

Statische Timinganalyse unter Berücksichtigung
kapazitiver Kopplung

Von der Fakultät für Elektrotechnik und Informatik
der Universität Hannover
zur Erlangung des akademischen Grades
Doktor-Ingenieur
genehmigte Dissertation

von Dipl.-Ing. Matthias Ringe

geboren am 16. April 1970 in Minden

2006

Statische Timinganalyse unter Berücksichtigung
kapazitiver Kopplung

Von der Fakultät für Elektrotechnik und Informatik
der Universität Hannover
zur Erlangung des akademischen Grades
Doktor-Ingenieur
genehmigte Dissertation

von Dipl.-Ing. Matthias Ringe

geboren am 16. April 1970 in Minden

2006

1. Referent: Prof. Dr.-Ing. Erich Barke
 2. Referent: Prof. Dr.-Ing. Claus-Eberhard Liedtke
- Tag der Promotion: 23. Mai 2006

Vorwort

Diese Arbeit entstand zu großen Teilen bei der Robert Bosch GmbH in Reutlingen.

Die universitäre Betreuung dieser Arbeit erfolgte durch Herrn Prof. Dr.-Ing. E. Barke. Ihm danke ich ganz besonders für die Erstellung des Erstgutachtens sowie seine hervorragende Betreuung und seine fachlichen Anregungen.

Ich danke Herrn Prof. Dr.-Ing. C.-E. Liedtke für die Übernahme des Korreferates und Herrn Prof. Dr. R. Parchmann für den Vorsitz der Prüfung.

Mein Dank gilt Herrn T. Lindenkreuz für die Betreuung dieser Arbeit seitens der Robert Bosch GmbH, insbesondere für seine fachliche und persönliche Hilfe.

Ebenso danke ich Herrn Dr. P. van Staa für die technischen und organisatorischen Möglichkeiten, die er mir als Abteilungsleiter eingeräumt hat. Ich danke allen Mitarbeiterinnen und Mitarbeitern der Abteilung K8/DIC der Robert Bosch GmbH in Reutlingen für die Unterstützung und die freundliche Aufnahme.

Matthias Ringe

Kurzfassung

Durch die Verkleinerung der Strukturen auf integrierten Digitalisierungen in CMOS-Technologie kommt den Auswirkungen der kapazitiven Kopplung erhöhte Bedeutung zu. Hierzu zählt insbesondere die vergrößerte Verzögerung bei entgegengesetztem Schalten benachbarter Leitungen. Das Zeitverhalten einer Digitalisierung wird heute meist durch statische Timinganalyse verifiziert.

In dieser Arbeit werden zwei Verzögerungsmodelle für kapazitive Kopplung vorgeschlagen. Diese Modelle sind in ein statisches Timinganalyseprogramm auf Transistorebene integriert. Ebenfalls integriert ist ein häufig verwendetes, in der Literatur beschriebenes Modell.

Die Validierung der vorgeschlagenen Verfahren erfolgt durch Vergleich mit Analogsimulationen. Dazu werden Schaltungen in einer 0,5 μm - und 130 nm-Technologie der statischen Timinganalyse und Analogsimulationen mit Kopplung unterzogen. Das Ziel der Analogsimulationen ist, durch Variierung der Ankunftszeiten der Aggressoren die maximale Gatter- beziehungsweise Pfadverzögerung zu ermitteln. Im Falle der Validierung der Gatterverzögerung werden einzelne Gatter in den oben genannten Technologien unter typischen Betriebsbedingungen verwendet. Für die Pfadverzögerung werden verdrahtete und extrahierte Schaltungen untersucht.

Es wird gezeigt, dass beide neuen Verzögerungsmodelle für Kopplung bessere Vorhersagen der Verzögerung als das zum Vergleich herangezogene Modell liefern. Ebenfalls wird gezeigt, dass der Rechenaufwand des einen Modells geringer als des Vergleichsmodells ist.

Schlüsselwörter: CMOS, statische Timinganalyse, kapazitive Kopplung

Abstract

Due to the reduction of feature sizes in integrated digital circuits in CMOS technology the impact of capacitive coupling becomes increasingly important. This applies in particular to opposite mode switching of adjacent wires. Today, the temporal behavior of a digital circuit is mostly verified by static timing analysis.

In this work two delay models for capacitive coupling are proposed. Both models are integrated in a transistor-level static timing analysis program. A popular timing model described in the literature is also incorporated.

The validation of the proposed methods is carried out through comparisons to analog simulations. For this purpose designs in a 0.5 μm and 130 nm technology with coupling are analyzed using both the static timing analysis and analog simulations. The target of the analog simulations is to maximize the path and the gate delay, respectively by varying the arrival time of the aggressors. In case of the validation of the gate delay single gates are used using typical operating conditions. Wired and extracted designs are used for validating the path delay.

It is demonstrated that both new delay models for coupling predict the delay more accurately than the reference model. Moreover, it is also shown that the computational run time of one of the new models is smaller than the one of the reference model.

Keywords: CMOS, static timing analysis, capacitive coupling

Inhaltsverzeichnis

1	Einleitung	1
2	Statische Timinganalyse	3
2.1	Einführung	3
2.2	Mathematische Grundlagen	4
2.2.1	Graphentheorie	4
2.2.2	Boolesche Algebra	5
2.2.3	Komplexitätstheorie	6
2.3	Grundlagen der Timinganalyse	8
2.4	Abbildung einer Schaltung	12
2.5	Suche der kritischen Pfade	15
2.5.1	Problemstellung	15
2.5.2	Breitensuche	16
2.5.3	Anwendung der Breitensuche in der statischen Timing- analyse	19
2.6	Problem der falschen Pfade	21
2.7	Grenzen der statischen Timinganalyse	22
3	Verzögerungsmodelle	25
3.1	Einführung	25
3.2	Verzögerung	25
3.3	Verzögerungsmodelle auf Gatterebene	28
3.3.1	Aufgabenstellung	28
3.3.2	Modellierung der Ein- und Ausgangssituation	30
3.3.3	MOS-Transistoren	32
3.3.4	Transistorbasierte Verzögerungsmodelle	35
3.3.5	Empirische Gattermodelle	37
3.4	Leitungsmodelle	39
3.5	Modellierung der Verzögerung bei kapazitiver Kopplung	41

3.5.1	Problemstellung	41
3.5.2	Passive Modelle	43
3.5.3	Aktive Modelle	46
3.6	Kapazitive Kopplung in der statischen Timinganalyse	47
4	Kurvenformbasierte statische Timinganalyse	51
4.1	Einführung	51
4.2	Modellierung der Schaltungselemente	51
4.3	Verbesserte Propagierung von Spannungskurven	55
5	Verzögerung bei kapazitiver Kopplung	59
5.1	Theoretische Betrachtungen	59
5.2	Verzögerungsmodelle	69
5.3	Empirische Validierung	73
5.3.1	Erstes Verzögerungsmodell	73
5.3.2	Zweites Verzögerungsmodell	76
5.3.3	Bewertung	78
5.4	Berücksichtigung kapazitiver Kopplung in der statischen Timinganalyse	79
6	Ergebnisse	83
6.1	Voraussetzungen	83
6.2	Validierung der statischen Timinganalyse und der Verzögerungsmodelle	85
6.2.1	Passive Verzögerungsmodellierung	85
6.2.2	Erstes Verzögerungsmodell	87
6.2.3	Zweites Verzögerungsmodell	88
6.2.4	Verzögerungsmodell nach Chen, Kirkpatrick und Keutzer	92
6.2.5	Bewertung	93
6.3	Vergleich der Verzögerungsmodelle	94
6.3.1	Kopplung ohne Berücksichtigung der Aggressorsituation	94
6.3.2	Kopplung unter Berücksichtigung der Aggressorsituation	95
6.3.3	Bewertung	97
6.4	Verbesserte Spannungskurvenpropagierung	99
6.5	Grenzen der vorgeschlagenen Verfahren	99
7	Zusammenfassung	101

Literaturverzeichnis	103
Index	111

Formelzeichen

\oplus	Exklusiv-Oder-Operator
A, B, \dots	Bezeichner von Primärein- und -ausgängen
a, b, \dots	Bezeichner von Netzen
C	Kapazitätswert
C_C	Koppelkapazität
C_{Ersatz}	Ersatzkapazität
C_{GND}	Kapazität, an Masse angeschlossen
δ	Grad eines Knotens
d	Tiefe eines Knotens
E	Menge der Kanten eines Graphen
f	Funktion, boolesche Funktion
f_{nc}	Nicht-kontrollierende Funktion
G	Graph
h, h_n	Zeitschrittweite des Analogsimulators
$h(t)$	Impulsantwort linearer, zeitinvarianter Systeme
I	Strom, nicht zeitabhängig
$I(s)$	Laplacetransformierter Strom
I_D	Drainstrom
I_G	Gatestrom
$i, i(t)$	Strom, zeitabhängig
i_{aus}	Gatterausgangsstrom
i_V	Strom des Opfergatters
k	Miller-Faktor
k_C	Konstante, die im K-Faktor-Modell die Abhängigkeit der Verzögerung von der Ausgangskapazität modelliert
k_R	Konstante, die im K-Faktor-Modell die Abhängigkeit der Verzögerung von der Anstiegszeit modelliert
l	Länge eines Pfades

NP	Menge der nichtdeterministisch in polynomialer Zeit lösbarer Probleme
P	Pfad
P	Menge der deterministisch in polynomialer Zeit lösbarer Probleme
R	Widerstand
t	Zeit
T, T_n	Zeitpunkte
T_ϕ	Taktperiode
T_{AT}	Ankunftszeit
T_C	Kopplungszeitpunkt
T_D	Verzögerung, Elmore-Verzögerung
$T_{D,i}$	Last- und ansteigszeitunabhängige Verzögerung
T_{Max}	Endzeitpunkt eines Kopplungsereignisses, $T_{Max} = T_C + T_R$
T_R	Anstiegs- bzw. Abfallzeit
T_{RAT}	Notwendige Ankunftszeit
T_{Slack}	Slack
U, U_n	Spannung, zeitunabhängig
$U(s)$	laplacetransformierte Spannung
U_D	Drain-Substrat-Spannung
U_{DD}	Versorgungsspannung
U_{DS}	Drain-Source-Spannung
U_G	Gate-Substrat-Spannung
U_{GS}	Gate-Source-Spannung
U_P	Extremwert einer Spannung
U_S	Source-Substrat-Spannung
$U_{th,n}, U_{th,p}$	Schwellspannung von N- und P-MOS-Transistoren, $U_{th,p} < 0$ V
$U_{th,u}, U_{th,o}$	Untere und obere Schwellspannung zur Berechnung der Anstiegszeit bzw. zur Definition des Verzögerungsmodells bei vorhandener Kopplung
$U_{th}, U_{th,m}$	Schwellspannung zur Berechnung bzw. zum Vergleich von Verzögerungen
ΔU	Spannungsdifferenz
$u, u(t)$	Spannung, zeitabhängig
$\dot{u} = \frac{du}{dt}$	Zeitliche Ableitung

u_A	Spannung eines Aggressorschaltungsknotens
u_{ein}	Spannung an einem Gattereingang
u_{aus}	Spannung an einem Gatterausgang
u_V	Spannung eines Opferschaltungsknotens
V	Menge der Knoten eines Graphen
v, v_n	Bezeichner für Knoten eines Graphen
Y	Admittanz



1 Einleitung

Integrierte Digitalschaltungen werden heute überwiegend in statischer CMOS-Schaltungstechnik realisiert. Die Aufgabe des Entwurfs einer digitalen Schaltung besteht darin, eine vorgegebene Spezifikation in einen Schaltkreis zu überführen, der diese Spezifikation erfüllt. Da die Fertigung sehr kostenintensiv ist, ist es sinnvoll, vorher Prüfungen anzuwenden, die die Erfüllung der Spezifikation sicherstellen. Vor der Fertigung muss dazu auf Modelle der Schaltung zurückgegriffen werden. Diese Modelle können verschiedenen Sichtweisen und Abstraktionsebenen angehören [1].

