrent satellite TCAD simulators [JS05].

The integration of ageing effect models into circuit component models is another approach to reliability simulation. A *circuit ageing simulator* essentially simulates a circuit as a whole, with ageing effects, whilst the circuit is performing its duty over a lifetime. For quite some time, this has been the only access to reliability investigations in the analogue domain. Circuit ageing simulation is sometimes referred to as reliability analysis with *updated conditions*, and incorporated into commercial tools [Cad]. Without updated conditions and based on Eq. 2.3, curiously wrong results have been computed, see Section 5.5.2. Modelling circuit level ageing effects that permit updating conditions is nontrivial in general c. f. Chapter 4.

A key issue with reliability simulation of analogue circuits is a missing concept of suitable analogue parameter degradation models. According to [Cao+14], a BTI model that is suited for circuit level simulation does not exist. This is most credible – but also, a literature search reinforces the statement. Reading the fine print and references of publications about analogue circuits involving ageing effects reveals digital levels and single-use empirical models under the hood [GD08; Yil+13]. These are valid under prescribed custom conditions, and sometimes backed by measurements. Yet they can not work for analogue circuits in general, c. f. Section 4.2.3.

Ageing effects simulation in full semiconductor based technology-CAD models, are a viable path to predictions of parameter deviations and ranges [Pfä+12]. Simulation of these models however is orders of magnitude more complex than circuit simulation based on the usual semiconductor compact models. Efforts have been made to couple circuit simulation with lower level technology models [JS05]. Still, this approach impedes fast simulation on the circuit end and restrict the signal levels on the technology side. For example, efforts to evaluate the RD model have been successful for rectangular signal patterns [Gra+06; Wit07]. Symbolic solutions exist for the constant stress case [OS95].

2.3.1 Ageing Simulation Algorithms

Simple ageing effect models permit ageing simulation. Under appropriate conditions, the harm function from Section 2.1.3 can be computed by a circuit simulator using a corresponding transistor model. The parameter deviation feeds back into the circuit simulator. Under the assumption that average harm level is constant, the ageing component state may be computed at any age in constant time c. f. Section 5.5.2. Even if this average is not constant, evaluating this model in the context of a circuit has a simple run time complexity. To some extent, algorithms have been implemented in commercial tools [LMM06; Cao+14]. Ageing parameters that do not fit into Eq. 2.1, e. g. due to relaxation properties, are harder to cope with, c. f. Section 4.1.2. In digital circuits a parameter shift such as ΔV_{th} does not depend much on the operating point. Hence, repeated simulation (Fig. 1.2) using static degraded models (Fig. 2.3b) can be considered accurate.

Algorithmically, there are two approaches to circuit ageing simulation. One is integrating ageing effect models into ordinary circuit models and simulate a long transient. *Warping* the relative rate of ageing effects can reduce simulation times [Mar+06] by a constant factor, see Section 5.2.2 for side effects. A second is repeatedly simulating short transients and extrapolate

ageing component states in between. This variant is referred to as *two times* simulation, more generically as *multi rate* simulation. Two times simulation requires models that support explicit extrapolation between simulations (c. f. Section 4.3.5) and involves step control on the ageing time axis. The ageing time step width is constrained by the model evaluation algorithms as well as by circuit operation mode, see Section 5.5 for some examples. Current state of the art is user input based step control. Commercial tools [Cad] let the user set the number of steps to be computed. A more advanced approach [MG13] is based on circuit performances. From these, a step size can be derived. A similar problem is known from step control in conventional transient simulators [Vla94; Dav03]. These solve a differential equation mainly using implicit methods. The step size is chosen such that the intrinsic integration error is below a tolerable bound.

Simulator run time is a critical constraint for applicability. An evident bottleneck is ageing state evaluation, particularly with respect to circuit state. The ageing state evolves over time depending on the operation a circuit performs. Hence, analogue circuit ageing simulation requires a model that permits fast evaluation such as Eq. 2.1 or Eq. 2.3 while at the same time approximates the behaviour of a BTI mechanism such as Equations 4.18 and 4.19. Trap-charge based models [Sch+10; Yil+13] are ready for use in digital simulations, see Section 4.2.3. See Section 4.3 for a generic view on ageing model evaluation.

2.3.2 Hardware Description

Block level models of semiconductor circuits abstract the essential behaviour from an electrical network in order to reduce simulation complexity. Such models usually maintain the external connections of a network whilst replacing the internal functionality by arithmetic relations and abstracted state machines [SZ04; OSC05]. Ageing effects on block level are common practice in the digital domain. No different to the transistor parameter shifts, the ageing impact on delays in digital logic circuits are described in terms of rectangular periodic stress patterns [DS09]. The structure of these ageing parameters are independent of the circuit implementation. In the analogue domain the most advanced block ageing models incorporate transistor netlists [Mar+06]. Here, a usual behavioural model of an amplifier is backed by a transistor netlist that hosts the ageing mechanisms. The advanced approaches to BTI in transistor compact models [LQB08] are based on precomputed age dependent parameter shifts following some power-law (Eq. 2.3).

Hardware description languages do not currently support ageing parameter semantics or facilitate modelling. This applies to transistor compact modelling as well as behavioural block level modelling. In the conclusion of [DPD10], the authors ask for a solution for this. See Section 5.6 for an approach.

CHAPTER

THREE

ROBUSTNESS OF SYSTEMS

3.1 Preliminaries

Lately, efforts have been made to formalize the otherwise subjective term of *robustness* in the context of machines, technical equipment and gadgets. Robustness is meant to formally define a property that is close to its intuitive connotation, which is related to determinism, reliability, immunity, longevity and endurance. The overall consensus is, that robustness of a system needs to be defined with respect to the purpose, or a *mission* it is meant to carry out [Nir+14; Bar+12]. Desirable criteria on a robustness term for systems should in our opinion look like the following. It turned out to be important that the list does not include too specific properties.

- Robustness is a numerical property of a system model.
- Robustness is independent of how the specification of that system is written.
- Restricting the purpose of a system increases determinism, hence robustness should increase.
- Robustness takes into account the specification of the system.
- A Robustness should induce a *partial order*. In particular being of higher robustness should be transitive.
- This should specialize to a *total preorder*(Germ. "Präferenzordnung", lit. "preferring order") when restricted to implementations of the same specification. This property makes robustness useful for making choices.
- Anything that "feels" less robust must be ranked as less robust.

We suggest that a general approach to finding a formal notion for robustness should be guided by some simple rules. We shall refer to these rules as the **roadmap**.

1. Find a numerical property that may be related to robustness.

2. Check well-definedness.
3. Check scope of applicability.
4. Check correlation with subjective experience in definite cases.
5. Discard property if it does not comply (and rethink).

To get started, we reconcile a simplified interpretation of the essentials from [Rad+10; ROB09]. We shall not differentiate between real existing entities and their respective models.

Definition 3.1. *A **system** F is a map taking inputs to outputs. **In- and outputs** are time dependent signals of any kind directed into or out of the system, respectively. F may also be a random variable with the corresponding semantics.*

Since the *purpose* of a component is entangled with robustness, we need a suitable model for such a purpose which is not too restrictive.

Definition 3.2. *A **mission profile** M is a measurable set of inputs and run time conditions equipped with a* probability distribution. *A set of missions with its probabilty then constitutes a purpose.*

Some care needs to be taken when defining probability distributions on function spaces [Doo47], hence we chose this very basic innocuous notion.

We may now take a mission $m \in M$, a function over time, and run the system F under m. The output $F(m)$ is another time dependent function. We need properties to characterize inputs and to specify the system. System properties can be described by permitted output signals or by a transfer characteristic of some sort. The former approach is less restrictive and easier to deal with.

Definition 3.3. *A property space is a set. A **property** or **characteristic** is a map c that takes a signal to an element of a property space.*

A set of voltages, Ξ is a property space. The mean voltage characterizes a signal. The choice of the property space is application dependent. If for example, the peak voltage and period width is of interest as well, then, a cartesian product $\Xi \times \Xi \times \mathbb{T}$ is the property space to choose.

Finally we need to decide whether a device *works*. Commonly, a device is said to work if it behaves as specified. While a device is modeled in terms of signals, the specification can only be formulated in terms of properties. For a map $f\colon A \to B$, we write $\operatorname{dom}(f) := A$ for the **domain** of f. A **partial** map $g\colon A \to B$ is a map $a \to B$ for a subset $\operatorname{dom}(g) := a$ of A.

A **specification** S of a system is a partial map from the set of input properties to sets of output properties. In other words, a specifiction declares that under some conditions, the system should work in a certain way. We say that

Definition 3.4. *a pair of signals (s_{in}, s_{out}) is **covered** by S, if $c(s_{in}) \in \operatorname{dom}(S)$ and $c(s_{out}) \in S(c(s_{in}))$. A system F **meets** S, if for all signals s that satisfy $c(s) \in \operatorname{dom} S$, the pair $(s, F(s))$ is covered by S.*

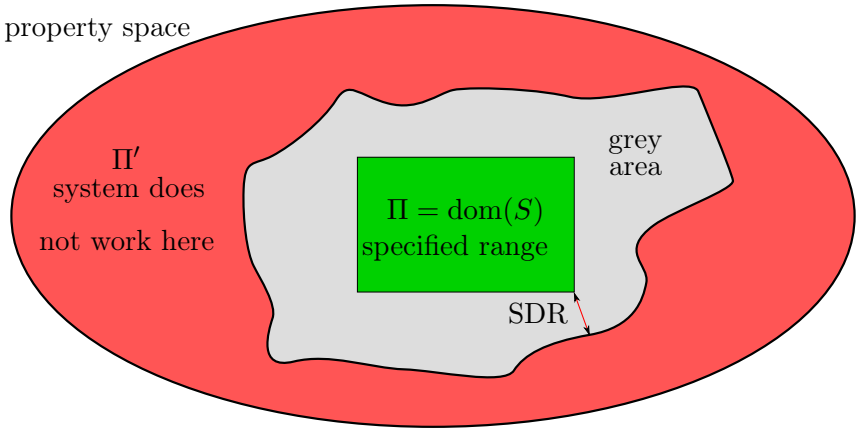


Figure 3.1: For some inputs, a system is working. For others, it is not. The specification workaround tries to capture the grey area in between.

3.2 Existing Notions of Robustness

We demonstrate an existing concept for robustness and its interpretations. These can be applied to a simple physical system and to produce actual numbers. Here we experience effects that challenge the well-definedness. This exercises the understanding of the boundary conditions, and sheds some light into technical details.

3.2.1 Robustness from Distances

With the terminology introduced above, defining robustness for a system F, specified by S, in a mission M ought to be possible. In [Rad+10], it is assumed that F meets S in the following definition of a robustness. Set $\Pi := \mathrm{dom}(S)$ and denote by

Definition 3.5. Π' *the set of input properties p such that a signal s with this property does not yield the output the specification describes for another signal. Writing $\bar{S}$ for the extension of S by the empty set, we have*

$$\Pi' := \left\{ p \,\middle|\, \exists s \colon \forall s' \colon c(s) \notin \bar{S}(s') \right\}. \tag{3.1}$$

Π' is supposed to "simply" express the set of conditions under which a system "works", although it has not been specified to (see Fig. 3.1). In the following, we will make use of this definition for Π', which we shall refer to as **specification workaround**. There are certainly other options. Here, and in [Rad+10], we tried to deduce Π' from the specification, so a robustness does not require any further input. We leave to the reader, whether this interpretation is valid or coincides with what its inventors had in mind.

A **metric space** (D, d) is a set D together with a *distance map* or *metric* $d \colon D \times D \to \mathbb{R}_{\geq 0}$. A **distance** between two nonempty subsets A and B of a metric space can be defined as

$$d'(A, B) := \min_{a \in A, b \in B} d(a, b).$$

Definition 3.6. *The **stationary distances robustness** of F with respect to M is the distance* $\mathrm{SDR}(F, M)$ *between* Π *and* Π'*. The distance is measured using a metric on the property space induced by the distribution of M.*

Visually, the SPR characterizes the size of a grey area between the specification domain and the non-working area (Fig. 3.1).

In general, we do not know how M is supposed to induce a metric on the property space. In special cases, this looks easy to do. Assuming, that all properties embed into a finite dimensional real vector space, a basis induces a metric. A *canonical* embedding into such a space does not exist, so just choose any embedding. Then M induces probability distributions on axes and axes may be normalized using standard deviations. This metric is well defined, if we restrict to multivariate Gaussian distributions on a euclidean space. Otherwise, we cannot be sure.

Eventually, we have set up a robustness as a distance. This robustness is easy to motivate, as it describes how much headroom a design leaves. A larger headroom clearly feels more robust. However this robustness depends on the choice of a set of bad inputs Π' and of a metric on the property space. We have seen that a choice of Π' is possible, but our choice lacks intuition. More severely, there seems to be no way to define the metric on the property space. Obviously some metric exists and can be worked with, but searching for a well defined notion of robustness this is likely a dead end.

3.2.2 Robustness as Probability

To avoid inconveniences with a metric, there is an alternative straightforward robustness defined in [Rad+10] as follows.

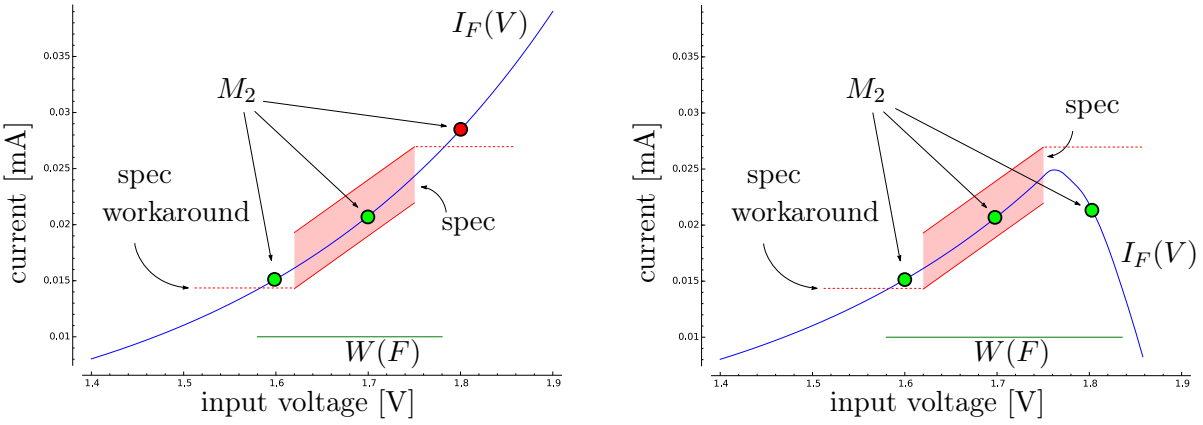
Definition 3.7. *The **stationary probabilistic robustness**,* $\mathrm{SPR}(F, M)$ *is the probability, that F works on $M \setminus \mathrm{dom}(S)$. This is a* conditional *probability.*

This definition still relies on the workaround in Definition 3.5, but it appears to be much more practical. We set up an example.

3.2.3 Example

An LED dissipates about 20 mA at 1.7 V. The specification says, at DC inputs characterized by voltages V between 1.6 V and 1.8 V the output is characterized by a constant current between $V/17\,\Omega - 0.081\,\mathrm{V}$ and $V/17\,\Omega - 0.076\,\mathrm{V}$. A real LED F dissipates $0.1\,\mathrm{mA} \cdot (\exp(V \cdot \pi/\mathrm{V}) - 1)$, see Fig. 3.2a.

If the mission profile M_1 contains only one run at 1.7 V, then clearly F will do the right thing during all missions. There is no mission outside of the specified range, so $\mathrm{SPR}(F, M_1) = 1$. Let M_2 be a mission profile that consists of runs at 1.6 V, 1.7 V and 1.8 V, each with probability 1/3. Two of these runs are out of specification, and SPR is the probability, that the LED will work under these conditions. We figure, that F dissipates about 15 mA at 1.6 V which is OK, as



(a) An LED with a characteristic of an expected shape.

(b) An LED with an unconventional characteristic.

Figure 3.2: An LED specification, instance and mission profile. The extension workaround evaluates out-of-spec runs.

0.015 mA is permitted, even at 1.637 V. At 1.8 V, we get a current of 28.4 mA which is clearly out of range. Eventually, one of two cases does not work as specified and we get $\mathrm{SPR}(F, M_2) = 1/2$.

We find that F imposes an permissive interval W on the supply voltage *in addition* to the lower and upper bound to the supply voltage that we deduce from the specification domain. Thus, on a mission M, the stationary probabilistic robustness $\mathrm{SPR}(F, M)$ is the conditional probability $\Pr(M \in W | M \notin S)$.

If we narrow down the specification interval, making sure that F stays compliant, the SPR will drop. This is intuitively correct, since a tougher specification is expected to be harder to implement robustly. Further, we may also variate F, and observe changes in robustness. Notably, the SPR is negatively correlated to the steepness of the dissipation wrt. supply voltage within the specified range. This does not instantly make sense. Take for example another LED Fig. 3.2b. Then, W becomes larger on the right end. Thus, the third run counts as successful and the robustness increases to 1. Arguably, this does not feel correct. The presumed reason is the specification extension workaround, which does not do what we hoped it would do. Also if we replace the condition on the power dissipation by a condition on the conductivity. We end up with another specification that describes the exact same, but leads to a different SPR, due to the workaround. These corner cases indicate that SPR, as we set it up, is not well defined and not suited as a robustness measure. We take the LED case as evidence that the idea of a specification workaround should be discarded. To overcome this, we either need to write a new specification that covers previously undefined regions, or we do not try to specify whether an out-of-spec operation is to be considered successful. We do not feel entitled to write or extend specifications, thus we continue with the second option.

3.3 A New Notion of Robustness

We have demonstrated how specification extension is a main issue in robustness definition. We need to find a notion of robustness, that is not based on the behaviour of a system outside of the specified conditions. For this we revisit the semantics of a specification. The specification S_F of a system F is a description of its function in terms of input and output signal properties. The specification is bound to a condition on the input. Clearly, if the input is out of range, the specification *does not* impose conditions on the output. A system that is ran out-of-specification *can not* violate its specification. Still, it is possible to harden against unspecified influences. Unspecified inputs may lead to transient faults or even permanent damage such as faster ageing, which can well be observed using the specification.

In this section we define a robustness that does not require a specification extension. For this, a well defined model of the systems behaviour is needed. The specification gives rise to

Definition 3.8. *the **time of failure** T_F. This is the smallest time, at which F violates its specification and F **breaks** at T_F. T_F depends on the mission profile M, we write T_F^M if the mission is not clear from the context. Usually this time is hard to predict and T_F is modelled as a random variable. The **mean time to failure**,* $\mathrm{MTTF}(F)$*, is the expected value of T_F (assuming that it exists) [08].*

A high MTTF does not imply robustness. For this, assume that a bogus system B either breaks after 1 yr with probability 1/2, or otherwise $T_B = 5\,\mathrm{yr}$. We have $\mathrm{MTTF}(B) = 3\,\mathrm{yr}$, and a feeling that B is not particularly robust, no more robust than a component that will work for exactly three years.

We recall the **variance** of a random variable, is the squared expectation of the deviation from the expected value (if it exists).

$$\operatorname{Var} X := \mathrm{E}\left((\mathrm{E}\, X - X)^2\right) = \mathrm{E}\left(X^2\right) - (\mathrm{E}\, X)^2 = \sigma_X^2.$$

This measures how much the expected value can be trusted. With this in mind, we construct another notion for robustness.

Definition 3.9. *The **σ-robustness** of F with respect to M is*

$$R_\sigma(F, M) = \frac{\mathrm{E}(T_F^M)}{\sqrt{\operatorname{Var} T_F^M}} = \frac{\mathrm{MTTF}_M(F)}{\sigma_{T_F^M}}. \tag{3.2}$$

If the mission is clear from the context, we just write $\mathrm{R}_\sigma(F)$.

The σ-robustness measures how *reliable* the MTTF is. The corner cases give some intuition.

- If $\mathrm{R}_\sigma(F)$ is high, the time to failure is precisely determined. That means no mission affects the lifetime of F a lot.
- If $\mathrm{R}_\sigma(F)$ is low, the system may fail anytime. This is close to the opposite of subjective robustness.

At a fixed time, the σ-robustness is related to the idea that lead to stationary probabilistic robustness. Under the assumption that no ageing takes place, the system F does not change, and a failure is a mere consequence of the mission. To provoke a failure, a mission needs to move the system into an unintended state. If this is hardly possible, R_σ is high. This is what SPR was intended to express. There are more properties that appear to be useful.

- The σ-robustness is unitless.
- The σ-robustness describes a property of a system different from MTTF.
- The σ-robustness is not stationary. It includes information on the whole life-cycle. Particularly, ageing effects are covered.
- It has a canonical time evolution, like the subjective robustness. This makes it applicable to used cars (Section 3.3.3).
- Burn in increases σ-robustness significantly (Section 3.3.4).
- Redundancies and doubling resources increases R_σ (Sections 3.3.1 and 3.3.2).
- The σ-robustness is computable in various cases.

3.3.1 Redundancies

Redundancies improve R_σ. Let G_1 be a system that is built from two (independent) instances of F and some (infallible) redundancy mechanism. One instance of F runs first and the second takes over in case of failure. Due to the independence of the instances, the model for the failure time of G_1 equals $2 \cdot T_F$. We have $\mathrm{R}_\sigma(G_1) = \sqrt{2}\,\mathrm{R}_\sigma(F)$, if T_F is Poisson-distributed or (close to) normal-distributed. If T_F is log-normal, which is typical for failure times, there exist approximations [CRS] for the resulting distributions, so the σ-Robustness remains accessible.

3.3.2 Doubling Resources

Assume the failure time of a single system X is normally distributed with a variance of 10 % the MTTF(X). A system G_2 that consists of two independent instances of X, which are both working in parallel and both required for functionality, has a time to failure of 94.36 % that of X. Three systems in one G_3 yield $\mathrm{MTTF}(G_3) \approx 91.54\,\% \cdot \mathrm{MTTF}(X)$. However, the standard deviations decrease to $\sigma_{G_2} \approx 8.26$ and $\sigma_{G_3} \approx 7.48$ (see Fig. 3.3) and the sigma-robustness increases from 10 to 11.4 to 12.24. So, the robustness increases when using more (independent) resources – with a penalty on MTTF.

In comparison to the redundancy calculation in the previous Section 3.3.1, doubling resources scores badly. We may assert that sigma-robustness correctly discriminates redundancy measures from putting together random parts.

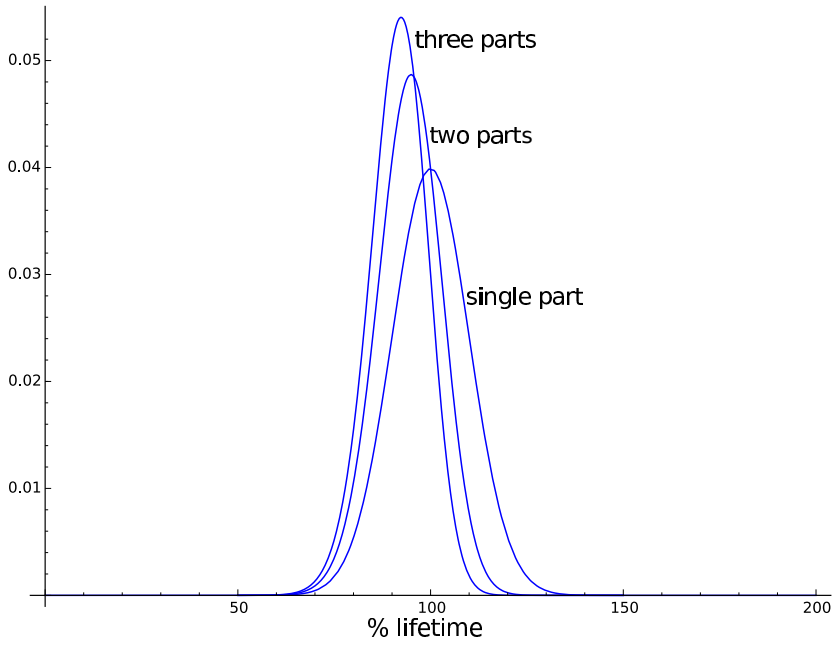


Figure 3.3: Failure time distributions of parallel systems.

3.3.3 Ageing

Robustness is meaningful only for a working instance, hence the expected value for the failure time of a system is a conditional probability and the condition is, that it is still working. The time to failure for a system with failure time model T that is still working at t is the conditional expectation $\mathrm{MTTF}_t(T) := \mathrm{E}(T - t | T > t)$. As example choose a system with failure time T uniformly distributed between 0 and $t_{\max}$. For times before $t_{\max}$, we get $\mathrm{MTTF}_t = (t_{\max} - t)/2$. Similarly, the conditional standard deviation at that time is $\sigma_t = \sqrt{1/12 \cdot (t_{\max} - t)^2}$. Thus, the σ-robustness at time $t < 1$ is constantly $4\sqrt{3} \approx 6.9$. See Section 4.2.5 for a more detailed example.

3.3.4 Burn in

Consider a system S with an simplified expected-to-fail distribution. S is expected to fail within its first few hours with probability 10 %. If it does not, its failure time shall be uniformly distributed between 9 yr and 11 yr. The expected failure time without burn in is $\mathrm{MTTF}(S) =$ 9 yr. The burn in (assuming that it takes long enough to catch all faulty instances), will increase MTTF to about 10 yr.

The burn in has a much more significant impact on the σ-robustness. Before the burn in, we have $\sigma(S) = \sqrt{0.3 \cdot 16 + 8.1}$, afterwards it amounts to about $\sqrt{2/3}$. After all, the burn-in increases the σ-robustness from about 2.5 to about 12.2. Intuitively, robustness increases a lot during burn in, so this constitutes a match.

3.4 Conclusion

We have set up the basic preliminaries to approach a formal robustness property. Based on this, we have discussed existing ideas that pointed to a definition of notions for robustness. This lead into a problem related to specification, that we tried to work around, retaining the underlying intuition. However the attempt was unsuccessful, and counter example has emerged. Restarting, we suggest another notion of robustness that hopefully is correlated to intuitive properties. The *σ-robustness* reuses the terms related to functionality from the previous attempt. It is easy to define, and can be applied to example systems to check its expressive power. Our examples show that σ-robustness plays well with mission profiles, redundancies, ageing and burn-in procedures. Also, this new notion can be easily seen to fulfil the desired properties listed in Section 3.1.

Computing the robustness of a complex design is a different issue. As for the computation of life times, ageing effects play an important role in σ-robustness. We strongly believe that a different – maybe a better – notion of robustness is similarly entangled in life time and failure time considerations. Hence, a better understanding of ageing effects will be necessary to leverage robustness exploration. The rest of this work is dedicated to modelling and simulating these effects and their respective impact.

CHAPTER

FOUR

AGING EFFECT MODELLING

4.1 From Stress to Parameters

Ageing effects cause phenomena that can be roughly divided into two categories. One are the effects that slowly change the parameters of a device or material, the others are effects that are hard to notice until something remarkable happens. To get an idea, take as example a bicycle with brakes and tyre. The brakes will wear out slowly, while the grip decreases noticeably. In contrast, the wearout of the tyre will (for the sake of the example) progress unnoticed until it cannot withstand a shard. The cyclist will probably not have noticed before the tyre is perforated. Modelling these effects generally means modelling the phenomena, hence we get two effect model types, the (time-)*continuous* models, that control a property that slowly, and significantly changes according to how a device is being used, and the *event* models that predict incidents, often based on empirical data. In either case, we have a stress influence that affects a parametrical property of a system.

This discrimination is applicable to the list of ageing effects in electronics, see Section 2.2. Here, we have HCI and BTI as specimen of the ageing effects that may be considered continuous in time. Dielectric breakdown and the consequences of electromigration may be modelled as events, and often are [HL99; RT08]. From an arithmetic or algorithmic point of view, there are a lot of possible ageing effect models, most of which will not be of any use. A formal base to ageing effect models helps clarify the properties of the existing physical effects and also characterizes the models. This enables us to propose classes of models for various parametric ageing effects. We propose an algorithm that is suited for data based hierarchical ageing modelling. This algorithm learns an element from a selected class of models that correlates best with the data.

4.1.1 Terminology

The first ingredient to an ageing effect model is a model for the very cause of ageing. We separate out some related notions, and fix some conventions. Let $\mathbb{T}$ denote the set of times, which is the set of nonnegative reals multiplied by 1 s.

Definition 4.1. *Let n be a positive integer. A* ***stress level function***

$$L\colon \mathbb{T} \to \mathbb{R}^n \tag{4.1}$$

is a (possibly multivariate) real valued time dependent variable. A ***stress wave*** *w of duration T is a stress level function restricted to times before $\mathrm{T}_w := T$, a partial stress level function*

$$w\colon [0, \mathrm{T}_w] \to \mathbb{R}^n. \tag{4.2}$$

We write $\mathrm{dom}(w)$ *for the set of these times, the domain of w.*

As a notational convention we write w_t for the restriction of w to times before a time $t \in \mathrm{dom}(w)$. Also for stress waves $w\colon [0, \mathrm{T}_w] \to \mathbb{R}^n$ and $w'\colon [0, \mathrm{T}_{w'}] \to \mathbb{R}^n$ we write $w \odot w'$ for the stress wave

$$t \mapsto \begin{cases} w'(t) & \text{if } t < \mathrm{T}_{w'} \text{ and} \\ w(t - \mathrm{T}_{w'})) & \text{otherwise.} \end{cases} \tag{4.3}$$

Definition 4.2. *The* ***(energetic) state space*** *$\mathbb{S}$ of a system, is the set of all configurations for the energetic potentials. The* ***ageing (state) space*** *$\mathbb{A}$ of such a system is the set of all constitutions of the materials involved. An* ***ageing state*** *is an element of $\mathbb{A}$.*

For example, the energy state space of a system model is the set of variable assignments that are physically meaningful. Usually, we are considering discrete models of electrical circuits and it is safe to think of $\mathbb{S}$ as a finite dimensional space. In this case an energetic state is determined by the node voltages and branch currents. Similarly, for practical considerations, $\mathbb{A}$ is of finite dimension. The **system (state) space** is the Cartesian product $\mathbb{X} := \mathbb{A} \times \mathbb{S}$. If the context is clear, we just write **state** for an element of $\mathbb{X}$, $\mathbb{A}$ or $\mathbb{S}$, respectively. A **transient** is a (typically continuous) function $x\colon \mathbb{T} \to \mathbb{X}$.

Let $n \in \mathbb{N}$ be a nonnegative integer. A *stress level* L is a function $\mathbb{S} \to \mathbb{R}^n$. Then for admissible times t, $t \mapsto L(x(t))$, is the stress wave associated to a transient x. As a device operates, and material wears out, it changes some of its behavioural properties. These depend on the constitution of a system.

Real numbers are ordered. By choosing coordinate axes, we partially order ageing states accordingly, greater states will correspond to more damaged or older devices. An ageing state valued *random variable* $A\colon \Omega \to \mathbb{A} \subset \mathbb{R}^m$ on a probability space Ω is an **indeterministic ageing state**. If there is no risk of confusion, we refer to it as just an ageing state. An indeterministic state $A \in \mathbb{A}$ can be considered greater than a state B, if (for a choice of axes) the *cumulative distributions* of A are below the cumulative distributions of B, i. e.

$$\forall z\colon \mathrm{cdf}_A(z) \leq \mathrm{cdf}_B(z).$$

This extends the order of deterministic ageing states introduced above. A variable ageing state that is subject to change with respect to time or other influences may then be defined as follows.

Definition 4.3. *An* ageing process Q *is a map taking stress waves to random ageing states,*

$$Q\colon \{ \textit{ stress waves } \} \to \{\Omega \to \mathbb{A}\},$$
$$w \mapsto Q(w).$$

Bound to the following ***locality condition****. Let* L *and* L' *be stress waves. If* $Q(L) = Q(L')$*, then for all* K*,* $Q(K \odot L') = Q(K \odot L)$*. If the image of* Q *consists of* degenerate *distributions, the ageing process is called* ***deterministic****.*

The condition on concatenating stress expresses the property that characterizes physical processes we eventually intend to model. Intuitively it means, whatever we have done to a device to get it into a certain state, its future behaviour will be determined by that state (which is a local property), and not by what we did. For simulation complexity (Chapter 5) this will be of importance. Keeping track of a device state can be simple, while keeping track of the whole history is often unfeasible. With this, an ageing process can be seen as a map that turns stress level function L into a state function $t \mapsto Q(L_t)$. Usually, stress waves with zero duration map to a state with zero expectation, the **initial** or **fresh** state.

With a metric on stress waves such as

$$\{ \text{ stress waves } \}^2 \to \mathbb{R}_{\geq 0}$$
$$(L, N) \mapsto |\mathrm{T}_L - \mathrm{T}_N|/\mathrm{s} + \max_{t \in \operatorname{dom} L \cap \operatorname{dom} N} |L(t) - N(t)|, \tag{4.4}$$

where (L, N) denotes a tuple of stress waves, and T_N, T_L their respective durations, we may distinguish between *continuous* and *discontinuous* ageing processes.

Another objective is

Definition 4.4. *an* ageing effect parameter p *is a real parameter that depends on an ageing state,*

$$p\colon \mathbb{A} \to \mathbb{R}.$$

While p may well take random states to ageing effect parameter valued random variables, it is not intended to introduce randomness by itself. Put together, this yields an ageing effect model.