Grundsätzlich ist jeder Entwurfsschritt auf Korrektheit zu prüfen, was auch die Äquivalenz der Schaltungsbeschreibungen vor und nach dem Entwurfsschritt einschließt. Ist die angewendete Prüfungsmethode erschöpfend, das heißt, werden alle möglichen Betriebsarten und Eingangsmuster der Schaltung erfasst, spricht man von *Verifikation*. Werden nur bestimmte, repräsentative Muster berücksichtigt, handelt es sich um eine Validierung. Eine erfolgreiche Validierung ist für die korrekte Umsetzung eines Entwurfsschrittes nicht hinreichend.

Der Prüfung des physikalischen Entwurfs, also der Layoutgenerierung, kommt besondere Bedeutung zu, da hier nicht erwünschte aber unvermeidliche Schaltungselemente, die so genannten »Parasiten«, auftreten. Mittels der Parasitenextraktion können die Leitbahnen der physikalischen Sicht durch Kapazitäten und Widerstände modelliert werden. Die Parasiten haben wesentliche Auswirkungen auf die Verzögerung und die Leistungsaufnahme.

Die Validierung der Layoutgenerierung kann durch eine ereignisgesteuerte Digitalsimulation erfolgen. In diesem Fall werden Funktion und Zeitverhalten gleichzeitig geprüft [2][3]. Da dieses Validierungsverfahren sehr rechenintensiv ist, werden meist Funktion und Zeitverhalten getrennt geprüft. Die Prüfung des Zeitverhaltens erfolgt durch die statische Timinganalyse.

Die statische Timinganalyse hat zwei Schlüsseigenschaften:

1. Sie ist ein Verfahren zur Verifikation des Zeitverhaltens einer Schaltung.
2. Ihre Komplexität ist linear abhängig von der Schaltungsgröße.

In statischer CMOS-Schaltungstechnik ist die Gatterverzögerung, das ist die Zeit, die ein Gatter nach Änderung eines Eingangs benötigt, um einen veränderten stabilen Ausgangswert einzustellen, von der zu treibenden Kapazität abhängig [2]. Die klassische Verzögerungsmodellierung erfordert, dass die Kapazität an den Masseschaltungsknoten angeschlossen ist. Durch die fortschreitende Miniaturisierung steigt der Anteil der Koppelkapazität, also der Kapazität, die nicht an den Masseknoten, sondern an einen anderen Leitungsknoten, angeschlossen ist. Hierfür gibt es folgende Gründe:

1. Die Anzahl der Metallisierungslagen steigt. Eine $1\ \mu\text{m}$ -Technologie hat zwei, eine $130\ \text{nm}$ -Technologie hat bis zu acht Lagen. Der Abstand zum Substrat höherer Lagen steigt. Die Massekapazität verringert sich.
2. Der Abstand zweier benachbarter Leitungen derselben Lage ist verringert, die Höhe bleibt im Wesentlichen unverändert.

Die Koppelkapazität kann daher erheblichen Einfluss auf die Verzögerung des einzelnen Gatters und somit auf das Zeitverhalten der gesamten Schaltung haben. Die vorliegende Arbeit hat folgende Ziele:

- Die Entwicklung von Verzögerungsmodellen, die den Einfluss von Koppelung auf das Zeitverhalten gut wiedergeben.
- Die Integration dieser Modelle in eine statische Timinganalyse.
- Die Validierung der Verzögerungsmodelle und der statischen Timinganalyse durch Analogsimulationen.

2 Statische Timinganalyse

2.1 Einführung

In diesem Kapitel soll der Stand der Technik algorithmischer Aspekte der statischen Timinganalyse dargestellt werden. Dazu wird vorausgesetzt, dass die Verzögerung ein bereits berechneter, konstanter Zeitwert ist. Die Verzögerung beeinflussende physikalische Größen, wie zum Beispiel Anstiegszeit oder der Verlauf der Spannungskurve, werden in diesem Kapitel nicht behandelt. Alle Verzögerungen sind entweder ein fester Zeitwert oder sichere zeitliche Ober- und Untergrenzen. Die Darstellung der Berechnung dieser Werte erfolgt in Kapitel 3.

Die Obergrenze der Verzögerung eines Gatters ist die Zeit, die höchstens vergeht, bis der Ausgang nach einer Änderung an einem Eingang den neuen Wert stabil erreicht hat. Die Untergrenze der Verzögerung ist die minimale Zeit, die der Gatterausgang den alten Wert behält. Verzögerung von Gattern meint die Zeit, die zum Erreichen des neuen Zustandes des Ausgangs erforderlich ist, Verzögerung von allgemeinen kombinatorischen oder sequentiellen Schaltungen die Zeit, in der sich Primärausgänge und synchrone Eingänge von Registern nach einer Änderung auf einen stabilen Wert eingestellt haben.

Grundsätzlich sind die in diesem Kapitel angestellten Betrachtungen unabhängig von der Abstraktionsebene und dem der Analyse zugrunde liegenden Verzögerungsmodell. Im Folgenden werden Gatter und Leitungen als verzögerungsbehaftete Elemente verwendet. Transistoren, Makroblöcke oder Verhaltensmodelle mit Verzögerungsinformation sind jedoch ebenso möglich. Zur Diskussion der Abstraktionsebenen siehe [2] und [4].

2.2 Mathematische Grundlagen

2.2.1 Graphentheorie

Ein *Graph* [5] ist Paar $G = (V, E)$, wobei V eine Menge von *Knoten* und E eine Menge von *Kanten* ist. Man unterscheidet gerichtete und ungerichtete Graphen. Der ungerichtete Graph G ist durch die Angabe der Menge seiner Knoten V und Kanten E vollständig beschrieben. Eine Kante ist eine Menge von ein oder zwei Knoten und stellt die Verbindung zwischen zwei Knoten p und q dar, wobei $p = q$ ebenfalls zulässig ist. In diesem Fall $p = q$ heißt die Kante *Schlinge*.

Für einen *gerichteten Graphen* wird darüber hinaus verlangt, dass die Kanten eine Ordnung beinhalten, die die Richtungsinformation darstellt. Eine gerichtete Kante $E_1 = (p, q)$ ist bezüglich der gerichteten Kante $E_2 = (q, p)$ unterschiedlich, falls $p \neq q$. Der *Grad* δ eines Knotens ist die Anzahl der inzidenten Kanten, wobei Schlingen doppelt gezählt werden. In gerichteten Graphen unterscheidet man Ein- und Ausgangsgrad.

Ein *Pfad* ist eine Abfolge von Knoten (v_0, \dots, v_{n-1}) , die durch Kanten miteinander verbunden sind, also $(v_{i-1}, v_i) \in E$ mit $i \in \{1, \dots, n-1\}$.

Ein *Kreis* ist ein Pfad (v_0, \dots, v_{n-1}) , für den $v_0 = v_{n-1}$ gilt. Für einen *Zyklus* gilt darüber hinaus, dass außer v_0 und v_{n-1} kein Knoten mehrfach im Pfad vorhanden ist. Ein Kreis enthält somit mindestens einen Zyklus. Graphen, deren Menge aller Pfade keine Zyklen enthält, heißen *azyklische Graphen*. Die Methoden der statischen Timinganalyse setzen die Abbildung einer Schaltung auf einen *gerichteten, azyklischen Graphen* voraus.

Knoten eines gerichteten Graphen mit dem Eingangsgrad null heißen *Startpunkte*. *Endpunkte* sind entsprechend Knoten mit dem Ausgangsgrad null. Die maximale Anzahl Knoten über alle Pfade von jedem Startpunkt zu einem Knoten wird als *Tiefe* bezeichnet. Startpunkte haben die Tiefe 0. Ebenfalls werden der erste und letzte Knoten eines Pfades oder Teilpfades als Start- und Endpunkt bezeichnet.

Bei einem gewichteten Graphen sind die Kanten mit einem reellen Zahlenwert, dem *Gewicht*, versehen. In gewichteten Graphen kann es mehrere Kanten unterschiedlichen Gewichtes zwischen zwei Knoten geben. In diesem Fall ist die Angabe der Kante durch deren Knoten nicht eindeutig, weshalb die Kanten mit

denselben Knoten enumeriert werden. Die *Länge* l eines Pfades (v_0, \dots, v_{n-1}) ist die Summe der Gewichte der Kanten (v_{i-1}, v_i) , die diesen Pfad bilden.

Ein gerichteter, azyklischer Graph beinhaltet *Rekonvergenz*, wenn es zwischen zwei Knoten mehrere, unterschiedliche Pfade gibt.

2.2.2 Boolesche Algebra

Eine *boolesche Variable* x ist ein Platzhalter und kann die Werte 0 und 1 annehmen. Ein *Literal* ist eine boolesche Variable x oder ihr Komplement \bar{x} . Auf booleschen Variablen ist die boolesche Algebra definiert. Näheres dazu findet sich in [5], [6].

Eine boolesche Funktion bildet ein Tupel aus mehreren booleschen Variablen auf eine boolesche Variable ab. Eine *Konjunktion* ist die logische Und-Verknüpfung, eine *Disjunktion* die logische Oder-Verknüpfung. Boolesche Funktionen lassen sich unter anderem als *Summenform*, der Disjunktion von Konjunktionen, oder als *Produktform*, der Konjunktion von Disjunktionen von Literalen darstellen. Mit der zusätzlichen Forderung, dass in jedem Konjunktionsterm der Summenform bzw. in jedem Disjunktionsterm der Produktform ein Literal höchstens einfach auftritt, erhält man die *disjunktive* (DNF) bzw. *konjunktive Normalform* (KNF). Die Notation einer Funktion in konjunktiver Normalform ist:

$$y = f(x_0, \dots, x_{n-1}) = \prod_{i=0}^{m-1} \sum_{j=0}^{n_i-1} l_j, \text{ mit } l_j \in \{x_j, \bar{x}_j\} \quad (2.1)$$

Dabei ist m die Anzahl der Summenterme. Die Bildung der disjunktiven Normalform erfolgt analog, mit Produkt- und Summenbildung miteinander vertauscht. Die Darstellung einer booleschen Funktion mit Hilfe der oben genannten Normalformen ist im Allgemeinen nicht kanonisch, da dieselbe Funktion mehrere Darstellungsformen haben kann. Dies gilt insbesondere für die Darstellung der booleschen 1 und 0 in disjunktiver beziehungsweise in konjunktiver Normalform.

Die Suche nach einem Variablen-tupel, das eine boolesche Funktion in KNF-Notation auf den Funktionswert 1 setzt, ist als *Erfüllbarkeitsproblem* bekannt.

Die *nicht-kontrollierende* Funktion f_{nc} bezüglich einer booleschen Funktion $f(a, r_0, \dots, r_{n-1})$ und einer Funktionsvariablen a ist definiert als:

$$f_{nc}(f, a) = f(a = 0, r_0, \dots, r_{n-1}) \oplus f(a = 1, r_0, \dots, r_{n-1}) \quad (2.2)$$

Beispiel:

$$f(a, b, c) = abc \quad (2.3)$$

$$f_{nc}(f, a) = f(a = 0, b, c) \oplus f(a = 1, b, c) \quad (2.4)$$

$$= 0 \oplus bc \quad (2.5)$$

$$= bc \quad (2.6)$$

Ist die nicht-kontrollierende Funktion als Und-Verknüpfung der Restvariablen r_0, \dots, r_{n-1} darstellbar und gilt $f_{nc} = 1$, so ist es üblich zu sagen, dass die Restvariablen einen nicht-kontrollierenden Wert einstellen müssen [7] [8]. Im obigen Beispiel gilt $b = 1$ und $c = 1$. $f_{nc} = 0$ bedeutet, dass eine oder mehrere Restvariablen einen kontrollierenden Wert haben.

2.2.3 Komplexitätstheorie

Die *Ordnung* O ist ein qualitatives Maß für den Verlauf einer Funktion in Abhängigkeit einer oder mehrerer Eingangsvariablen. Eine Funktion f hat die Ordnung von g , wenn es Konstanten $c, n_0 \in \mathbb{N}$ gibt, so dass für alle $n \in \mathbb{N}$, $n \geq n_0$ gilt: $|f(n)| \leq c|g(n)|$.

Die Ordnung eines Algorithmus ist ein Kriterium zur Bewertung seiner Laufzeit oder seines Speicherbedarfs. Wünschenswert sind Algorithmen mit konstanter, linearer oder zumindest polynomialer Laufzeit, also $t(n) = O(n^p)$, $p \in \mathbb{R}$, $p \geq 0$. Entscheidungsprobleme, die durch Algorithmen polynomialer Laufzeit lösbar sind, bezeichnet man als *P-Probleme*. Hat die Laufzeit eines Algorithmus exponentielle Ordnung $t(n) = O(e^n)$, so sind schon für kleine n hohe Rechenzeiten erforderlich. Probleme dieser Art nennt man *unzugänglich* [4]. Ein tabellarischer Vergleich der Rechenzeit von Algorithmen in [4] verdeutlicht die Problematik exponentieller Ordnung.