Definition 4.5. *An* ***ageing effect model*** A *is the composition* $p \circ Q \circ L$ *of a parameter* p*, a process* Q *and a stress level* L*. It takes a transient* x *on* $\mathbb{T}$ *to the function* p *on* $\mathbb{T}$*, or* x_t *to* $p(t)$*, the value of* p *under stress* l*, caused by* x *at time* t*, see Fig. 4.1. The model is called* ***nominal****, if* Q *is deterministic.*

We intend to reduce confusion and tell apart ageing processes and ageing effect models. An ageing process is more general, and typically involves a physical constitution that is responsible for ageing effects. The ageing effect model is explicit. It is meant to carry details on what the stress level is, how this level affects an ageing state and how an ageing state controls a

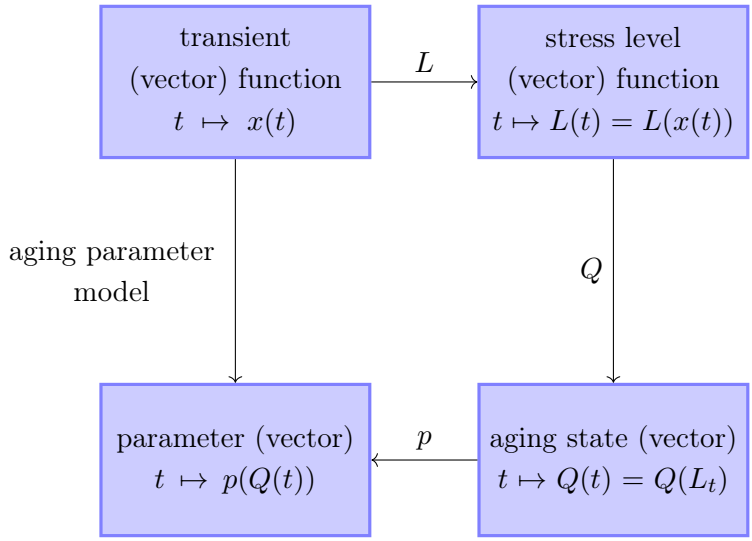


Figure 4.1: An analogue ageing parameter model consists of a parameter function p, an ageing process Q and a stress level L. These are implicitly time dependent.

parameter. From the outside, an ageing effect model constitutes a univariate ageing process. Time-invariance is a common notion in behavioural modelling. For ageing processes, it may be expressed as follows.

Definition 4.6. *An ageing process Q is called **time-invariant**, if for any stress wave L and $t \in \mathbb{T}$ we have $Q(L) = Q(L \odot 0_t)$, where 0 denotes the constant zero stress level function.*

4.1.2 Monotonicity and Memory

Recall that a function f between partially ordered sets is **monotonic** if $f(x) \leq f(y)$ whenever $x < y$ or whenever $y < x$. Let L, L' be Stress Waves on $[0, T]$. We write $L < L'$, if $L(t) < L'(t)$ for all $t \in [0, T]$. This expands to the following definition.

Definition 4.7. *An ageing process Q is **monotonically increasing** and **monotonically decreasing**, if for all stress level functions L, L' and times $T \in \mathbb{T}$ we have*

$$\big(\forall t \leq T\colon L(t) \leq L'(t)\big) \Rightarrow Q(L_T) \leq Q(L'_T), \tag{4.5}$$

or

$$\big(\forall t \leq T\colon L(t) \leq L'(t)\big) \Rightarrow Q(L'_T) \leq Q(L_T), \tag{4.6}$$

respectively.

If an ageing process models the damage to a device, the monotonically increasing condition expresses that a higher stress wave causes more damage.

For a fixed stress level function F, the map $t \to Q(F_t)$ is a stochastic process in the usual sense. Similarly, in probability and statistics the *Markov property* helps separate future and past. We transfer this notion to ageing processes.

Definition 4.8. *The (smooth, nominal) process Q is called* ***memoryless****, if for all stress level functions L and all times t, the time derivative $\mathrm{d}Q(L_t)/\mathrm{d}t$ is a function of $Q(t)$ and $L(t)$.*

Memoryless processes are time-invariant, the converse is not true, see Section 4.3.2 for the construction of a time-invariant process with memory. Note that this notion is unrelated to stateless combinatoric circuits. A memoryless process may well have a state.

Lemma 4.9. *Let p be a smooth ageing parameter, f a real valued function on stress levels. An ageing process Q built from f and p as in*

$$Q\colon L \mapsto p \circ \int_0^{\mathrm{T}_L} f\left(L(t')\right) \mathrm{d}t'$$

is memoryless, if p is monotonic.

Proof. If p is strictly monotonic, it admits a left inverse p^{-1}, such that $p^{-1} \circ p$ is the identity on $\mathbb{R}$. Denoting here by p' the differential of p, we may write

$$\begin{aligned}\frac{\mathrm{d}Q(L_t)}{\mathrm{d}t} &= p'\left(p^{-1} \circ p \circ \int_0^t f \mathrm{d}t'\right) \cdot f(L(t)) \\ &= p'(p^{-1}(Q(t) \cdot f(L(t)). \end{aligned} \tag{4.7}$$

Thus the time derivative is a function of Q and L. If p is just monotonic, a partial inverse will do, because in $\mathrm{d}p(p^{-1})$ the choice of the preimage does not matter. □

On the converse, note that (in case of a nontrivial function f) monotonicity of p is required. Otherwise p is not injective and we find $t \neq t'$ and a constant level C such that $Q(C_t) = Q\left(C_{t'}\right)$, which implies a memory. In particular, memoryless nominal processes are monotonic.

Definition 4.10. *An ageing process Q is called* ***separable****, if there exist functions f and g such that*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q(L_t) = f(Q(t)) \cdot g(L(t))$$

for all stress level functions L.

Note that separable implies memoryless. Separable effects have an integral representation that will turn out to be useful for simulation purposes.

Theorem 4.11. *For practical considerations, a monotonic and separable univariate ageing process Q is a process*

$$Q\colon L \mapsto p_Q \circ \int_{\mathrm{dom}\, L} f_Q(L(t))\mathrm{d}t, \tag{4.8}$$

with a function f_Q and a function p_Q.

Proof. Choose a constant nonzero stress level function U and define $p(t) := Q(U_t)$. We may assume that p is nonzero and $U(t) = 1$, otherwise we rescale if Q is nontrivial. The function p is monotonic, because Q is, so there is an inverse p^{-1}. For other constant stress levels < 1, let

C be the corresponding stress level function and define $f(c) := p^{-1}(Q(U_{1\,\mathrm{s}}))$. We need to check that $Q(L_t) = p \circ \int_0^t f(L(t))\mathrm{d}t$. At zero time, $Q(L_0)$ and $p(0)$ coincide. We show that

$$\frac{\mathrm{d}Q(L_t)}{\mathrm{d}t} = p'\left(\int_0^t f(L(t))\mathrm{d}t\right). \tag{4.9}$$

Due to memorylessness we have

$$\begin{aligned}\frac{\mathrm{d}Q(L_t)}{\mathrm{d}t} &= F(Q(L_t), C(t)) \\ &= F(Q(C_1), L(t)) \\ &= F(f(c), L(t)).\end{aligned} \tag{4.10}$$

□

Most memoryless processes can be written as an integral of a (harm) function of a stress level. This will turn out to be useful when modelling effects in general, c. f. Section 5.6. Theorem 4.11 gives a canonical representation for some memoryless ageing effects. To model physical effects with memory, we will have to introduce a different approach.

4.1.3 Frequency Dependence

Measurements in [Ala03] suggest that below 10 MHz, the BTI effect can be characterised as "frequency independent". If it really is, it would be advantageous to decide which models are frequency independent as well. In contrast, recent research [Gra+12] classifies BTI as frequency dependent at higher frequencies. Modelling this case is expectedly more complex, and frequency dependent models need to be examined. Eventually, frequency independence is a crucial property that we need to examine with more detail. Some terminology and notation is required.

We manipulate the frequency of a stress wave. The binary *modulo operator* % maps a pair of nonnegative reals (a, b) to $a \,\%\, b := a - \lfloor a/b \rfloor \cdot b$, where $\lfloor x \rfloor$ denotes the greatest integer not greater than x.

Definition 4.12. *Let L be a stress wave and $[0, T]$ be its domain. Choose a real number $\alpha \geq 0$ and set*

$$L^\alpha \colon t \mapsto L((\alpha t) \,\%\, T),$$

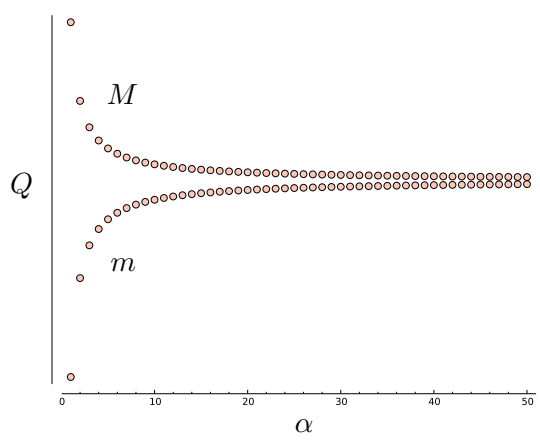
the stress level function L ***warped by*** *α.*

We declare $L_t^\alpha := (L^\alpha)_t$, the restriction of the manipulated stress wave. Note that L^α is defined on all $\mathbb{T}$, and so is $(L_t)^\alpha$.

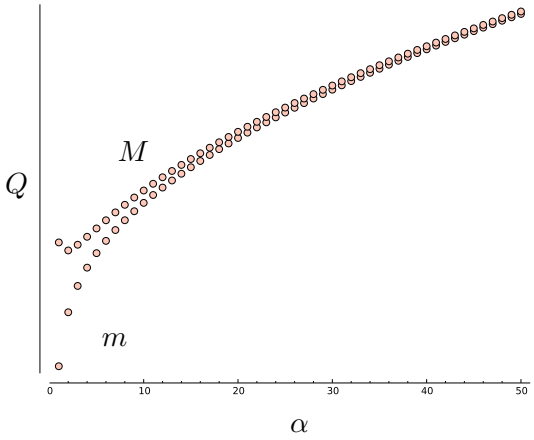
Definition 4.13. *Let Q be a process and L a smooth stress wave. Changing frequencies, we get sequences*

$$M_{L,n} := \max_{\alpha \in [n, n+1]} Q(L^\alpha) \text{ and } m_{L,n} := \min_{\alpha \in [n, n+1]} Q(L^\alpha)$$

for $n \in \mathbb{N}$. If for any L, the sequence M is falling and m is rising, we say Q is ***frequency independent****.*



(a) Frequency independence. An SRC process (Section 4.3.4).



(b) Frequency dependent process. Data inspired by PTM BTI, see Section 6.1.4.

Figure 4.2: Frequency dependency. Sequences M and m for example processes and stress level.

Essentially, the frequency independence condition states that the range of values $Q(L)$ takes does not grow or move with increasing frequency of L (Fig. 4.2a).

Definition 4.14. *If for a stress wave L this range shrinks to a point p, or formally, the sequences M_L and m_L both converge to ℓ, then $\ell = \ell_Q(L)$ is the **frequency limit** of Q under L.*

This condition guarantees that parameter fluctuations are related to stress level changes, and *not* caused by the higher frequency, as in Fig. 4.2b. A simple, nontrivial example for a frequency independent process is the **integrator**

$$Q_I \colon L \mapsto \int_{\operatorname{dom} L} L(t)\mathrm{d}t. \tag{4.11}$$

Take a stress wave L on $[0, T]$. Let the maximum of $Q_I(L_t)$ be at t_0. With $\alpha_n := (t_0 + nT)/T$, we have $M_{n,L} = Q_I(L^{\alpha_n})$. We compute

$$M_{n,L} = \int_0^T L(((nT + t_0)/T \cdot t) \,\%\, T)\mathrm{d}t = \frac{n \cdot \int_0^T L(t)\mathrm{d}t + \int_0^{t_0} L(t)\mathrm{d}t}{n + (t_0/T)},$$

and these form a falling sequence, it converges to $Q_I(L)$. Similarly, $(m_{L,n})_n$ is a rising sequence with the same limit. The integrator process can be found in HCI models, see Section 4.2.2.

Clearly, the example still works when adding functions p and f, and we get

Lemma 4.15. *An ageing parameter*

$$L \mapsto p \circ \int_{\operatorname{dom} L} f(L(t))\mathrm{d}t.$$

is frequency independent.

While it is not possible to observe a frequency limit directly using measurements (except for constant stress conditions) it is still possible to find meaningful limits in some cases. A frequency independent model is suitable to model effects that are *believed* to be frequency independent, or frequency independent within a targeted frequency range. Frequency limit is an effective tool to identify (un)suitable models (c. f. Section 4.2.3) and also to fit model parameters, see Section 4.4.

4.2 Ageing Effects as Processes

The common ageing effects and models can be seen as parameters controlled by processes in the sense of Definition 4.4. Then the notions introduced above apply. In this section, we describe and characterize known ageing effects and investigate some of their properties.

4.2.1 Electromigration (EM)

There is a high level but simple mean time to failure law for Electromigration. A conductor C conducting a constant current density J (in $\mathrm{A\,m^{-2}}$) starting from time 0 will expectedly fail at $\mathrm{MTTF}(C)$. According to *Black's law* [Bla69][1], there are constants A_{Bl} and φ such that at a temperature ϑ, we have

$$1/\mathrm{MTTF}(C) = A_{\mathrm{Bl}} J^2 \cdot \exp\left(-\frac{\varphi}{k_{\mathrm{B}}\vartheta}\right). \tag{4.12}$$

Here, k_{B} denotes the *Boltzmann constant*.

The underlying mechanism can be expressed in terms of a current controlled ageing process. Denote by $j(t)$ the (always nonnegative) current through a conductor. Let c be some nonnegative parameter. We define a stress level

$$L\colon j \mapsto c \cdot (j/\mathrm{A})^n. \tag{4.13}$$

The ageing state of the migration process at a time $t \in \mathbb{T}$ is a two-valued random variable

$$X_t\colon \Omega = \{0,1\} \subset \mathbb{R}.$$

We represent it as a number $a_t \in [0,1]$, the probability for X_t to be in state 1, the broken state. We choose $\sigma > 0$ and set

$$a_t = a(L_t) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{L_t} \frac{L\,(t')}{\int_0^{t'} L(t'')\mathrm{d}t''} \cdot \exp\left(-\frac{\left(\ln\left(\int_0^{t'} L(t'')\mathrm{d}t''/\mathrm{s}\right) + \sigma^2/2\right)^2}{2\sigma^2}\right) \cdot \mathrm{d}t', \tag{4.14}$$

Clearly, a_t increases with t if and only if $L > 0$. Fix a current $j_0 > 0$ and set $L(t) = L(j_0)$ for all $t \in \mathbb{T}$. The function a_t then simplifies to

$$a_t = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \frac{1}{t'} \cdot \exp\left(-\frac{\left(\ln(t'/\mathrm{s}) + \ln(L) + \sigma^2/2\right)^2}{2\sigma^2}\right) \mathrm{d}t'. \tag{4.15}$$

[1] In [Bla69], MTTF (denoted MTF) is defined as the *median time to failure*. We consider this a typographic error, or maybe a concurrent naming convention. The figures in the same work support that the law is actually about the usual mean time to failure.

This is the cumulative distribution of a time valued random variable T_C that models the event of failure. The distribution of T_C is *log-normal* with *location* $\mu = -\ln L - \sigma^2/2.$ and *scale* σ, thus $a_t = \mathrm{P}(T_C < t)$. We recover the mean time to failure of this process as

$$\mathrm{MTTF}(C) = \mathrm{E}(T_C) = \int_0^\infty t \cdot \frac{\partial a_t}{\partial t} \mathrm{d}t = \exp\left(\mu + \sigma^2/2\right) = 1/L = 1/c \cdot (j_0/\mathrm{A})^{-n}.$$

This coincides with Black's law (Eq. 4.12) for $n = 2$, $J = j_0 \cdot \mathrm{m}^{-2}$ and $c = A_{\mathrm{Bl}} \cdot \exp(-\varphi/k_{\mathrm{B}}\vartheta)$. Similarly the process can be made temperature dependent and suited to model failures for conductors under varying temperature conditions. In particular, we were able to express electromigration as an ageing effect as defined in Definition 4.5.

4.2.2 Hot-Carrier Injection (HCI)

Hot-carrier injection models usually express an operation dependent nominal parameter shift. Filled with concrete data [WWS86], the threshold voltage shift in forward operation can be expressed as

$$\Delta V_{\mathrm{th}}\left((I_{\mathrm{sb}}, I_{\mathrm{ds}})_t\right) \sim \left(\int_0^t \frac{(I_{\mathrm{sb}}/\mathrm{A})^3}{(I_{\mathrm{ds}}/\mathrm{A})^2} \mathrm{d}t\right)^{0.3} \mathrm{V} \tag{4.16}$$

If the currents I_{ds} and I_{sb} are positive.

Eq. 4.16 decomposes into a function p of the **integrator process**

$$Q_I \colon L \mapsto \int_{\mathrm{dom}(L)} L(t) \mathrm{d}t \tag{4.17}$$

that integrates a stress level $L = L(I_{\mathrm{sb}}, I_{\mathrm{ds}})$. Thus the HCI model qualifies as ageing effect model in the sense of Definition 4.5. Evidently, this model is defined for operation under arbitrary conditions. The physical effect is close to continuous and deterministic. Evidence suggests that HCI is memoryless by all practical means. In particular, the HCI model is frequency independent by construction, see Lemma 4.15. Non-determinism may be added by randomizing the initial conditions in terms of the transistor parameters, these will then propagate deterministically to the ageing state and the parameter drift.

A more expressive ageing process model that accounts for single events involving transient parameter discontinuities and randomness is currently not known or required. This is based on the fact, that single HCI events have a minor impact. While BTI induced noise already plays an important role in SRAM cells [Sim+11; Goe+11], there is no indication that HCI causes similar issues.

In most digital circuits the positivity assumption on the currents is valid. Particularly in analogue circuits, the model 4.16 may become invalid [Cho12]. Section 6.1.1 for an extension to analogue use cases.

4.2.3 Bias Temperature Instability (BTI)

There are three concurrent approaches to BTI modelling, two of which can be seen as controlled ageing processes. We outline these approaches and point out some properties.

The Reaction-Diffusion (RD) Model describes a state of an invertible chemical reaction process at a transistors channel and the distribution of related particles withing the gate oxide. The strength of the electric field controls the chemical reaction and can be seen as a stress level. The reaction is bidirectional and, simplified turns the amount of bound protons P into an amount of free hydrogen H with forward rate k_f and reverse rate k_r respectively. The reaction model factors in the respective phase concentrations N at the interface,

$$\frac{\mathrm{d}N_H}{\mathrm{d}t} = k_f \cdot N_P - k_r \cdot N_H \cdot (N_P - N_{P,0}). \tag{4.18}$$

The forward rate k_f is boosted by the electrical field. Hydrogen released from the reaction diffuses into the gate oxide. With a positive m and a diffusion coefficient D, the *porous medium equation*

$$\frac{\partial H_x}{\partial t}(t) = D\nabla^2 H_x^m \tag{4.19}$$

models the movement of the particles trough the oxide. Together with an expression for the parameter shift as a function of the reaction state, in the sense of Definition 4.4, this gives an ageing effect parameter model Definition 4.5. For example, the threshold voltage shift ΔV_th is related to N_H. However to obtain more accuracy, a more complicated reaction process with more phases is required [Sch06].

This model imitates a physical process, that accounts for a parameter shift that is continuous in time and monotonic. Unlike the HCI model, it supports relaxation and is not easy to fit to measurements. The evaluation, particularly in analogue context, is really hard and can be considered infeasible.

The predictive technology model (PTM) for BTI [Bha+06] models a threshold voltage shift ΔV_th. The stress level is the vector $L = (V_\mathrm{gs}, V_\mathrm{ds})/\mathrm{V}$. The threshold is updated according to a rule that we interpret as follows.

$$|\Delta V_\mathrm{th}| \leftarrow \begin{cases} \sqrt{K(L)^2 \Delta t^2 + {\Delta V_\mathrm{th}}^2} + \delta & \text{after stress } L \text{ lasting for } \Delta t, \\ (|\Delta V_\mathrm{th}| - \delta) \cdot \left(1 - \sqrt{\eta \Delta t / t}\right), & \text{if } L = 0 \text{ between } t \text{ and } t + \Delta t \end{cases} \tag{4.20}$$

Here K is a voltage slew rate function, δ is an offset Voltage and η is a positive constant. This model does not designate a state variable (note the use of t in the relaxation branch of Eq. 4.20), and the assumption that a process model explains Eq. 4.20 leads to difficulties including the following.

- If the device was in fresh state s_0 at time 0, and s_0 corresponds to $\Delta V_\mathrm{th} = 0$, then after no stress between 0 and Δt, the substitution rule Eq. 4.20 yields $\Delta V_\mathrm{th}(\Delta t) \neq 0$. Experimental evidence suggests that BTI is time-invariant, hence it should be $\Delta V_\mathrm{th}(\Delta t) = 0$ in case of no stress.
- The recovery effect on ΔV_th depends on t. Modelling ΔV_th based on a state, it is not possible to reach a state at a different time. This again violates time-invariance, but it is even harder to fix.

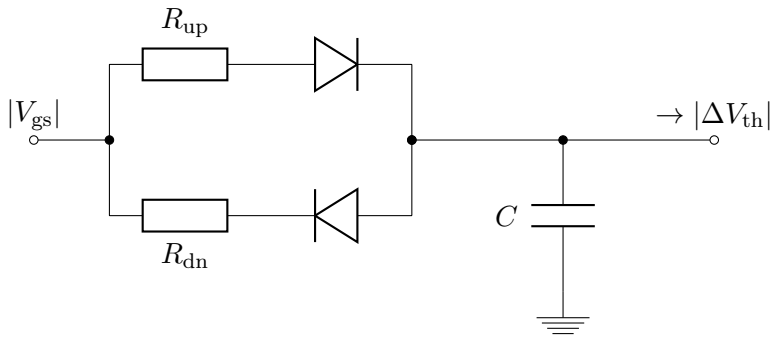


Figure 4.3: Charge trap model. Analogue BTI model? Not yet.

- Constant stress L applied for Δt has a different impact from L applied twice for $\Delta t/2$ (even if $\delta = 0$!), according to Eq. 4.20. An ageing process can not take different values on L_t and $L_{t/2} \odot L_{t/2}$ because they are equal.
- It is hard to interpret Eq. 4.20 considering variable stress levels. An ageing process is meant to deal with any time dependent stress level. This is a general problem related to digital models, c. f. Section 6.4.

The PTM BTI model gives a rule of thumb for the impact of a specific set of rectangular stress patterns on the threshold voltage. Our assertions are not in conflict with the validity in these cases. Yet, the model is incomplete and fundamentally unsuited for (analogue) circuit ageing simulation. In Section 6.1.4 we will evaluate this model at carefully chosen sample points. We will then set up a process based model that orients itself on these landmarks.

Often, the BTI effects is modelled with digital applications in mind. Therefore, most formulas found in the literature make sense for rectangular stress patterns only. As another example, [Bha+06; Wan+07] propose formulas such as

$$\Delta V_{\text{th}}(t) = \left(\frac{\sqrt{K_v \alpha T_{\text{clk}}}}{1 - \beta_t^{1/8}} \right)^{1/2} \tag{4.21}$$
$$\beta_t = \beta_t(T_{\text{clk}}, t, \alpha)$$

involving a duty cycle α and a clock period T_{clk}. These indicate stress and recovery cycles similar to Eq. 4.20. No matter how accurate these models are, they only work for periodic rectangular stress.

The charge trap perspective yields a different class of BTI models that emerged from digital circuit modelling [Gra+11]. It involves an ageing process that can be expressed as a netlist Fig. 4.3. This netlist represents an ageing state as the charge on a capacitor.

This is a fundamentally different and practical perspective on modelling the BTI mechanism [Sch+10] for AC conditions. Looking closer, AC means constant stress with periodic interrup-

Table 4.1: Properties of ageing effects and ageing effect models.

	effect/ parameter model	continuous, deterministic	monotonic	memoryless	time-invariant	stress pattern	frequency dependent
HCI	(real) ΔV_{th}	✓[2]		probably	✓	any	✗
	integral model	✓	✓	✓	✓	any	✗
BTI	(real) ΔV_{th}, $\Delta\mu$	?[3]	✓	✓	(✓)	any	maybe[4]
	RD-model	✓	✓	✗	✓	any	?[5]
	PTM model	✓	✗[6]	–	–	rectangle	?[5]
	charge traps	✓	✓	✗	✗	rectangle[7]	✗
EM	(real) connectivity	✗	✓	?	✓	any	?[5]
	statistic model	✗	✓	✓	–	fixed	✗
	failure model	✗	✓	✓	–	fixed	✗
elementary	power-law	✓	✓	✓	✗[8]	fixed	–
	integrator	✓	✓	✓	✓	any	✗[9]
	RCD	?[10]	✓	✓	✓	any	✗[9]
	weighted RCD sum	?[10]	✓[11]	✗	✓	any	✗[9]

tions. This is a bit disturbing, since the very netlist from Fig. 4.3 is often referred to in the context of reliability studies in the analogue domain [GD08; Yil+13].

Clearly, the diode circuit mounted into a BTI model can only work (as intended) for one single fixed AC stress amplitude. Assuming the contrary, we may for example apply twice the voltage for V_{gs}. Doing so retains the time constants which only depend on R and C while at the same time doubles the number of defect sites. However it is widely believed, that the time constants of the BTI effect highly depend on the stress voltage, while the number of defect sites is a geometric invariant.

Leaving out the diodes [MG13] does not seem to amend this issue. We shall construct and examine an unambiguous defect state model in Section 4.3.4 (Fig. 4.7). See [SH11] and Section 4.3.3 for how to make charge traps ready for stress levels of the analogue kind.

4.2.4 Overview

Table 4.1 displays some properties of the introduced ageing effects and corresponding effect models We also include the *elementary* ageing effect models such as the integrator, the power-law and RCD-based ageing processes we will introduce later.

[2] for practical considerations
[3] application dependent.
[4] for high frequencies, [Gra+12]
[5] unknown
[6] see Fig. 6.8
[7] using fixed time constants, c. f. Fig. 4.7.
[8] see stress pattern
[9] can be made frequency dependent using a filter on the stress level.
[10] the expected value is deterministic

4.2.5 Lifetime Hand Calculations

4.2.5.1 Electromigration

We compute the mean time to failure of a conductor that suffers from EM. It is supposed to carry the supply current between a battery and a system. We assume that the system is not influenced by ageing effects otherwise, and it can be seen as an Ohmic resistor with resistance R. The battery voltage model is $V(t) = V_0 \exp(-t/\tau)$ for $t \in \mathbb{T}$, and the system be functional for battery voltages above $k \cdot V_0$. Black's law works with constant currents. With an arbitrary average voltage $\overline{V}$, we may state that

$$\mathrm{MTTF}\left(\overline{V}\right) = A \cdot \left(\overline{V}/R\right)^n \cdot \exp(\varphi/k_\mathrm{B}\vartheta),$$

using Eq. 4.12 and the associated constants. This time is in between $\mathrm{MTTF}(k \cdot V_0)$ and $\mathrm{MTTF}(V_0)$, if we neglect that the system will expectedly stop working after $\ln(k^{-\tau})$. Even if k is above .7, we still get an uncertainty of about 2. The process model (Eq. 4.14) permits variable currents. Under the assumptions above, we get the probability for a functional conductor at time t, which is

$$\mathrm{P}(X < t) = a_t(V)$$

from Eq. 4.14. With this, we can express the precise mean time to failure as

$$\mathrm{E}(T_X) = \int_{\mathbb{T}} t \cdot \mathrm{d}a_t(V(t)). \tag{4.22}$$

Moreover, we can compute the σ-robustness (Definition 3.9) from the lifetime model. The fixed-voltage model results in $\mathrm{R}_\sigma = \mathrm{MTTF}/\sigma$ for the conductor alone, where sigma coincides with standard deviation from Black's law. The system MTTF is lower, as the battery might run out first. The standard deviation of the system failure time is lower as well. The determination of the failure time increases when taking into account the battery run-out, similar to the observation in Section 3.3.2.

4.2.5.2 Failure Statistics

Sometimes, redundant parts are used to increase the expected life time. As an example consider a conductor in parallel to the weak spot in the example above. Now, two failures are required to bring down the system. Let X_1 and X_2 be the random variables that model the respective failure times. Let Y be a model for the failure time of the whole system. For simplicity we assume the supply voltage is constant. Again, Black's law is not applicable. At the time one of the conductors tears, the load on the other almost doubles. If we ignore this, X_1 and X_2 are independent, and the model will surely last longer. We denote by X^{lo} and X^{hi} the individual systems under constant low (half) and high (full) stress respectively. We get an upper bound

$$\mathrm{MTTF}(Y) \leq \mathrm{E}\left(\max(X_1^{\mathrm{lo}}, X_2^{\mathrm{lo}})\right) \leq \mathrm{E}\left(X_1^{\mathrm{lo}} + X_2^{\mathrm{lo}}\right) = 2\,\mathrm{E}\left(X^{\mathrm{lo}}\right) \sim 2/J^2 \tag{4.23}$$

[11] If all weights are positive or all weights are negative.

Assuming that the load on each conductor is twice as high from the beginning (and does not change in between), X_1 and X_2 become independent, and we find a lower bound

$$\mathrm{MTTF}(Y) \geq \mathrm{E}(\max(X_1^{\mathrm{hi}}, X_2^{\mathrm{hi}})) \geq \mathrm{E}(X_1^{\mathrm{hi}}) \sim 1/(2J)^2 = 1/4 \cdot 1/J^2, \tag{4.24}$$

one eighth of the upper bound. The actual MTTF is the expected value of Y. We determine Y from the processes behind X_1 and X_2. Clearly, the system lasts longer than its second conductor. The breakdown may happen in two orders, either X_1 or X_2 comes first. Due to symmetry, these events are both equally probable, and we get

$$\begin{aligned}\mathrm{P}(Y < t) &= 2 \cdot \mathrm{P}(Y < t \text{ and } X_1 < X_2) \\ &= 2 \cdot \int_0^t \mathrm{P}(t' < X_2 < t) \cdot \mathrm{P}(X_1 = t')\mathrm{d}t'.\end{aligned}$$

Note that the corner case $X_1 = X_2$ has zero probability. With this, we can express the exact MTTF as

$$\mathrm{E}(Y) = \int_{\mathbb{T}} \frac{\mathrm{d}\,\mathrm{P}(Y < t)}{\mathrm{d}t}\mathrm{d}t. \tag{4.25}$$

This example demonstrates that computing failure times for compound systems from expected values alone is inaccurate even in simple cases. Knowledge of the underlying processes enables the computation of MTTF.

4.3 Modelling Ageing Processes

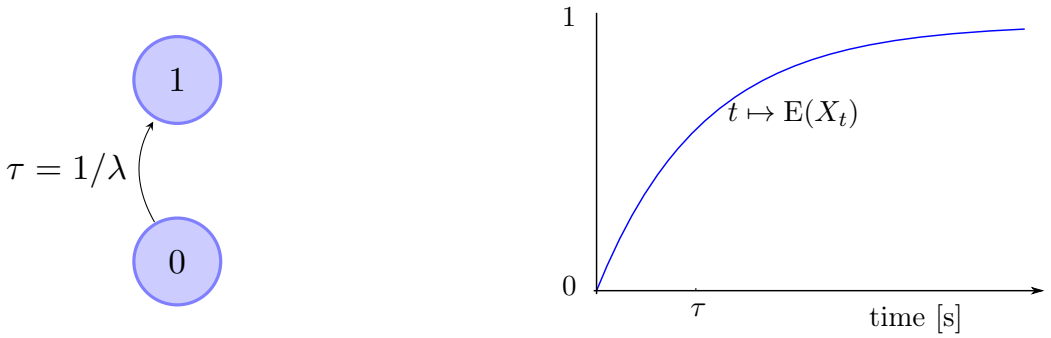
One important aspect of ageing models is computability. We have modeled the value of an ageing parameter during operation or simulation as a function of the state of an ageing process evaluated at the corresponding stress wave that models the operation. This makes the efficient evaluation of processes inevitable for two tasks,

- *finding* process models for ageing parameter, and
- *simulating* systems involving ageing effects.

We are investigating feasibility and efficiency of algorithms or formula that evaluate the map

$$\{\text{processes}\} \times \{\text{stress waves}\} \to \{\text{process states}\} = \mathbb{A}$$

on a class of processes and functions. Restrictions on both, the processes or the stress waves typically improve the evaluation efficiency. For example, any stress wave can be computed on a constant process easily. Integration processes (Eq. 2.1) are particularly simple on constant stress levels, but also harmless on rectangular or sine waved stress waves. Restricting the stress waves to rectangles, there is some success in simulating the RD process, see [Wit07]. With this simplification, a high sensitivity to the number of sampling points has been observed [Gra+06]. Even under constant stress it turns out to be very difficult [OS95]. A circuit level model that



(a) Decay process with constant transition rate λ.

(b) Probability of state 1 increases over time. Probability 0.5 is reached at τ.

Figure 4.4: A simple process model.

mimics the mechanics of the RD process under analogue operating conditions does not seem feasible. Next, we will introduce another class of processes and examine its evaluation properties.

From this perspective, a process which has good evaluation properties for various types of stress waves is needed. We propose the *Reversely Coupled Decay (RCD)* process as a good candidate. We will show that constant time evaluation of RCD processes is possible for rectangular periodic waves and for arbitrarily shaped periodic high-frequency waves. This is particularly useful for fitting purposes. Evaluation on arbitrary stress waves is more involved and will be refined in Section 4.3.

4.3.1 Decay Processes

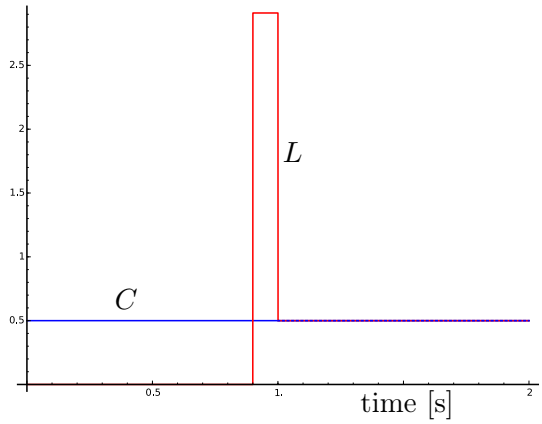
An *exponential process* – known from radioactive decay – models an entity with two states $\Omega = \{0, 1\}$ and a transition from 0 to 1, see Fig. 4.4a.The state of the entity forms a stochastic process X, which at time t is a random variable X_t. The event of the transition is Poisson distributed. This distribution traces back to a *constant* **decay rate** λ. Starting at time 0 in state 0, the probability for $X_t = 1$ is $1 - \exp(-\lambda \cdot t)$. This is also the expected value $\mathrm{E}(X_t)$. We write $\tau := 1/\lambda$ for the **time constant** of the process. It is related to the **half-life** $T_{0.5} = T_{1/\mathrm{e}} \ln(2) = \tau \ln(2)$ of the process.