Ein unzugängliches Problem ist das Erfüllbarkeitsproblem (Abschnitt 2.2.2). Hierfür ist bisher noch kein Algorithmus, der eine polynomiale Ordnung besitzt, gefunden worden. Gleiches gilt für zahlreiche andere Probleme und führte zur Definition *nichtdeterministisch* arbeitender Verfahren. Es werden zunächst ausschließlich Entscheidungsprobleme behandelt.

Ein nichtdeterministischer Algorithmus wählt zufällig ein Variablen-tupel, auch Konfiguration genannt, des Problems aus und prüft, ob diese Konfiguration das Entscheidungsproblem löst. Ist die zeitliche Abhängigkeit von der Problemgröße der Prüfung der Lösung polynomial, so handelt es sich um ein *NP*-Problem, also ein *nichtdeterministisch* in *polynomialer* Zeit lösbares Problem.

Ein nichtdeterministischer Rechner ist ein Modell zur Abgrenzung von deterministischen und nichtdeterministischen Algorithmen ohne praktische Bedeutung. Er ist als Rechner vorstellbar, der ein Lösungstupel des Problems rät und den Beweis, dass es sich um die Lösung handelt, in polynomialer Zeit erbringt.

Es ist einsichtig, dass deterministisch zugängliche Probleme auch nichtdeterministisch zugänglich sind, also $P \subseteq NP$. Umformuliert bedeutet das, dass bestimmte NP-Probleme auch deterministisch in polynomialer Zeit lösbar sind. Die interessante Frage ist, ob dies für alle NP-Probleme gilt. Da zahlreiche Probleme bisher nichtdeterministisch in polynomialer Rechenzeit gelöst wurden, hat man die Klasse der *np-vollständigen* Probleme definiert. Die Definition der NP-Vollständigkeit kann in [9] nachgelesen werden. Der Beweis, dass ein neues Problem np-vollständig ist, erfolgt in zwei Stufen: 1. Das Problem muss in *NP* liegen. 2. Ein bekanntes, np-vollständiges Problem muss sich durch polynomiale Transformationen auf das neue Problem abbilden lassen.

Probleme, die keine Entscheidungsprobleme sind, deren Komplexität jedoch mindestens so groß wie die eines np-vollständigen Problems sind, heißen *np-harte* Probleme.

Viele in der Realität auftretende Probleme sind np-vollständig oder np-hart[4][5]. Dadurch, dass ihre Konfigurationen leicht prüfbar sind, eröffnen sie Möglichkeiten, gerichtete Suchverfahren anzuwenden. So gibt es für das Erfüllbarkeitsproblem effiziente Heuristiken [10], deren Laufzeit im Allgemeinen deutlich unter der durch $O(e^n)$ zu erwartenden Laufzeit liegt. Allen diesen Heuristiken ist jedoch gemein, dass in jedem Fall Probleminstanzen angebar sind, die eine exponentiell von der Anzahl der Eingangsvariablen abhängige Rechenzeit erfordern.

2.3 Grundlagen der Timinganalyse

Diese Arbeit beschränkt sich auf synchrone Digitalschaltungen in statischer CMOS-Technik [2]. Die Beschreibung der Schaltungen erfolgt durch eine Netzliste. Eine Netzliste enthält Schaltungselemente und Schaltungsknoten. Der Begriff »Schaltungsknoten« wird zur Unterscheidung von den Knoten eines Graphen verwendet.

Schaltungsknoten tragen eine zeitabhängige Spannung $u(t)$, die zwischen 0 V und der Versorgungsspannung U_{DD} liegt, diese Grenzen auch kurzzeitig und mit kleinen Beträgen über- oder unterschreiten kann.

$U_{th,u}$ und $U_{th,o}$ sind Schwellspannungen, für die gilt: $0 < U_{th,u} < U_{th,o} < U_{DD}$. Ein Ereignis ist eine Abbildung $t \rightarrow u(t)$, für die gilt:

$$T_S \leq t \leq T_E \quad (2.7)$$

$$U_{th,u} \leq u(t) \leq U_{th,o} \quad (2.8)$$

$$u(t = T_S) = U_{th,u}, u(t = T_E) = U_{th,o} \quad \text{für steigende Ereignisse} \quad (2.9)$$

$$u(t = T_S) = U_{th,o}, u(t = T_E) = U_{th,u} \quad \text{für fallende Ereignisse} \quad (2.10)$$

Der Begriff »Flanke« wird synonym zu Ereignis gebraucht. Die Spannung $u(t)$ repräsentiert eine boolesche Variable x . Es gilt:

$$x = 0 \quad \text{für} \quad u(t) \leq U_{th,u} \quad (2.11)$$

$$x = 1 \quad \text{für} \quad u(t) \geq U_{th,o} \quad (2.12)$$

Im Bereich $U_{th,u} < u(t) < U_{th,o}$ ist der boolesche Wert von u undefiniert.

Ein Gatter ist die strukturelle Realisierung einer booleschen Funktion. Für jedes Gatter existiert eine Transistornetzliste. Ein Gatter hat mindestens jeweils einen Ein- und Ausgang. Zur Vereinfachung wird angenommen, dass es keine Gatter mit mehreren Ausgängen gibt, da diese sich auf Gatter mit einem Ausgang abbilden lassen. Es werden kombinatorische und sequentielle Elemente unterschieden.

Ein Ereignis an einem Eingang eines kombinatorischen Gatters führt entweder zu keiner Reaktion am Ausgang (in diesem Fall ist einer der weiteren Eingänge mit einem kontrollierenden Wert belegt) oder zu einem Ausgangereignis nachdem eine Verzögerungszeit verstrichen ist. Alle Eingänge eines kombinatorischen

Gatters sind asynchrone Eingänge, da Eingangsereignisse kausale Ursache von Ausgangsereignissen sein können.

Ändern sich zwei Gattereingänge kurz hintereinander, so können am Ausgang zwei Ereignisse mit unterschiedlicher Monotonie entstehen. Dieses Phänomen bezeichnet man als Glitch oder Hazard [11].

Sequentielle Gatter, auch Speicher oder Register genannt, lassen sich in zwei Gruppen einteilen: Flanken- und pegelgesteuerte Speicher. Erstere werden üblicherweise als »Flipflops«, letztere als »Latches« bezeichnet [2]. Beide Varianten können zwei Gruppen von Eingängen, asynchrone und synchrone Eingänge, besitzen. Es ist mindestens ein asynchroner Eingang, der Takteingang, vorhanden.

Die synchronen Eingänge haben keinen kausalen Bezug zum Ausgang. Das heißt, ein Ereignis an einem synchronen Eingang löst kein Ausgangsereignis aus. Ereignisse an synchronen Eingängen können jedoch den Speicherzustand modifizieren.

Im Gegensatz dazu kann ein Ereignis am Takteingang zu einer Ausgangsänderung führen. Der Takteingang hat jedoch keinen funktionalen Einfluss auf den Ausgang, da dieser vom Speicherzustand bestimmt wird. Der Takteingang hat die zweite Funktion, den Zustand der synchronen Eingänge in den Speicher zu übernehmen.

Beide bisher behandelten Eingänge unterscheiden sich von den Eingängen kombinatorischer Gatter, bei denen jeder Eingang sowohl funktionalen als auch kausalen Einfluss auf den Ausgang hat. Es gibt Bauformen von Registern mit regulären asynchronen Eingängen. Diese dienen dazu, Ausgang und Speicherzustand gleichzeitig auf einen definierten Wert zu setzen, zum Beispiel Set- oder Reset-Eingänge.

Das Unterscheidungsmerkmal zwischen Flipflops und Latches ist die Art der Steuerung durch den Takteingang. Bei Flipflops führt ein Ereignis am Takteingang zur Änderung des Speicherzustandes und zur Ausgangsänderung. Dies hat zur Konsequenz, dass die synchronen Eingänge eines Flipflops zu keiner Zeit Ausgangsänderungen hervorrufen. Bei Latches gibt es zwei Phasen, die vom Zustand des Taktes abhängen: In der ersten sind die Eingänge vom Rest des Gatters abgekoppelt. Der Ausgang ist identisch dem Speicherzustand. In der zweiten Phase wirken die Eingänge gleichzeitig auf den Speicherzustand und auf den Ausgang. In dieser Phase verhalten sich die Eingänge wie die kombinatorischer

Gatter. Man bezeichnet das Latch in dieser Phase als »transparent«. Ein Flipflop lässt sich aus zwei Latches und einem Inverter konstruieren [2].

Während Eingänge an kombinatorischen Gattern ohne Beeinträchtigung der Gatterfunktion zu beliebigen Zeiten Ereignisse aufweisen können, ist dies an Eingängen von Registern nicht der Fall. Ein Problem entsteht dadurch, dass die interne Speicherzelle des Registers sicher gesetzt werden muss. Dieser Vorgang kann fehlschlagen, wenn sich ein Eingang während des Setzprozesses ändert. Im Falle von Flipflops müssen die synchronen Eingänge eine Mindestzeit vor und nach der übernehmenden Taktflanke stabil sein. Die erstgenannte Zeit heißt *Setup*-, die zweite *Holdzeit*. Daneben muss auch der Takt selbst bestimmte Anforderungen erfüllen. Insbesondere müssen die zeitlichen Abstände zwischen zwei aufeinander folgenden Ereignissen hinreichend groß sein. Dieser Abstand wird als *Pulsweite* bezeichnet. Zur Erläuterung der angesprochenen Zeiten siehe Abbildung 2.1.

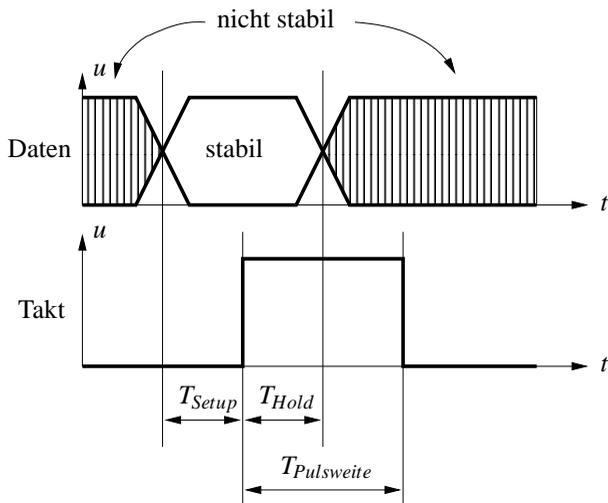


Abbildung 2.1: Setup-, Hold- und Pulsweitenzeit bei einem positivereignisgesteuerten Flipflop

Die Zeiten sind auch für Latches einzuhalten. Sie beziehen sich in diesem Fall auf das Taktereignis, das das Latch von transparent auf nicht-transparent schaltet, da hierdurch der Speichervorgang beendet wird.

Eine der Aufgaben der statischen Timinganalyse ist die Prüfung der Einhaltung dieser Zeiten.

Eingänge der Schaltung, die der externen Kommunikation dienen, heißen *Primäreingänge*. Ereignisse an Primäreingängen sind keine Schaltungseigenschaft sondern müssen spezifiziert werden. Analog werden Schaltungsausgänge zur externen Kommunikation *Primärausgänge* genannt.

Parasiten sind unerwünschte, durch die physikalische Realisierung der Schaltung jedoch unvermeidbare Schaltungselemente. Parasiten sind zum Beispiel die Kapazität und der Widerstand von Leitungen. Die Parasiten werden aus dem Layout der Schaltung durch Extraktion gewonnen.

Ein *Netz* ist die strukturelle Sicht einer Leitung. In einer parasitenfreien Netzliste wird ein Netz durch einen Schaltungsknoten repräsentiert. Dies gilt auch, wenn die Netzparasiten durch ihre Summenkapazität und ohne Widerstand modelliert werden.

Die behandelte Definition des Zeitverhaltensgraphen orientiert sich an der in [12] gegebenen. Es handelt sich um einen gerichteten, azyklischen Graphen.

Die Knoten des Zeitverhaltensgraphen speichern die Ereignisse der zugrunde liegenden Schaltung. Im Allgemeinen hat jeder Knoten einen zugeordneten Schaltungsknoten. Es können jedoch auch Hilfsknoten erzeugt werden, die keine solche Entsprechung haben. *Verzögerungskanten* (»timing arcs« in [12], Propagierungssegmente in [13][14]) sind Kanten nach der Definition in Abschnitt 2.2.1. Die Gewichte der Kanten repräsentieren die Verzögerung zwischen den Knoten der Kanten. Ereignisse auf durch eine Verzögerungskante verbundenen Knoten müssen kausal voneinander abhängig sein. Zwei Knoten werden dann durch mehrere Kanten verbunden, wenn die Verzögerung für verschiedene Ereignisse unterschiedlich ist. Beispiel: Die Verzögerung für ein steigendes Eingangsereignis ist unterschiedlich zu der Verzögerung für ein fallendes Eingangsereignis.

Neben der Verzögerungskante ist die Definition einer *Testkante* oder eines *Tests* (»constraint arc« in [12]) sinnvoll. Testkanten stellen den Zusammenhang zwischen jeweils zwei Knoten, an denen bestimmte zeitliche Anforderungen erfüllt

sein müssen, her. Die Anforderungen sind im Allgemeinen benutzerdefiniert. Sie ergeben sich in vielen Fällen aus den Eigenschaften eines Speicherelementes (zum Beispiel Setup- und Holdzeiten), sind aber meist nicht automatisch erkennbar. Testkanten sind nicht Teil des Graphen nach Definition in Abschnitt 2.2.1 und werden bei der Bestimmung von beispielsweise Tiefe, Zyklen und Pfaden ignoriert.

Die Länge eines Pfades in einem Zeitverhaltensgraphen, das heißt, die Summe der Gewichte seiner Kanten, ist die Verzögerung dieses Pfades.

Man unterscheidet *Daten-* und *Taktpfade*. Datenpfade beginnen an Primäreingängen oder an Ausgängen von Registern der Schaltung. Sie enden an synchronen Eingängen von Registern oder an Primärausgängen. Taktpfade beginnen an Taktquellen, meist Primäreingänge oder Oszillatorausgänge. Sie enden an den Takteingängen von Registern.

2.4 Abbildung einer Schaltung

Zur Abbildung einer Schaltung auf einen Zeitverhaltensgraphen muss jeder Gattertyp selbst durch einen Zeitverhaltensgraphen repräsentiert werden. Dieser Graph soll folgende Eigenschaften haben:

- Jeder Gattereingang wird einem Knoten zugeordnet.
- Jeder Gatterausgang wird einem Knoten zugeordnet.
- Jede Kombination aus einem Eingang und einem Ausgang, bei der ein Eingangereignis ein Ausgangereignis verursachen kann, wird durch eine Verzögerungskante modelliert (asynchrone Eingänge).
- Der Graph kann benutzerdefinierte Testkanten enthalten. Dies sind insbesondere Setupzeit- und Holdzeit-Tests von Speichern.
- Der Graph kann zusätzliche benutzerdefinierte Kanten und Knoten enthalten.

Die Abbildung der Schaltung erfolgt in dieser Weise:

- Jeder Primäreingang wird einem Knoten zugeordnet. Diese Knoten sind Startpunkte.

- Jeder Primärausgang wird einem Knoten zugeordnet. Diese Knoten sind Endpunkte.
- Jeder Gattereingang wird einem Knoten zugeordnet.
- Jeder Gatterausgang wird einem Knoten zugeordnet.
- Jeder Primäreingang und Gatterausgang wird mit jedem Primärausgang und Gattereingang desselben Netzes durch eine Kante verbunden.
- Jedes Gatter wird durch seinen Gatter-Zeitverhaltensgraphen ersetzt.
- Benutzerdefinierte zeitliche Abhängigkeiten werden durch Kanten modelliert.
- Benutzerdefinierte Prüfungen werden durch Testkanten modelliert.

Anhand eines Beispiels soll gezeigt werden, wie eine Schaltung mit einem Takt und benutzerdefinierten Tests auf einen Zeitverhaltensgraphen abgebildet wird. Die Schaltung ist in Abbildung 2.2 dargestellt.

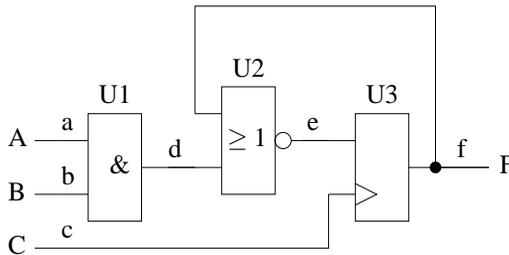


Abbildung 2.2: Beispielschaltung

Die folgenden zeitlichen Annahmen und Beschränkungen werden getroffen:

1. Der Primäreingang C ist die Taktquelle. Der Takt hat eine Periode von T_ϕ . Das steigende Ereignis erscheint zur Zeit 0, das fallende zur Zeit $\frac{1}{2}T_\phi$.
2. Jedes Eingangereignis der Primäreingänge A und B ist mit dem Takt synchronisiert. Die Eingänge ändern ihren Wert im Intervall $[\frac{T_\phi}{2} - T_{\delta 1}, \frac{T_\phi}{2} + T_{\delta 1}]$.

3. Jedes Netz ist verzögerungsbehaftet.
4. Jedes Gatter ist verzögerungsbehaftet.
5. Der Ausgang der Schaltung soll in einem Intervall $[\frac{T_\phi}{2} - T_{\delta 2}, \frac{T_\phi}{2} + T_{\delta 2}]$ stabil sein.

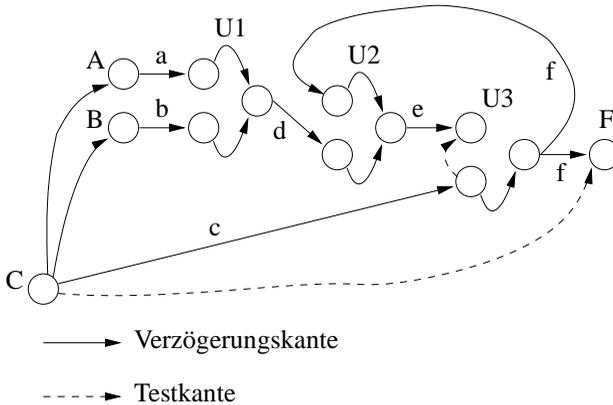


Abbildung 2.3: Zeitverhaltensgraph

Der zugehörige Zeitverhaltensgraph ist in Abbildung 2.3 dargestellt. Die Bezeichnungen a...f sind Netznamen der Schaltung und nicht Bezeichner der Kanten. Das Netz f resultiert in zwei Kanten im Zeitverhaltensgraph. Obwohl kein Netz vorhanden ist, gibt es eine Kante zwischen dem Taktknoten C und den Primäreingängen A und B, da es eine benutzerdefinierte zeitliche Abhängigkeit gibt.

Die Testkante zwischen C und Primärausgang F wird durch die benutzerdefinierte Prüfung des Zeitverhaltens an F hervorgerufen, da hier ein Fenster bezüglich des Taktes vorgegeben wurde. Die Testkante zwischen dem Takt- und Dateneingang des Flipflops stellt die notwendige Prüfung der minimalen Setup- und Holdzeiten dar. Der Dateneingang des Flipflops unterbricht Pfade. Der Graph enthält keine Zyklen.

Nimmt man für jedes Netz und jedes Gatter eine Verzögerung von eins an, so hat der Pfad von C über die Netze b, d, e zum Dateneingang des Flipflops die maximale Länge $l = 5 + \frac{1}{2}T_\phi + T_{\delta 1}$.

2.5 Suche der kritischen Pfade

2.5.1 Problemstellung

Die Aufgabe der Timinganalyse ist die Prüfung der Einhaltung aller Tests. Durch die Abbildung der Schaltung auf einen Zeitverhaltensgraphen wird das Problem auf die Suche nach *kritischen Pfaden* abgebildet. Kritische Pfade sind dadurch gekennzeichnet, dass sie das geforderte Zeitverhalten verletzen oder mit dem kleinsten Betrag einhalten.

Kurze Pfade sind von Bedeutung, da sie möglicherweise mit den späten Grenzen der Zeitfenster für stabile Datenereignisse kollidieren, zum Beispiel mit der minimalen Holdzeit eines Registers. In diesem Fall würde durch das Taktereignis ein Wert eingelesen, der durch dasselbe Taktereignis ausgelöst wurde. Die korrekte Funktion wäre, dass nur Ereignisse eingelesen werden, die vom vorhergehenden Taktereignis verursacht wurden. Probleme dieser Art können im Allgemeinen durch Einfügen von Puffern, das sind Gatter mit der Identität als Schaltfunktion, behoben werden.

Lange Pfade können die frühen Grenzen der Stabilitätszeitfenster verletzen, also die Setupzeit eines Registers. Das Register würde dann einen ungültigen alten Wert des Eingangs einlesen. Lange Pfade zwischen zwei Registern betreffen grundsätzlich verschiedene Taktereignisse. Das heißt, der Test des Pfades wird von einem späteren Taktereignis, als dem, mit dem er gestartet wurde, ausgelöst. Üblich ist, dass eine Taktperiode vergeht. Lange Pfade kann man nicht ohne Weiteres verkürzen. Siehe dazu [13] und die Referenzen darin.

Die *Ankunftszeit* T_{AT} (engl. »arrival time«, abgekürzt *AT*) ist die Zeit, ab der ein Schaltungsknoten nach Änderung der Primäreingänge oder Taktquellen einen neuen Endwert angenommen hat. Da das Taktnetzwerk selbst verzögerungsbehaftet ist und die Ankunftszeiten der Taktereignisse an unterschiedlichen Speicherelementen verschieden sein können, wird die Propagierung von Ereignissen in zwei Modi durchgeführt: Im *Late-Mode* werden nur späte Ereignisse propagiert, im *Early-Mode* nur frühe. Im Allgemeinen benötigen Tests die Propagierung sowohl im Early- und Late-Mode. Der Zeitraum zwischen frühester und spätester Ankunftszeit an einem Knotens heißt *Schaltfenster*.

Die *notwendige Ankunftszeit* (engl. »required arrival time«, abgekürzt *RAT*) ist die Ankunftszeit, die notwendig ist, damit der korrekte Wert in ein Register übernommen werden kann oder sich an einem Primärausgang einstellt. Die notwendige Ankunftszeit errechnet sich aus der Ankunftszeit des Taktereignisses und charakterisierter Eigenschaften des Registers oder aus benutzerdefinierten Anforderungen. Für den Setupzeit-Test eines Flipflops gilt für die notwendige Ankunftszeit am synchronen Eingang:

$$T_{RAT,Daten} = T_{AT,Takt,min} - T_{Setup,min} \quad (2.13)$$

$T_{AT,Takt,min}$ ist die frühestmögliche Zeit, an der ein Ereignis am Takteingang des Registers auftreten kann. Für den Holdzeit-Test gilt analog:

$$T_{RAT,Daten} = T_{AT,Takt,max} + T_{Hold,min} \quad (2.14)$$

Im Falle einer einzigen Taktquelle und einer Setup-Prüfung beinhaltet $T_{AT,Takt}$ mindestens eine Zykluszeit, im Falle einer Holdzeit-Prüfung jedoch nicht. Die genaue Bestimmung der Anzahl der enthaltenen Taktzyklen wird als *Adjust* bezeichnet und verlangt im Falle mehrerer Taktquellen ein aufwändiges Regelwerk [14].

Slack (deutsch etwa »Spiel«) ist ein Maß für die Einhaltung der zeitlichen Bedingungen eines Ereignisses. Der Slack ist derart definiert, dass ein negativer Wert die Verletzung zeitlicher Bedingungen anzeigt. Es gilt für Setupzeit-Tests:

$$T_{Slack} = T_{RAT} - T_{AT} \quad (2.15)$$

und für Holdzeit-Tests:

$$T_{Slack} = T_{AT} - T_{RAT} \quad (2.16)$$

Abbildung 2.4 zeigt die Ankunftszeiten und den Slack für einen Setupzeit-Test.