The expected value $\mathrm{E}(X_t)$ is a deterministic ageing effect model in the sense of Definition 4.5. It is continuous and memoryless and independent of a stress level.

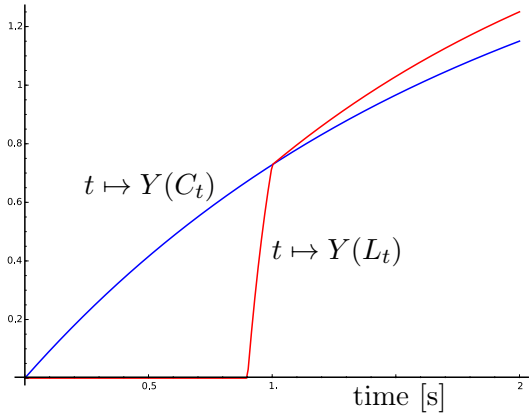
4.3.2 Controlled Decay

Coupling λ to a (positively valued) stress level, e. g. $\lambda = \lambda(L)$ yields another Process X_λ with an expected value of

$$\mathrm{E}(X_{\lambda,t}) = 1 - \exp\left(-\int_0^t \lambda(L(t'))\mathrm{d}t'\right) \tag{4.26}$$



(a) Stress level input.



(b) Response, state coincidence at 1 s.

Figure 4.5: One ageing process with memory, two input stress waves.

at time t. This is a separable ageing effect model, since

$$\frac{\mathrm{d}\,\mathrm{E}(X_{L,t})}{\mathrm{d}t} = \exp\left(-\int_0^t \lambda(L(t'))\mathrm{d}t'\right) \cdot \lambda(L(t)),$$

is a function of $\mathrm{E}(X_{\lambda,t})$ and $L(t)$. In particular, it is memoryless.

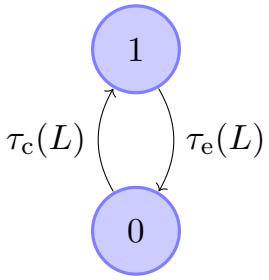
Remark 4.16. *We also use the term* time constant *for the inverse decay rate $\tau(L) = 1/\lambda(L)$ of a controlled process. It is not constant at all, and particularly not w. r. t. to time. With this in mind, we will not need to introduce another notion for the exact same.*

Using these processes, we can construct ageing parameters that are not memoryless. Let $\lambda_1(L) := \sqrt{|L|}$ and $\lambda_2(L) := L^2$. The sum $Y = X_{\lambda_1} + X_{\lambda_2}$ is a stress wave dependent random variable, and the map $Y\colon L \mapsto Y(L) = \mathrm{E}(Y_L)$ is an ageing parameter. Let C be the stress wave that is constantly 1 and

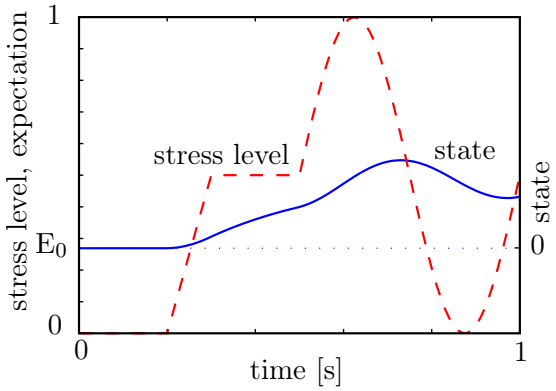
$$L(t) = \begin{cases} 0 & \text{if } t < 0.9\,\mathrm{s}, \\ 2.9103 & \text{if } 0.9\,\mathrm{s} \leq t < 1\,\mathrm{s} \\ 1 & \text{else} \end{cases}$$

The parameters $Y(C)$ and $Y(L)$ coincide at time 1s (see Fig. 4.5). After 1 s, they evolve differently, although the levels are equal, which contradicts Definition 4.8. This essentially happens in BTI, where a parameter drift is a compound of multiple independent variables, such as Y.

A memory in an ageing process has a subtle consequence that is crucial for simulation purposes later on in Chapter 5. Consider the value y of the process Y in the example at 1 s. The time evolution after 1 s does *not only* depend on just y and the subsequent stress level. This can be seen as the initial motivation for introducing ageing parameter models (Definition 4.5) as an access to ageing processes.



(a) Example RCD. We set $\tau_c = \exp(1 - L)\,\mathrm{s}$ and $\tau_e = \exp(L)\,\mathrm{s}$.



(b) A stress signal (dashed) and its effect on the state of the process.

Figure 4.6: Stress level controlled coupled process. The initial state 0 is the equilibrium E_0 at $L = 0$.

4.3.3 Coupled Decay Processes

The **Reversely Coupled Decay** process (RCD) models a generic ageing mechanism based on stress level controlled stochastic processes. In a controlled decay process, the decay rate depends on a stress level which may increase the transition probability density wrt. time. Two such processes acting on the same set of states in opposed directions form a new process (c. f. Fig. 4.6a). The compound process is suited to model time continuous, arbitrary stress level controlled and recovery aware ageing parameters [SH11].

Definition 4.17. *An **RCD** is a stochastic process defined by two controlled decay processes on an entity with the states* $0, 1 \subset \mathbb{R}$*. It is determined by a tuple* (τ_e, τ_c) *of monotonic, nonnegative functions of the stress level. One of them,* τ_e*, is increasing, the other one,* τ_c*, is decreasing with the level. These turn a stress level into time constants of the corresponding processes, see Fig. 4.6a.*

The following essential equations can be easily deduced from the differential equation (see [SH11]). At a constant stress level L (if $\tau_e(L)$ or $\tau_c(L)$ is nonzero), the expected value of an RCD r converges to

$$\mathrm{E}_L := \tau_e(L)/(\tau_e(L) + \tau_c(L)), \tag{4.27}$$

for increasing time t. The expected value $\mathrm{E}_r(L)$ is the **equilibrium** at L. The centered expected value $\mathrm{E} - \mathrm{E}_0$ can be considered the **state of the RCD** r, as the process is memoryless. Unless stated otherwise, the state at time zero is zero. The equilibrium state at a stress level L then amounts to the **limit state**

$$\ell(L) := \mathrm{E}_r(L) - E_r(0) = \frac{\tau_e(L)}{\tau_e(L) + \tau_c(L)} - E_r(0). \tag{4.28}$$

The time constant of a transition to the equilibrium is

$$\tau(L) = \tau_{\mathrm{e}}(L) \cdot \tau_{\mathrm{c}}(L)/(\tau_{\mathrm{e}}(L) + \tau_{\mathrm{c}}(L)). \tag{4.29}$$

The RCD r in a state a_0 at time zero then has a simple time evolution at a constant stress level function L,

$$r\colon t \mapsto r(L_t) = a_0 + \left(1 - \exp\left(\frac{t}{\tau(L)}\right)\right)(\ell(L) - s_0). \tag{4.30}$$

Lemma 4.18. *An RCD is constantly zero iff τ_e and τ_c are linearly dependent, i. e. if one is a multiple of the other.*

Proof. Clearly ℓ is constant in case of linear dependency. Conversely, the state of a constant RCD is fixed at $\ell(0)$. For any Level L, the transition rate $1/\tau(L)$ is positive if $\tau_{\mathrm{e}}(L)$ and $\tau_{\mathrm{c}}(L)$ are nonzero. Hence, for these levels, the *limit state* $\ell(L) = \tau_{\mathrm{e}}(L)/(\tau_{\mathrm{e}}(L) + \tau_{\mathrm{c}}(L))$ must equal $\ell(0)$. Thus $(\ell - 1)\tau_{\mathrm{e}} = \tau_{\mathrm{c}}$ for all these levels. If for some L, one of $\tau_{\mathrm{e}}, \tau_{\mathrm{c}}$ is zero, then the other one must also be zero to maintain constness. □

If L is a stress wave on $[0, \mathrm{T}_L] \subset \mathbb{T}$ and r is an RCD, we write $r(L)$ for the state of r under L at T. In turn $r(L_t)$ is the state at time t. Let s be a real number, we define $s \cdot r$ as $L \mapsto s \cdot r(L)$. With this, the RCDs span a subspace S_{RCD} in the vector space of ageing processes.

4.3.4 Rectangular Stress

An RCD process controlled by rectangular stress behaves in a particularly simple and arithmetically accessible way.

Definition 4.19. Rectangular stress *of duration T is a stress wave on $[0, T]$ that is parameterized by levels ℓ_{lo}, ℓ_{hi} a period $p > 0$, and a duty-cycle $d \in [0, 1]$. Its value at t is*

$$R_{(\ell_{lo},\ell_{hi},p,d)}(t) := \begin{cases} R_{(\ell_{lo},\ell_{hi},p,d)}(t-p), & \text{if } t > p, \\ \ell_{hi}, & \text{if } t > dp, \\ \ell_{lo} & \text{else.} \end{cases}$$

With rectangular stress, we may control a simpler class of processes. Pick two times τ_{hi} and τ_{lo}.

Definition 4.20. *Let $\mathbf{L}$ be the set of stress waves that take levels ℓ_{hi} and ℓ_{lo}. We define a deterministic process controlled by a level $L \in \mathbf{L}$ with a state x and a time evolution defined by*

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \begin{cases} (x(t) - \ell_{hi})/\tau_{hi}, & \text{if } L(t) = \ell_{hi}, \\ (x(t) - \ell_{lo})/\tau_{lo}, & \text{if } L(t) = \ell_{lo} \text{ and} \\ \textit{anything} & \textit{otherwise.} \end{cases}$$

This process may be seen as a netlist of switches, resistors and a capacitor that holds the state Fig. 4.7. We call it SRC process.

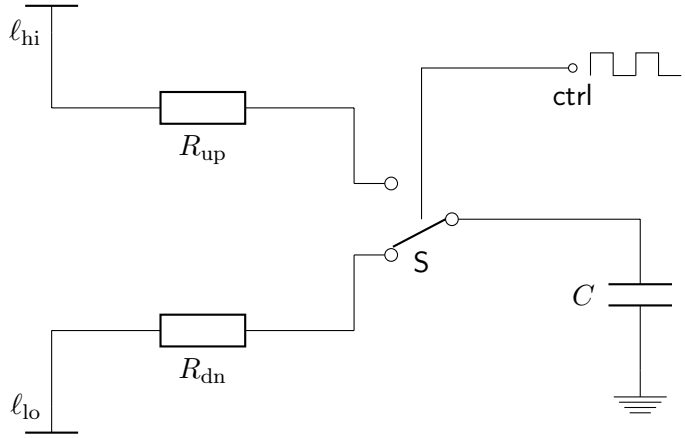


Figure 4.7: An SRC process expressed as a circuit. The switch state is controlled by a two-valued stress level.

The SRC purposefully may do anything in the third case, to emphasise the ill-definedness for stress levels $L \notin \{\ell_{\mathrm{hi}}, \ell_{\mathrm{lo}}\}$. The following observation has been published in [SH11], with a slightly different wording. There we show that a new class of analogue BTI models extends a model for BTI that only works for digital operation and fixed amplitude.

Lemma 4.21. *For any RCD process r and a rectangular stress wave R, there exists an SRC process r' and rectangular stress R' such that for times $t \in \mathrm{dom}(R)$ we have $r(R_t) = r'(R'_t)$.*

The following property is essential.

Theorem 4.22. *An SRC is frequency independent and there is an explicit expression for the frequency limit at a time t.*

Proof. Let R be a rectangular stress wave with duty cycle d, duration T and period p. We may assume $0 < d < 1$, as the assertion is trivial for the corner cases. Also, without restriction, $p = T$. Let S be an SRC and $\lim_{\alpha\to\infty} S(R^\alpha) = x$. In particular, if the SRC is in state x, then

$$\lim_{t\to 0} S\left(R_t^{1/t}\right) = x. \tag{4.31}$$

For positive t, the argument for lim on the left hand side is

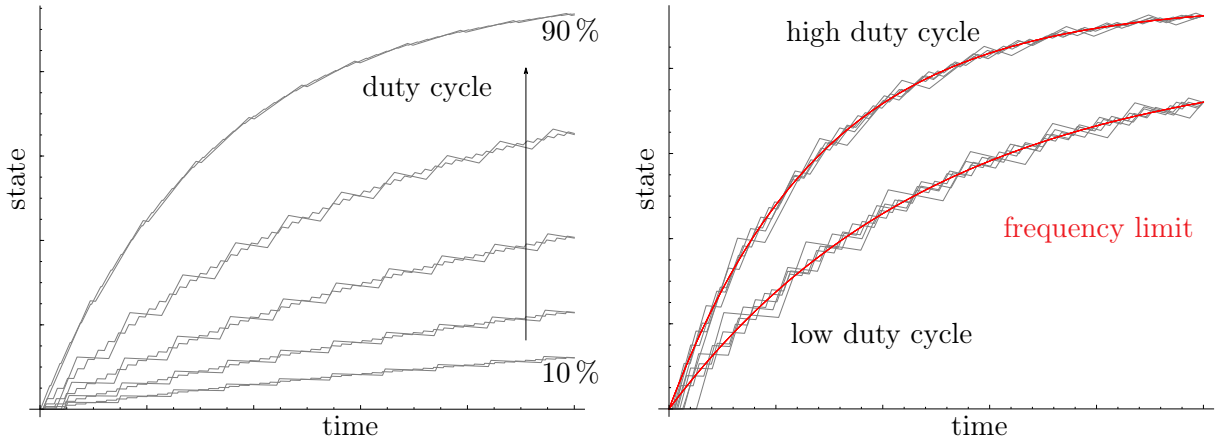
$$C := l + (x - \ell_{\mathrm{hi}}) \cdot \exp(td/\tau_{\mathrm{hi}})$$

at $t = td$, as $R = \ell_{\mathrm{hi}}$ between 0 and t. Similarly, we have

$$\ell_{\mathrm{lo}} + (C - \ell_{\mathrm{lo}}) \cdot \exp(t(1-d)/\tau_{\mathrm{lo}})$$

at t. With this, taking the limit $t \to 0$ is easy, and we get

$$x = \frac{-d\tau_{\mathrm{lo}}}{((d-1)\cdot\tau_{\mathrm{hi}} - d\cdot\tau_{\mathrm{lo}})}. \tag{4.32}$$



(a) Low and high frequency trajectories for five different duty cycles.

(b) Two duty cycles. With increasing frequency, the trajectories converge pointwise.

Figure 4.8: Rectangular stress in an SRC process.

With similar considerations we get the effective time constant for the limit $\tau := ab/(a+b)$ with $a = \tau_{\text{hi}}/d$ and $b = \tau_{\text{lo}}/(1-d)$. Then

$$\lim_{t\to 0} S\left(R_T^{1/t}\right) = \ell_{\text{hi}} + \exp(t/\tau)(\ell_{\text{lo}} - \ell_{\text{hi}}). \tag{4.33}$$

□

We illustrate this result as follows. Let S be an SRC. Feeding in admissible rectangular stress, the SRC state will go up and down. And it will go up and down faster as the input frequency increases (Fig. 4.8a). However at any fixed time, the frequency increase makes the state converge. The range of states for frequencies above f at such a fixed time is shrinking with increasing f, it converges to a singleton, the frequency limit Fig. 4.8b.

In particular

Theorem 4.23. *any RCD process has an explicit expression for its frequency limit when operated under rectangular conditions.*

For an RCD process operated under rectangular conditions only the two input levels are relevant. With the input level fixed, the RCD converges to the equilibrium with a fixed rate – independent of the current state. Hence the expected value of the RCD behaves like an SRC, and the frequency limit translates back. □

Similarly, we may characterize the upper and lower bounds of an RCD process under rectangular conditions using the accompanying SRC model. Starting from the frequency limit, the trajectory for one period of the input stress wave can be written down explicitly. Hence the maximum and minimum are explicit. The global sequence of local maxima and minima respectively converge to these, exponentially with the same rate the frequency limit does. Thus, For a given rectangular stress wave, the upper and lower extrema sequences M and m have an explicit representation, see the illustration in Fig. 4.9.

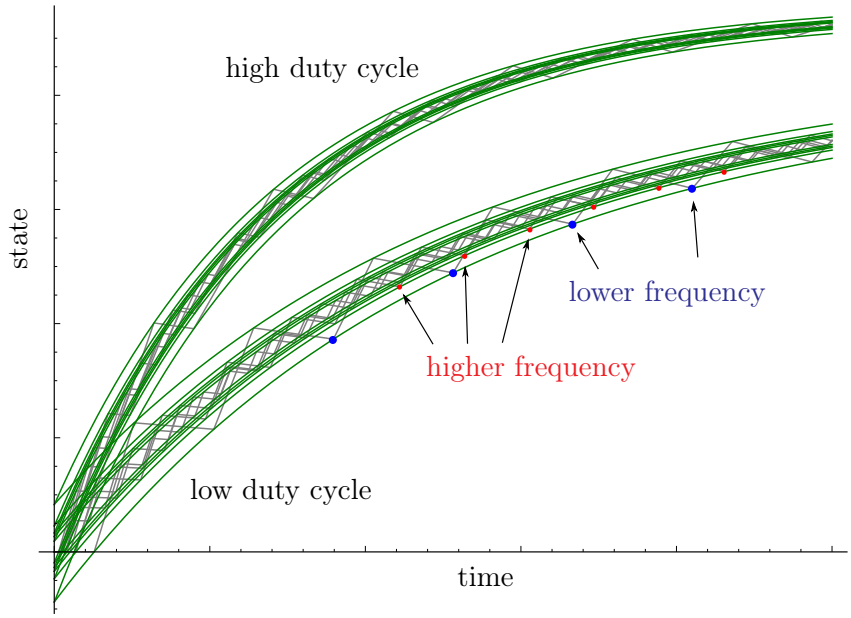


Figure 4.9: Upper and lower bounds of an SRC process unter rectangular stress (with different duty cycles), converge to each other with increasing frequency. The plotted envelopes are exponential curves with explicit known formulas.

In particular we have the following.

Theorem 4.24. *The effect of a periodic rectangular shaped stress wave on an RCD can be computed in constant time, if the time constants of the RCD permit.*

Computing the time constants is not really an issue, as these appear as ordinary explicit arithmetic expressions.

4.3.5 Generalisations and Average Stress

The explicit frequency limit from Theorem 4.22 has a generalization to RCDs. Choose a stress wave Q that takes finitely many values l_i, each for a fraction $d_i \leq 0$, and $\sum_i d_i = 1$, i. e. within it domain, Q takes the level l_i for $d_i \cdot T$ the time. These values induce limit points $\ell_i = \ell(_i)$ that are approached with time constants $\tau_i = \tau(l_i)$. Generalizing the technique for two-valued stress levels in SRCs, the frequency limit of an RCD under L can be expressed as

$$\ell_Q = \frac{\sum_i \ell_i d_i \tau_i}{\sum_i d_i \tau_i}.$$

The effective time constant for the frequency limit amounts to

$$\tau_Q = \frac{1}{\sum_i d_i / \tau_i}.$$

From this formula, we find that there is not necessarily a stress level C, s. th. $\tau(C) = \tau_Q$ and $\ell(C) = \ell_p$. Hence it is not possible do describe the impact of arbitrary periodic stress wave

by a constant stress level. The impact of only one stress wave however can be translated to the impact of a constant stress, which gives rise to the following definition from [SHM13].

Definition 4.25. *Let r be an RCD and L be a stress wave. Denote by r_a the RCD r in state a at $t = 0$. The* ***locally equivalent constant stress*** *of L for r_a is a constant stress level function C, that, applied for the duration T_L, has the effect on r_a that L has. To put it short,*

$$r_a(L) = r_a\left(C_{\mathrm{T}_L}\right). \tag{4.34}$$

This locally equivalent stress does *not* account for the impact of a stress wave alone, but only for the impact of a stress wave on an RCD in a particular state. This is somehow logical, as a stress wave holds more information as a constant does. However this notion will turn out to be highly useful for simulation purposes, see Section 5.5.1. Indeed, a locally equivalent stress exists in various cases.

Theorem 4.26. *For an RCD r with continuous time control the equivalent constant stress exists for any stress wave L.*

Proof. The RCD r and the level L induces a continuous map $e\colon \mathbb{R} \to \mathbb{R}$. Denote by $C(c)$ the constant stress wave with level c and duration T_L. With this, define

$$e\colon c \mapsto r(C(c)). \tag{4.35}$$

Monotonicity ensures that

$$e\left(\min_{t\in\operatorname{dom} L} L(t)\right) \leq r(L) \leq e\left(\max_{t\in\operatorname{dom} L} L(t)\right). \tag{4.36}$$

Using *Bolzano's theorem*, the function $c \mapsto r(L) - e(c)$ has a zero c_0 between $\min_{\operatorname{dom} L} L$ and $\max_{\operatorname{dom} L} L$. The level c_0 qualifies as equivalent constant stress of L. □

4.3.6 Exponential Control

Restricting to a subspace of S_{RCD} we will identify models with several useful properties.

First, we need a standard fact from linear algebra about linear dependency [Fis05]. The set of functions $\mathbb{R} \to \mathbb{R}$ is a vector space over $\mathbb{R}$ by means of pointwise scalar multiplication defined by $s \cdot f\colon x \mapsto s \cdot f(x)$.

Lemma 4.27. *Let n be a positive integer and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be mutually different. The functions $f_i\colon x \mapsto \exp(\lambda_i x)$ are linearly independent as functions $\mathbb{R} \to \mathbb{R}$.*

Proof. Differentiation is a linear map from the real vector space of smooth functions $\mathbb{R} \to \mathbb{R}$ onto itself. The function f_i is eigenvector of the differential operator D with eigenvalue λ_i, i. e. $\mathrm{D}f_i = \lambda_i$. Eigenvectors to different eigenvalues are linearly independent. □

Definition 4.28. *An **exponential RCD, (ERCD)** is an RCD (τ_c, τ_e) where*

$$\tau_c \colon L \mapsto \exp(k_c - m_c L) \text{ and}$$
$$\tau_e \colon L \mapsto \exp(k_e + m_e L),$$

with real coefficients, k_c, k_e and nonnegative m_c, m_e.

An ERCD is zero iff $m_\mathrm{c} = m_\mathrm{e} = 0$, this follows from Lemma 4.27 and Lemma 4.18. An ERCD S in fresh state at $t = 0$ has the following time evolution upon aplying a constant stress L.

$$S(L_t) = (\mathrm{E}_L - \mathrm{E}_0) \cdot (1 - \exp(-t/\tau)) - E_0$$

where

$$\begin{aligned} E_L &= \frac{1}{1 + \exp(k_\mathrm{c} - k_\mathrm{e} - (m_\mathrm{c} + m_\mathrm{e}) \cdot L)}, \text{ and} \\ \tau &= \frac{1}{\exp(m_\mathrm{e} L + k_\mathrm{e}) + \exp(-m_\mathrm{c} L + k_\mathrm{c})} \end{aligned} \tag{4.37}$$

Theorem 4.29. *The non-zero ERCDs with mutually different parameters (m_e, m_c, k_e, k_c) are linearly independent.*

Proof. Choose ERCDs $r_1, \ldots, r_n$ with mutually different parameters and the time constants $\tau_1, \ldots, \tau_n$. We need to show that all weights $w_1, \ldots, w_n$ are zero, if

$$\sum_{i=1} n w_i r_i = 0.$$

For $1 \leq i, j \leq n$ define the sets $B_{i,j} = \{L \,|\, \tau_i(L) = \tau_j(L)\}$. Let a be the sequence $(a_1, a_2, ...)$ with $a_n := 1/n$. First we show that $\{a_1, a_2 ...\}$ is not a subset of $U := \bigcup_{i<j} B_{i,j}$. Assuming the contrary, a subsequence $(a_{n_1}, a_{n_2} ...)$ is contained in one of the $B_{i,j}$, say in $B_{1,2}$. Then, we have $\tau_1(a_{n_k}) = \tau_2(a_{n_k})$ for all k. Using the Identity Theorem for entire functions [FB05] we get $\tau_1 = \tau_2$. That means, the linear combination of exponential functions in L,

$$\begin{aligned} L \mapsto &\exp(k_{\mathrm{e},1}) \exp(m_{\mathrm{e},1} \cdot L) + \exp(-k_{\mathrm{c},1}) \exp(m_{\mathrm{c},1} \cdot L) \\ &- \exp(k_{\mathrm{e},2}) \exp(m_{\mathrm{e},2} \cdot L) - \exp(-k_{\mathrm{c},2}) \exp(m_{\mathrm{c},2} \cdot L) \end{aligned}$$

equals zero. Due to Lemma 4.27, we first get $m_{\mathrm{e},1} = m_{\mathrm{e},2}$ and then immediately $k_{\mathrm{e},1} = k_{\mathrm{e},2}$. A contradiction, the ERCDs are mutually different by premise. We have showed that the sequence a has elements outside of U.

Set L_0 to an element of a that is not contained in U. Summing up the trajectories from Eq. 4.37, we get

$$0 = \sum_i w_i c_i \exp(t/\tau_i(L_0))$$

with all $c_i \neq 0$. By construction, the $\tau_i(L_0)$ are mutually different and due to the independence of the $t \mapsto \exp(t/\tau)$ the coefficients w_i must be all zero. □

In the above proof, we have only used constant stress. We conclude that two non-constant ERCDs coincide on all constant stress functions if and only if they are equal. Moreover, using the above linear independence, we may say the vector space S_{ERCD} spanned by the ERCDs is as free from relations as it can be.

Corollary 4.30. *Two weighted sums of mutually different non-zero ERCDs, possibly shifted by a constant are equal iff all involved parameters are equal.*

4.3.7 Practical Considerations

We define $S_{\mathrm{ERCD}} := \mathrm{span}_{\mathbb{R}}(\mathrm{ERCDs})$, the real vector space spanned by the ERCDs. With this we can model stress dependent parameters from data involving constant stress conditions. A **data set** D is a set of tuples with a stress wave and a parameter value. For example, constant stress waves are determined by a level c and a duration t. A triplet $(c, t, p) \in D = \mathbb{R}\times\mathbb{T}\times\mathbb{R}$ describes the impact of a constant stress wave as a parameter p. We are looking for an element $R \in S_{\mathrm{ERCD}}$, such that $R(C_t) \approx p$ whenever C is a constant stress function and $(C(0), t, p) \in D$. Such a model R, operated under the respective constant stress levels, will reproduce the parameters accordingly. Constant stress is particularly easy to cope with, as an RCDs state can be easily evaluated on constant stress waves. This will be of importance during the search for R. We have introduced more such admissible stress waves in Section 4.3.4, and we shall include rectangular periodic stress waves to the set D later.

An ageing process $m \in S_{\mathrm{ERCD}}$ has a set of properties we know about. A priori, a deterministic ageing process is a kind of map

$$Q\colon \{\mathbb{T} \to \mathbb{R}\} \to \{\mathbb{T} \to \mathbb{R}\}.$$

Data originating from simulations or measurements can be seen as (approximations for) evaluations of this process at some chosen points. Searching for a model $m \in S_{\mathrm{ERCD}}$ for Q makes sense if that process agrees in properties with m. This makes

- locality,
- time invariance,
- determinism,
- continuity and
- frequency independence

prerequisites for Q. It would be advantageous to know whether any process with these properties can be approximated by elements of S_{ERCD} (in the sense a real number can be approximated by rationals). However, we do not. From this perspective, a process should pass the implied check, before we even try to model it. Also, with extra knowledge about Q, restrictions on the elements of S_{ERCD} can reduce the search space. For example, models with positive weights are monotonically increasing.

4.4 Fitting Process Models

We present an algorithm that finds a configuration of ageing processes to fit a set of measurement data. Measurement data is organized as a map from stress waves to ageing parameters (**values**). Our approach is is looking for a linear combination of simplistic ageing processes (**cells**) that explains the measurements. The objective is to find a weighted sum of cells that fits the measurements at the sample stress waves and produces reasonable states for the others.

It is necessary that a cell can be evaluated quickly for a given stimulus. Hence the stress waves types in the measurement set constrain the type of cells that can be used, and vice versa. A cell that just integrates a stimulus is probably the simplest nontrivial example. Such a cell can be evaluated symbolically for about all practically relevant stress waves. We chose RCD processes as the cells, as these still admit constant stress, as well as pulses with either few transitions or arbitrarily high frequency.

If a weighted sum of cells M is found to fit a set of measurements for stress waves, the so obtained model M may still misbehave. Consider the simplified case where two measurements on two constant stress waves of duration 1 and level 0.1 and 1 yield a value of .28 and 1 respectively. We find a model $M = 3g - 2f$ built from cells $f\colon L_t \mapsto \int_0^t L^2 \mathrm{d}t$ and $g\colon L_t \mapsto \int_0^t L \mathrm{d}t$. This model fits the given data points. However, issues arise when changing the stress waves. For example stress with level 0.75 of duration 1 results in a value above both measurements, although arguably the applied stimulus is in between known measure points.

To amend this, we may restrict the cells to monotonically increasing ones and only permit positive coefficients. As a consequence, we can only fit to measurements that (are believed to) originate from monotonically increasing ageing processes. ERCDs are positively monotonic and good candidates. Note that positively weighted sums of such processes are monotonic increasing.

4.4.1 Optimizing Weights

Fix a given a set of cells $z_1, ..., z_n$, such as ERCDs. Given a set M of stress waves L_j that cause p_j we are looking for a vector w of nonnegative weights $w_1 ... w_n$ such that – in some sense – the sample errors $e_i := \sum_i w_i z_i(L_j) - p_j$ are small for all j. With this, the model $M = \sum_i w_i z_i$ reproduces the data points of D well. This condition is met, if

$$\mathcal{E} := \sum_j \left(\sum_i w_i z_i(L_j) - v_j \right)^2, \tag{4.38}$$

the sum of squares, is zero. If the sum is small, the samples are approximated. Clearly, the e_i are bounded from above by $\sqrt{\mathcal{E}}$. $\mathcal{E}$ is a quadratic function in w and of the form

$$\mathcal{E}(w) = (Aw + b)^\mathsf{t} \cdot (Aw + b) = w^\mathsf{t} A^\mathsf{t} A w + 2 b^\mathsf{t} A w + b^\mathsf{t} b, \tag{4.39}$$

for some matrix A and a vector b. The matrix $Q := A^\mathsf{t} A$ is positive semidefinite, hence the minimum of $\mathcal{E}$ is easy to compute. We may be looking for weights within a certain range, for

example nonnegative ones or not too big ones, maintaining numerical stability. In that case the minimum of Eq. 4.39 needs to be found under an additional linear constraint defined by a matrix G and a vector h in $Gx \leq h$, where the inequality is implied component-wise.

After evaluating the cells on all stress waves, finding weights turns out to be a prototype linearly constrained quadratic programming problem. It has been extensively studied and implemented before. We rely on the correctness of the `qp` algorithm and its implementation that is part of cvxopt [ADV12] package. Altogether, we can evaluate the function

$$\begin{aligned} F\colon \{\text{Sets of ERCDs}\} &\to \mathbb{R}_{\geq 0} \\ \{z_1, ..., z_n\} &\mapsto \mathcal{E}(\operatorname{argmin}(\mathcal{E})) = \min_w \sum_i e_i^2 \end{aligned} \tag{4.40}$$

efficiently, *if* we can compute $z_i(L_j)$ efficiently. The function F evaluates the quality or **fitness** of a set of cells with respect to measurements. This perspective on model fitting has been explained in [Ros+15] and presented at a workshop associated to ZuE12.

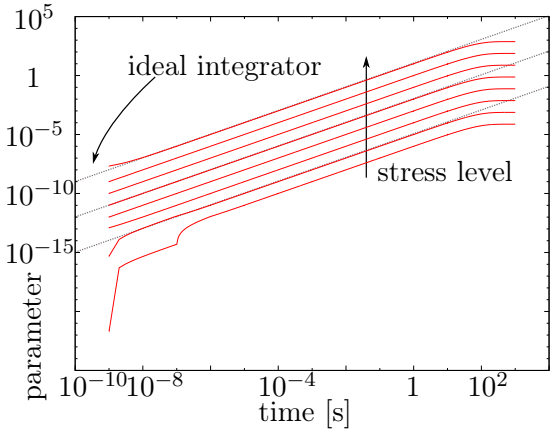
4.4.2 Optimal Cell Selection

An optimal solution for Eq. 4.40, under the given constraints, is a set of cells. For monotonicity reasons, we start with ERCD cells, these have four real parameters each. and permit quick evaluation for a fairly large class of input stress waves (Theorem 4.24). Rectangular periodic waveforms with arbitrary low and high levels and arbitrary frequencies (including 0 and ∞) can be applied in $O(1)$. Other types of cells can be added later on. *Evolutionary algorithms* [Ash06] work with a set of sets of cells (a **population**) and a fitness function such as F in Eq. 4.40. The population consists of *individuals* that randomly *mutate* and *crossover*, focusing on the individuals that F maps to small values. These algorithms are a good choice for the search for sets of cells, as these have canonical crossover methods. Clearly two equally sized sets which are implemented as lists of four-tuples give rise to randomly chosen sublists. Mutations are random parameter shifts in cells with variance controlled by their range. The mutation rate of a cell can be aligned to the size of its *range*, the set of ageing states that are reached from the stress waves in the data set. Cells with a small range generally are of no use and made to mutate faster. We have implemented the learning strategy using pyevolve [Per09], it works well with about 100 stress waves and up to 20 cells. We then use the derivative-free optimizer COBYLA [PCP92; Pow98], which is implemented for nlopt [Joh], on the same objective. This can improve the fitness of the best by a few percent.