2.5.2 Breitensuche

Die *Breitensuche* (engl. »breadth first search«, abgekürzt *BFS*, Breitendurchlauf in [5]) ist eine Vorschrift zum Besuch aller Knoten eines Graphen. Es werden zunächst alle Knoten derselben Tiefe und dann die Knoten der nächstgrößeren Tiefe

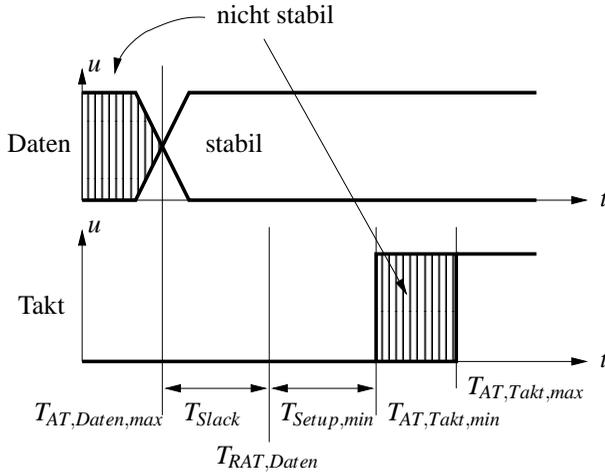


Abbildung 2.4: Ankunftszeiten und Slack für einen Setupzeit-Test

besucht. Die Breitensuche wird zur Bestimmung der kritischen Pfade verwendet werden. Das Verfahren wird beispielhaft für die Suche nach dem längsten Pfad behandelt. Zur Anwendung der Breitensuche ist es sinnvoll, zunächst die Tiefe (siehe Abschnitt 2.2.1) aller Knoten zu bestimmen. Ein rekursives Verfahren funktioniert wie folgt. d_v ist die Tiefe des Knotens v :

Setze die Tiefe jedes Knotens auf $d_v = -1$.

Besuche jeden Startpunkt s .

 setze $d_s = 0$

Besuche jeden Endpunkt e .

 Rufe L mit $v := e$ auf

L:

 Besuche jeden Vorgänger \tilde{v} von v

 Falls $d_{\tilde{v}} = -1$

 Rufe L mit $v := \tilde{v}$ rekursiv auf

 Falls $d_{\tilde{v}} > d_v$

$d_v := d_{\tilde{v}}$

$d_v := d_v + 1$

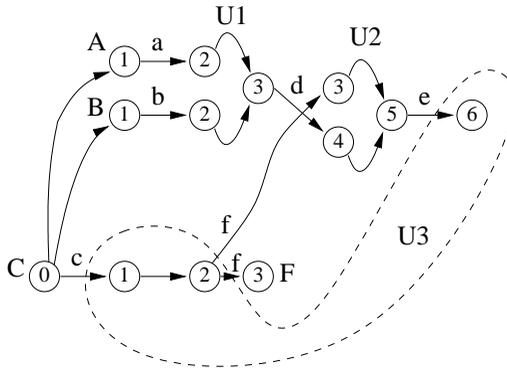


Abbildung 2.5: Tiefe der Knoten des Zeitverhaltensgraphens aus Abbildung 2.3

Dieses Verfahren hat die Komplexität $O(|V| + |E|)$, wobei $|E|$ die Anzahl der Kanten und $|V|$ die Anzahl der Knoten sind. In Abbildung 2.5 ist der Graph von Abbildung 2.3 dargestellt, wobei jeder Knoten mit seiner Tiefe versehen ist.

Zur Anwendung der Breitensuche wird vereinbart, dass jeder Knoten des Graphen einen Pfad oder Teilpfad P symbolisch speichern kann. Die Länge des Pfades wird als $P.l$ notiert, das Gewicht einer Kante mit $w(v_1, v_2)$ bezeichnet. Die Notation $v.P$ bezeichnet den aktuell in Knoten V gespeicherten Pfad, $v.P.l$ dessen Länge. Der folgende Pseudocode zeigt die Suche nach dem längsten Pfad.

```

Für jeden Startpunkt  $s$ 
     $P = \{s\}$ 
Für jedes  $d := 1 \dots d_{max}$ 
    Für jeden Knoten  $v$  der Tiefe  $d$ 
         $v.P = \{\}$ 
        Für jeden Vorgängerknoten  $\tilde{v}$  von  $v$ 
            Falls  $\tilde{v}.P.l + w(\tilde{v}, v) \geq v.P.l$ 
                 $v.P = \tilde{v}.P + v$ 
 $P_{max} = \{\}$ 
Für jeden Endpunkt  $e$ 
    Falls  $e.P.l > P_{max}.l$ 
         $P_{max} = e.P$ 
    
```

P_{max} enthält schließlich den längsten Pfad des Graphen. Die Idee des Verfahrens ist, jeden Knoten mit dem längsten Pfad bis genau zu diesem Knoten zu markieren. Damit das möglich ist, müssen zu jeder Zeit alle Vorgängerknoten berechnet sein. Dies ist durch die iterative Berechnung in der Reihenfolge der Tiefe sichergestellt. Neben dem hier gezeigten iterativen Verfahren kann auch ein rekursives Verfahren angewendet werden, das mit den Endpunkten beginnt. Beim Besuch eines Knotens wird geprüft, ob jeder Vorgängerknoten bereits berechnet wurde. Falls das nicht der Fall ist, wird der Vorgänger rekursiv berechnet. Die Rekursion ist beendet, wenn der betrachtete Vorgängerknoten bereits berechnet wurde oder ein Startpunkt erreicht wird.

Die Komplexität der Breitensuche ist $O(|V| + |E|)$, also linear, da jeder Knoten und jede zu diesem Knoten führende Kante genau einmal besucht wird. Diese lineare Komplexität ist eine Schlüsseleigenschaft der statischen Timinganalyse. Es ist möglich, selbst Schaltungen von mehreren hunderttausend Gattern Größe in Minuten zu verifizieren. Dies ist mit anderen Verfahren wie der ereignisgesteuerten Simulation nicht möglich.

2.5.3 Anwendung der Breitensuche in der statischen Timinganalyse

Die Übertragung des Verfahrens auf die statische Timinganalyse erfordert im Allgemeinen die Propagierung von Ereignissen sowohl im Late- als auch im Early-Mode. Die Auswahl des kritischen Pfades der Gesamtschaltung erfolgt nicht nach der Länge eines Pfades, sondern anhand des Slacks (siehe Abschnitt 2.5.1) eines Datenpfades.

Die Anwendung der Breitensuche zur statischen Timinganalyse impliziert eine wesentliche Voraussetzung: Es muss sichergestellt sein, dass das zu Knoten größerer Tiefe propagierte Ereignis den ungünstigsten Fall darstellt. Diese Forderung wird durch das in diesem Kapitel angenommene einfache Verzögerungsmodell sichergestellt. Im Fall anspruchsvollerer Verzögerungsmodelle müssen zur Einhaltung des ungünstigsten Falles unter Umständen mehrere Ereignisse propagiert werden. Bei Einhaltung des oben genannten Kriteriums ist garantiert, dass das letzte Ereignis jedes Knotens des Verzögerungsgraphen und somit der kritische Pfad der gesamten Schaltung gefunden werden. Als Korollar ergibt sich aus der

oben genannten Forderung, dass Ereignisse eines Knoten keine Auswirkungen auf Knoten kleinerer Tiefe haben dürfen.

Kann die Forderung, dass an einem Knoten das ungünstigste Ereignis bestimmt werden kann, nicht erfüllt werden, müssten zur Verifikation des Zeitverhaltens alle existierenden Pfade berechnet werden. Die Anzahl aller existierenden Pfade kann nur bei Kenntnis des Zeitverhaltensgraphens angegeben werden. Es gibt Klassen von Graphen, deren Anzahl der Pfade exponentiell mit der Anzahl der Knoten wächst: Gegeben sei ein gerichteter, azyklischer Graph $G = (V, E)$ mit je einem Start- und Endpunkt. δ sei der Ausgangsgrad jedes Knotens mit Ausnahme des Endpunktes. Jeder Knoten mit Ausnahme des Endpunktes hat genau einen Nachfolger und ist mit diesem durch δ Knoten verbunden. Die Anzahl der Pfade beträgt $\delta^{|V|-1}$. Abbildung 2.6 zeigt ein Beispiel für $\delta = 2$ und $|V| = 3$.

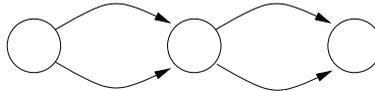


Abbildung 2.6: Beispiel zur Anzahl der Pfade eines Graphen mit $\delta = 2$ und $|V| = 3$

Als weitere Besonderheit der Anwendung der Breitensuche in der Timinganalyse soll die Behandlung des Taktes herausgestellt werden. Im Falle einer Taktquelle wiederholt sich das zeitliche Verhalten mit jedem Zyklus. Alle Taktzyklen sind zeitlich äquivalent. Es ist daher ausreichend, genau einen Zyklus zu analysieren. Setzt man für statische Timinganalyse und Simulation dasselbe Verzögerungsmodell voraus, so entspricht die statische Timinganalyse einer erschöpfenden Simulation innerhalb eines Taktzyklusses.

Ein Pfad in der statischen Timinganalyse ist die Abfolge kausal voneinander abhängiger Ereignisse. Ein Pfad hat die folgende Datenstruktur:

$$P = \{\{v_0, T_{AT,0}, e_0\}, \dots, \{v_{n-1}, T_{AT,n-1}, e_{n-1}\}\} \quad (2.17)$$

v_i bezeichnen Knoten und $T_{AT,i}$ die Ankunftszeiten. e_i indiziert, ob das Ereignis steigend oder fallend ist. Die Länge eines Pfades ist identisch mit der Ankunftszeit $T_{AT,n}$ am letzten Knoten. Im Falle komplexerer Verzögerungsmodelle sind $T_{AT,i}$ und e_i durch weitere oder andere Zeitwerte oder das tabellierte Ereignis $t \rightarrow u(t)$ zu ersetzen.

Die Anwendung der Breitensuche in der statischen Timinganalyse fand Verwendung in [12][15][16]. Die Verfahren in [17][18] sind Beispiele für die statische Timinganalyse ohne Breitensuche.

2.6 Problem der falschen Pfade

Die statische Timinganalyse ist, soweit bisher behandelt, ein Werkzeug zur Verifikation des Zeitverhaltens einer Schaltung, das funktionale Zusammenhänge außer Acht lässt. Diese funktionalen Zusammenhänge können allerdings dazu führen, dass die statische Timinganalyse kritische Pfade ermittelt, die kein Ereignis am Endpunkt hervorrufen. Ein Beispiel für einen solchen Pfad ist in Abbildung 2.7 gegeben.

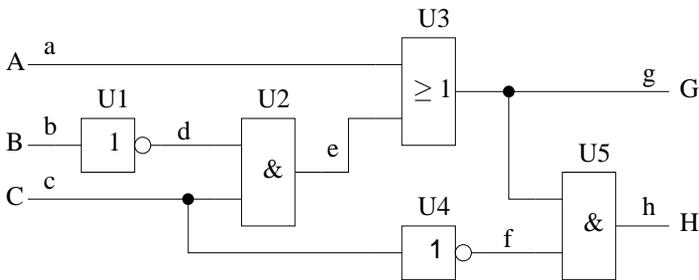


Abbildung 2.7: Kombinatorische Schaltung mit falschem Pfad

Es wird vorausgesetzt, dass alle Gatter die Einheitsverzögerung und alle Netze keine Verzögerung haben. Alle Primäreingänge seien zur Zeit 0 ns stabil. Die Ausgänge aller Gatter haben den Bezeichner »Z«. Als Knoten des Verzögerungsgraphen werden die primären Ein- und Ausgänge sowie die Gatterausgänge verwendet. Die Notation b bezeichnet den booleschen Wert des Netzes b . Die Unterscheidung steigender und fallender Ereignisse wird weggelassen, da sie hier ohne Bedeutung ist. Gleiches gilt für die Bezeichner der Gattereingänge.

Der Pfad $P_1 = \{\{B,0\},\{U1/Z,1\},\{U2/Z,2\},\{U3/Z,3\},\{U5/Z,4\},\{H,4\}\}$ ist mit der Länge vier der topologisch längste Pfad. Es können jedoch keine kausal abhängigen Ereignisse entlang dieses Pfades propagiert werden.