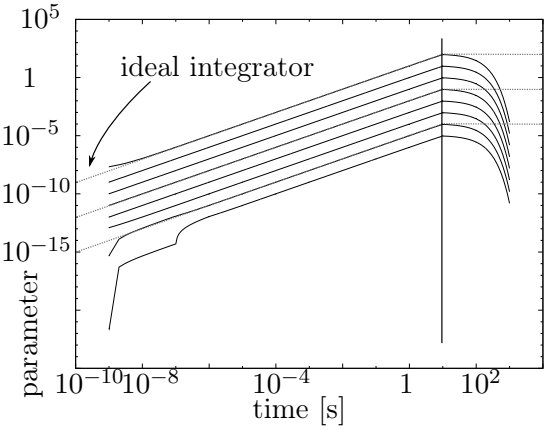
4.4.3 Illustration

An integrator is a simple univariate ageing process. It integrates the stress level. A constant stress of level L, applied for a time t will yield an output of $L \cdot t$. We demonstrate, that a sum of ERCDs approximates this behaviour. We have combined the times

$$t \in \{1, 2, 2.5, 3, 3.5, 4, 5, 0.1, 0.001, 0.0001\}\text{s}$$

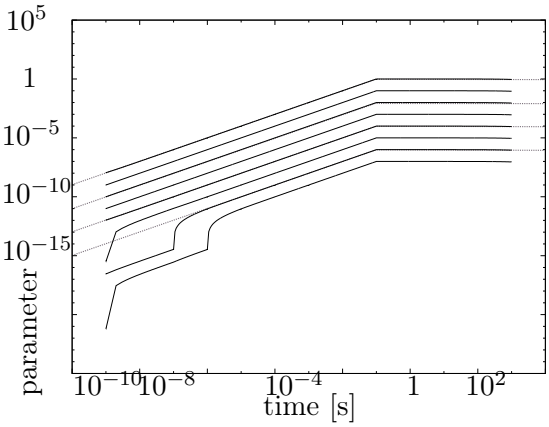


(a) Constant stress conditions are reproduced well

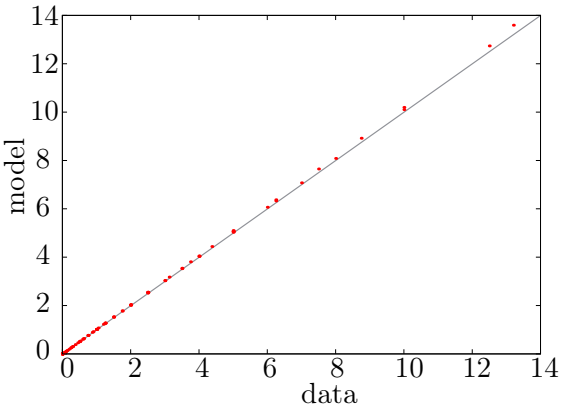


(b) Erratic relaxation after switching off the stress at 10 s.

Figure 4.10: An ERCD sum fitted against integrated constant stress.



(a) Exposed to different levels of stress, switching off at 0.1 s.



(b) Model quality histogram, one dot for each point in the learning set.

Figure 4.11: An ERCD sum fitted to integrated constant stress and integrated pulses.

with levels $L \in \{1, 2, 2.5, 0.5, .3\}$ to a set of 50 stress waves, and the respective outputs of the integrator. From this data set we have found a sum of eight ERCDs, using the optimization algorithm described above. This sum produces the output shown in Fig. 4.10a and Fig. 4.10b. In the first plot, we have applied constant stress of levels $\{0.001, 0.01, \cdots, 10\}$ throughout the simulation time, for the second plot, we have switched of stress after 10 s. We observe that the ERCD sum reproduces the samples and performs arbitrarily outside the learning set.

To get a more appropriate ERCD based integrator model, more input is required. So, for each pair (t, L) from above, we add the pulse of width $t/2$ and level L followed by $t/2$ zero level with their respective outcomes to the learning set. With this, we get a model that much more resembles an integrator, see Fig. 4.11a.

With growing learning sets, it becomes increasingly cumbersome to compare transient results for model verification purposes. Also, with real ageing effects, transients are only sparsely available. Talking about data based models, everything that can be possibly asserted is the model properties and the fitness of the model for the given data points. The learnt integrator model is a frequency independent and monotonic controlled ageing process that coincides with an integrator on a few sample points. While the former property can be formally proved (see Lemma 4.15), the latter coincidence is subject to numerical disturbance and may be visualized by a model quality histogram Fig. 4.11b.

4.4.4 Multivariate Stress Levels

In the above examples, a single stress level controls a process that models the drift of a single parameter. In practice, multiple parameters describe the behaviour of a component, and ageing effects have an impact on many of them. Likewise, components are sensitive to a variety of stress levels that depend on their operating point. Consider for example an amplifier with a gain and an offset and a cut-off frequency. These depend on the characteristics of the built-in transistors with their ageing processes which in turn are controlled by their individual stress levels.

This example gives rise to a multivariate variant of the ERCD sum model. A **selector** is a linear form $\mathbb{R}^m \to \mathbb{R}$ that picks a coordinate. Formally it is an element of the dual of the standard basis of $\mathbb{R}^m$, or a *row vector* $(0, ..., 1, ...0)$.

Given set of cells $z_1, \ldots, z_k$, and sets of weights $w_1^1, w_1^2, ...w_N^k$, and selectors π_i, we get ageing parameters

$$Q_m \colon L = (L_1, \ldots, L_m) \mapsto \sum_i w_N^i z_i(\pi_i L) \tag{4.41}$$

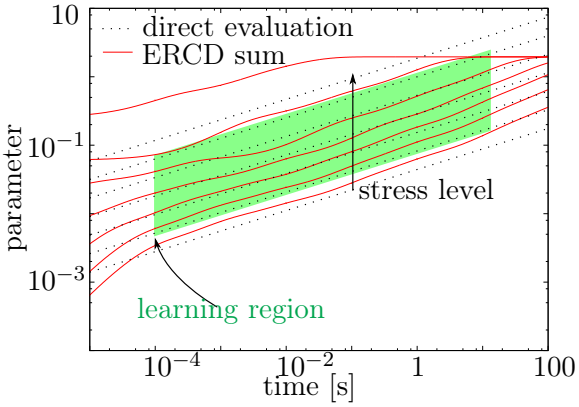
that depend on the state of the ageing processes z_i. Similarly to the univariate case, if we know a set of assignments for the stress waves $L_1, \ldots, L_m$ for which the impact on the parameters is known, we can – under some circumstances – find cells and weights in Eq. 4.41.

The weight optimizer (Section 4.4.1) needs to take into account multiple parameters, but this does not change the problem statement. Clearly, finding a vector v minimizing $|A_1 v - b_1^\mathsf{t}|^2 + |A_2 v - b_2^\mathsf{t}|^2$ under linear constraints on v is the same as finding the minimum of

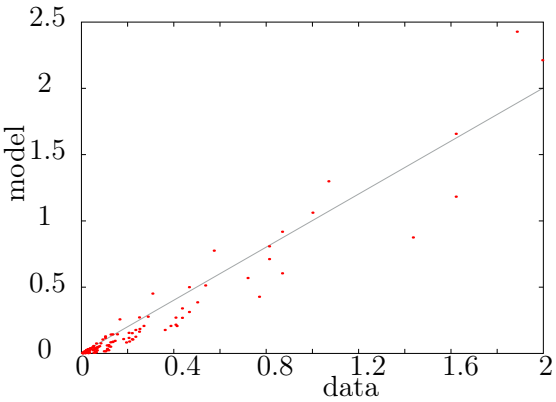
$$\left| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} v - (b_1, b_2)^\mathsf{t} \right| \tag{4.42}$$

under these constraints on v. The cell optimizer (Section 4.4.2) needs to take care of the selectors π_i. This means, that the dimension of the search space increases exponentially with m, the number of independent stress level functions. For $m \leq 5$, this is not yet an issue.

Finding accurate stress levels is challenging, particularly if the component model is intended to be free from the transistor netlist. A potentially similar approach has been described in [Mar+06]. Here, a lower level model of the transistor netlist is retained to keep track of stress impacts on individual transistors. Also, the ageing impact is restricted to HCI. Similarly, only



(a) Responses to constant stress levels $2^{-5}, 2^{-3}, \ldots, 2^{1}$. Direct evaluation vs. simulation of an ERCD sum model.



(b) Corresponding model histogram. The discrepancy is suspicious.

Figure 4.12: Fitting Eq. 4.43.

integrator cells are used. From the perspective of ageing processes this appears to be a trivial corner case.

4.4.5 Fitting Caveats

Fitting ageing models becomes accessible with the right set of tools, still its not trivial. The above examples are chosen to demonstrate the semantics of the fitting process algorithm. The shape of the measurement data points plays an important role for fitting. As an example, consider the model

$$M : L \mapsto \left(\int_{\mathrm{dom}\, L} L(t)^3 \mathrm{d}t \right)^{.3}, \tag{4.43}$$

which is related to HCI modelling. Probing this model, we generate a set of samples for an example learning set. We evaluate Eq. 4.43 on stress pulses, and the frequency limit for stress of levels $2^{-4}, 2^{-3}, \ldots, 2^{0}$, durations within $[1 \times 10^{-4}, 10.0]$ and various duty cycles including 1. We then tried to find an optimal fit of a sum of 20 ERCD cells to 120 such sampling points. The result is dramatically worse in comparison with the integrator case. The reached minimal error square sum is above 10.0, clearly indicating that something went wrong, see Fig. 4.12b. The example plot in Fig. 4.12a reveals that the ERCD sum does not even reproduce the response to constant stress very well.

This example points out that a reinterpretation of stress levels and parameters helps fitting. If we replace L by L^{-3} and M by $M^{-.3}$ in Eq. 4.43, we get back the integrator that we managed to fit perfectly. Reinterpretation of stress levels in principle improve the fitting methodology. Note that this also affects the class of the stress waves and may complicate or break the evaluation scheme.

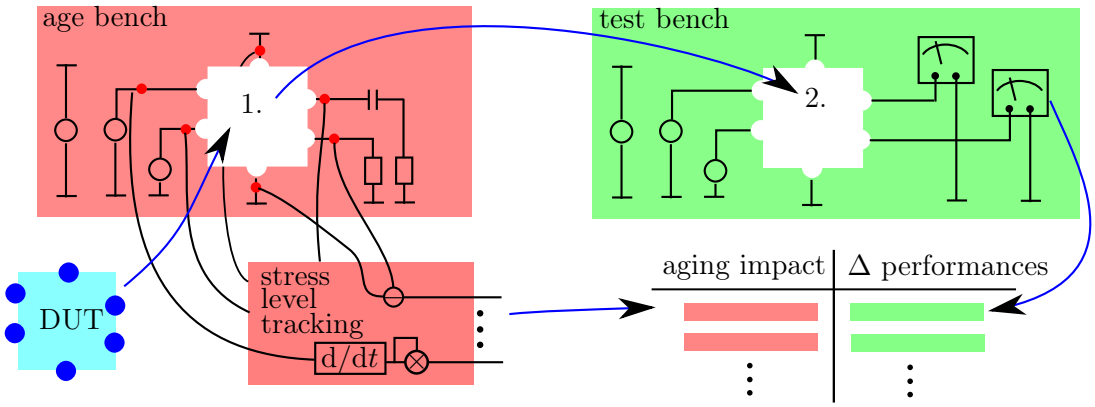


Figure 4.13: Ageing data set extraction. Filling a table with stress waves and their respective impact.

Multiple level dependencies introduce another challenge. In practice, it cannot be avoided that a property is affected by multiple influences. Consider an ageing parameter that is pulled up by one ageing mechanism and pulled down by another. These ageing mechanisms and their controlling levels may be independent even in applications. Consider the offset voltage of the comparator in Section 6.2.2 for an example. Obviously, a positively weighted sum of monotonically increasing cells cannot model such a relation. Instead, a model, such as

$$Q(L) = w_1 z_1(L) - w_2 z_2(L) \tag{4.44}$$

with monotonically increasing cells z_1, z_2 and positive weights w_1, w_2 is required, and this contradicts monotonicity. The unconstrained learning algorithm for Eq. 4.44 may in this case come up with huge weights and very similar processes z_1, z_2, and also fit measurements very well, which used to indicate success. Due to the lack of monotonicity there is no guarantee, that the model is useful outside of the learning set. In practice, the huge coefficients make the model highly sensitive to level changes and unphysical behaviour. To amend this, further restrictions to the weights and processes, such as upper bounds and range criteria need to be added to the learning strategy. The resulting model needs to be examined with care before use.

4.5 Hierarchical Ageing Modelling

Ageing models are required to simulate the consequences of ageing effects. Modelling ageing effects for analogue applications is more complex than in the digital domain, as the set of possible operating points is much larger.

From ongoing work on reliability modelling [MG13, 136ff] an approach to predict the ageing impact on larger circuits has evolved. This work is based on the idea that a *performance* $\mathbf{P}$, which is the same as a behavioural parameter, of a circuit block is a function of an *input* $\mathbf{u}$ and the time t. The input is implicitly but inevitably restricted to a constant. With u assumed to be constant, the performance $\mathbf{P}$ becomes a function that depends on u and t. Such functions can

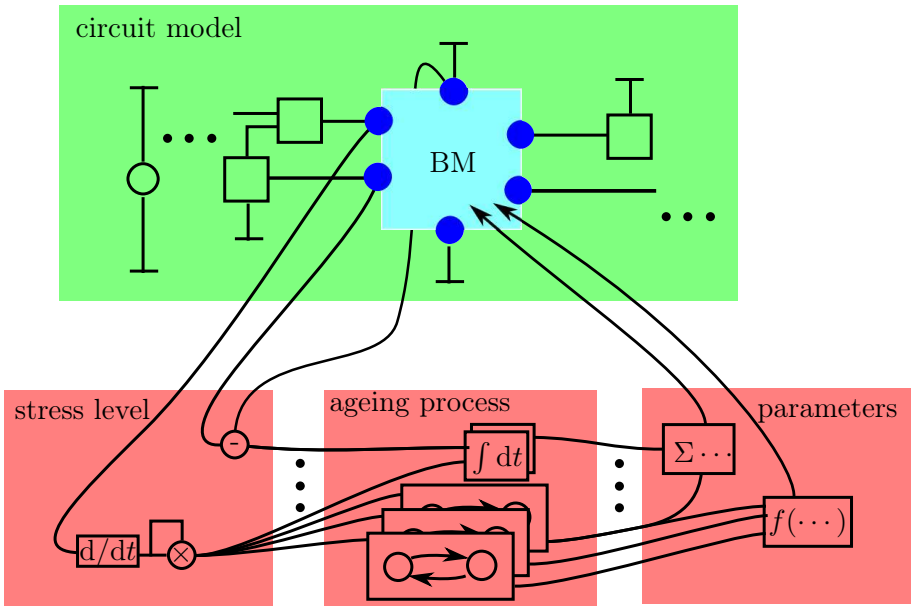


Figure 4.14: Behavioural model with embedded ageing effect.

be learnt and then used to speed up simulations of bigger circuits, possibly containing multiple instances of one and the same component that is operated under similar conditions. In contrast, the proposed ageing effect models based on ageing processes can be used to model performance shifts that turn transient waves into (ageing) parameters.

A transistor in a digital circuit has a few different operating modes (like *on*, *off*, *switching* or *powerdown*) and the behavioural impact manifests as a threshold voltage shift which is good enough to explain performance degradation. An analogue ageing effect model not only needs to support arbitrary modes of operation, that cause damage in various ways, and also must account parameter shifts that are specific to the current operating point. We restrict to nominal modelling, as this is most important for transient simulation purposes. Our focus is a faithful model for the state and the parameter deviations that is not limited to a fixed mission profile and permits fast evaluation.

We explain how to create and evaluate ageing models for behavioural parameters of analogue components. Our approach is based on data, and unrelated to physical mechanisms. An analogue component has ports, states and behavioural parameters. The ageing model is supposed to model the relation between transients and behavioural parameters. The model is based on nominal ageing processes from Section 4.2, these are controlled by stress levels as in Definition 4.5. Altogether, it is made of three stages (Fig. 4.1). These stages need to be addressed individually.

This ageing effect model extends the behavioural model and is intended to be simulated as a part thereof (Fig. 4.14). The circuit surrounding the behavioural model is – to some extent – arbitrary, and we need to choose test benches to study the age impact. We use one test bench where the circuit is exposed to stress, and one to compute its performances afterwards (Fig. 4.13). While the device is ageing, we record characteristics of the transients and the corresponding

outcome for the parameter measurements. These fill a table, one row per parameter extraction. The fitting algorithm (Section 4.4.4) looks for a compound process model realising this data. Note that to make the fitting work, the stress level test patterns need to be supported by the used ageing process cells (see Section 4.4.1).

One particularly crucial step is the choice of stress levels. In the case of BTI, we know that $|V_{\mathrm{gs}}|$ and $|V_{\mathrm{gd}}|$ are good candidates, and related to the physical cause of the effect. In general, a good choice is hard, and a proper understanding of the underlying mechanism helps. An automatic approach would fill the missing link to fully automated modelling. See Section 6.4 for example applications and data.

4.6 Conclusion

We have found a common denominator for different sorts of ageing effects in terms of level controlled ageing processes. Using this notion of an ageing process, we could describe and characterize stress dependent defect statistics as well as stress dependent nominal parameter ageing. We have identified a subclass of nominal ageing process models, that permits constant time evaluation on practically relevant stress waves. We have presented a two-staged learning algorithm that identifies ageing models with specific properties that fit to a given data set. These models do not impose restrictions to the operating mode, such as counters or periodicity, and can be used in analogue environments for analogue waveforms.

CHAPTER

FIVE

CIRCUIT SIMULATION

5.1 Problem Statement

5.1.1 Reliability Simulation

The simulation of a circuit with aged components makes ageing effects visible on circuit level. If aged component models are available they can be plugged into a circuit simulator. For quite some time, *reliability simulation* has been employed to analyse the impact of ageing effects on circuits [Tu+91]. How a circuit ages depends on the task it performs, and how it performs a task. The latter depends on its current ageing state. Thus reliability simulation often refers to *ageing simulation*, covering both sides, the simulation of the task and the simulation of the ageing. Until today, the state of the art is to run repeated ordinary transistor level *transient* simulations – sometimes *periodic steady state* analyses – to estimate the ageing impact of the task. This gives rise to a modification of the transistor parameters before the next simulation [Tu+93; Cad; MG13]. Some approaches take into account the transient stress on a transistor to compute its *age*. Early approaches assumed unchanged operation and stress conditions throughout the life of a transistor, which lead to inaccuracies (Section 5.5.2). While some aspects have improved, a proper solution for ageing simulation is still actively worked on.

Simulating ageing effects in analogue circuits requires both, efficient ageing component models and a simulator that is capable of ageing evaluation during long term transient simulation. The commercial approaches of the last decade are restricted to memoryless ageing effect models. With these, modelling the constitutive BTI effect in analogue circuits is hardly possible. Most attempts to incorporate the ageing impact on transistor level have been found to mispredict parameter shifts by factors up to 100 % [Sch+10]. Ageing effects in non-transistor components have hardly ever been addressed. In addition, the efficiency of the existing simulation approaches is unclear. Sometimes ageing parameters originate from accompanying physics based simulators [Gra+06], sometimes step control is not implemented and left to the user [Cad], often relevant details on model evaluation are missing [MG13]. See Section 5.2 for a more detailed view.

5.1.2 The Circuit Model

A circuit model is usually expressed as a **netlist**, a set of component models with interconnections. Modified nodal analysis (MNA), turns a circuit model into a time dependent differential equation (DAE) [VS94]. With a state $u_0 \in \mathbb{S}$ this becomes an initial value problem for a transient $u\colon \mathbb{T} \to \mathbb{S}$,

$$\begin{aligned} f\left(u(t), \frac{\partial u(t)}{\partial t}, \ldots, t\right) &= 0. \\ u(0) &= u_0. \end{aligned} \tag{5.1}$$

Adding ageing states, the DAE naturally extends to an initial value problem on $\mathbb{X} = \mathbb{S} \times \mathbb{A}$. However, the ageing effects we are interested in are time independent and stress level controlled. Hence we get a simpler DAE for a transient

$$x = (u, a)\colon \mathbb{T} \to \mathbb{S} \times \mathbb{A}.$$

This DAE splits into functions f_u and f_a,

$$\begin{aligned} x(0) &= x_0 \\ f_\mathrm{u}\left(u(t), \partial u(t)/\partial t, \ldots, t, a(t)\right) &= 0 \\ f_\mathrm{a}\left(u(t)\right) &= \partial a(t)/\partial t. \end{aligned} \tag{5.2}$$

The function f_a models the ageing processes. With $f_\mathrm{a} = 0$, we retain Eq. 5.1, the equation for the fresh device. In analogy to stress level functions and stress waves, we use transients and transient waves. For a transient $x\colon \mathbb{T} \to \mathbb{X}$ and a time $T \in \mathbb{T}$ we write $x_T\colon [0, T] \to \mathbb{X}$ for the **transient wave**, that equals x for times before T.

5.1.3 Repetitive Transients

Ageing effects tend to have larger time constants than usual circuit operations. Hence ageing effect simulation requires the knowledge of circuit operation over long times. Explicit computation of transients over long times is generally not a viable approach to ageing simulation.

Long term transients often involve repetitive patterns. For example, when simulating an oscillator, the *periodic steady state* is of interest. In ageing simulation, repetitive tasks are implied by the targeted range of time, since a typical use case of a device in general involves repetitive tasks even without oscillators. We characterize such transients with the following definitions.

Definition 5.1. *A transient (wave) x is **quasi-periodic** with period $\zeta > 0$ if, for admissible times t, the difference $|x(t) - x(t + \zeta)|$ is uniformly small. It is **locally quasi-periodic** if there exists a continuous function $\zeta\colon \mathbb{T} \to \mathbb{T}$ such that*

$$t \mapsto |x(t) - x(t + \zeta(t))|$$

is uniformly small. A transient wave is ***almost locally quasi-periodic****, if x is the concatenation of locally quasi-periodic transient waves $x_1 \odot x_2 \odot \cdots x_r$ for a reasonably small r.*

Note that the exact meaning of "small" is hard to specify. In general, "not so small" will increase the simulation run time accordingly. The applications presented in Chapter 6 raise evidence that the transients in practically relevant cases are exactly the almost locally quasi-periodic ones. In this chapter, we shall discuss algorithms that are intended to solve Eq. 5.2. We introduce an adaptive long term transient algorithm that focuses on the efficient computation of almost locally quasi-periodic transients.

5.2 A Survey

5.2.1 TCAD

TCAD (Technology CAD) simulation refers to simulating electronic circuits as a physical system involving silicon and dopants. This method gives the most accurate results amongst all practical simulation techniques. On this level it is possible to directly incorporate ageing effects by means of their impact on material properties or geometry. In [JS05], whole circuits have been simulated to demonstrate the impact of ageing on higher-level behavioural properties. This approach involves a long path from circuit simulation, extracting transient voltage levels, over reaction-diffusion simulation, computing interface trap densities, and back to circuit simulation after extracting semiconductor model parameters from a structural model. Clearly, simulation cost is immense and is the bottleneck of this approach. Simulation over a lifetime of a circuit does not look close to practical using these methods.

5.2.2 ATSF

In [Mar+06], the use of an Ageing Time Scale Factor (ATSF) is proposed to turn a conventional circuit simulator into an ageing simulator. The factor warps the ageing effect speeds relative to the circuit time. With this concept, arbitrary ageing effects can be modeled as parts of ordinary compact models and simulated in linear time, relative to lifetime. Linear run time is grossly suboptimal, and deeply entangled into the approach in question. From our experience it is close to impossible to formulate ageing processes, long term extrapolation, ageing time step control or just dynamic warping in a traditional hardware description language.

When it comes to recovery effects – which were not part of the elaboration – we face substantial accuracy problems. Consider an FET, modelled with BTI enabled, exposed to pulsed stress (Fig. 5.1 top). Warping the ageing time by a factor > 1, the falling edge will be simulated through a device that is much more damaged than in reality. The rising edge will face a much fresher device, as in the meantime, the transistor will have recovered more than it should have (Fig Fig. 5.1, bottom). A larger ATSF ($\pi \times 10^7$ in [Mar+06]), might even erroneously destroy the transistor during the first pulse.

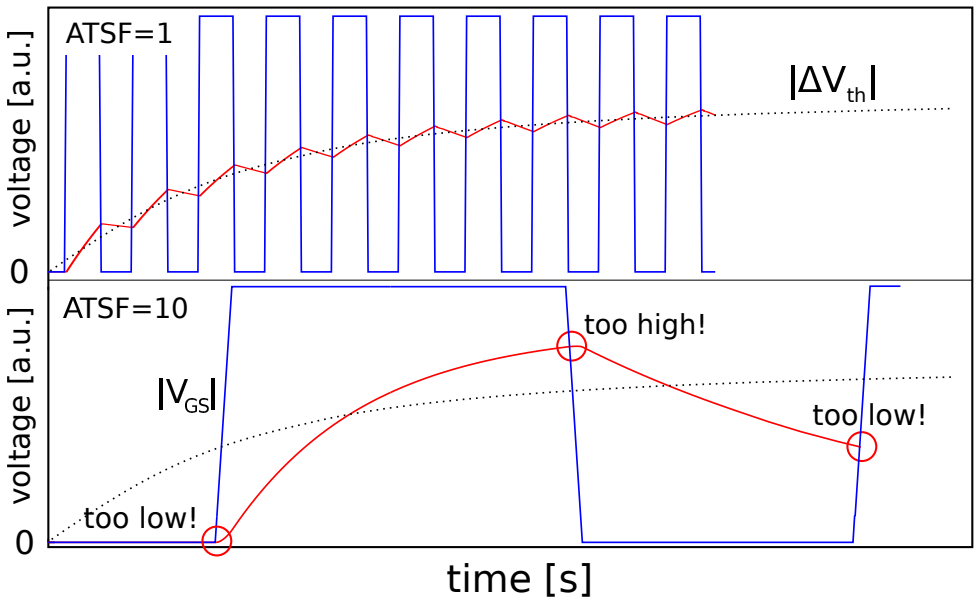


Figure 5.1: Scaled ageing time during BTI simulation. A warping factor unequal to one will lead to unpredictable simulation faults. Originally, ATSF was intended to be as huge as $\pi \times 10^7$.

A method similar to ATSF is explained in [MG13]. There, warping the time is used to speed up the evaluation of ageing effects, particularly the BTI effect, under periodic conditions. They argue that due to "frequency independence" [Ram+09] of the physical effect, their model is immune to the error in Fig. 5.1.

5.2.3 Multirate Simulation

Multirate simulation computes transients of oscillating systems [BB15; Pek+04; Wan+11]. We quickly explain the multirate method without getting too formal. An *oscillation period*, a smooth map $\zeta\colon \mathbb{T} \to \mathbb{T}_{>0}$, induces a smooth $\mathbb{Z}$ operation on the Cartesian product $\mathbb{T}^2 := \mathbb{T} \times \mathbb{T}$ defined by $z\colon (t, T) \mapsto (t + z \cdot \zeta(T), T)$. This operation induces an equivalence relation $\sim_\zeta$ on $\mathbb{T}^2$,

$$(T, t) \sim_\zeta (T', t') \text{ if } T = T' \text{ and } t - t' = z \cdot \zeta(T) \text{ for a } z \in \mathbb{Z}$$

and a sort of cylinder R with circumference $\zeta(T)$ at coordinate T (see Fig. 5.2). It represents the equivalence classes of the elements of $\mathbb{T}^2$.

With this, we may rephrase Eq. 5.1 as a partial differential equation $\hat{f}$ in $\hat{x}$, which becomes a function on R,

$$\hat{f}\left(\hat{x}, \frac{\partial \hat{x}}{\partial T}, \frac{\partial \hat{x}}{\partial t}\right) = 0.$$

The solution for Eq. 5.1 can be recovered from a solution $\hat{x}$. For this, denote by w the canonical map $\mathbb{T}^2 \to R$ induced by the above construction. This map "wraps" $\mathbb{T}^2$ around R, and also embeds the straight timeline into R by

$$t \mapsto (t, t) \mapsto w(t, t)$$

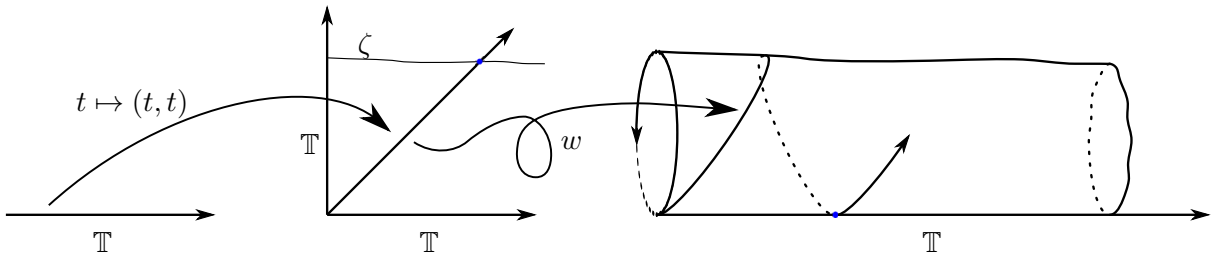


Figure 5.2: Multirate Domain Construction

and we retrieve $x(t)$ as $\hat{x}(w(t,t))$. Solving the multirate equation involves finding the solution on the **envelope** $\mathbb{T} \times \{0\} \subset R$.

Typical multirate applications can reduce simulation times for oscillating systems by factors of about three [Wan+11; Pek+04]. Here, an explicit integration method on the state space is applied to extrapolate between transients. We shall reach higher speedups due to the relative slow progression of the ageing effects.

5.2.4 Performance based Step Control

The long term ageing simulation approach presented in [MG13] involves a two-time evaluation scheme with a step control mechanism. This step control mechanism is based on a set of *circuit performance parameters* P and their respective errors defined as

$$\epsilon := \left| \frac{\mathrm{P} - \mathrm{P}_{\text{previous}}}{\min(\mathrm{P}, \mathrm{P}_{\text{previous}})} \right|.$$

The time step is then chosen by the *Step Size Algorithm* basically such that ϵ is not greater than a tolerance $\epsilon_{\max}$.

It seems hard to reproduce this step control mechanism. There are indefinitely many performance parameters and most of them cannot be observed from a simulated short transient. For one of them, the error ϵ depends on the representation of the performance parameter. On the one hand, we do not know in which way stress levels, which should be in charge of step control, qualify as performance parameters. For example, if a stress level changes or deviates from predictions, then step size should be decreased, as otherwise device ageing states will be miscomputed. On the other hand, these choices appear to be user input, and the approach probably works with the right choices, which we do not know.

5.2.5 Commercial Implementations.

Simulation tools (from Cadence and others) incorporate a sort of long term simulation. Here, circuit operation, and ageing effects are treated separately. Transient results turn a transistor parameter set into a (more) aged transistor parameter set. The user is free to repeat this step at different times of ageing and may investigate the properties of the so aged circuit. The actual inner workings of this implementation is opaque, but has been reported to consistently produce

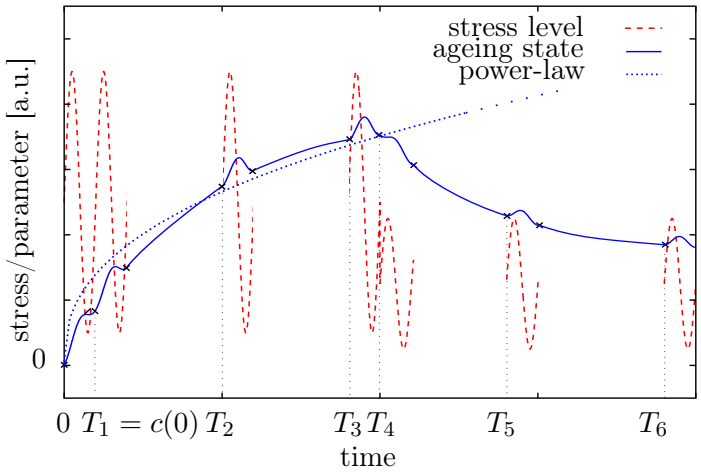


Figure 5.3: Two-times ageing simulation.

problems. If stress patterns (stress level functions) change during the lifetime of a circuit there are many ways to get aged device parameters wrong. See Section 5.5.2 for an early issue triggered by conceptual issues, or [MG13] for a detailed review of the used algorithms.

Some tools also provide an ageing effect modelling interface (URI, Unified Reliability Interface). However closer investigation reveals, that only memoryless ageing effects in MOSFET models are supported. Within this framework we were not able to mitigate time step related issues or model effects such as BTI, which is not memoryless by nature (c. f. Section 4.1.2).

Sure, commercial products sometimes improve over time. Yet the internals of the current tools and models are kept secret, and it is safe to expect that future versions will not reveal their internals. A standardised ageing component description could resolve this.

5.3 Two-Times Simulation

Ordinary multirate methods target systems described by generic differential equations such as Eq. 5.1. These do not differentiate between energy states and ageing states and thus cannot take advantage of the refined structure of the equations. For example Eq. 5.2 provides an ordinary differential equation for the ageing states. The dependency of the system model f_{m} on the ageing state a can be assumed to be smooth, and hardly nonlinear. In particular, the ageing state dependency induces a homotopy of transients for one ageing state to a transient for another, if they are not too far off. With this in mind, we formulate a specialized multirate algorithm, that we refer to as **two-times** simulation.

We compute transients of systems of type of Eq. 5.2. As in the multirate setting Fig. 5.2 We factor absolute time into $T + t$, where T is the circuit **age** and t is some sort of virtual local time, that we call the **transient time**. Transient simulations, **frames**, are executed at increasing ages T_i, these are of duration $\zeta(T_i)$

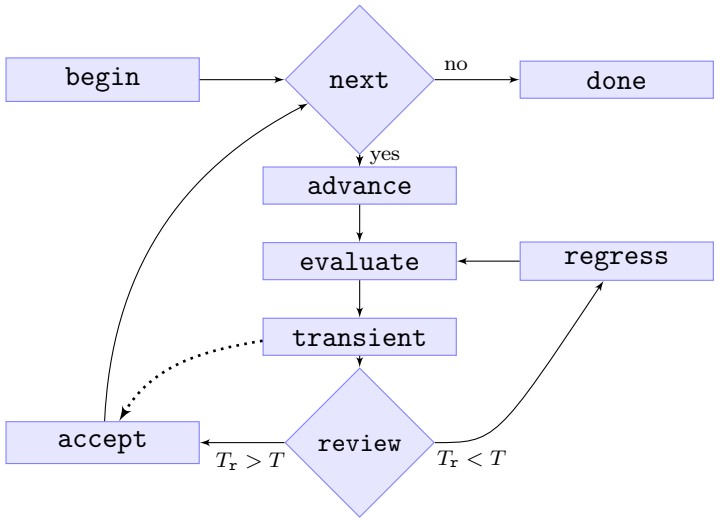


Figure 5.4: Adaptive ageing simulation algorithm, naming adopted from Gnucap transient simulation. Dotted bypass: conventional simulation, effectively.

The times are controlled by the respective `next` routines. The second-time T is fixed to the begin of the frame that is currently simulated. The ageing components collect stress levels at each transient step the simulator accepts. The last accept within a frame computes $\overline{L}$. The algorithm that controls the second-time is close to a transient simulation loop (Fig. 5.4). This involves a `review` procedure, where the components determine their individual maximum tolerable second-time step or schedule events. The discrepancy between the extrapolated average $\overline{L}_\mathrm{e}$ to $\overline{L}_z$, the average computed from the extrapolated ageing state, is relevant. In the second order case, $\overline{L}_\mathrm{e}$ is obtained using quadratic extrapolation, the `review` demands

$$T_\mathrm{r} \leq \mathrm{tol} / \left(\overline{L}_\mathrm{e}(T) - \overline{L}_z(T)\right)^{1/2} \tag{5.3}$$

for the entries L of L.

`begin` Initialize ageing simulation. Reset to fresh state. Alternatively, `restore` may be called instead to restore state from values left behind by previous two-time simulations.

`advance`, `regress` Update implementation specific variables such as internal times and extrapolation method or integration order for the current second time step. These functions are used exclusively, depending on whether T has grown (accepted step) or not (rejected step). Unchanged T may be used to control corrector methods.

`evaluate` Before a transient simulation is ran, a component computes its state by extrapolating from the data collected from the previous transient simulations.

`review` Just after a transient simulation, this function will be called to reconcile what has happened to the states related to ageing. For example an average stress level can be stored for

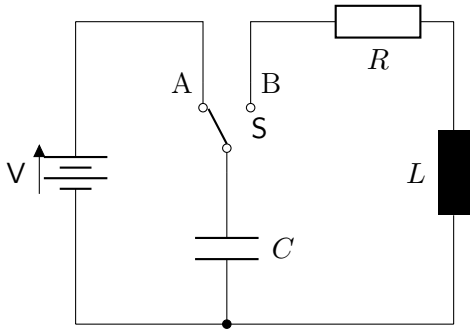


Figure 5.5: A simple oscillator.

extrapolation purposes. The current states may be corrected with respect to most recent data. Additionally, this function proposes a new second time to the simulator to control extrapolation errors and may register events to schedule transient simulations during expected changes in operation mode.

`accept` After a time step has passed review, this function is called. A component may now store data that will be considered history in subsequent steps.

Transient simulation mandates a `tr_accept` method. It is called whenever a transient time step is accepted. Here the ageing component collects stress information. Similarly, `tr_review` may be used to control errors originating from stress level processing. These functions were already defined within Gnucap [Dav03].

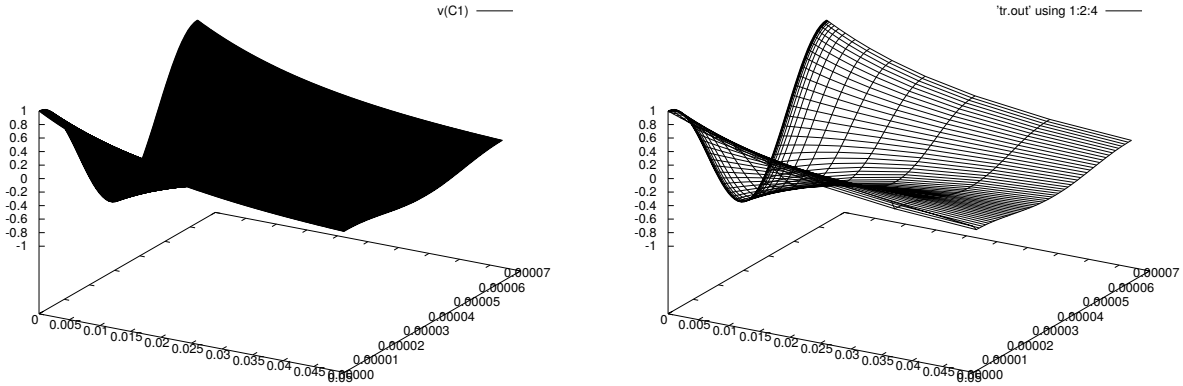
5.4 Quasi-periodic Simulation

5.4.1 Circuit State Extrapolation

Consider the oscillator made from a coil, a capacitor and a resistor as in Fig. 5.5. The transient after switching over S to B is quasi-periodic for period $\zeta = \sqrt{LC}$. Clearly, two neighboring periods are close to each other, see Fig. 5.6b. Under these conditions, the simulation time may be drastically reduced, as only the period width and the conditions at the start of each period are relevant. These conditions perform a long term drift, that we locally approximate using a tilted saturation curve such as

$$s\colon t \mapsto \alpha \cdot (1 - \exp(t/\tau)) + \beta t \tag{5.4}$$

parameterized by α, β and a positive τ. The parameters can be found using linear regression on the differential of s. The differential of s corresponds to the discrepancy of the initial conditions between periods. This is a very basic "multirate" strategy that certainly admits refinement. But it completely fits our needs w. r. t. ageing simulation.



(a) Full transient of LRC oscillator over two times.

(b) Bypassing transient evaluation.

Figure 5.6: Computing quasi periodic transients.

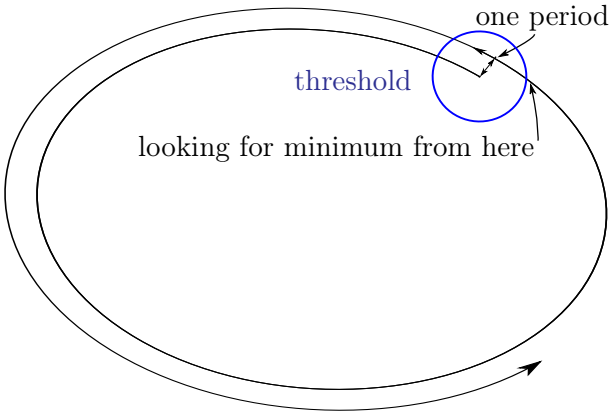


Figure 5.7: Period detection. One period is reached where the transient is closest to the initial point within a threshold.

5.4.2 Period Detection

The linear oscillator is particularly simple, as the period width is constant. In other cases, this gets more complex, particularly because the period width is unknown. The period width of a transient $x\colon \mathbb{T} \to \mathbb{S}$ can be detected using a suitable metric d on $\mathbb{X}$. We use a heuristic that finds a local minimum of $y\colon t \mapsto d(x(t), x(0))$ in a range where y is small (see Fig. 5.7).

For example, consider a voltage controlled oscillator (VCO) during an input transition as in Fig. 5.8. With period detection, a two-times simulation becomes possible. Extrapolating the capacitors states, we can skip 13994 out of 14070 transient simulations, see Fig. 5.9.

5.4.3 Evaluating Ageing Effects

Given a quasi-periodic stress level and an ageing process, we are interestested in the time evolution of the state of the process. The time evolution of the process state a during a transient

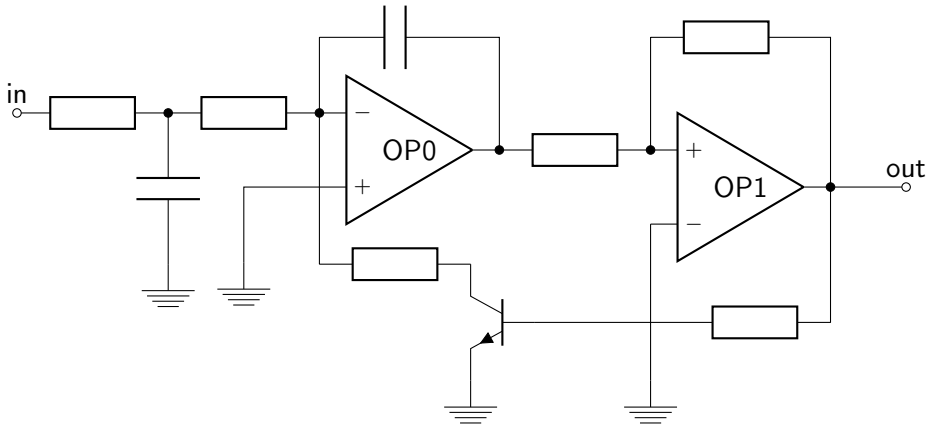
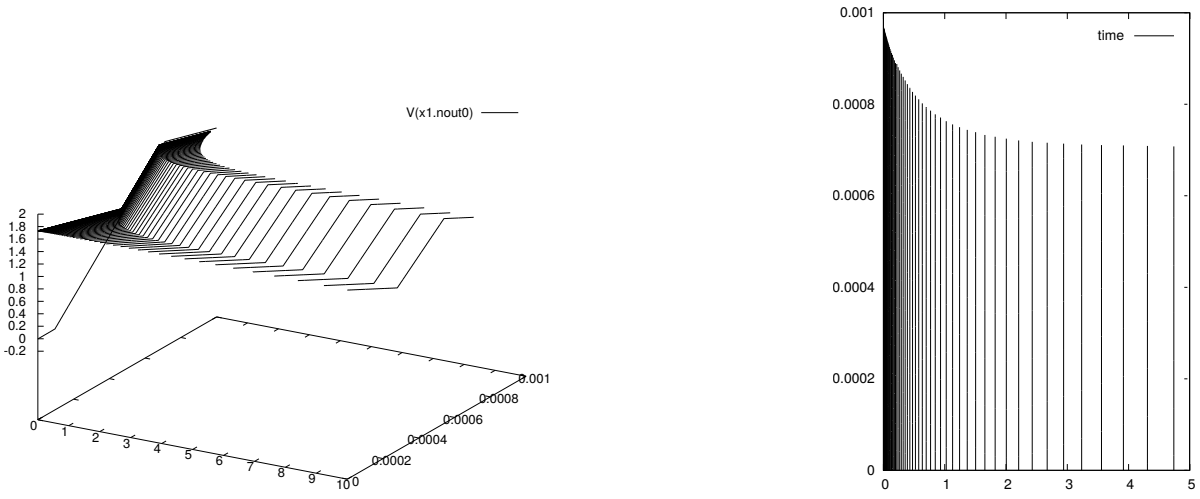


Figure 5.8: A simple VCO.



(a) Two-times simulation result.

(b) Top view.

Figure 5.9: Transient of VCO during frequency transition.

$x \in \mathbb{X}$ can be obtained by integrating the *ordinary differential equation* defined by f_{a} in Eq. 5.2. Here, in the spirit of Section 4.1, we highlight the stress level that controls the rate and obtain

$$f_{\mathrm{a}}\left(L(u(t))\right) = \partial a(t)/\partial t, \tag{5.5}$$

for a circuit model involving ageing effect models in the sense of Definition 4.5.

For integrating models such as

$$p(L_T) = p(Q_{\mathrm{I}}(L_T)) = p\left(\int_0^T L(t)\mathrm{d}t\right) \tag{5.6}$$

there is an *average stress level* that equals the average integrand

$$\overline{L}(T) = \int_T^{T+\zeta(T)} L(t)\mathrm{d}t/\zeta(T). \tag{5.7}$$

averaged over a frame. The level $\overline{L}$ is the constant stress level function that has the same duration and the same effect as the actual stresswave has. Here, we can easily verify that

$$p(Q_{\mathrm{I}}(L_{\zeta(T)})) = p(Q_{\mathrm{I}}(\overline{L}_\zeta(T))). \tag{5.8}$$

In this situation, we can extrapolate p for $T > T_0 + \zeta(T_0)$ from the current frame at T_0,

$$p(L_T) \approx p(\overline{L}(T_0)_{T-T_0} \odot L_{T_0}). \tag{5.9}$$

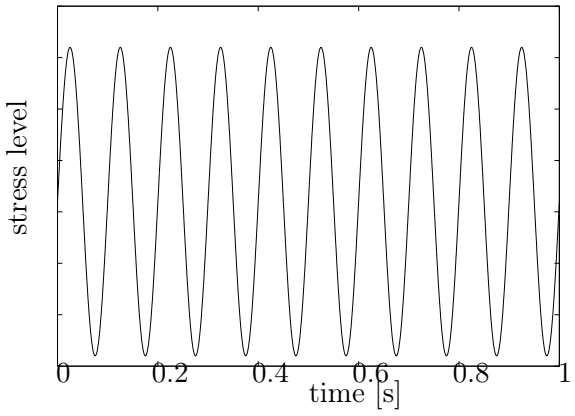
In the previous chapter, we prove that RCD based ageing processes permit a *locally equivalent stress* that has the same property, see Section 4.3.5 and Theorem 4.26. We can efficiently compute this level using a binary search on the evaluation function Eq. 4.36.

Given an RCD based ageing process Q that permits an average, and a transient wave x, we may in turn compute $a = Q(L(x))$, the average stress C and then extrapolate the state of a on the envelope as $Q_a(C_{\delta t})$. However, now, the locally equivalent constant stress is only a local average and depends on the ageing state. For example, controlling an ERCD with a sine wave results in an increasing ageing state and a decreasing local average. Using a fitting function, we may approximate and extrapolate the locally equivalent stress and still use it to extrapolate the ageing state. in the RCD case we can approximate the extrapolated stress by step functions and compute the RCD state using the ordinary trajectory Eq. 4.30. This is still much more efficient than computing the ageing state from a periodic sine wave.

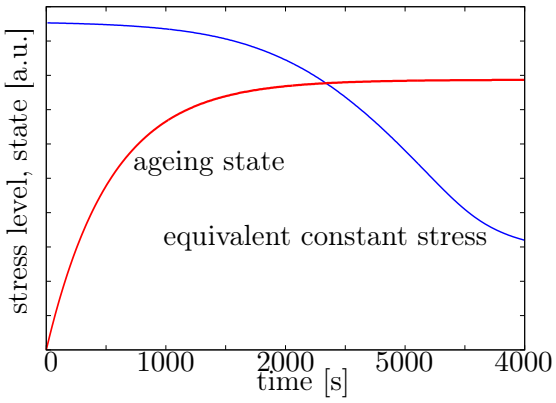
5.5 Step Control

5.5.1 Average Stress

An estimate for the extrapolation error gives rise to step control. Assume that the transient simulator computes an average stress level $\overline{L}(0)$ of an integrating model with an (absolute) accuracy of ε_L. Then, integrating this average level to some time T implies an intrinsic error



(a) A periodic sine stress wave.



(b) The locally equivalent constant stress depends on the ageing state and the stress level.

Figure 5.10: Locally equivalent constant stress

$\varepsilon_L \cdot T$. With this approximation, we may compute the aged parameters and compute an average stress level $\overline{L}(T)$ from a simulation starting at time T. The deviation

$$\varepsilon_{\int} = T \cdot \left(\overline{L}(0) - \overline{L}(T)\right)/2 \tag{5.10}$$

is the *truncation error* of the *explicit integration* method. Choosing time steps to limit this error, as well as higher order integration is business as usual. We are interested in how this scheme transfers to nontrivial ageing processes, such as RCDs. Averaging the stress levels for integrators is straightforward. In particular, the average is constant throughout all simulations, if the operation is periodic. For RCDs we get a weaker stability property.

Definition 5.2. *Choose a time $T \in \mathbb{T}$, let f be a map defined on functions $[0,T] \to \mathbb{R}$ and let L be a function on $\mathbb{T}$. For a time $x \in \mathbb{R} \cdot \mathrm{s}$, denote by $s_x(L)$ the function L shifted by x, i. e.*

$$s_x(L) \colon t \mapsto L(x+t).$$

*The function f is **stable** on L, if the sequence*

$$n \mapsto f\left((s_{nT}(L))_T\right)$$

converges.

As an example, let e_0 be the map $L \mapsto L(0)$. Then e_0 is stable on L, if L converges. The following lemma excludes crossings of trajectories. It should be included with every course on ordinary differential equations. We give an elementary proof.

Lemma 5.3. *Let r be an RCD. For ageing states $a > b$ and bounded stress waves L, we have $r_a(L) > r_b(L)$.*

Proof. Set $\gamma_a \colon t \mapsto r_a(L_t)$, the trajectory of r under L starting at a. Let $\Gamma = \gamma_a - \gamma_b$, we have $\Gamma(0) = a - b > 0$ and

$$\Gamma(t) = a - b + \int_0^t \frac{\partial \Gamma}{\partial t} \mathrm{d}t' \tag{5.11}$$

$$= a - b - \int \frac{\Gamma(t')}{\tau(L(t))}. \tag{5.12}$$

Let $\tau_{\min} = \min_{t \in \operatorname{dom} L} \tau(L(t))$, and let $\Gamma_{\min}$ be the solution of

$$\Gamma_{\min}(t) = a - b - \int_0^t \frac{\Gamma_{\min}(t')}{\tau_{\min}} \mathrm{d}t', \tag{5.13}$$

particularly $\Gamma_{\min}(t) > 0$. Assume that $U := \{t \,|\, \Gamma(t) < \Gamma_{\min}(t)\}$ is nonempty. Choose t_0 maximal, such that $[0, t_0) \subset \mathbb{T} \setminus U$. Both Γ and $\Gamma_{\min}$ are continuous, hence $t_0 \notin U$. We get

$$\left.\frac{\partial(\Gamma - \Gamma_{\min})}{\partial t}\right|_{t_0} = \Gamma_{\min}/\tau_{\min} - \Gamma/\tau(L(t_0)) \tag{5.14}$$

$$\geq \Gamma(t_0)/(\tau_{\min} - 1/\tau) \geq 0. \tag{5.15}$$

Thus, there exists $t_1 > t_0$ s. th.

$$\Gamma(t_1) \geq \Gamma_{\min}(t_1) \text{ on } [t_0, t_1). \tag{5.16}$$

A contradiction to the construction of U. After all $\Gamma > 0$ at least until T_L, the duration of L, and the assertion follows. □

Using this lemma, we are able to prove the following on locally equivalent constant stress.

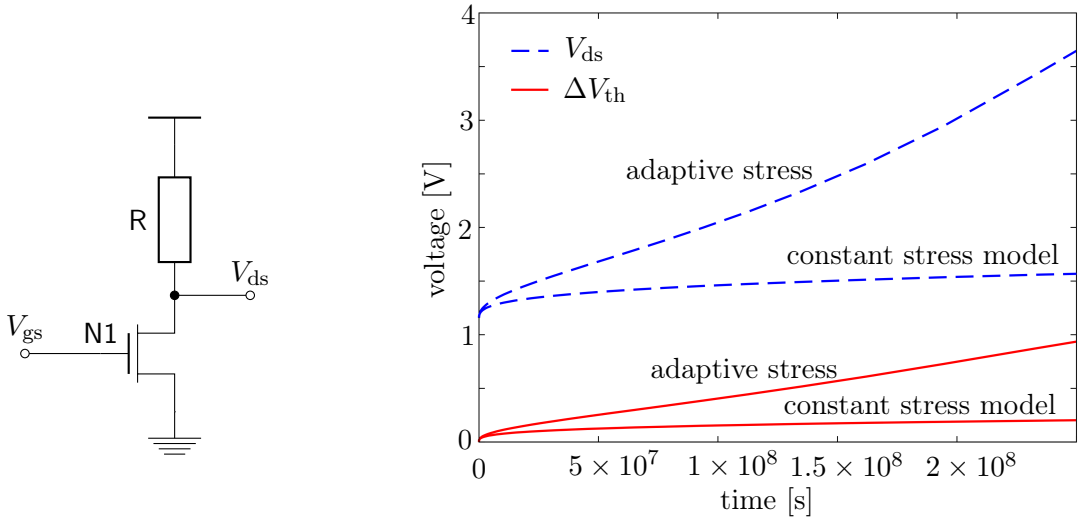
Theorem 5.4. *Let r be an RCD with continuous time constants and let L be a stress wave. The locally equivalent constant stress of r is stable on L.*

Proof. From Section 4.1.3 recall L^1, the periodic continuation of L from Section 4.1.3. Consider the sequence

$$n \mapsto r((L^1)_{n \cdot \mathrm{T}_L}.$$

This sequence is bounded by ± 1. According to Lemma 5.3 the sequence is also monotonic, hence it converges to ℓ_L. Due to continuity, the equivalent constant stress c of L in r at ℓ_L must fix this limit and we get $\ell_r(c) = \ell_L$. The limit state function ℓ_r of r is continuous by premise, and hence the equivalent constant stress converges to c. □

With this, the locally equivalent constant stress $\overline{L}$ has been identified as well-behaved on periodic signals. Thus, during a long term simulation we may use polynomials to extrapolate the evolution of $\overline{L}$. The discrepancy between the predicted $\overline{L}_{\mathrm{p}}$ and the computed $\overline{L}$ is caused by ε_L, the accuracy of the transient simulator, the deviation from periodicity ε_P and the error of the polynomial fit ε_F. The integral between the assumed stress level (that has been used to



(a) n-FET in common source mode. More damage increases V_{ds}, which increases the stress level.

(b) Simulation requires adaption to stress level.

Figure 5.11: HCI can amplify itself.

extrapolate the ageing process) and the a posteriori fit is a good measure for the simulator error. at extrapolation order k, this error scales with $\left|\overline{L}_p - \overline{L}\right| \cdot \partial^{k+1}\overline{L}/\partial^{k+1}T^{k+1}$, and we get

$$\Delta T < \varepsilon_{tol,k} \cdot \frac{\partial^{k+1}\overline{L}}{\partial^{k+1}} \cdot \left|\overline{L}_p - \overline{L}\right|^{1/(k+1)} \tag{5.17}$$

for the step control. In particular, if the stress level deviates from periodic unprecedentedly, this may lead to a change in $\overline{L}$. In that case the error is much bigger and Eq. 5.17 mandates smaller time steps. Conversely, if a deviation from periodicity does not affect $\overline{L}$ – for example a phase change – the ageing processes are unaffected, and there is no need to reduce time steps.

5.5.2 HCI and step control

An n-FET in a common source setup Fig. 5.11a suffers from HCI. With the gate voltage V_{gs} a little above the threshold, the potential at V_{ds} is low, and so is the stress level associated to HCI. As the time flies, the threshold of N1 will rise due to degradation, with a negative impact on its conductivity. This in turn will increase the potential at the drain and increase the stress level. After all, the degradation speed increases during operation. This came as a surprise for early reliability simulators, who captured the stress levels at the initial operating point and did not update them later on.

This example illustrates that a model for the stress level is advantageous, and that the stress level needs to be taken into account for the step control algorithm. In [SH15] we compare simulator run times and precision for different stepping methods wrt. ageing time.

5.5.3 Changing Operating Modes

Using the locally equivalent stress extrapolation and step control, we can control the error of $\overline{L}$. For this, we need to reject simulation results, if the step control mandates a smaller step. Sometimes, a circuit changes its operating mode during long term simulation, or more formally, the transient solution is not locally quasi periodic. Assume that we have two locally quasi-periodic transient waves x_1 and x_2 and the wave $x_2 \circ x_1$ is the transient of a circuit model we want to compute. Assuming further that x_1 and x_2 are sufficiently different, so we need to compute a frame close to $T_\mathrm{s} = \mathrm{T}_{x_1}$. A simulation based on the step control scheme Eq. 5.17 will compute several transients before T_s. The first attempt on a transient after T_s will result in a large error, as the predicted stress level $\overline{L}_\mathrm{p}$ does not coincide with the computed $\overline{L}_\mathrm{e}$. The simulation step must be repeated at a smaller time T'. However, T_s and T' are unrelated and we can only say that the above might repeat arbitrarily (but finitely) many times.

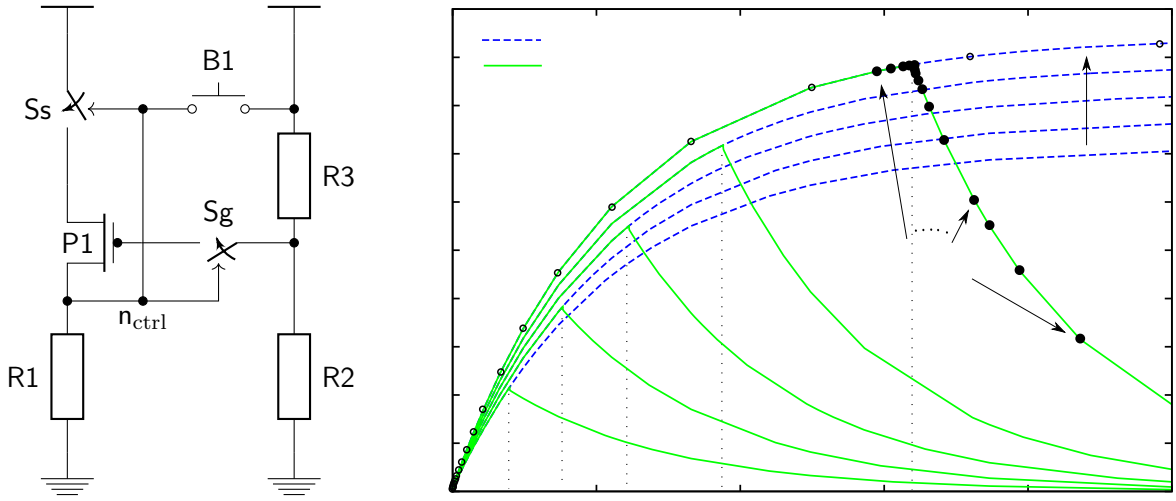
This issue can be partly amended using *behavioural models* for switches. A switch can keep track of the upper and lower bounds of its input signal during a transient frame. With an extrapolation such as Eq. 5.4, it can predict an estimate for T', and schedule a transient simulation just before T' and reduces the number of discarded transient simulations. A similar technique has been used before to reduce rejects in transient simulations [Dav03]. To our knowledge, nothing like this has been applied to ageing simulation, our approach has been used in [SH15].

The **Halting Problem** circuit in Fig. 5.12a Illustrates this approach. The switches Sg and Ss are voltage controlled, are on at high voltage and have the same threshold. Pushing button B1 enables P1. P1 and R1 are chosen, such that ageing causes the voltage at n_ctrl drift below the switching threshold, shutting down the circuit. The time of shutdown depends on the degradation state of P1. Its ageing state depends on the stress history. We activate B1 once, at time 0. During the simulation P1 switches itself off, the circuit *halts*. The halting time depends on the supply voltage. A higher supply voltage causes the transistor to age faster, but then also more damage is required to reach the threshold of Sl and Sg.

5.6 Behavioural Modelling

The inner workings of an ageing process is beyond the expressive power of widespread hardware description languages. We intend to not compute aged component models individually or set up models for each ageing state they might be in. Rather an ageing effect model is supposed to be part of a component model. Usually, a simulator reads in component models, for our purpose these models include ageing effects.

We add the semantics of ageing effects into behavioural modelling languages by means of subdevices and a specially crafted node type (see Fig. 5.13a). The ageing state then corresponds to a potential, and the stress level is communicated as a flow. The computation of the state from the level is done by ageing process models that look like ordinary component models. This exploits the structure of the ageing effect model from Definition 4.4. The stress level L as well as the parameter model p can be easily expressed as an arithmetic expression assignment in



(a) The voltage at n_{ctrl} controls the switches Sg and Ss.

(b) $|\Delta V_{th}|$ in P1 varies over time, and causes a shutdown. Circles mark the computed second-time steps, those at filled circles are introduced by the shutdown.

Figure 5.12: The Halting Problem.

any language. Just the ageing process Q needs to be taken care of separately. In particular, this representation expands to Eq. 5.5, the precondition for the simulation algorithm. From the examination of ageing processes in Chapter 4 there is evidence that only a few ageing process models are sufficient for modelling the relevant ageing effects. We have good faith that linear combinations of integrators and ERCDs are sufficient to model the frequency independent processes.

The computation of stress levels from transient states or parameter shifts from states x_a does not require language extensions. To represent stress `Level` and ageing `State`, we stretch the notion of conservative signal flow, defining natures with the corresponding accesses (see Fig. 5.13a). Model compiler and simulator need to be extended to cope with nodes of the so defined discipline. A behavioural model then instanciates and connects subdevices for the ageing processes as in Fig. 5.13b. The component models for the subdevices are hand-coded and implement the ageing processes with an interface matching the simulator algorithm described in Section 5.3.

5.7 Simulator Output Assessment

Different approaches to simulation often yield different results, at least in a numerical sense. A comparison of run times that does not take into account the numerical precision of the simulator results is insignificant. The trade-off between precision and run time is hard to establish and difficult to compare with a different trade-off. A simplified comparison of run times is still feasible if a useful valuation of precision is available. Comparing transients with reference transients gives rise to such a valuation. A similar approach is presented in [MG13]. Here a reference simulation

```
nature Damage
  units = "1";
  access = State;
endnature
nature Stress
  units = "1";
  access = Level;
endnature
discipline degradational
  potential Damage;
  flow Stress;
enddiscipline
```

(a) Disciplines used to model ageing effects. Stress levels and states are unitless.

```
module ageing_component(a, b, c);
  electrical a, b, c;
  degradational d1, d2;
  [ parameters, variables, functions .. ]
  integrator int0(d1);
  [..]
  ercd #(..) rcd0(d2);
  analog begin
    param = f_p(State(int0),
          + [..] + State(rcd0));
    I(a,b) <+ f_bm(param,V(a,b),V(b,c));
    Level(int0) <+ f_L1(V(a), V(b), V(c));
    [..]
    Level(rcd0) <+ f_L2(V(a), V(b), V(c));
  end
endmodule
```

(b) Verilog-A module with ageing effect submodules. Here, an `integrator` and an `ercd` are connected to evaluate the ageing processes.

Figure 5.13: Semantics for ageing effect models in behavioural modelling application.

is used to compute the *normalized mean square error*, NMSE. It remains unclear how a statistical error can assess the quality of a nominal data set, hence we used a much simpler metric in [SH15]. Similar metrics have been discussed in [Ohl+08].

A **metric space** M is a set with a symmetric map $d_M \colon M \times M \to \mathbb{R}_0^+$ that complies with the *triangle inequality*. From [Hau14] we take the definition of an asymmetric distance in the space of subsets of a metric space.

Definition 5.5. *Let M be a metric space. Let A and B be subsets of M. The* ***directed Hausdorff distance*** *is*

$$\operatorname{dhd}(A, B) := \max_{a \in B} \min_{b \in B} d(a, b). \tag{5.18}$$

Let f be a sampled real valued function on a bounded subset of the real numbers. The sampling of f gives rise to the set of sample points $S_f = \{(x, f(x)) \mid f \text{ sampled at } x\} \subset \mathbb{R}^2$. The space $\mathbb{R}^2$ is a metric space, the euclidean metric is a good choice. Assume S_f has nonzero height and width. An affine transformation A_f scales and translates S_f into the unit square such that it intersects with all its edges.

Definition 5.6. *Let g be another function. We define the* normalized directed discrete Hausdorff-distance $\operatorname{ddhd}(f, g)$ *to be the directed Hausdorff distance* $\operatorname{dhd}(A_f(S_f), A_g(S_g))$.

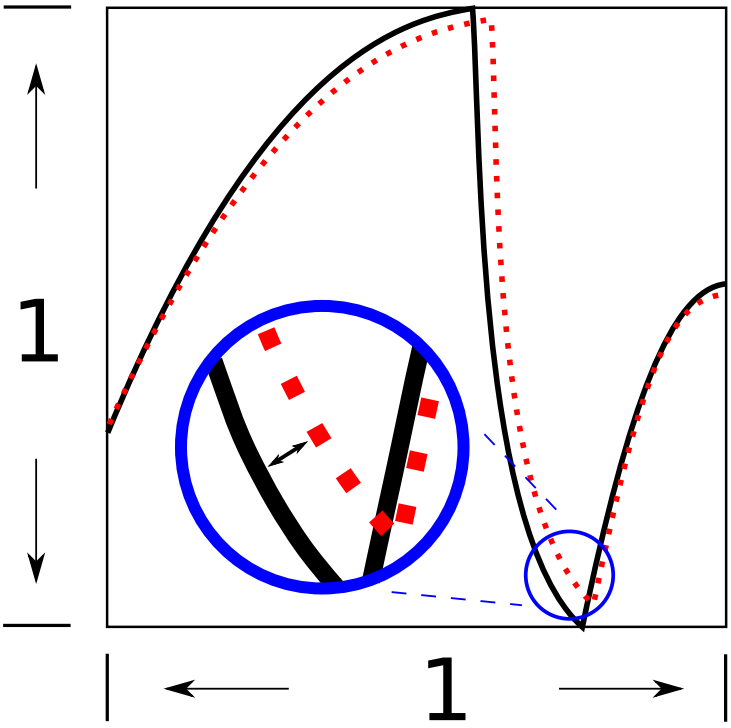


Figure 5.14: Normalized Discrete directed Hausdorff distance measures accuracy of a sampled waveform.

Less formal, $\mathrm{ddhd}(f, g)$ is the distance of the sampling point of f, that is furthest from g, to g – after rescaling and moving f and g to fit the unit square. This defines a measure that is suited to compare simulation results. Transients are considered close to each other not only if they are close pointwise, but still if they are (moderately) delayed relative to each other. The motivation for the rescaling is the otherwise incompatible units.

An efficient scan-line algorithm computes ddhd, making use of the totally ordered structure of the sample point data. For comparison, there's an algorithm that computes the directed Hausdorff distance for sets of n and m disjoint line segments [Alt09] in $O(n+m)$. In fact it can be adopted to work with piecewise linear functions. However, if the function samples are not meant to represent a piecewise linear function, this does not help.

After all, we chose the ddhd to an excessively fine sampled precise simulation result as a measure for the quality. With this we may compare run times – to some extent. A simulator or algorithm that produces a transient result with no lower quality in shorter time than another is considered faster. With this, we can objectify run times and simulation results for variants of step control schemes in ageing simulation [SH15].

We compare results obtained using different sampling strategies with an excessively finely sampled reference. Adaptive sampling is the proposed strategy. Equidistant and exponentially growing time steps are chosen to produce comparable results. Table 5.1 lists run times (on a 2.2 Ghz CPU) and $\mathrm{ddhd}(p_s, p_{\mathrm{fine}})$ for probes p and strategies s. The common source n-FET

Table 5.1: Run times and directed discrete Hausdorff distances to fine sampling of the example circuits.