An den Gattern U2 und U5 müssten die Netze c und f der Seiteneingänge jeweils den nicht-kontrollierenden Wert 1 haben ($c = 0$ erzwingt $e = 0$ unabhängig von d), damit ein Ereignis propagiert werden kann. Dies ist wegen $f = \bar{c}$ jedoch nicht möglich. Ein tatsächlich kritischer Pfad ist zum Beispiel $P_2 = \{\{B, 0\}, \{U1/Z, 1\}, \{U2/Z, 2\}, \{U3/Z, 3\}, \{G, 3\}\}$ mit einer Länge von 3. Pfade wie P_1 , die topologisch vorhanden sind, jedoch keine zeitlich abhängigen Ereignisse transportieren können und daher das zeitliche Verhalten der Schaltung nicht beeinflussen, bezeichnet man als *falsche Pfade*.

Dieses Beispiel soll Folgendes verdeutlichen:

1. Statische Timinganalyse ohne Prüfung auf falsche Pfade kann nur eine Untergrenze für den Slack liefern. Der tatsächlich vorhandene Slack kann größer als ermittelt sein.
2. Damit ein Ereignis zu einem Gatterausgang propagiert werden kann, muss das Gatter die nicht-kontrollierende Funktion einstellen können.
3. Im Falle eines falschen Pfades müssen mehrere Seiteneingänge boolesch voneinander abhängen. Der Zeitverhaltensgraph muss also rekonvergent (siehe Abschnitt 2.2.1) sein. Rekonvergenz ist eine notwendige Voraussetzung für die Existenz falscher Pfade.

Timinganalyse, die falsche Pfade erkennt und ausschließt, wird als funktionale Timinganalyse bezeichnet. Ein Pfad, der kausal voneinander abhängige Ereignisse transportieren kann, heißt »sensibilisierbar«.

Das Problem der falschen Pfade wird in dieser Arbeit nicht weiter behandelt. Statt dessen wird auf die Literatur [7][8][19][20][21][22][23] und die Referenzen darin verwiesen. Die Prüfung, ob ein Pfad falsch ist, ist np-vollständig[21]. Die Suche nach dem längsten, nicht falschen Pfad ist np-hart.

2.7 Grenzen der statischen Timinganalyse

Statische Timinganalyse kann nur dann angewendet werden, wenn Ereignisse an den Primäreingängen einer Schaltung einen zeitlichen Bezug zu einem Takt haben, wobei es sich bei diesem Takt auch um einen virtuellen Takt handeln kann. Kann sich der boolesche Wert von Primäreingängen zu beliebigen Zeiten

ändern, spricht man von asynchronen Primäreingängen. Die statische Timinganalyse kann in diesem Fall nicht zur Verifikation des Zeitverhaltens verwendet werden, da es keine sicheren zeitlichen Ober- und Untergrenzen für Ereignisse gibt. Dies gilt insbesondere, wenn Pfade zu den Takteingängen von Registern an asynchronen Primäreingängen beginnen.

Das Problem der Transparenz von Latches wurde schon angesprochen (siehe Abschnitt 2.3). Die Schwierigkeit der Timinganalyse liegt darin, den kombinatorischen Zyklus zu unterbrechen. Die statische Timinganalyse könnte eine funktionierende Schaltung möglicherweise als inkorrekt einschätzen, da der Zyklus nicht optimal aufgebrochen wird. Auch auf diesem Gebiet gibt es Lösungsvorschläge [24] [25].

Statische Timinganalyse ist nur solange sinnvoll, wie sie auf Breitensuche basiert (siehe Abschnitt 2.5.3). Eine weitere Restriktion ist die Existenz falscher Pfade, die in Abschnitt 2.6 schon behandelt wurde. Hier liegt das Problem darin, dass mit der Suche nach falschen Pfaden die lineare Komplexität der Breitensuche aufgegeben wird.

3 Verzögerungsmodelle

3.1 Einführung

In diesem Kapitel soll der Stand der Technik bei der Berechnung der Verzögerung von Gattern und Leitungen dargestellt werden. Die Berechnung der Verzögerung setzt die Extraktion oder, für die Prä-Layout-Abschätzung, eine hinreichend genaue Modellierung der parasitären Schaltungselemente voraus.

Im Folgenden werden nacheinander Verzögerungsmodelle für Gatter und Leiter behandelt. Den Abschluss dieses Kapitels bildet die Behandlung der Modellierung des Einflusses kapazitiven Übersprechens auf die Verzögerung.

Die behandelten Verzögerungsmodelle entstammen nicht nur dem Gebiet der statischen Timinganalyse, sondern wurden teilweise zuerst in Timingsimulatoren wie MOTIS [26] eingesetzt.

3.2 Verzögerung

Verzögerung wird als ein singulärer Zeitwert T_D angesehen. In diesem Sinne gibt es zwei sinnvolle Anwendungen:

- Die Verzögerung ist eine Messnorm.
- Die Verzögerung ist Bestandteil eines Verzögerungsmodells.

Im ersten Fall ist es sinnvoll, als Verzögerungsmaß die zeitliche Differenz des Erreichens der halben Versorgungsspannung zweier kausal abhängiger Ereignisse

zu bilden [2]. Sei u_2 das kausal von u_1 abhängige Ereignis und seien T_1 und T_2 die Zeitpunkte des Erreichens von $\frac{1}{2}U_{DD}$:

$$u_1(T_1) = \frac{1}{2}U_{DD} \quad (3.1)$$

$$u_2(T_2) = \frac{1}{2}U_{DD} \quad (3.2)$$

so ist die Verzögerung:

$$T_D = T_2 - T_1 \quad (3.3)$$

Abbildung 3.1 zeigt die Verläufe der Ein- und Ausgangsspannungen eines Inverters und die Verzögerung zwischen Ein- und Ausgang.

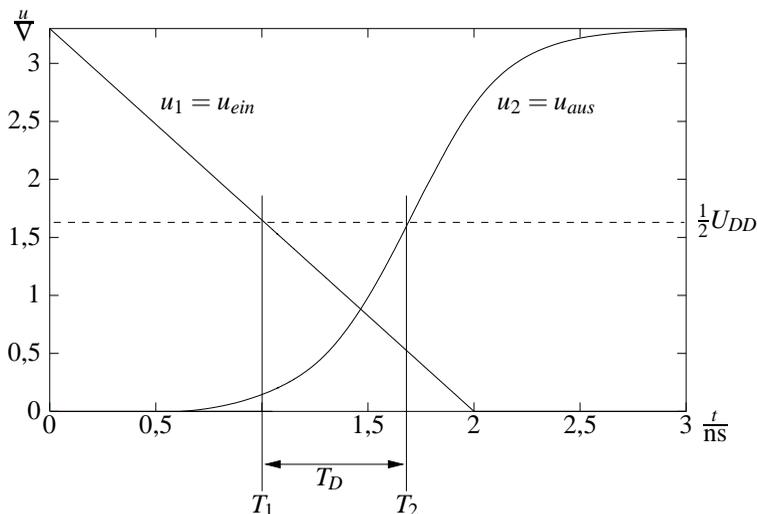


Abbildung 3.1: Verzögerung als Vergleichsmaß

Im Folgenden wird diese Verzögerungsdefinition als »50%-Verzögerung« bezeichnet. Die 50%-Verzögerung kann negativ sein [27]. Dies ist kein Widerspruch zur Kausalität.

Der Begriff »Verzögerung« im Rahmen eines Verzögerungsmodells ist definitionsabhängig. Wie im Folgenden dargestellt, gibt es Verzögerungsmodelle, die die 50%-Verzögerung als Teil der Modelldefinitionen beinhalten.

Da in CMOS-Schaltungen die Spannungen sich asymptotisch an 0 V und U_{DD} annähern, die Werte aber theoretisch nie erreichen, ist die Festlegung weiterer Spannungsschwellwerte $U_{th,u}$ und $U_{th,o}$ sinnvoll. Es gilt:

$$0 < U_{th,u} < U_{th,o} < U_{DD} \quad (3.4)$$

Ereignisse $t \rightarrow u(t)$ sind im Spannungsbereich durch diese Schwellwerte beschränkt:

$$U_{th,u} \leq u(t) \leq U_{th,o} \quad (3.5)$$

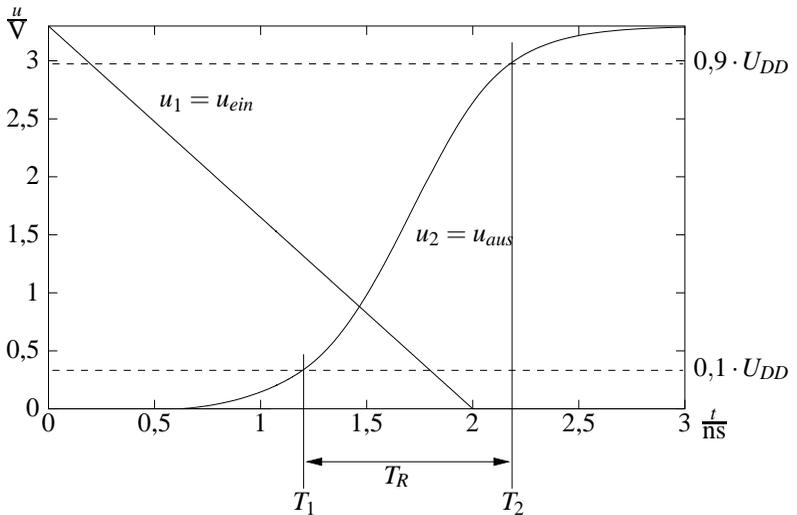


Abbildung 3.2: Anstiegszeit als Vergleichsmaß

Die Spannungsschwellen können auch zur Definition einer Anstiegszeit T_R verwendet werden. Ebenso wie beim Verzögerungsbegriff kann die Anstiegszeit

Messnorm oder Teil eines Verzögerungsmodells sein. Wird Anstiegszeit als Messnorm verwendet, so sind die Spannungsschwellen

$$U_{th,u} = 0,1 \cdot U_{DD} \quad (3.6)$$

$$U_{th,o} = 0,9 \cdot U_{DD} \quad (3.7)$$

üblich. Abbildung 3.2 zeigt die Anstiegszeit der Ausgangsspannungskurve u_{aus} eines Inverters. Die Anstiegszeit ist nicht negativ.

Die Werte $0,1 \cdot U_{DD}$ und $0,9 \cdot U_{DD}$ sind nach Wissen des Verfassers willkürlich gewählt.

3.3 Verzögerungsmodelle auf Gatterebene

3.3.1 Aufgabenstellung

Ziel eines Verzögerungsmodells ist es, das zeitliche Verhalten des zu modellierenden Schaltungsteils zu beschreiben. Die Modellierung anderer physikalischer Größen wie Stromaufnahme kann gegenüber dem Zeitverhalten zurücktreten.

Abbildung 3.3 stellt die Aufgabe der Berechnung der Gatterverzögerung in einer abstrahierenden Form dar. Betrachtet wird ein Ereignis an einem Gattereingang, das zu einem Ereignis am Gatterausgang führt. Die physikalischen Gegebenheiten, die die Verzögerung beeinflussen, werden in *Eingangssituation*, *Ausgangssituation* und Konfiguration unterteilt. Ein- und Ausgangssituation werden in Abschnitt 3.3.2 behandelt. Das Verzögerungsmodell hat die Aufgabe, abhängig von Eingangs- und Ausgangssituation sowie der Konfiguration, das Verhalten der Ausgangsspannung $u_{aus}(t)$ mit der erforderlichen Genauigkeit (zum Beispiel bezüglich Fehlergrenzen) wiederzugeben. Dabei ist der exakte Spannungsverlauf in vielen Fällen nicht interessant. Vielmehr können u_{ein} und u_{aus} stark vereinfacht in das Modell einfließen. Wichtig ist jedoch die gute Voraussage des Zeitverhaltens.

Die Konfiguration beschreibt die Eingangsbelegung der nicht betrachteten Eingänge oder bei Speicherelementen den Speicherzustand. Die Konfiguration ist bei Komplexgattern von Interesse, die bei gleichem Eingangsereignis unterschiedliche Ausgangsereignisse haben können.