Sampling	Common Src. n-FET			Halting Circuit			Monitored Amp			Redundancy		
Reference	4.3 h	ΔV_{th}	V_{gs}	16 h	ΔV_{th}	$V(\mathsf{nc})$	23 h	ΔV_{th} (N2)	$V(\mathsf{nc})$	14h	V(nc1)	V(nc2)
Adaptive	5 ms	0.42 %	0.18 %	1.7 s	0.01 %	0.1 %	44.3 s	0.41 %	1.2 %	7.5 min	0.24 %	0.18 %
Equidistant	7 ms	2.95 %	1.70 %	1.5 s	1.3 %	4.5 %	16.8 min	0.23 %	3.3 %	3.2 h	4.9 %	8.3 %
Exponential	11 ms	13.96 %	7.99 %	1.6 s	4.9 %	6.3 %	50.4 s	4.85 %	10.2 %	N/A	N/A	N/A

and the halting circuit have been discussed in this chapter, the *monitored amplifier* and the *embedded redundancy circuit* are part of the next chapter. The proposed adaptive sampling strategy reduces run times significantly and increases accuracy, in particular the accuracy of the ageing parameters.

5.8 Outlook

As the influence of ageing effects is growing with decreasing structure size, accurate nominal simulation and modelling techniques are inevitable. This involves proper algorithms such as two-times simulation of arbitrary ageing effects as well as integration of suitable effect models into existing modelling infrastructure such as hardware description languages. In this chapter, we have deployed the common denominator of ageing effects, behavioural modelling and circuit simulation. This becomes possible with the proper efficient ageing effect models. On top of this, we have developed a stress level aware adaptive step control scheme to simulate circuits involving ageing effects.

The efficiency of this approach is due to the restriction on ageing processes. Thus, the approach can only work for models with stable average stress levels. The integrator, as the prototypical ageing process, evidently fits into this category, and integrators have already be used to model and simulate ageing effects. However, most ageing effects are more complex. With the ERCD process we manage to build a bridge from data sets describing ageing parameters to efficient simulation. In the next chapter we shall simulate real world circuit examples.

CHAPTER

SIX

APPLICATIONS

6.1 Ageing Effect Models

In this section we model ageing effects using parameter models and ageing processes. These are meant to be used for analogue simulation purposes. The usual unidirectional HCI model does not necessarily work in analogue applications. We extend this model for our purposes. Overlaying unidirectional models is reasonable, since the defect locations can be considered as independent. The BTI parameter case involves a defect spread over the channel. This makes the observable parameter shift strongly operating point dependent. We propose a model for an effective parameter dependency based on local damage that we deduce from corner case ordinary BTI models with fixed operating point. Also we create a BTI process based on decay processes that extends the digital model. Altogether, we describe the first HCI and BTI effect models that work under arbitrary analogue conditions and incorporate a realistic impact on behavioural parameters.

Thanks to the separation into levels, processes and parameters introduced in Section 4.1, the resulting ageing effect models are representable in a hardware description language. The goal of this section is a transistor compact model with integrated ageing effects, such as Fig. 6.1.

6.1.1 An Analogue HCI Model

HCI degradation affects the threshold voltage of a transistor device. Commonly available models reduce the effect to just the ΔV_{th}, assuming that the transistor is operated under ideal conditions Eq. 2.1. These conditions reflect the behaviour in digital circuits, where ΔV_{th} is the most relevant parameter.

The underlying conceptual problem of HCI in analogue operation is outlined in [Cho12]. Here an n-FET has been stressed so electrons have been injected at the drain side. Measurements then indicate highly asymmetric device characteristics. Switching drain and source after the stress phase enlarges the HCI impact, as the damage has more effect on the source side. The observation suggests that the threshold voltage deviation should be modeled as a function of a

```
module ageing_mosfet(d, g, s, b);
  electrical d, g, s, b;
  degradational hci, bti0, .. btik;
  [ parameters, variables, functions .. ]
  tt_int hci_integrator(hci);
  rcd_exp #(..) rcd0(bti0), .., #(..) rcdk(btik);
  analog begin
    vth = ..;
    vth *= ( 1 + State(hci)**hci_n
               + State(bti0) + .. + State(btik));
    I(d,s) <+ K * (V(g,s) - vth)**2;
    Level(hci) <+ hci_fct(V(d),V(g),V(s),V(b));
    Level(bti0) <+ V(g,s);
    ..
    Level(btik) <+ V(g,s);
  end
endmodule
```

Figure 6.1: Verilog-A implementation of an FET affected by BTI and HCI, simplified for readability. The subdevices used for integration (`tt_int`) and decay process evaluation (`rcd_exp`) are not specific to defect mechanisms.

defect state and the current operating point. However, [Cho12] does not elaborate on how such a function might look like. Instead empirical worst-case models for constant ΔV_{th} and $\Delta_{\text{rel}} I_{\text{ds}}$ are declared precise enough and even ready to use as part of some commercial software.

We suspect that the empirical static defect model is good enough for applications where damage is restricted to just one side of the channel, and wherever reverse currents are impossible. Empirical static models can also incorporate a proportion of reverse stress impact, but this fixes the models to the chosen special use case. In case these workarounds are not adequate or feasible, we provide a flexible and symmetric model implementing operating-point dependency.

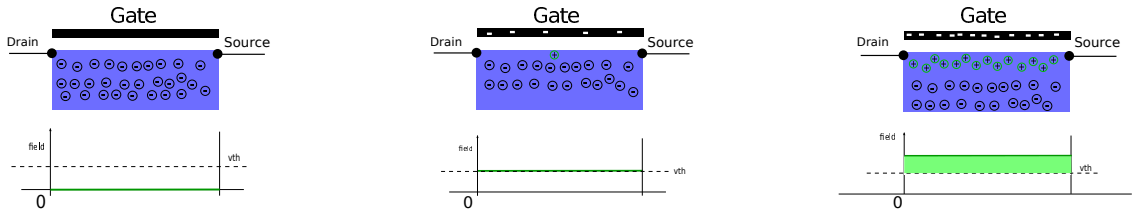
A first step to a parameter model is choosing a defect state. The defect state for the positive polarity model from Eq. 4.16 is

$$a_{\text{hci}} := \int_0^t \frac{(I_{\text{sb}}/\text{A})^m}{(I_{\text{ds}}/\text{A})^{m-1}}. \tag{6.1}$$

This damage is known to be physically located at the drain. Hence it makes sense to extend to a two dimensional damage state that also includes the defect at the source. We define the defect state $a \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ to

$$a = (a_+, a_-) := \left(\int_0^t \Theta(p(t)) \cdot \frac{(I_{\text{sb}}/\text{A})^m}{(I_{\text{ds}}/\text{A})^{m-1}}, \int_0^t \Theta(p(t)) \cdot \frac{(I_{\text{db}}/\text{A})^m}{(I_{\text{sd}}/\text{A})^{m-1}} \right),$$

where $p \in \{\pm 1\}$ denotes the polarity of the operating point and Θ denotes the Heaviside function on $\mathbb{R}$ that evaluates to 1 if the argument is positive and to 0 otherwise. The threshold voltage

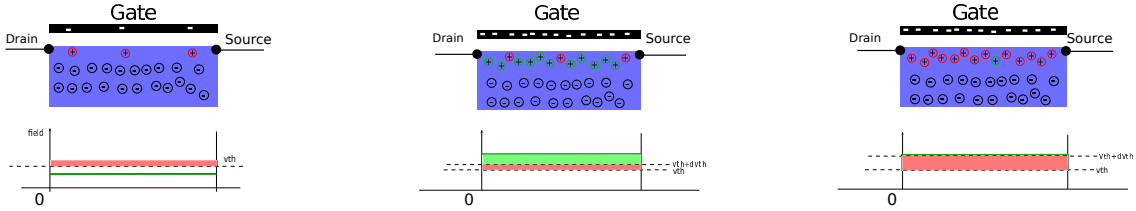


(a) No BTI, transistor off.

(b) Transitor at the threshold.

(c) BTI interferes with conductivity.

Figure 6.2: Simplified pMOS transistor. The gate voltage controls the electric field, that activates carriers. Here, drain and source are shortened, hence the field is homogeneous.



(a) For gate voltages below the threshold, BTI does not make any difference.

(b) Charge traps cancel out carriers.

(c) The resulting threshold voltage corresponds to the number of traps.

Figure 6.3: BTI traps catch carriers. Reducing the electric field has approximately the same effect.

shift is proportional to a_+^n in the positive case, and the reverse threshold is about $\gamma \cdot a_+^n$ for a $\gamma \geq 1$ after positive stress. Assuming that both impacts on ΔV_{th} add up, we get a total ΔV_{th} proportional to $a_+^n + \gamma \cdot a_-^n$ during forward operation and $\gamma \cdot a_+^n + a_-^n$ during reverse operation. Clearly, if the transistor is only operated in one direction, one of the states is degenerate and we retain the behaviour of the original model.

Such a model is simple to calibrate, once measurements of source adjacent damage impact is available, and if the hypothesis on additivity holds. We continue without these measurements and choose an arbitrary value for $\gamma \geq 1$. Nevertheless, we get an appropriate and symmetric HCI model for use in analogue circuits.

6.1.2 Physical Background of Digital BTI Model

The parameter shift induced by BTI can be explained using a simplified physical transistor model (c. f. [Sze81; Tsi99] for a fully detailed view). Basically, a field effect transistor is a capacitor with one semiconductor electrode. Consider the P-channel FET cross sections in Fig. 6.2. The blue region is a negatively doped semiconductor. Negative dopants move away from the top electrode upon increasing the electrical field, i. e. upon placing electrons to the top electrode (Fig. 6.2b). A higher electrical field causes a layer of mobile positive charge carriers (Fig. 6.2c). The conductivity of the channel is roughly proportional to the number of these carriers. The

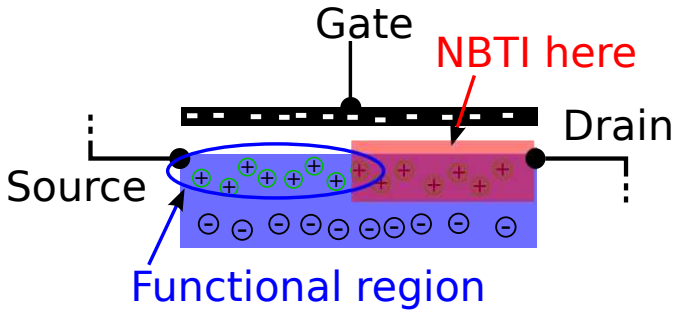


Figure 6.4: A p-FET in corner case ageing state and forward operation.

(negative) BTI effect installs positive dopants (**charge traps**) to the interface. These act in a way similar to the mobile carriers – except they can not move (Fig. 6.3). Then, if the electrical field required to maintain a first mobile positive charge is higher (Fig. 6.3c), the threshold voltage of the transistor has increased.

Counting charge carriers is discrete and cumbersome, and a continuous view is more practical, see Fig. 6.2. Here the electrical field is abstracted to the effective corresponding voltage, the threshold voltage V_{th} is a constant, and the conductivity corresponds to the area between the effective voltage and $V_{\text{th}} + \Delta V_{\text{th}}$. This view will turn out to be useful in Section 6.1.3.

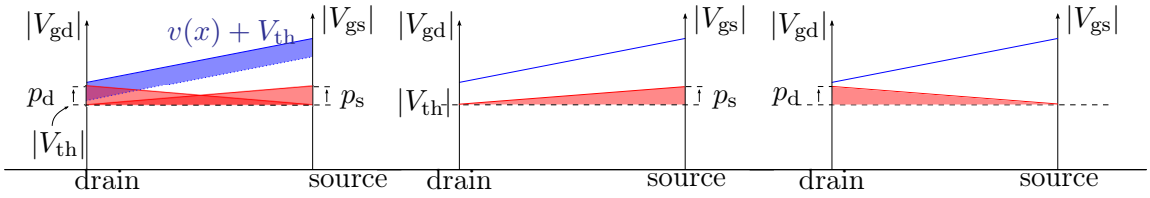
6.1.3 An Analogue BTI Parameter Model

Various existing BTI models are not suited for ageing simulation, particularly not for ageing simulation of analogue circuits. We will establish a list of criteria for a BTI process model Section 6.1.4 and reconcile the properties.

The voltage V_{gs} between gate and source is correlated to the homogeneous electrical field through the oxide – if the transistor is part of a digital circuit. In digital circuits the set of operating points is limited, and a threshold voltage shift ΔV_{th} gives a good characterization of behaviour change. This is due to the fact that in common source configuration (Fig. 2.2b), charge traps close to the source are the most influential. This is where the electric field is highest at the moment the transistor is switched on. In analogue circuits, the field is not as homogeneous as in Fig. 2.2b, and V_{gd} becomes equally important. In this case, the static ΔV_{th} model is inaccurate. Obviously, interchanging the roles of drain and source of the FET in Fig. 2.2b will decrease ΔV_{th}. This may actually happen in a transmission gate. It is still practical to have a notion of threshold voltage change. In the analogue case it depends on the distributions of the charge traps and the electric field density on the surface of the interface. We define

Definition 6.1. *the **effective** threshold voltage shift $\overline{\Delta V_{th}}$ is a change in threshold voltage that explains the channel conductivity deviation for a single operating point and a given trap distribution.*

Note that the effective ΔV_{th} depends on the operating point. A fixed ΔV_{th} does not factor in the complete device ageing state, such as interface trap locations. It thus cannot describe the



(a) $\overline{\Delta V_{th}} = (p_d + p_s)/2$ (b) $p_d = 0$, $\overline{\Delta V_{th}} = p_s/2$ (c) $p_s = 0$, $\overline{\Delta V_{th}} = p_d/2$

Figure 6.5: Symmetric BTI model in a digital circuit. In triode mode, all traps contribute to $\overline{\Delta V_{th}}$.

analogue behaviour of a transistor very well. A corner case illustrates this. Consider a p-FET that suffers from trapped charges close to the drain (Fig. 6.4). The impact on the conductivity depends on whether that region is active or not. In (forward) saturation, the charge carriers might originate from close to the source, unaffected by the damage.

With this in mind, we set up an analogue BTI ΔV_{th} model as follows. Let Q_{BTI} be a process that generates and takes charge traps based on an electric field controlled by V_{gs} as in Fig. 2.2a. Choose a parameter p such that $p \circ Q((-V_{gs}/\mathrm{V})_t)$ is the resulting threshold voltage change at time t, *as observed in Fig. 2.2b*. We will construct such a process in Section 6.1.4.

We choose the stress level $L = (-V_{gs}/\mathrm{V}, -V_{gd}/\mathrm{V})$, which controls the doubled process $Q := (Q_{BTI}, Q_{BTI})$. Set $p_s = p \circ Q_1$ and $p_d = p \circ Q_2$. We now hypothesise that the charge traps responsible for p_s are distributed with density $p_s \cdot (1 - x)$, where x is the coordinate of a one dimensional channel model with drain at 0 and source at 1. Symmetrically, the distribution of the other traps be $p_d \cdot x$. The electrical field effective voltage be modelled linearly as

$$v(x) := V_{gs} \cdot (1 - x) + x \cdot V_{ds} - V_{th}. \tag{6.2}$$

This gives rise to the intersection

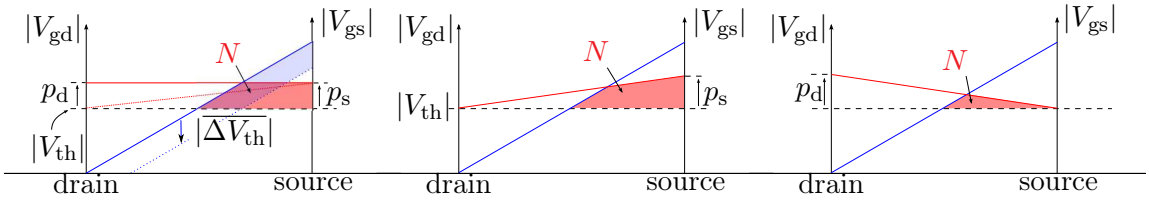
$$N := \int_0^1 \Theta(-v(x)) \cdot \max(v(x), p_s \cdot (1 - x) + p_d \cdot x)\mathrm{d}x, \tag{6.3}$$

where Θ is the Heaviside function. Note the sign of the voltages, and that $\max(v_1, v_2)$ equals $-\min(-v_1, v_2)$. Then, the zero of

$$w \mapsto \int_0^1 \Theta(-v(x)) \cdot v(x)\mathrm{d}x - \int_0^1 \Theta(w - v(x)) \cdot (v(x) - w) \cdot \mathrm{d}x - N \tag{6.4}$$

is $\overline{\Delta V_{th}}$, the shift in V_{th}, that corresponds to the loss of carriers modelled by N. Eventually, this effective shift models the operating point dependent loss of gate potential that imitates the BTI impact on the number of active charge carriers. We illustrate the operating point dependency of this model. In digital mode, Fig. 6.5, the drain and source potential is similar. The whole channel provides carriers, and the model just sums them up. In analogue mode, Fig. 6.6, the drain and source potentials may be distinct, then the defect location does matter. Defects below the green potential graph count.

More formally, we may assert that



(a) $\overline{\Delta V_{th}} = p_d = p_s$ (b) $p_d = 0$, $|\overline{\Delta V_{th}}| < |p_s|$ (c) $p_s = 0$, $|\overline{\Delta V_{th}}| < |p_d|$

Figure 6.6: Symmetric BTI model in an analogue circuit. In saturation, the defect location determines the contribution to $\overline{\Delta V_{th}}$.

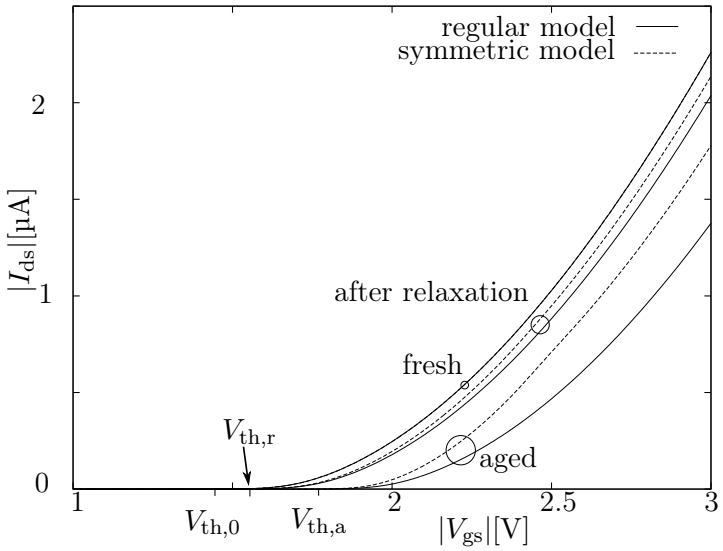


Figure 6.7: Symmetric BTI parameter model. In saturation, the characteristic is different from the regular fixed ΔV_{th} model.

Theorem 6.2. *The model Eq. 6.4 extends the underlying digital BTI model. Or, in other words, when used in Fig. 2.2a or in a digital circuit,* $\overline{\Delta V_{th}}(t)$ *simplifies to* $p \circ Q_{BTI}((V_{gs}/\mathrm{V})_t)$.

Proof. In the common source circuit, the effective threshold $\overline{\Delta V_{th}}$ does not depend on p_2, and thus $\Delta V_{th} = \overline{\Delta V_{th}}$. □

The effective ΔV_{th} model has the required properties for use in analogue circuits.

- Symmetry is faithfully modelled. Turning around a transistor does not invalidate the model.
- The model has a realistic first order dependency of ΔV_{th} on the operating point.
- The impact of negative V_{gd} stress during relaxation becomes visible.

We have implemented the symmetric BTI model, injecting the formula for $\overline{\Delta V_{th}}$ into a BSIM3.3 p-FET. The resulting current I_{ds} has a V_{gs} dependency different from the static ΔV_{th}

model presented in Fig. 2.3b. Fig. 6.7 shows the characteristic of an aged and relaxed device in comparison. For this simulation, the cross potential $|V_{\mathrm{ds}}|$ has been set to 1V.

Note that the assessment does not even take into account negative $-V_{\mathrm{gd}}$ relaxation, which will increase the asymmetry of the charge trap distribution. For example in an inverter with high input, the gate voltage of the p-FET is above the drain voltage. This may result in a faster relaxation of the traps close to the drain relative to those at the source, and eventually increase the operating point dependence of ΔV_{th} further. For an accurate model, a trap model with negative stress support is certainly required.

6.1.4 Modelling a BTI Process

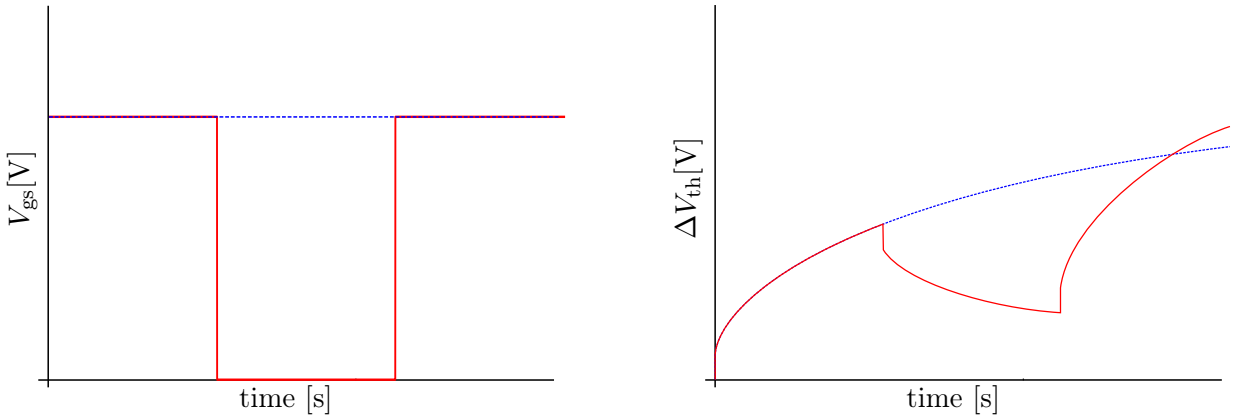
In theory, the RD model permits arbitrary electric fields at the interface. Leaving aside the controversy about its validity and the controversial range of interpretation [OS95; KA86; Bha+06; Kac+08], we still do not know about an efficient evaluation scheme that does not involve precomputation based on fixed stress patterns. The precomputed formulae for ΔV_{th} do not make sense for stress levels that vary in time. For example an empiric power-law relation $\Delta V_{\mathrm{th}}(t) \sim t^n$ is invalid as soon as a device is shut down (see also Section 5.5).

The PTM BTI model [VWC06] provides a formula for ΔV_{th} for transistors operated under the conditions depicted in Fig. 2.2. The PTM model formula is defined for rectangular waveforms It seems to be a good idea to evaluate the PTM model on a set of admissible stimuli and then find a process model that reproduces these data, interpolating for arbitrary analogue signals. However, there are a few obstacles related to the process incompatibility of the PTM BTI model we have discussed in Section 4.2.3 Some of these can be worked around to some extent.

Cross potential. There is a dependency on V_{ds}, the potential through the channel. Yet, there is sparse indication [VWC06] on how this dependency is meant to either model the impact on a damage-state or the resulting effective observable ΔV_{th}. As it seems, the formula tries to empirically fit some ΔV_{th} measurements under normalized conditions, after stressed with selected pulses and fixed V_{ds}. We avoid this by setting V_{ds} to zero, and retain the case of symmetric damage on the channel. Addressing asymmetric damage and effective ΔV_{th} in analogue applications requires a different approach (see Section 6.1.3) and leads back to the symmetric case.

Monotonicity. The ΔV_{th} shift predicted by the PTM BTI model is not monotonic. As a certificate, apply constant stress with $|V_{\mathrm{gs}}| > 0$ to a fresh device. To another device apply constant stress interrupted by zero stress. The latter has certainly suffered less, but shows a higher ΔV_{th} according to Eq. 4.20. This violates monotonicity (Fig. 6.8), against all evidence that BTI is monotonic. We are prepared that a monotonic model can not reproduce this behaviour, but only approximate selected data points.

Boundedness. With higher frequencies and duty cycles, the formula Eq. 4.20 diverges (see also Section 4.1.3). Yet the expression for ΔV_{th} under constant stress should be an upper bound



(a) Two stress waves ...

(b) ... lead to unexpected results.

Figure 6.8: PTM BTI model monotonicity.

for ΔV_{th} caused by stress of lower level, no matter which frequency. We retain boundedness by replacing

$$\Delta V_{th}(L) = \min\left(\Delta V_{th}(L), \Delta V_{th}(\max L)\right).$$

This will filter out some (fast) dynamics, but there seems to be no reasonable alternative.

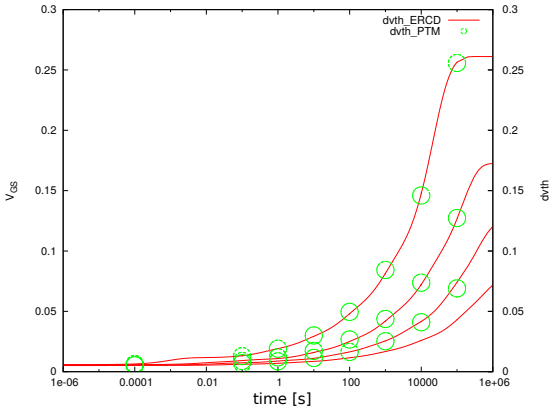
Negative stress. The PTM BTI model does not support negative values for the stress imposed by V_{gs}. However, the model we are looking for must operate under negative stress conditions as well. We will demonstrate the effect and then carefully augment the data set to get a physically meaningful response to negative stress.

Interestingly, another formula is provided by the authors of [VWC06]. With constants K and η, it predicts a shift of

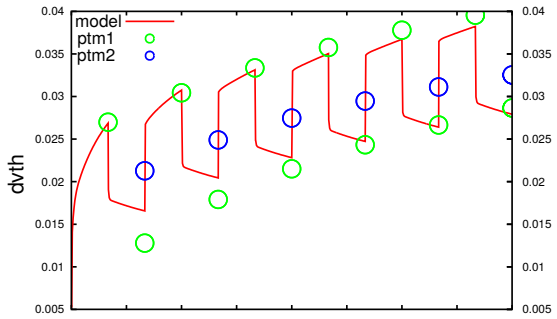
$$|\Delta V_{th}(t)| = K\beta^{1/4}(t/n)^{1/4}\left(\frac{1-(1-\sqrt{\eta(1-\beta)/n})^{2n}}{1-(1-\sqrt{\eta(1-\beta)/n})^{2}}\right)^{1/2} + \delta \tag{6.5}$$

n cycle of duty β. However increasing the frequency and duty, the expression diverges, while a signal with duty cycle β below 1 should probably not cause a shift higher than a signal with $\beta = 1$.

Under these preliminaries we have extracted a data set with 201 stimuli from the PTM model. An ERCD sum with 14 cells has been found to approximate the data set. This model reproduces the response to constant stress pretty well, see Fig. 6.9a. Rectangular Waveforms give good results, see Fig. 6.9b, matching the values from the iterative PTM BTI formula Eq. 4.20. Longer pulse-shaped stress patterns give results that deviate from the PTM BTI values Fig. 6.10. Here, the different PTM BTI formulas contradict, see above. We cannot expect that our model fits both variants. Also we do not know in which way the workaround is consistent with other sample points. Triangular stress, as presented in [MG13] can be simulated as well, see Fig. 6.11a. The



(a) Constant $|V_{gs}|$ stress at 1/4, 1/2, 1 and 2 V. Within the learning region (below 1×10^6 s) the ERCD model (red) reproduces the data points.



(b) Rectangular stress pattern. The green circles are produced using Eq. 4.20, some of them are in our learning set for the ERCD model (red response). The blue circles were computed by Eq. 6.5.

Figure 6.9: Selected transients of the ERCD BTI model and corresponding PTM BTI evaluations.

response is consistent with the other PTM BTI sampling points, which we believe do not properly reflect short-term effects. Note that the PTM BTI model cannot be evaluated on this pattern at all.

The PTM BTI model is not meant to work for negative stress. Hence, we have not included patterns involving negative stress in the learning set. This leads to potentially bogus results once the electric field switches directions. The effect can be seen in Fig. 6.12. Here a pulsed stress with a slightly negative lower value leads to excessive over-relaxation. Inputs that are not covered by the learning set still give arbitrary results. This effect is similar to the over-relaxation observed in Section 4.4.3.

One way to amend this is to cap the stress level at zero. Then, the relaxation rate will not depend on the stress level if the stress level is negative. However, there is evidence that it should. Clearly, the RD model incorporates an equilibrium for the reaction at the interface, with a dependency on the electric field. By physical means, this dependency is not restricted to positive electric fields, and likewise controls the reaction speed for negative stress. We add to the learning set a constant stress function with level -1 and duration 1 ks causing a parameter shift of almost zero. Due to monotonicity, this extra data point contradicts the outcome of negative stress in Fig. 6.12. If constant negative stress does not cause a negative shift, then no higher stress is supposed to. We then find a model that fits the previous set properly and meets the additional point as well. Comparing the results from the previous fit under truncated stress with the outcome of the new fit (Fig. 6.13a), we find that the model has become more physical. The overall physical correctness of this approach is questionable, it is based on the assumption that a fresh device has the smallest threshold voltage, or – in other words – all parts involved in BTI are in the freshest possible state at time zero. In any case, this assumption will help us to

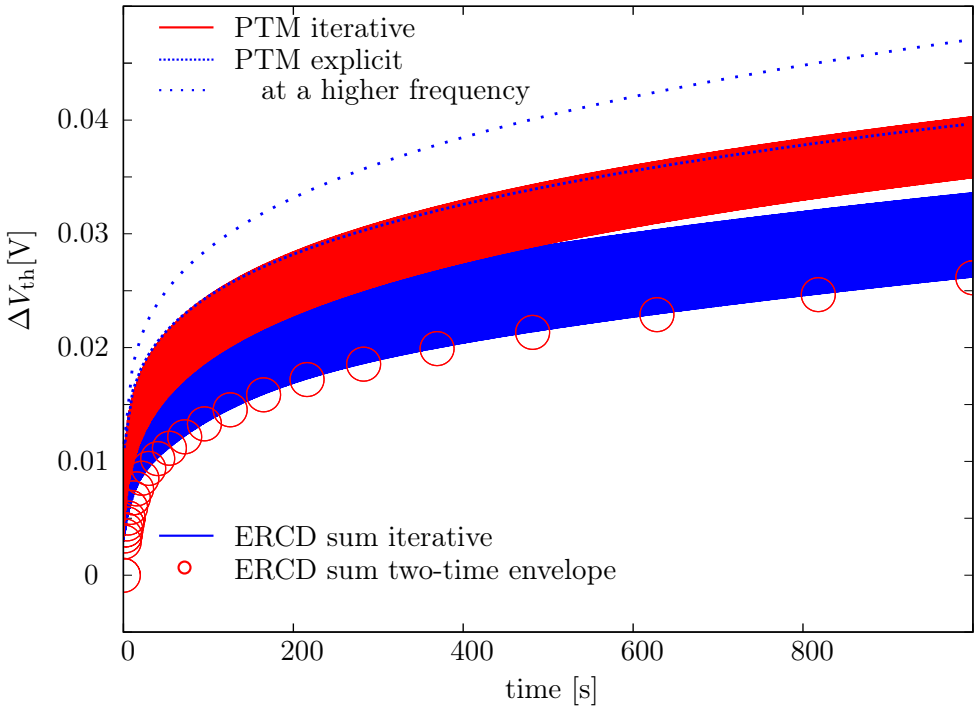
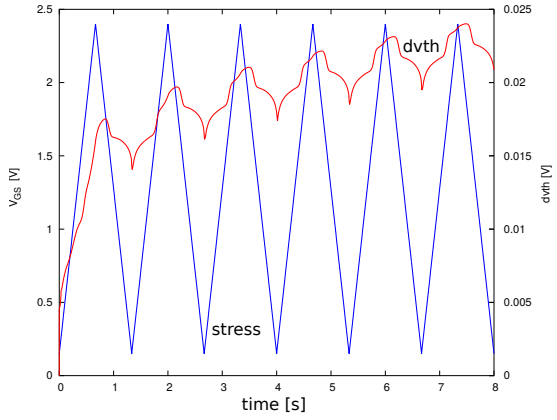
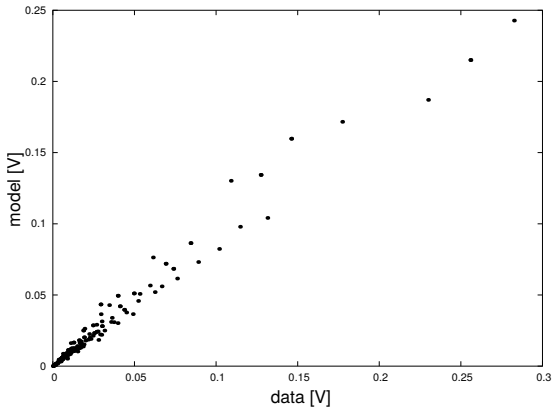


Figure 6.10: Pulsed PTM fit at 1 kHz and duty cycle 1/2.



(a) Triangular stress in ERCD BTI model.



(b) Model quality histogram, based on learning set. Taking into account the inconsistencies of the input data, the coincidence is surprising.

Figure 6.11: More interesting model property assessment.

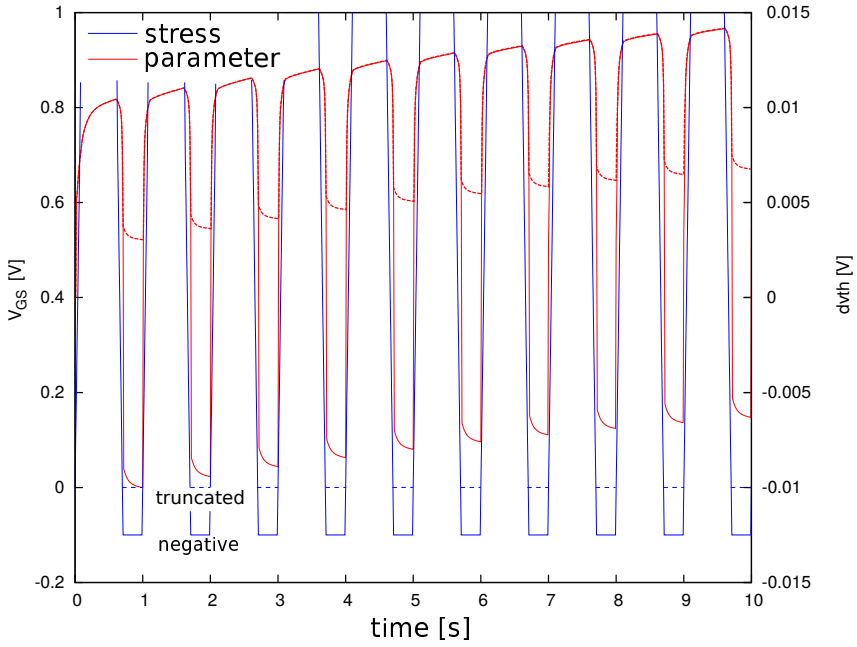


Figure 6.12: Slightly negative stress causes excessive over-relaxation. For comparison, the impact of the truncated stress wave (dashed).

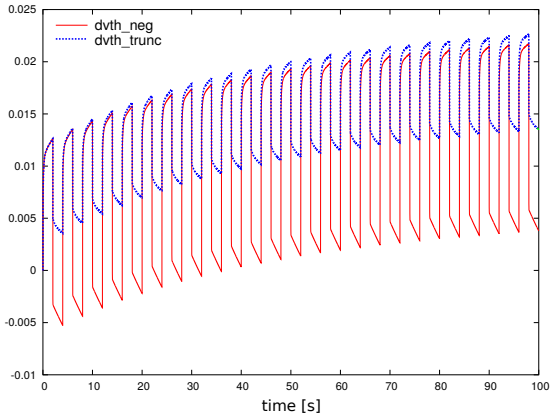
demonstrate that ERCDs are suited to model BTI. In case of emerging evidence we could easily assume otherwise.

We have created a monotonic ΔV_{th} model that is valid for V_{gs} close to V_{th} (Fig. 2.2b) based on an interpretation of the PTM BTI model. The assumed measurement points are reproduced properly, and the model processes stimuli beyond rectangular waves in a consistent way. It would have been nice to have more realistic and consistent measurement data. Anyway, we have a model that behaves a lot like BTI and can be used representatively for simulation purposes.

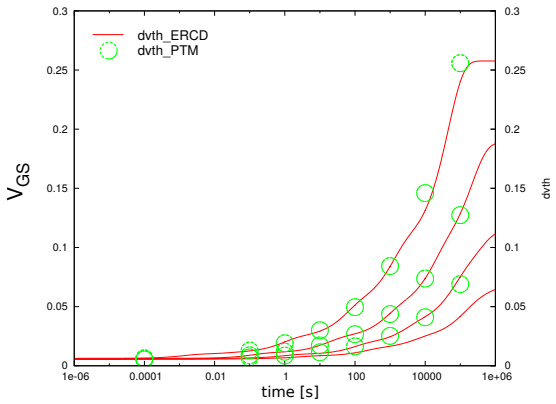
6.2 Circuit Simulation

6.2.1 Free-Running Oscillation

An *astable multivibrator* (Fig. 6.14b) constitutes a free running oscillator. These, or similar oscillators are often used to monitor the ageing state of a circuit indirectly. While the circuit in question carries out some task, the oscillator is switched on and gets older. The frequency of the oscillation then accounts for the overall state of the circuit that is believed to wear out in a similar fashion. Ageing simulation of such oscillators as analogue systems with ageing effects becomes possible with the appropriate ageing models. In Fig. 6.15, we have plotted the effective ΔV_{th} evolution and the period width during two-times simulation.

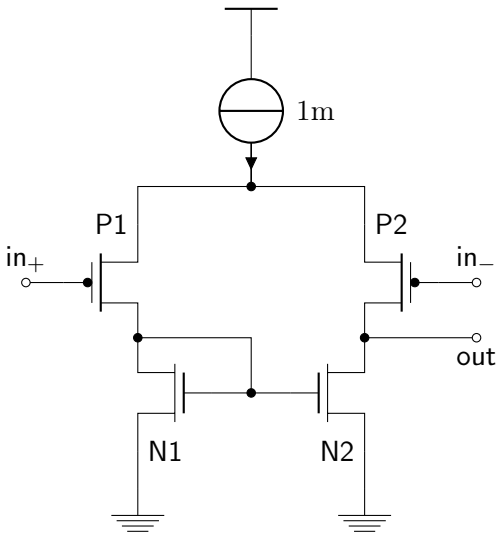


(a) A pulsed stress wave with negative lower value results in a ΔV_{th} (red) different from the ΔV_{th} (blue) caused by the truncated pulse.

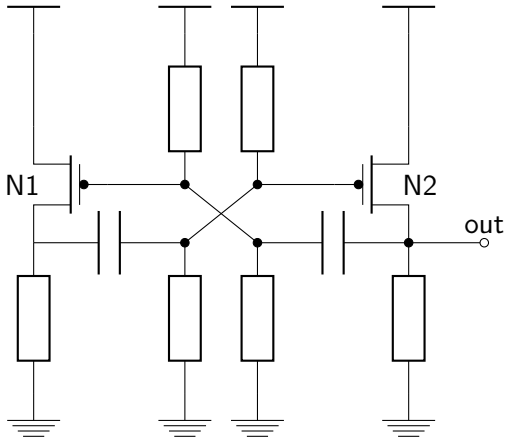


(b) Constant stress impact of the modified model remains acceptable.

Figure 6.13: Extending BTI model for negative stress, resulting in more physical behaviour.

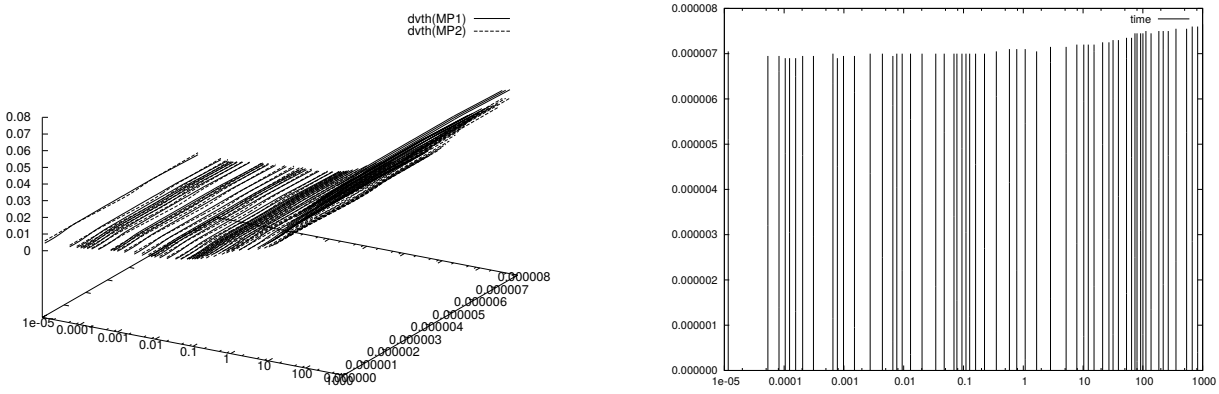


(a) A comparator is susceptible to ageing. The impact depends on the operation mode.



(b) An astable multivibrator, p-FET based implementation.

Figure 6.14: Circuits affected by ageing effects.



(a) $|\Delta V_{\mathrm{th}}|$ during two-times simulation.

(b) The oscillation frequency decreases with increasing $|\Delta V_{\mathrm{th}}|$

Figure 6.15: Simulation of p-FET astable multivibrator.

6.2.2 A Comparator

An early motivation for ageing simulation has been a *comparator*. Simulating a comparator with generic or precomputed average ageing parameter drift models will hide behavioural asymmetries. In [SH11] we were the first to simulate BTI ageing in analogue circuits under different conditions without precomputed parameters. This became possible with an early version of the BTI process model explained in Section 6.1.4. We have used the BTI model DC parameters from [Sch+10] to calibrate the analogue model. Here, the dynamics outside the digital DC operating points was purely hypothetic. Still, we had asserted that a refinement would be possible with more data available. Later experiments strongly support this, see Section 6.1.4.

The comparator (Fig. 6.14a) takes two input voltages (at in_- and in_+) and outputs a high or low voltage, depending on which of the input voltages is higher. If the transistors are affected by different V_{th} shift, the comparison trivially is biased by the difference of these shifts. We applied a constant input u to in_+ and a sine wave of amplitude 0.2 V at an offset u_0 and a period of 10 ms to in_-. We monitor the ageing indirectly by measuring the time of the output crossing $V_{\mathrm{dd}}/2$ for the first time in each cycle. We observe that with our model parameters, the threshold voltage for the p-FET exposed to the sine is bigger than the one exposed to DC stress. Also the effect depends on u_0.

6.2.3 Monitor Circuits

Monitor circuits can be used to detect circuit failures [DB10]. Long-term ageing simulation enables the development of analogue on-line monitors. The circuit in Fig. 6.17 implements an offset monitor connected to a power stage built from source followers. The power stage is symmetric, except for the p- and n-FETs interchanged. These are subject to HCI and NBTI stress. As the n-FETs age, the offset voltage lowers, and higher damage on the p-FET side causes a higher offset. A lower levelled signal has more effect on the n-FETs, especially on $\mathsf{N2}$. At the

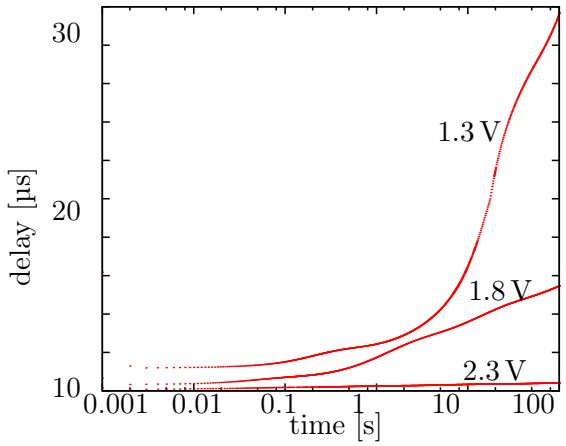


Figure 6.16: Delay over time of the crossing of $v_{\mathrm{dd}}/2$ at $u_0 = 1.3, 1.8, 2.3\mathrm{V}$

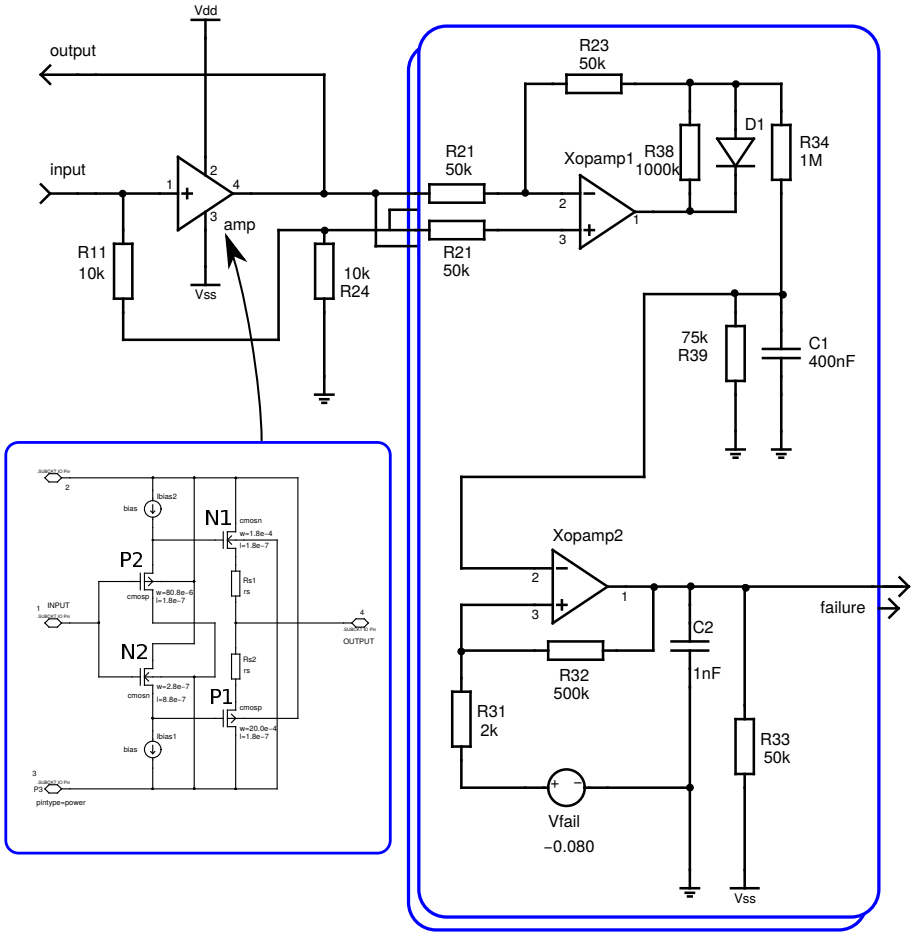


Figure 6.17: A power stage with a peak-error window detector.

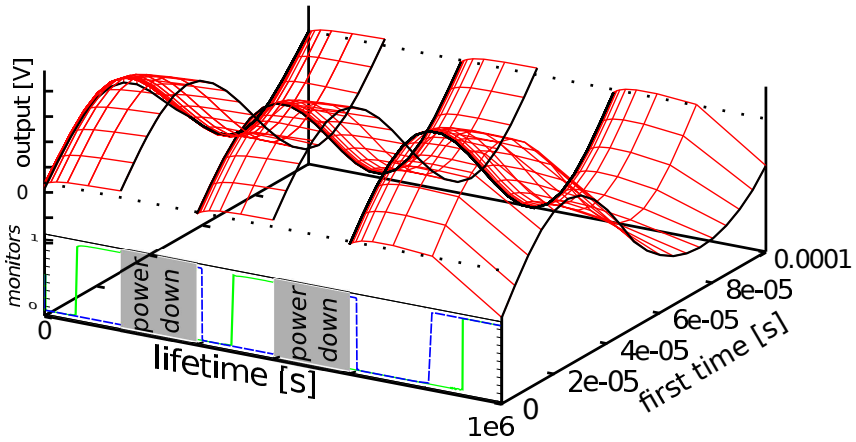


Figure 6.18: Monitoring the power stage: The failure signals (front) indicate too high (green) and too low (blue, dashed) offset peaks.

beginning, the output offset is rising due to NBTI in P1 and P2. later, HCI catches up. Two monitors track the offset deviation comparing output with the input voltage. Inside the monitor, a trigger is connected to a control node that holds a low-passed peak-difference. Two monitors with interchanged input polarity report too high and too low peak-level offset. In Fig. 6.18 the circuit is simulated with a sine input signal to the power stage. Power down phases, such as prescribed by a mission profile, interrupt the operation.

6.2.4 Ageing and Relaxation

The conversion characteristic of converters is affected by ageing effects. For example, a simple DA-converter (Fig. 6.19) relies on precise current mirroring. Ageing effects bend the transfer characteristic, but also a power-down phase attenuates the deviation. As examples, we have simulated a few long term transients. During these transients, the input signal is relevant. A low input has a higher impact on the OTA hence the characteristics deviate, see Fig. 6.20a. After converting cyclic binary input over several months, we have simulated a cold power down Fig. 6.20b.

Neither before nor after ageing or relaxation, this converter is particularly linear. The circuit has been presented at the robustness tutorial of the conference ZuE 2012 to motivate ageing simulation that supports analogue stress levels and relaxation effects. This example is particularly interesting, because the p-FETs P0 to P4 are subject to different voltage levels at their sources. While the source of P4 is always above $V_{\mathrm{dd}}/2$, the source of P0 can take any potential, hence P0 has more space to relax.

6.2.5 Redundancy Simulation

Cross coupling two instances of the Halting Circuit we obtain a *redundant circuit* (see Fig. 6.21). The instance on the right side uses an n-FET. Pulling up the control node nc1 (via pON), loads

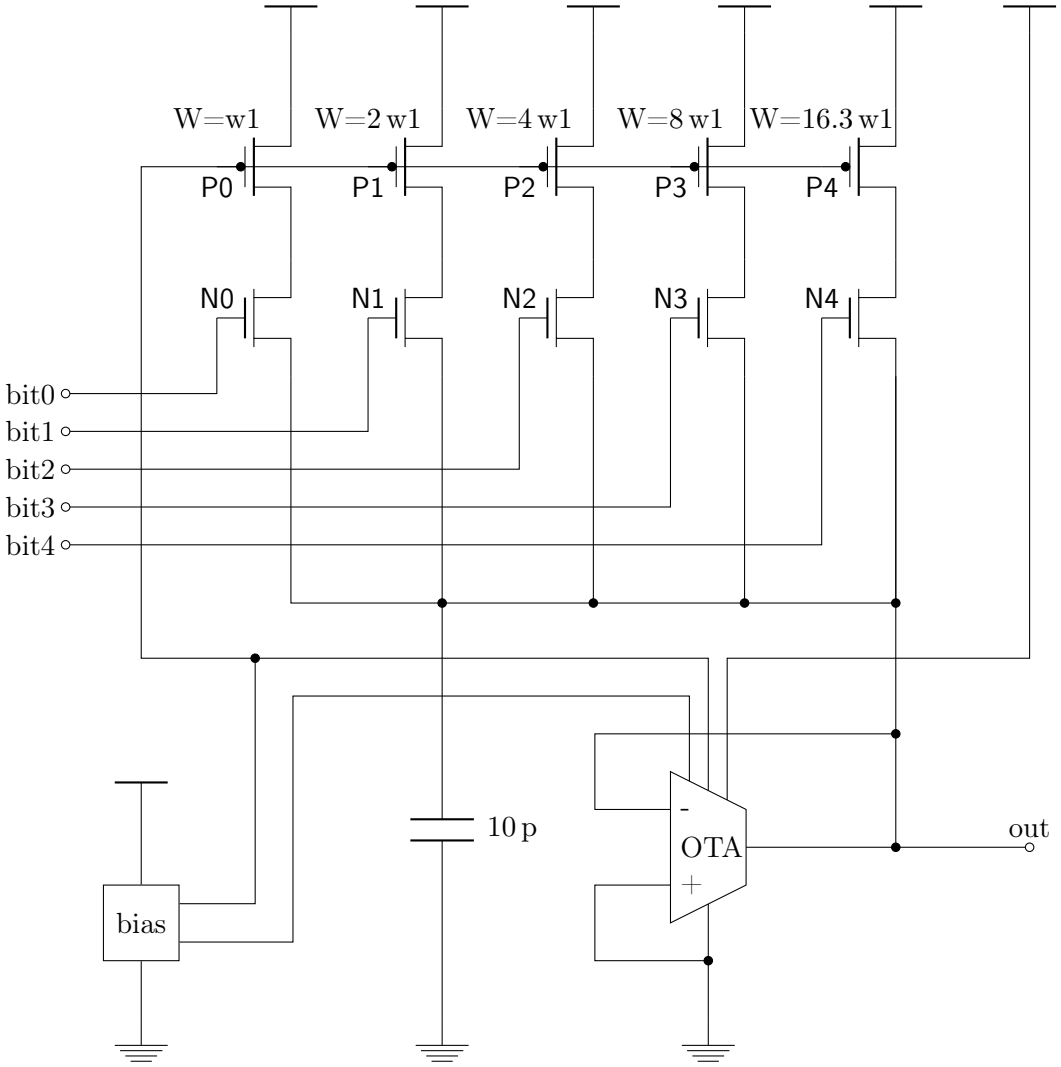
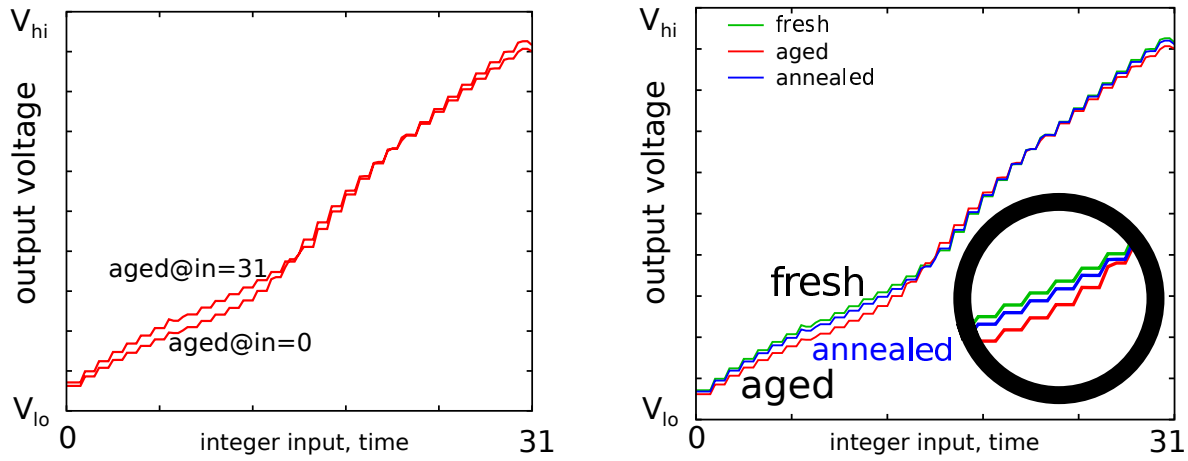


Figure 6.19: A digital-analogue converter based on a current adder.



(a) The ageing impact depends on the input signals.

(b) Ageing and Relaxation.

Figure 6.20: Converter ageing effects simulated.

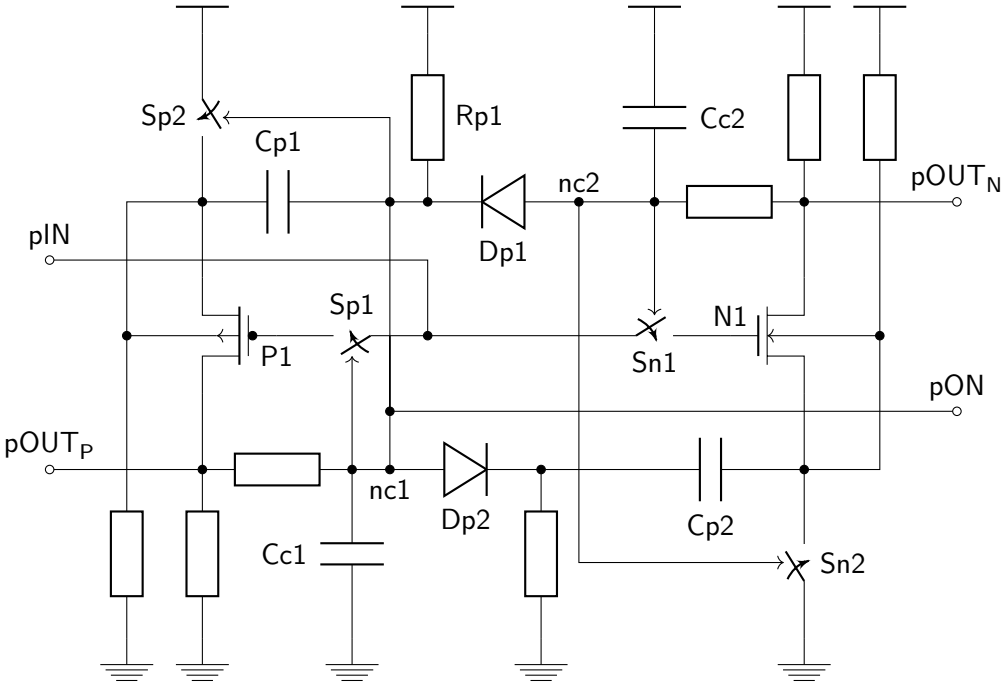


Figure 6.21: A simplistic embedded redundancy circuit. This circuit is symmetric except for the different ageing effects in N1 and P1.

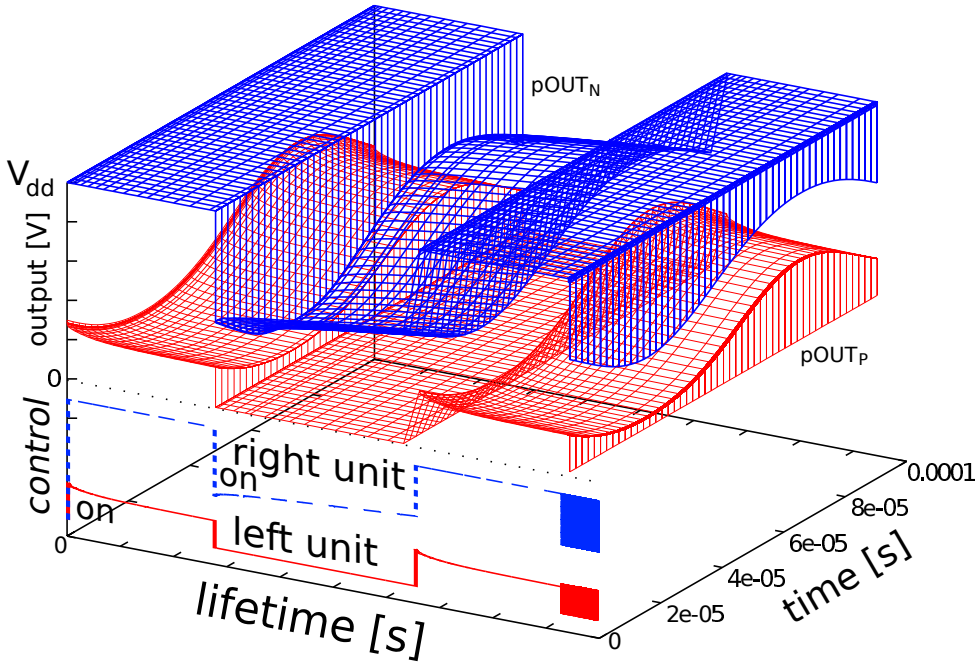


Figure 6.22: Output and control voltages during simulation. The circuit is processing a sine-waved signal at pIN.

the controlling capacitor Cc1 and turns on the unit on the left. The input applied to pIN is an analogue signal chosen such that the fresh FET P1 keeps the voltage at nc1 above the power-off threshold of Sp1 and Sp2. This voltage, the low-passed output voltage, will drift down as P1 ages. Consequently, the left unit will switch off at some time, pulling down Dp1 through a high pass (Cp1, Rp1), activating the complementary unit on the right. During the operation on the right side, P1 is off and recovering from NBTI. Thus, at the time the right unit breaks down, the left unit will take over. The takeover works similar, with directions interchanged. As N1 does not recover from HCI, the third breakdown will be definite. Here, the voltages controlling the switches do not solely depend on the ageing state. Due to the alternating voltages, the charges on the capacitors Cc1 and Cc2 are volatile. Over- and underestimating these charges during second-time extrapolation will distort the simulation results. The trend of a capacitors charge becomes a component of the ageing state of the circuit. We extrapolate the charges on capacitors from the data collected by transient simulations as described in Section 5.4.1. Fig. 6.22 shows the two-times simulation output.

6.3 Amplifier Screening

The search for circuits matching a given specification is referred to as synthesis. Automated synthesis usually means creating lots of candidate circuits and examining their various properties. Checking different in a well-considered order reduces the amount of candidates, see [MH15].

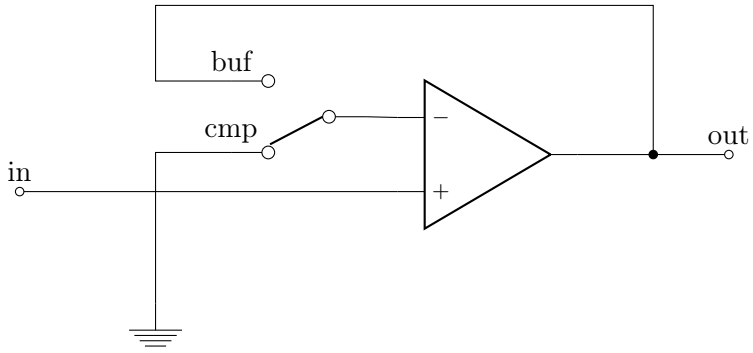


Figure 6.23: Two ageing scenarios

Usually, ageing effects are not considered, or considered indirectly by means of safety margins [McC+09]. However, ageing effects and their impact depend on the use-case of the circuit. Thus it is possible to improve synthesis, once a mission profile for the sought circuit is available.

We have screened ageing performances for a set of operational amplifiers to be used in different applications and operating modes [SHM13]. Our experiment demonstrates the feasibility of a simulational approach and the capabilities of the ageing models and simulation techniques. The quantitative results are subject to imperfections of the models and simulator accuracy, which has improved in the meantime. The impact of structual differences of the different circuits on the performances in different tasks however is unchallenged.

The amplifiers have been automatically synthesized from building blocks without ageing considerations [Mei+12]. We have compared performances of these amplifiers measured in a simple closed loop testbench, after a variety of ageing experiments. Of many possibilities for choices for testbenches, performances, missions and optimality criteria we have selected a few we consider interesting. As our amplifiers differ in performances, we use the performances of each fresh device as reference for ageing effects. Relative ageing of a performance P means

$$\Delta_{\text{rel}} P = P_{\text{aged}} - P_{\text{fresh}}/P_{\text{avg}},$$

where P_{avg} is the average performance over all devices in fresh state.

The testbench in Fig. 6.23 provides two modes of operation, that we use to age the devices. Performance measurements take place in the open-loop setup, i. e. with the switch in cmp position. First of all, we run the amplifiers in buffer mode, simply buffering $V_{\text{dd}}/2$ for some time. The change in performances reveals, that the variation in ageing impact is noteworthy. In Fig. 6.24, we have plotted the relative change for dc-gain, unity gain frequency, dc-range and slew rates. Roughly, the total parametric change of a device can be measured by $\prod_P (|\Delta_{\text{ref}} P| + 1)$ (here, we exchange the zero offset performance by its absolute value). Considering this valuation, for example, the best candidate for buffering a sine is the one in Fig. 6.25.

Ageing as a comparator can have a different impact on the performances. We change the input signal to a sine. The relative parameter shifts for ageing in buffer mode are different from the shifts observed after comparator mode ageing (Fig. 6.26). Note that we have truncated the

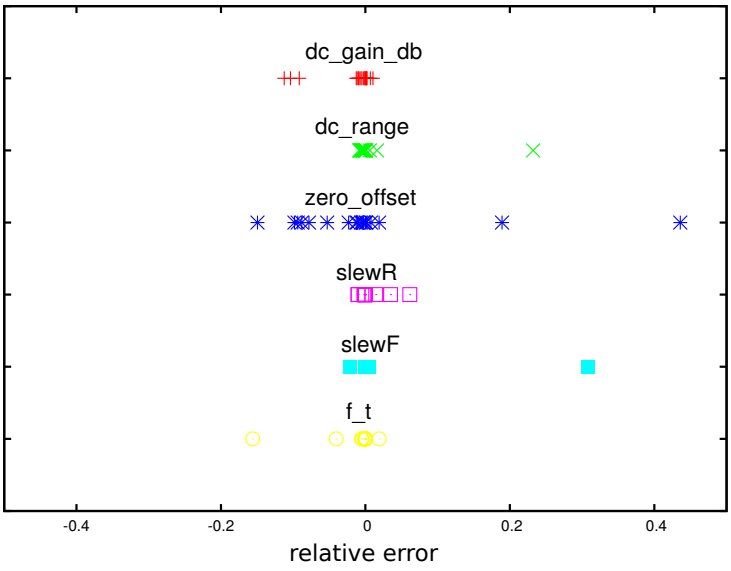


Figure 6.24: Relative ageing impact (buffering $V_{\mathrm{dd}}/2$).

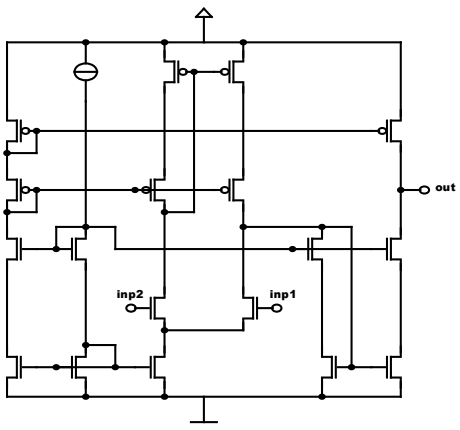
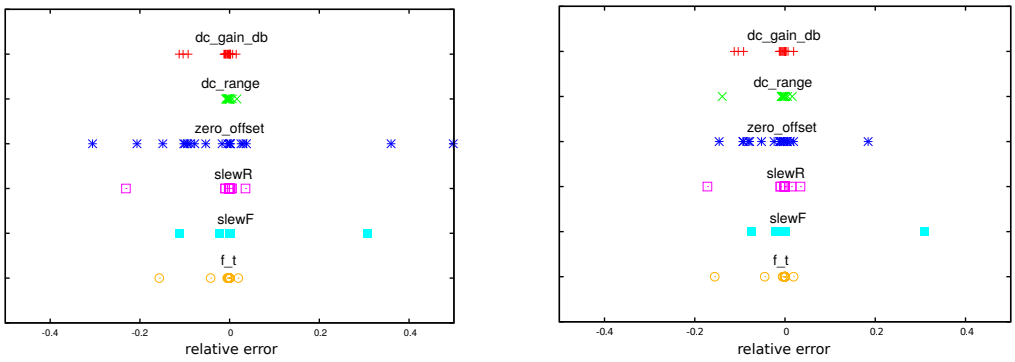


Figure 6.25: An example operational amplifier (op19_5), good candidate for buffering a sine.



(a) Comparing a sine.

(b) Buffering a sine.

Figure 6.26: Relative ageing impact.