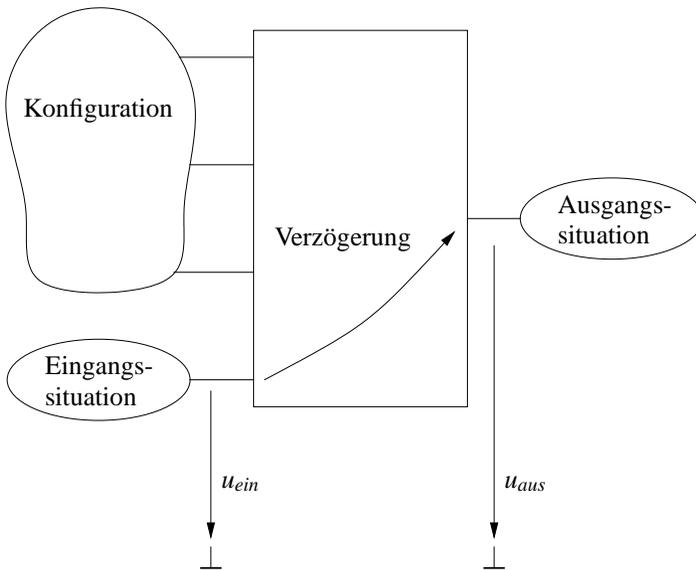


Abbildung 3.3: Verzögerung bei Gattern

Die Gewinnung der Parameter des Verzögerungsmodells in Abhängigkeit von Ein- und Ausgangssituation und Konfiguration nennt man *Charakterisierung*. Die Charakterisierung erfolgt derart, dass die Transistornetzliste des Gatters mit extrahierten Modellparametern und Parasiten sowie mit typischen Ein- und Ausgangssituationen und Konfigurationen belegt und in einem Analogsimulator simuliert wird.

Die *Treiberstärke* von Gattern ist ein qualitatives Maß zum Vergleich von Gattertypen. Eine größere Treiberstärke führt zu kleineren Anstiegszeiten bei derselben Lastkapazität.

3.3.2 Modellierung der Ein- und Ausgangssituation

Die Eingangssituation beinhaltet die physikalischen Eigenschaften am Gattereingang, die die Verzögerung beeinflussen. Dies ist grundsätzlich der Spannungsverlauf. Dieser Spannungsverlauf wird durch Berechnung der vorhergehenden Stufe bestimmt, wobei die elektrischen Eigenschaften des Eingangs Teil der Ausgangslast der vorhergehenden Stufe sind.

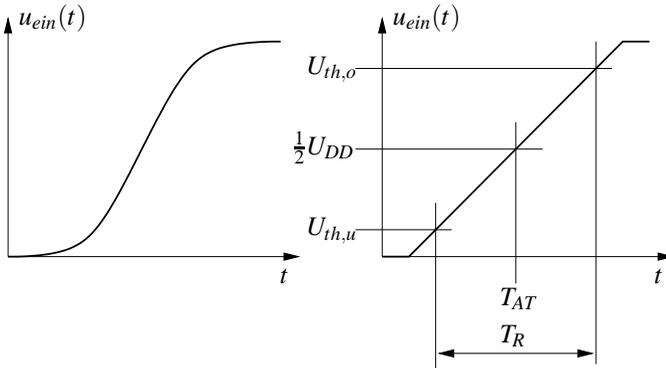


Abbildung 3.4: Eingangssituation

In Abbildung 3.4 sind typische Modelle für Eingangsspannungskurven eines steigenden Ereignisses dargestellt. Der linke Teil der Abbildung stellt den allgemeinen Fall, in dem ein Ereignis durch Paare von Zeit- und Spannungswert gespeichert wird, dar. In der rechten Abbildung ist die Spannungskurve durch eine lineare Näherung auf zwei Zeitwerte, der Anstiegszeit T_R und der Ankunftszeit T_{AT} , reduziert. Dieses Modell wird häufig in Timinganalysewerkzeugen, wie zum Beispiel in denen des Herstellers »Synopsys« [28], verwendet. Es ist die Modellierung der Eingangssituation des sogenannten K-Faktor-Modells [2].

Die Wahl einer Modellierung der Eingangssituation $u_{ein}(t)$ bedingt eine analoge Abstraktion des Ausgangsverhaltens $u_{aus}(t)$. Dies liegt daran, dass ein Gattereingang immer an einen Gatterausgang oder Primäreingang angeschlossen ist.

Die Ausgangssituation (Abbildung 3.5) ist eine Beschreibung des elektrischen Verhaltens der Schaltungsteile, die vom Gatterausgang umgeladen werden. All-

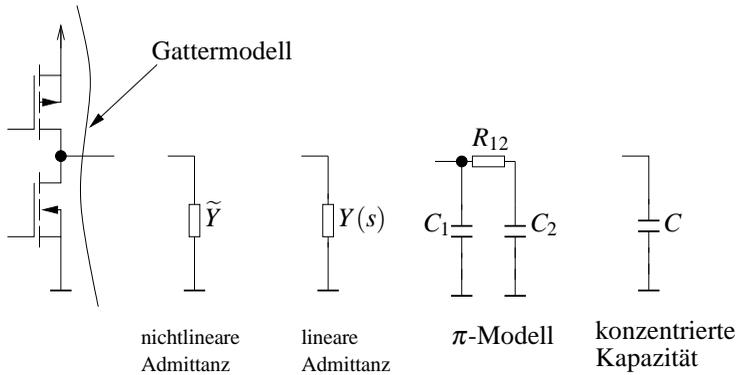


Abbildung 3.5: Ausgangssituation

gemein ist dies eine nichtlineare Admittanz. Die Nichtlinearität hat folgende Ursachen:

- Die Kapazitäten der Gates der Transistoren der folgenden Stufe sind nichtlinear [2][3].
- Die Sperrschichtkapazitäten der Transistoren sind nichtlinear. Für die Ausgangssituation müssen sie nur dann modelliert werden, wenn sie mit Gattereingängen verbunden sind, wie das zum Beispiel bei Pass-Gattern der Fall ist.

Das Leitbahnsystem kann ohne Einschränkungen linear modelliert werden.

Alle weiteren Modelle sind Vereinfachungen des Modells der nichtlinearen Admittanz. Üblicherweise wird die Gattereingangskapazität ebenfalls linear modelliert, so dass sich die Ausgangssituation zu einer linearen Admittanz reduziert. π -Modell und diskrete Kapazität sind Beispiele für lineare Admittanzen. Heute wird in den meisten Gatterbibliotheken für die statische Timinganalyse ein Verzögerungsmodell verwendet, das auf einer diskreten Lastkapazität basiert [2][28]. π -Modelle finden erst in jüngerer Zeit Eingang in die statische Timinganalyse [14]. Lineare, mehrstufige Admittanzen in Form von RC-Ketten fanden in der Software »Crystal« [18] Verwendung. Eine nichtlineare Modellierung der Ausgangssituation für die statische Timinganalyse ist dem Verfasser nicht bekannt.

Allen behandelten Modellen ist gemeinsam, dass sie einen Zweipol, der an Masse geschaltet ist, verwenden. Die folgenden Phänomene können nicht ohne Weiteres durch einen an Masse geschalteten Zweipol modelliert werden:

- Die Gatekapazität der Transistoren der folgenden Stufen. Da einstufige CMOS-Gatter invertierend schalten [3], entsteht eine Gegenkopplung zwischen Drain und Gate beziehungsweise zwischen Gatteraus- und -eingang. Dieses Verhalten wird üblicherweise implizit durch die lineare Eingangskapazität mit vergrößertem Wert modelliert. Die Charakterisierung kann zum Beispiel durch Messung der umgesetzten Ladung erfolgen.
- Die Koppelkapazität zwischen zwei Leitungen, wenn die Nachbarleitung mit entgegengesetzter Flanke schaltet.

3.3.3 MOS-Transistoren

Das Gleichstromverhalten des idealen MOS-Transistors wurde von William Shockley (siehe auch [2]) beschrieben. Es werden drei Operationsbereiche, die anhand eines NMOS-Transistors besprochen werden, unterschieden. Für einen PMOS-Transistor gilt, mit jeweils negativen Spannungen gegenüber Source, prinzipiell das Gleiche.

Im *Abschaltbereich* fließt kein Drain-Strom, da die Gate-Source-Spannung U_{GS} unterhalb der Schwellspannung U_{th} liegt.

Im *Widerstandsbereich* liegt die Gate-Source-Spannung über der Schwellspannung. Unterhalb des Gate-Oxids bildet sich der Kanal. Für den Drain-Strom gilt:

$$I_{DS} = \beta \left((U_{GS} - U_{th}) U_{DS} - \frac{U_{DS}^2}{2} \right) \quad \text{für } U_{DS} < U_{GS} - U_{th} \quad (3.8)$$

β ist die Zusammenfassung aller fertigungsabhängigen Konstanten. Details dazu finden sich in [2] und [29]. Obwohl dieser Bereich üblicherweise Widerstandsbereich genannt wird, ist annähernd lineares Verhalten nur für kleine Werte von U_{DS} gegeben. Der Widerstandswert wird durch U_{GS} gesteuert.

Im *Stromquellenbereich* oder Sättigungsbereich ist die Drain-Source-Spannung größer als die Gate-Source-Spannung. Dies führt zu einer Verarmung an La-

dungsträgern in der Nähe des Drains. Der Kanal ist in diesem Bereich unterbrochen (Abschnürung), der Stromfluss wird jedoch durch Injektion beweglicher Ladungsträger aufrecht erhalten. Für den Drain-Strom gilt:

$$I_{DS} = \beta \frac{(U_{GS} - U_{th})^2}{2} \quad \text{für} \quad U_{DS} \geq U_{GS} - U_{th} \quad (3.9)$$

Der MOS-Transistor ist also eine durch die Gate-Source-Spannung gesteuerte Stromquelle und im Idealfall unabhängig von der anliegenden Drain-Source-Spannung. Der Drain-Strom wächst quadratisch mit der Gate-Source-Spannung.

Neben dem beschriebenen Idealverhalten haben weitere Effekte Einfluss auf das Schaltverhalten. Bisher wurde angenommen, dass Source und Substrat auf gleichem Potential liegen. Dies ist bei der Serienschaltung während des Schaltvorgangs nicht der Fall: Das Source-Potential wird gegenüber dem Substrat angehoben. Dieser *Body-Effekt* hat zur Folge, dass sich die Schwellspannung erhöht, was den Drainstrom verringert. Dieser und weitere Effekte so genannter zweiter Ordnung werden in [2], [3] und [29] behandelt.

Für Kanallängen im Submikrometerbereich ($< 0,5 \mu\text{m}$) sind verschiedene Annahmen des idealen Transistors nicht zutreffend. Statt einer quadratischen Abhängigkeit des Drain-Stromes von der Gate-Source-Spannung stellt sich eine Abhängigkeit der Form

$$I_{DS} = k(U_{GS} - U_{th})^\alpha \quad \text{mit } \alpha \in [1, 2] \quad (3.10)$$

ein. Dies führte zur Propagierung des α -Power-Law-MOS-Modells [30], das sich insbesondere zur Timinganalyse eignet.

Des Weiteren ist die Stromquellen-Eigenschaft im Sättigungsbereich nicht so deutlich wie im Shockley-Modell gegeben [29]. Ein weiterer Kurzkanaleffekt ist die frühzeitige Entstehung einer schwachen Inversion, so dass es schon vor Erreichen der Schwellspannung zu einem geringen Stromfluss kommt [29]. Heute sind die BSIM3- und BSIM4-Modelle [31][32] die am häufigsten verwendeten MOSFET-Transistormodelle.

Neben den Gleichstrom-Eigenschaften haben die parasitären Effekte des MOS-Transistors Einfluss auf das Zeitverhalten der Schaltung. Zwischen Drain- und Source-Gebiet einerseits und Substrat beziehungsweise Wanne andererseits gibt

es PN-Übergänge, die in Sperrrichtung geschaltet sind. Diese können eine Ladung aufnehmen, die im Schaltfall umgeladen werden muss. Sie stellen nichtlineare Kapazitäten dar, deren Kapazitätswerte mit zunehmender Spannungsdifferenz zwischen P- und N-Gebiet geringer werden [2][29].

Die eigene Gate-Kapazität ist für das Schaltverhalten des Gatters relativ unbedeutend. Die Gates der nächsten Stufe stellen jedoch neben der Kapazität der Leitung einen erheblichen Anteil der umzuschaltenden Lastkapazität dar. Auch diese Kapazität ist nichtlinear. Ihr Wert erreicht für $U_{GS} = U_{th}$ ein Minimum [2][3], da sich in diesem Bereich unter dem Gate-Oxid eine maximal große Raumladungszone gebildet hat.