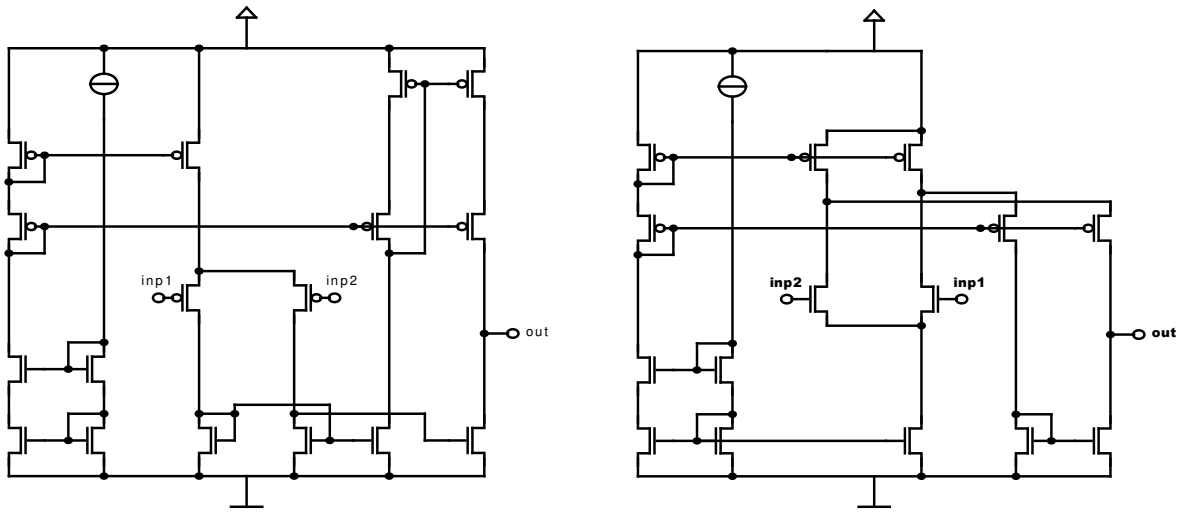
abscissa at ±0.5. The most obvious difference is the zero offset shift. Since in buffer mode, both inputs are approximately at the same level, the devices are stressed symmetrically. Hence, in more cases the zero offset shift is nearly unaffected. Now, the best candidate in buffer mode turns out to be Fig. 6.25, it is symmetric and has a pMOS input stage. The best in comparator mode is symmtric as well, but has an nMOS input stage (Fig. 6.27b).

Deeper inspection shows, for example, how the zero offset is affected differently. Fig. 6.28 shows the relative ageing effect on the individual amplifiers sorted by their respective fresh gain. We find that independent of the ageing mode, higher gain correlates with less impact on zero offset. The amplifier op67_14 has a pMOS input stage (Fig. 6.27a), its offset drifts a lot in comparator mode. In buffer mode, ageing does not affect the offset, as it is corellated. For the offset of another one, the ageing mode does hardly matter, although pMOS transistors are involved see Fig. 6.29a. Interestingly, in op19_5, the one that evaluates best in the buffer case, we get extremely high offset shifts.

Finally, after the sine-stressing, the amplifier op26_4 Fig. 6.29b turns out to recover well. In both the buffer and the comparator case it gets the best score.

6.4 Simulative Model Reduction

An inverter gate, built from two MOS-FETs is a device usually installed and operated in digital circuits. There, a characterization by its delay and transition times is sufficient for most purposes. Moreover, the life of an inverter gate can be described by the number of transitions n and the average input voltage or the *duty cycle* β. So the impact of ageing effects on the characteristics could be simply modeled as functions of n, β and the age t [Pau+05]. The validity of the model is then restricted to the use case that has been implicitly determined – rectangular waveforms. If the operation mode varies, the model is immediately useless. For example, we may want to operate an inverter at different supply voltages [KP11]. Here, BTI is particularly insidious, and n and β are not sufficient to model the use case. Also, variations in the shape of input transitions



(a) Here, we observe different impacts on zero offsets (op67_14).

(b) A good candidate for comparing with a sine (op1_2).

Figure 6.27: Two operational amplifier candidates.

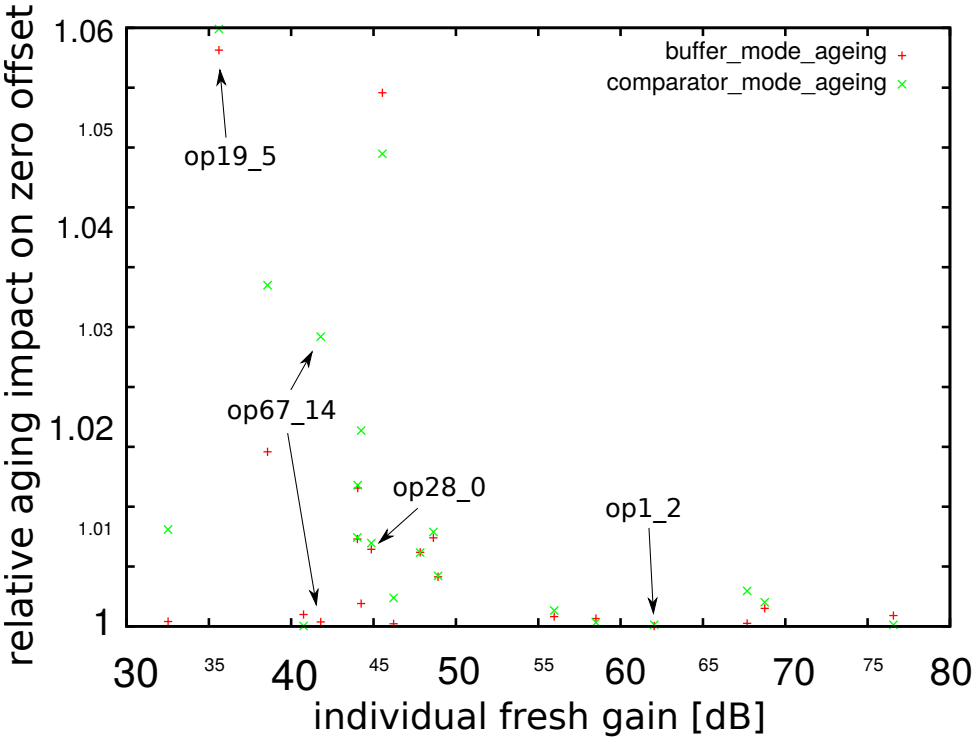
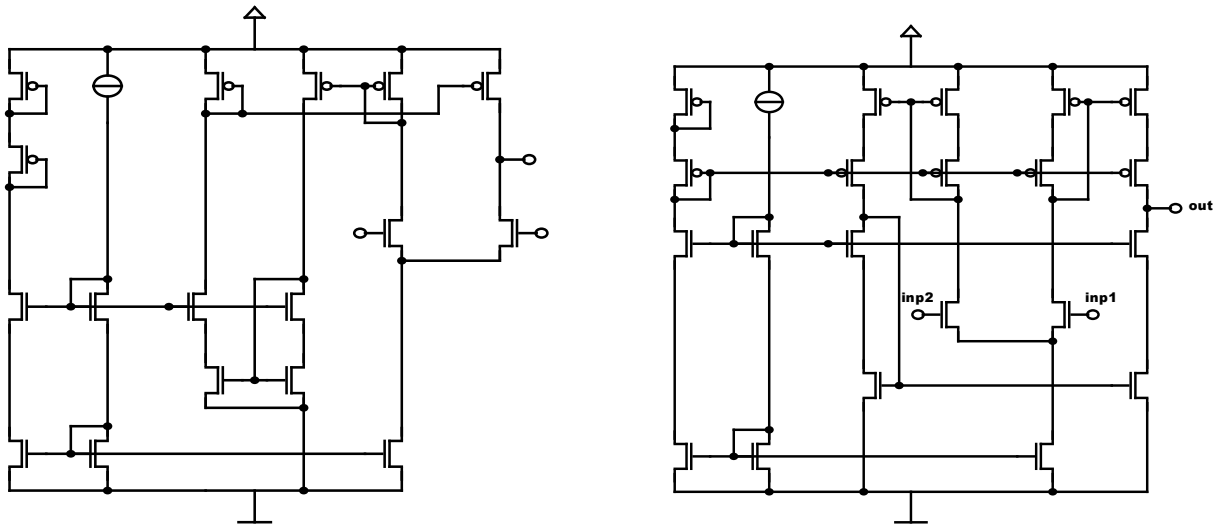


Figure 6.28: Relative change in absolute zero offset ageing for buffer (+) and comparator (×) mode ageing.



(a) Here (op28_0) we observe similar zero offset shifs in both operations.

(b) For op26_4, recovery plays an important role.

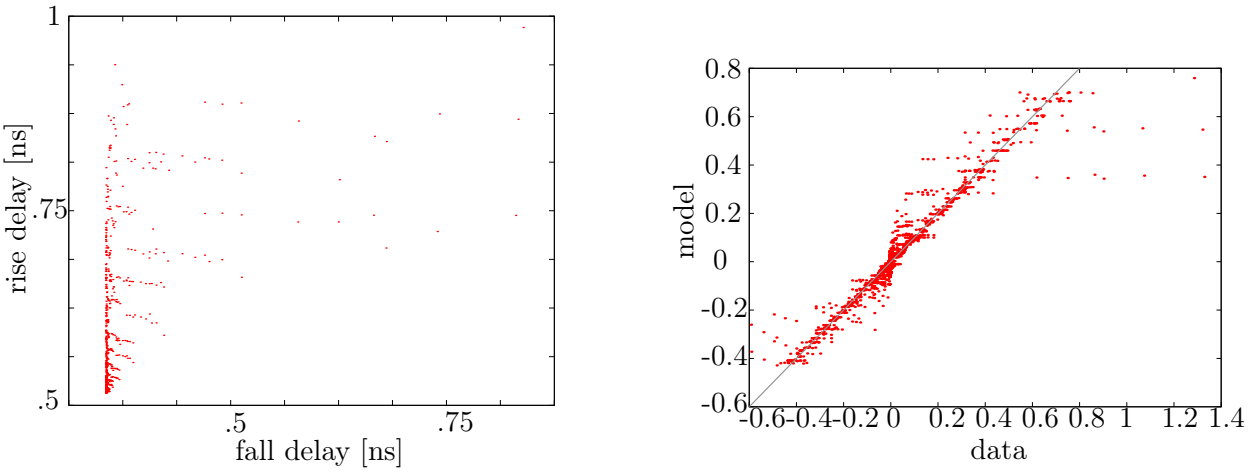
Figure 6.29: Two more operational amplifier candidates.

alter the ageing behaviour and are not covered. This is of particular importance if ageing affects the shape of transitions. Notably, relaxation during power down mode is not addressed at all by such simple models.

Forgetting about the limiting abstractions, the inverter gate is just an analogue circuit component. In the analogue domain, the operation is no longer constrained, and the transistors themselves need to be modelled for a full simulation. The characteristic times are still valid behavioural parameters, and are still subject to ageing effects. With ageing models for the transistors, we are able to compute the deviations of these parameters caused by arbitrary operation.

We will now elaborate on modelling ageing effects in analogue behavioural block models. An ageing model for an inverter gate as an analogue block is not particularly useful, but still a good demonstrator for the modelling workflow and model structure that targets analogue components Fig. 4.13. An analogue component has ports to the outside and inner nets or variables. The *state* of the model describes the assignments of values like voltages or currents to these entities. In the case of the inverter gate, the potentials $V = (V_{\mathrm{dd}}, V_{\mathrm{in}}, V_{\mathrm{out}})$ at the ports relative to ground sufficiently describe the state.

Using the transistor netlist and an appropriate test bench, we may now find out in which way an operation affects the behavioural parameters. An operation is a function x on a time interval $[0, T]$ into the state space. We use an ageing process to explain the parameter deviations. However, an ageing process requires a controlling stress level. Hence we need to turn the operations into stress level functions. Stress level functions that are suitable to control processes accounting for the parameter deviations are known, in case we start from a transistor model with process based ageing effect model. In [Mar+06] the stress levels from the transistor netlist are used to model the impact of HCI on the behaviour on an OTA. However, we do not consider



(a) Parameter combinations reached by ageing, delays.

(b) Relative parameter shifts, data set vs. model. All four parameters.

Figure 6.30: Inverter ageing data and model.

this approach fully appropriate. A model that does not reduce simulation complexity is not a proper behavioural model. Hence. it must not include a transistor netlist or stress levels modeled therein. Consider for example a device with a lot of transistors and a very high ageing resilience. The high level ageing model might be trivial, while tracking all stress levels in all transistors is still expensive.

Our approach is based purely on data that is relevant for and available within a behavioural model. We explain the impact on behavioural parameters using stress levels that are directly associated to the state of the behavioural model. These stress levels L_i may be the supply voltage or current and relative port values and their derivatives. These have been recorded during the simulation of the circuit in 756 cases. Together with the resulting behavioural parameter deviations these fill a table, one row per operation. We have chosen to model the delays and slew rates for the two transitions. Fig. 6.30a shows the simulated delay pairs for the chosen operation modes.

The model for the parameter deviations consists of a set of processes Q_j, each controlled by a level $L_{n(j)}$, and a weight matrix w,

$$\Delta p_i = \sum_{i,j} w_i^j Q_j(L_{n(j)}).$$

We use a learning algorithm that chooses ERCD processes for the Q_j and minimizes the sum of error squares $(p_i - m_i)^2$ as in Section 4.4.4. The resulting model consists of 14 ERCD processes that are controlled by stress levels derived from the node voltages, rather than from transistor models. Each of the four parameters is expressed as a linear combination of the states of these processes – the model can be simulated as a part of the behavioural model of the inverter gate. The fit (Fig. 6.30b) has some imperfections. These are likely related to the phenomena observed in Section 4.4.5, numerical issues of the simulation results and to the number of data points.

Nevertheless, we have created a model that works on arbitrary waveforms and mimics the low level relaxation effects on a behavioural level.

6.5 Conclusion

We have described a nominal ageing parameter model that supports analogue stress conditions as well as recovery. It can be directly fitted against established data points, which it reproduces by design. This way, the model allows to describe BTI not depending on a physical interpretation. A simulator has been implemented that evaluates this model in analogue real-world context, making use of an extrapolation scheme that exploits the simple structure of the models underlying differential equations. For example circuits we show, that ageing simulation reveals effects that are due to correlated ageing and recovery. Such effects should be taken into account during device synthesis, our experiments show a high variance in ageing impact on various operational amplifier topologies. We have improved modelling and simulation techniques to compute these effects in much more generality and much faster than previously accepted.

CHAPTER

SEVEN

DISCUSSION

Nowadays, analogue circuit design is moving away from manual tinkering to automated analysis and synthesis. State of the art workflows are still lacking proper means for modelling and simulation in the analogue domain. This ranges from missing concepts for modelling ageing effects at their very origin to missing concepts for modelling ageing effect on higher level. Simulators that are meant to simulate circuits are not suited well to simulate ageing effects. Simulating ageing effect without the circuit context does make sense in special cases (such as digital circuits), but in general does not answer how an actual circuit model will behave under particular conditions. It is widely known and accepted that BTI causes problems, particularly in analogue circuits, where no proper models exist.

This work raises strong evidence that circuit modelling and simulation with respect to ageing effects are tightly coupled problems. No different to ordinary component models, an ageing model needs to be evaluated during simulation, taking into account the current state of the simulated circuit. A continuous ageing effect model maps transient waves to parameter deviations. To be of any use, an ageing model must be computable in a reasonable time *and* behave in a way a physical device does. The power-law models are perfect at fast computation and the Reaction-diffusion models are perfect at imitating physical processes. We have presented a compromise model that is pretty good with respect to both criteria.

We have extensively studied the expressiveness of ageing models that are applicable in behavioural modelling contexts. These models are free from empirical approximations and assumptions, but rather turn properly modelled stress levels into the corresponding parameter deviations. With the tools developed in the context of this thesis it was first possible to create ageing component models from data sets that are suited for analogue operation. Such models are independent of physical interpretation and adjusted to the properties of the data.

With suitable and versatile ageing models at hand it starts to make sense to think about efficient and accurate long term ageing simulation. Long term simulation is simple for periodic operation. Lacking periodicity, it gets considerably more difficult. Multirate methods cover lots of the quasiperiodic simulations. Ageing effects are hardly ever periodic, thus efficient long term

ageing simulation not only requires efficient ageing effect models but also a variant of multirate analysis.

This thesis has presented a unified view on the the following topics related to integrated circuit reliability.

- The physical perspective on ageing processes
- Ageing parameter models and their properties
- Models for ageing effects in analog components
- Hardware description of ageing parameters
- Ageing simulation on circuit and system level
- Hierarchical ageing modelling

With new concepts, models and algorithms we simplify the analysis of ageing effects, and augment the applicability of ageing simulation to a wide range of analogue circuits.

BIBLIOGRAPHY

[08] *Sicherheit von Maschinen – Sicherheitsbezogene Teile von Steuerungen – Teil 1: Allgemeine Gestaltungsleitsätze*. Standard. Geneva, Switzerland: International Organization for Standardization, Dec. 2008.

[ADV12] M. S. Andersen, J. Dahl, and L. Vandenberghe. *CVXOPT: A Python package for convex optimization*. 2012. URL: http://abel.ee.ucla.edu/cvxopt.

[Ala03] M. Alam. "A Critical Examination of the Mechanics of Dynamic NBTI for PMOSFETs". In: *IEEE* (2003).

[Alt09] H. Alt. "The Computational Geometry of Comparing Shapes". In: *Efficient Algorithms*. Vol. LNCS 5760. Essays Dedicated to Kurt Mehlhorn on the Occasion of His 60th Birthday SpringerLink. Springer Berlin / Heidelberg, 2009, pp. 235–248.

[Ash06] D. Ashlock. *Evolutionary Computation for Modeling and Optimization*. 1st ed. Springer-Verlag New York, 2006.

[Bar+12] M. Barke, M. Kärgel, W. Lu, F. Salfelder, L. Hedrich, M. Olbrich, M. Radetzki, and U. Schlichtmann. "Robustness Validation of Integrated Circuits and Systems Asia Symposium on Quality Electronic Design". In: *ASQED* (2012).

[BB15] H. G. Brachtendorf and K. Bittner. "Initial Transient Response of Oscillators with Long Settling Time". In: *DATE '15: Design, Automation & Test in Europe*. 2015.

[Ber+06] J. B. Bernstein, M. Gurfinkel, X. Li, J. Walters, Y. Shapira, and M. Talmor. "Electronic circuit reliability modeling". In: *Microelectronics Reliability* 46 (2006), pp. 1957–1979.

[Bha+06] S. Bhardwaj, W. Wang, R. Vattikonda, and Y. Cao. "Predictive Modeling of the NBTI Effect for Reliable Design". In: *Custom Integrated Circuits Conference, 2006. CICC 2006. IEEE* (Sept. 2006), pp. 189–192.

[Bla69] J. R. Black. "Electromigration – A Brief Survey and Some Recent Results". In: *IEEE Transactions on Electron Devices* (1969), pp. 338–347.

[Cad] Cadence. *Users' Manuals for BSIMPro+/RelXpert/UltraSim tools*. URL: www.cadence.com.

[Cao+14] Y. Cao, J. Velamala, K. Sutaria, M. Shuo-Wei Chen, J. Ahlbin, I. Sanchez Esqueda, M. Bajura, and M. Fritze. "Cross-Layer Modeling and Simulation of circuit Reliability". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* (2014), pp. 8–23.

[Cho12] F. R. Chouard. "Device Aging in Analog Circuits for Nanoelectronic CMOS Technologies". In: *Dissertation* (2012).

[CRS] B. R. Cobb, R. Rumí, and A. Salmer. *Approximating the Distribution of a Sum of Log-normal Random Variables.*

[Dav03] A. T. Davis. "An overview of algorithms in Gnucap". In: *University/Government/Industry Microelectronics Symp.* 2003, pp. 360–361.

[DB10] H. Dagdour and K. Benerjee. "Aging-resilient design of pipelined architectures using novel detection and correction circuits". In: *DATE '01: Design, Automation & Test in Europe.* 2010.

[Doo47] J. L. Doob. "Probability in function space". In: *Bull. Amer. Math. Soc.* 53 (Jan. 1947), pp. 15–30.

[DPD10] G. Depeyrot, F. Poullet, and B. Dumas. *Verilog-A Compact Model Coding Whitepaper.* 2010.

[DS09] G. G. D. Lorenz and U. Schlichtmann. "Aging Analysis of Circuit Timing Considering NBTI and HCI". In: *IOLTS09* (2009).

[FB05] E. Freitag and R. Busam. *Complex Analysis.* Universitext (Berlin. Print). Springer, 2005. ISBN: 9783540257240.

[Fis05] G. Fischer. *Lineare Algebra.* Vieweg-Studium : Grundkurs Mathematik. Vieweg, 2005. ISBN: 9783834800312.

[GD08] G. Gielen and P. De Wit. "System-level Design of a Self-healing Reconfigurable Output Driver". In: *Analog Circuit Design* (2008), pp. 141–165.

[GHN66] A. Goetzberger, N. J. M. Hill, and H. E. Nigh. "Surface charge after annealing of Al-SiO2-Si structures under bias". In: *Proceedings of the IEEE* 54 (1966).

[Goe+11] W. Goes, F. Schanovsky, T. Grasser, H. Reisinger, and B. Kaczer. "Advanced Modeling of Oxide Defects for Random Telegraph Noise". In: *Noise and Fluctuations (ICNF), 2011 21st International Conference on* (2011).

[GR10] S. Ghosh and K. Roy. "Parameter Variation Tolerance and Error Resiliency: New Design Paradigm for the Nanoscale Era". In: *Proceedings of the IEEE International Symposium on Industrial Electronics* (2010), pp. 1718–1751.

[Gra+06] T. Grasser, R. Entner, O. Triebl, H. Enichlmair, and R. Minixhofer. "TCAD Modeling of Negative Bias Temperature Instability". In: *SISPAD* (Sept. 2006), pp. 330–333. DOI: 10.1109/SISPAD.2006.282902.

[Gra+09] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel. "A Two-Stage Model for Negative Bias Temperature Instability". In: *IEEE International Reliability Physics Symposium* (2009).

[Gra+10] T. Grasser, H. Reisinger, P.-J. Wagner, and F. Schanovsky. "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability". In: *IEEE International Reliability Physics Symposium* (2010), pp. 16–25.

[Gra+11] T. Grasser, B. Kaczer, W. Goes, and H. Reisinger. "The Paradigm Shift in Understanding the Bias Temperature Instability: From Reaction–Diffusion to Switching Oxide Traps". In: *IEEE Transactions on Electron Devices* (2011), pp. 3652–3666.

[Gra+12] T. Grasser, B. Kaczer, H. Reisinger, and P.-J. Wagner. "On the frequency dependence of the bias temperature instability". In: *IEEE International Reliability Physics Symposium (IRPS)* (2012).

[Hau14] F. Hausdorff. *Grundzüge der Mengenlehre*. Das Hauptwerk von Felix Hausdorff. Leipzig: Veit and Company, 1914.

[HL99] C. Hu and Q. Lu. "A unified Gate oxide Reliability model". In: *IEEE International Reliability Physics Symposium* (1999), pp. 47–51.

[IEC14] I. E. C. (IEC). *International electrotechnical vocabulary - Part 351: Control technology (IEC 60050-351:2013)*. Standard. International Organization for Standardization, Sept. 2014.

[Joh] S. G. Johnson. *The NLopt nonlinear-optimization package*. URL: http://ab-initio.mit.edu/nlopt.

[JS05] N. K. Jha and D. K. S. V. R. Sahajananda Reddy. "NBTI Degradation and Its Impact for Analog Circuit Reliability". In: *IEEE transactions on Electron Devices* (2005).

[JS77] K. O. Jeppson and C. M. Svensson. "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices". In: *Journal of Applied Physics* 48.5 (1977), pp. 2004–2014. DOI: 10.1063/1.323909. URL: http://link.aip.org/link/?JAP/48/2004/1.

[KA86] H. Küflüoglu and M. A. Alam. "A Generalized Reaction–Diffusion Model With Explicit H–H2 Dynamics for Negative-Bias Temperature-Instability (NBTI) Degradation". In: *IEEE Transactions on Electron Devices* 33.11 (1986), pp. 1754–1768.

[Kac+08] B. Kaczer, T. Grasser, P. J. Roussel, J. Martin-Martinez, R. O'Connor, B. O'Sullivan, and G. Groeseneken. "Ubiquitous relaxation in BTI stressing—-new evaluation and insights". In: *IEEE International Reliability Physics Symposium (IRPS)* (2008).

[KGA13] A. E. Khorasani, M. Griswold, and T. Alford. "A Fast $I-V$ Screening Measurement for TDDB Assessment of Ultra-Thick Inter-Metal Dielectrics". In: *Electron Device Letters, IEEE* 35.1 (Dec. 2013), pp. 117–119.

[KKS06] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar. "An Analytical Model for Negative Bias Temperature Instability". In: *ICCAD '06 Proceedings of the 2006 IEEE/ACM internationnal conference on Computer-aided design* (2006).

[KP11] P. K. Krause and I. Polian. "Adaptive Voltage Over-Scaling for Resilient Applications". In: *DATE '11: Design, Automation and Test in Europe* (2011).

[LMM06] Z. Liu, B. W. McGaughy, and J. Z. Ma. "Design Tools For Reliability Analysis". In: *DAC '06: Design Automation Conference* (2006), pp. 182–187.

[LQB08] X. Li, J. Qin, and J. B. Bernstein. "Compact Modeling of MOSFET Wearout Mechanism for Circuit-Reliability Simulation". In: *Device and Materials Reliability, IEEE Transactions on* 8.2 (Mar. 2008), pp. 98–121.

[Mar+06] F. Marc, B. Mongellaz, C. Bestory, H. Lévi, and D. Y. "Improvement of Aging Simulation of Electronic Circuits Using Behavioral Modeling". In: *Device and Materials Reliability, IEEE Transactions on* 6.2 (June 2006), pp. 228–233.

[McC+09] T. McConaghy, P. Palmers, M. Steyaert, and G. G. E. Gielen. "Variation-Aware Structural Synthesis of Analog Circuits via Hierarchical Building Blocks and Structural Homotopy". In: *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 28.9 (2009), pp. 1281–1294. ISSN: 0278-0070.

[Mei+12] M. Meissner, O. Mitea, L. Luy, and L. Hedrich. "Fast isomorphism testing for a graph-based analog circuit synthesis framework". In: *Design, Automation Test in Europe Conference Exhibition (DATE), 2012*. Mar. 2012, pp. 757–762.

[MG13] E. Maricau and G. Gielen. *Analog IC Reliability in Nanometer CMOS*. Springer, 2013.

[MH15] M. Meissner and L. Hedrich. "FEATS: Framework for Explorative Analog Topology Synthesis". In: *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 34.2 (Feb. 2015), pp. 213–226. ISSN: 0278-0070.

[Moo65] G. Moore. "Cramming More Components Onto Integrated Circuits". In: *Electronics* 38.8 (1965), pp. 114–117.

[Nir+14] T. Nirmaier, A. Burger, M. Harrant, A. Viehl, O. Bringmann, W. Rosenstiel, and G. Pelz. "Mission Profile Aware Robustness Assessment of Automotive Power Devices". In: *Proc. Design, Automation and Test in Europe DATE '14*. 2014.

[Ohl+08] O. Ohlendorf, S. Steinhorst, W. Hartong, and L. Hedrich. "Comparing Two Analog Waveforms - A Trivial Task?" In: *6. GMM/GI/ITG-Fachtagung für Zuverlässigkeit und Entwurf(ZuE 2008)*. Sept. 2008, pp. 153–154.

[OS95] S. Ogawa and N. Shiono. "Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-$SiO2$ interface". In: *Phys. Rev. B* 51.7 (Feb. 1995), pp. 4218–4230.

[OSC05] OSCI. *SystemC 2.1 Language Reference Manual*. 2005.

[Pau+05] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy. "Impact of NBTI on the Temporal Performance Degradation of Digital Circuits". In: *IEEE Electron Device Letter* 26.8 (2005), pp. 560–562.

[PCP92] J. Powell, U. of Cambridge. Department of Applied Mathematics, and T. Phyiscs. *A Direct Search Optimization Method that Models the Objective and Constraint Functions by Linear Interpolation*. University of Cambridge, Department of Applied Mathematics and Theoretical Physics, 1992. URL: http://books.google.de/books?id=rcw%5C_MwEACAAJ.

[Pek+04] S. D. Pekarek, O. Wasynczuk, E. A. Walters, J. V. Jatskevich, C. E. Lucas, N. Wu, and P. T. Lamm. "An Efficient Multirate Simulation Technique for Power-Electronic-Based Systems". In: *IEEE Transactions on Power Systems* 19.1 (2004).

[Per09] C. S. Perone. "Pyevolve: a Python open-source framework for genetic algorithms". In: *SIGEVOlution* 4.1 (Nov. 2009), pp. 12–20. ISSN: 1931-8499. URL: http://doi.acm.org/10.1145/1656395.1656397.

[Pfä+12] P. Pfäffli, P. Tikhomirov, X. Xu, I. Avci, Y.-S. Oh, P. Balasingam, S. Krishnamoorthy, and T. Ma. "TCAD reliability". In: *Microelectronics Reliability* 52 (2012), pp. 1761–1768.

[Pow98] M. J. D. Powell. *Direct Search Algorithms for Optimization Calculations*. 1998.

[Rad+10] M. Radetzki, O. Bringmann, W. Nebel, M. Olbrich, F. Salfelder, and U. Schlichtmann. "Robustheit nanoelektronischer Schaltungen und Systeme". In: *Zuverlässigkeit und Entwurf - 4. GMM/GI/ITG-Fachtagung* (2010).

[Ram+09] S. Ramey, C. Prasad, M. Agostinelli, S. Pae, S. Walstra, and J. H. S. Gupta. "Frequency and Recovery Effects in high-k BTI Degradation". In: *IEEE International Reliability Physics Symposium* (2009), pp. 1023–1027.

[ROB09] ROBUST. "Design of Robust Nanoelectronic Systems, Project funded by BMBF (label 01 M 3087), edacentrum". In: *http://www.edacentrum.de/robust* (2009).

[Ros+15] J. von Rosen, F. Salfelder, L. Hedrich, B. Betting, and U. Brinkschulte. "A highly dependable self-adaptive mixed-signal multi-core system-on-chip architecture". In: *Integration, the {VLSI} Journal* 48 (2015), pp. 55–71. ISSN: 0167-9260. URL: http://www.sciencedirect.com/science/article/pii/S0167926014000170.

[RT08] A. Roy and C. M. Tan. "Very high current density package level electromigration test for copper interconnects". In: *J. Appl. Phys.* (2008).

[Sch+10] C. Schluender, W. Gustin, H. Reisinger, and T. Grasser. "A new physics-based NBTI model for DC- and AC-Stress enabling accurate circuit aging simulations considering recovery". In: *Proc. Zuverlässigkeit und Entwurf (ITG-FB 231)* (2010), pp. 33–40.

[Sch06] D. K. Schroder. "Negative Bias Temperature Instability: What do we understand?" In: *Microelectronics Reliability* 47 (2006), pp. 841–852.

[SH11] F. Salfelder and L. Hedrich. "An NBTI Model for Efficient Transient Simulation of Analogue Circuits". In: *Proc. edaWorkshop 11*. VDE Verlag, 2011, pp. 27–32.

[SH15] F. Salfelder and L. Hedrich. "Ageing Simulation of Analogue Circuits and Systems using Adaptive Transient Evaluation". In: *Design, Automation Test in Europe Conference Exhibition (DATE), 2015*. Mar. 2015.

[She+87] B. J. Sheu, D. L. Scharfetter, P.-K. Ko, and M.-C. Jeng. "BSIM: Berkeley short-channel IGFET model for MOS transistors". In: *IEEE Journal Solid-State Circuits* SC-22 (1987), pp. 558–566.

[SHM13] F. Salfelder, L. Hedrich, and M. Meissner. "Evaluating NBTI in Synthesized Operational Amplifiers using an Accurate Ageing Model". In: *Proc. ANALOG 2013, Aachen, Germany* (2013).

[Sim+11] E. Simoen, B. Kaczer, M. Toledano-Luque, and C. Claeys. "Random Telegraph Noise: From a Device Physicist's Dream to a Designer's Nightmare". In: *ECS Transactions, 39* (2011), pp. 3–15.

[SZ04] K. S.Kundert and O. Zinke. *The Designer's Guide to Verilog AMS*. Springer, 2004.

[Sze81] S. M. Sze. *Physics of Semiconductor Devices*. Wiley, 1981.

[Tsi99] Y. Tsividis. "Operation and Modeling of The MOS-Transistor". In: *McGraw-Hill Book Company* (1999).

[Tu+91] R. Tu, E. Rosenbaum, C. Li, W. Chan, P. Lee, B.-K. Liew, J. Burnett, P. Ko, and C. Hu. "BERT - Berkley Reliability Tools". In: *University of California, Berkley, Memorandum No. UCBERL M91/107* (1991). URL: http://www.eecs.berkeley.edu/Pubs/TechRpts/1991/ERL-91-107.pdf.

[Tu+93] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami, K. Quader, P. K. Ko, and C. Hu. "Berkeley Reliability Tools–BERT". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 12.10 (1993), pp. 1524–1534.

[Vla94] A. Vladimirescu. "The Spice Book". In: *John Wiley & Sons, Inc., New York* (1994).

[VS94] J. Vlach and K. Singhal. "Computer Methods for Circuit Analysis and Design - Second Edition". In: *Van Nostrand Reinhold, New York* (1994).

[VWC06] R. Vattikonda, W. Wang, and Y. Cao. "Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design". In: *DAC6* (2006), pp. 1047–1052. URL: http://ptm.asu.edu/references/DAC_2006_NBTI.pdf.

[Wan+07] W. Wang, V. Reddy, A. T. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao. "Compact Modeling and Simulation of circuit Reliability for 65-nm CMOS Technology". In: *Device and Materials Reliability, IEEE Transactions on* 6 (Dec. 2007), pp. 517–517.

[Wan+11] L. Wang, B. M. A.-H. T. J. Kazimierski, A. S. Weddell, G. V. Merrett, and I. N. A. Garcia. "Accelerated simulation of tunable vibration energy harvesting systems using a linearised state-space". In: *Design, Automation Test in Europe Conference Exhibition (DATE), 2011.* 2011.

[Wit07] R. Wittmann. "Miniaturization Problems in CMOS Technology: Investigation of Doping Profiles and Reliability". In: *Dissertation, Technische Universität Wien* (2007).

[WWS86] W. Weber, C. Werner, and A. v. Schwerin. "Lifetimes and Substrate Currents in Static and Dynamic Hot-Carrier Degradation". In: *IEEE* (1986).

[Yil+13] C. Yilmaz, L. Heiß, C. Werner, and D. Schmitt-Landsiedel. "Modeling of NBTI-Recovery Effects in Analog CMOS Circuits". In: *IEEE International Reliability Physics Symposium* (2013).

[YN00] G. Yoh and F. N. Najm. "A statistical model for electromigration failures". In: *Quality Electronic Design (ISQED), 2000 First International Symposium on* (2000), pp. 45–50.