Gravierend für das Zeitverhalten ist jedoch nicht die Nichtlinearität der Gate-Kapazität der nächsten Stufe, sondern die Tatsache, dass sich während des Schaltvorgangs die Spannung von Drain und Source ändern kann. Wenn man annimmt, dass der betrachtete Transistor die Zustandsänderung am Gatterausgang auslöst, so tritt aufgrund des invertierenden Schaltverhaltens immer eine Gegenkopplung zwischen Drain und Gate ein. Lässt man zusätzlich gleichzeitiges Schalten mehrerer Transistoren zu, so kann es auch zu Gegenkopplung zwischen Source und Gate kommen.

Im Rahmen der statischen Timinganalyse ist es üblich, die Gate-Kapazität linear zu modellieren. Die Gegenkopplung durch die Gate-Drain- und Gate-Source-Kapazität wird in der Charakterisierung berücksichtigt. Dazu wird die Gate-Kapazität anhand der umgesetzten Ladung bei einer Spannungsänderung am Gate bestimmt, wobei gleichzeitiges Schalten mehrerer Transistoren im Allgemeinen ausgeschlossen wird.

Eine Berechnung der Verzögerung könnte unmittelbar das Shockley-Transistormodell verwenden und symbolisch durchgeführt werden. Diese Rechnung ist beispielhaft in [2] für einen Inverter durchgeführt. Zur Durchführbarkeit dieser Rechnung sind erhebliche Vereinfachungen notwendig. Unter anderem bleibt die Kurvenform der Eingangsspannung unberücksichtigt und die Ausgangskennlinie wird in einen Stromquellen- und einen linearen Widerstandsbereich unterteilt. Symbolische, auf analytischen Modellen basierende Berechnungen haben aufgrund dieser unerlässlichen Vereinfachungen keine praktische Bedeutung. Die numerische Berechnung basierend auf Transistormodellen ist üblich und wird im nächsten Abschnitt behandelt.

3.3.4 Transistorbasierte Verzögerungsmodelle

Mit transistorbasierten Modellen ist in der Timinganalyse die Modellierung der aktiven Schaltungsteile gemeint. Dieses ist zwar dem Kapitel Gattermodelle zugeordnet, die Transistormodelle können jedoch auch zur statischen Timinganalyse auf Transistorebene verwendet werden. Der Unterschied ist darin zu sehen, dass zur Timinganalyse auf Transistorebene vorab die Ermittlung des Signalfusses notwendig ist.

Das einfachste transistorbasierte Modell ist das so genannte *Schaltermmodell* (engl. »switch level model«) [18][33]. Transistoren werden durch die Reihenschaltung eines idealen Schalters mit einem Widerstand modelliert. Im einfachsten Fall ist der Widerstandswert konstant. Um zusätzlich die Anstiegszeit des Eingangsergebnisses berücksichtigen zu können, wurde der Ansatz durch eine Tabellierung des Widerstandswertes in Abhängigkeit von der Anstiegszeit erweitert [18]. Die parasitären Drain- und Source-Kapazitäten werden linear modelliert und berücksichtigt.

Der Vorteil dieses Modells ist die sehr einfache Berechnung, bei der die Ermittlung der Kurvenform am Gatterausgang bei Abschluss mit einer diskreten Kapazität rein analytisch oder sehr schnell numerisch erfolgen kann. Durch die Art des Ansatzes können sonst schwierig zu behandelnde Gattertypen wie Tri-State- und Pass-Gatter einfach als Serienschaltung von Widerständen aufgefasst werden. Ousterhout [18] hat 1985 mit der Software »Crystal« für Schaltungen in einem 4 µm-Prozess durchschnittlich höchstens zehn Prozent Abweichung gegenüber Spice-Simulationen erreicht.

Der Schaltermmodell-Ansatz beinhaltet zwei wesentliche Nachteile: Die Anstiegszeit am Gattereingang kann zwar berücksichtigt werden, ihre Ermittlung ist jedoch ungenau (bis zu 30 % Abweichung laut [18]). Gravierender ist die schlechte Modellierung des Body-Effektes bei der Serienschaltung von Transistoren. Die Verschiebung des Sourcepotentials gegenüber dem Substrat kann durch den Schaltermmodell-Ansatz nicht berücksichtigt werden.

Um die Nachteile des Switch-Level-Ansatzes zu überwinden, die Vorteile jedoch weitgehend zu erhalten, ist eine bessere Modellierung des Transistorverhaltens erforderlich. Dazu sollen zwei Ansätze dargestellt werden.

Das α -Power-Law-MOS-Modell [30] modelliert das Ausgangsverhalten von Transistoren als ideale Stromquellen beziehungsweise als Widerstände. Dazu werden Gleichungen angegeben, die diese Bereiche abhängig von U_{GS} und U_{DS} beschreiben. Die Parameter des Modells werden durch eine Charakterisierung des Transistors mit einem Analogsimulator gewonnen. Das Modell wurde für die Analyse von Serienschaltungen von MOS-Transistoren erweitert [34]. Das α -Power-Law-Modell wurde in [35] in einem Simulator zur Verzögerungsbeziehung verwendet. Die Steigerung der Geschwindigkeit gegenüber Spice lag nur bei Faktor zwei. Ein dem α -Power-Law-Modell ähnlicher Ansatz wird in [36] verfolgt, wobei die Parameter des Modells durch Methoden der Ordnungsreduktion [37] (siehe auch Abschnitt 3.4), gewonnen werden.

Vorteil dieses Modells ist die gute Erfassung des Einflusses der Eingangskurvenform auf die Gatterverzögerung. Nachteil, vor allem gegenüber dem im Folgenden beschriebenen Ansatz, ist der hohe Rechenaufwand zur Evaluierung der Modellgleichungen.

Der zweite Ansatz beschreibt das Gleichstromverhalten des Transistors in einer Tabelle [26]. Bei feiner Diskretisierung lässt sich ein nahezu beliebig kompliziertes Gleichstromverhalten der Transistoren erfassen. Wird der Drainstrom über die Variation von Drain-, Gate- und Source-Spannungen gegenüber Substrat tabelliert, so wird auch der Body-Effekt (siehe 3.3.3) korrekt berücksichtigt.

Vorteil des Tabellenmodells ist die sehr schnelle Evaluierung der Modellgleichungen, die sich auf das Auslesen von Tabellenwerten und gegebenenfalls eine anschließende Interpolation beschränkt. Die Interpolation muss auf sehr einfache Weise erfolgen oder entfallen, um die durch Tabellierung gewonnenen Vorteile nicht zu vergeben. Nachteilig ist der hohe Speicherbedarf der Tabelle. Die Gleichstrom-Charakterisierung der Transistoren mit einem Analogsimulator benötigt bei 50 Intervallen, also $51^3 = 132651$ Tabellenwerten, weniger als fünf Minuten. Häufig wird gegen Tabellenmodelle angeführt, dass für jede Parametervariation einschließlich der Temperatur eine komplette Tabelle angelegt werden muss. Für die Timinganalyse ist dies unbedeutend, da nur jeweils eine Tabelle für schnell und langsam schaltende Transistoren erforderlich sind.

Beide letztgenannten Modelle lassen sich bei beliebigen Eingangskurvenformen nicht mehr analytisch rechnen. Die Ermittlung der Ausgangskurvenform erfolgt durch einen numerisch arbeitenden Schaltungssimulator.

3.3.5 Empirische Gattermodelle

Empirische Gatterverzögerungsmodelle machen über die Struktur des Gatters keine Annahmen. Allen Ansätzen ist gemeinsam, dass jeder Gattertyp mittels eines Analogsimulators charakterisiert werden muss.

Die folgenden Gattermodelle zeichnen sich durch dieselbe Beschreibung der Ein- und Ausgangssituation aus. Die Eingangssituation nimmt eine Strecke als Spannungsverlauf an. Die Ausgangssituation ist ein diskreter Kapazitätswert.

Das bekannteste Gattermodell ist das so genannte *K-Faktor-Modell* [2] [28]. Der Name rührt daher, dass die Abhängigkeiten der Verzögerung von physikalischen Größen der Ein- und Ausgangssituation linear mit den Proportionalitätskonstanten k_i modelliert werden. Das einfache K-Faktor-Modell [2] beschreibt die Verzögerung T_D als:

$$T_D = T_{D,i} + k_C \cdot C \quad (3.11)$$

C ist die diskrete Lastkapazität am Ausgang, k_C der Proportionalitätsfaktor, der die Dimension des Widerstands besitzt. Unter der Annahme eines linearen Spannungsanstiegs am Ausgang wird das Verhalten einer Stromquelle modelliert. $T_{D,i}$ stellt die gatterinterne Verzögerung ohne Lastkapazität dar. Zur Bestimmung dieser Werte ist eine Charakterisierung erforderlich. Für dieses einfache Modell könnte jeweils eine Analogsimulation ohne und mit maximaler Ausgangslast durchgeführt werden.

Da dieses Modell die Eingangssituation nicht berücksichtigt, wurde es um eine weitere lineare Abhängigkeit erweitert [28]:

$$T_D = T_{D,i} + k_C \cdot C + k_R \cdot T_{R, \text{ein}} \quad (3.12)$$

In die Verzögerung geht ein Anteil ein, der von der Anstiegszeit am Eingang $T_{R, \text{ein}}$ linear abhängt. Die Anstiegszeit am Ausgang wird als

$$T_{R, \text{aus}} = 2 \cdot k_C \cdot C \quad (3.13)$$

definiert. Sie dient der Berechnung der Verzögerung der nächsten Stufe. Da dieses Modell in der so genannten »Slow-Ramp-Region« [38], in der die Anstiegszeit relativ groß, die Ausgangslast jedoch gering ist, die Verzögerung grob überschätzt, wurde in [38] eine weitere Stützstelle in genau diesem Operationsbereich eingeführt.

Um den empirischen Ansatz weiter zu verbessern, können die Verzögerung und Ausgangsanstiegszeit tabelliert werden [28]. Dazu sind viele Analogsimulationen zur Gewinnung der Tabellenwerte notwendig, die Abschätzungen von Verzögerung und Anstiegszeit sind unter den gemachten Annahmen gut. Die Verzögerungsberechnung mit dem genannten Tabellenmodell kann als heutiger Standard angesehen werden.

Alle bisher besprochenen empirischen Modelle haben folgende Gemeinsamkeiten:

- Die Spannungsverläufe am Ein- und Ausgang sind Strecken zwischen 0 V und U_{DD} (beziehungsweise zwischen $U_{th,u}$ und $U_{th,o}$).
- Die Charakterisierung der Parameter setzt die Definition von Spannungsschwellwerten, anhand derer die Anstiegszeit und die Verzögerung gemessen werden können, voraus.
- Die Ausgangssituation ist eine diskrete Kapazität.

Der Vorteil dieser Modelle liegt im geringen Rechenaufwand zur Evaluierung der Modellgleichungen, was eine hohe Ausführungsgeschwindigkeit der statischen Timinganalyse sicherstellt. Die Nachteile sind bereits im Ansatz impliziert und gelten daher für alle Modelle.

Die Annahme des linearen Verlaufs der Ereignisse ermöglicht zwar eine einfache Charakterisierung, ist jedoch physikalisch nicht haltbar. Ein Problem der Charakterisierung liegt darin, eine sinnvolle Definition der Anstiegszeit zu geben. Diese kann erheblichen Einfluss auf die Verzögerung des nächsten Gatters haben.

Ebenfalls ein Nachteil ist die Annahme einer diskreten Kapazität am Gatterausgang. Diese Annahme führt zu einer Überschätzung der Gatterverzögerung für Leitungen mit hohem Widerstandswert am Gatterausgang. Durch die Definition einer *effektiven Kapazität* in [39] kann dieser Nachteil überwunden werden. Andere Ansätze [40][41] schlagen empirische Modelle vor, die mit diskreten Kapazitäten charakterisiert, jedoch auch mit π -RC-Gliedern berechnet werden können.

3.4 Leitungsmodelle

Das bekannteste Verzögerungsmodell für Leitungen ist die *Elmore-Verzögerung* [42]. Sie ist nach ihrem Entwickler W. C. Elmore benannt, der diese Verzögerungsmetrik bereits 1948 vorschlug. Die Elmore-Verzögerung T_D ist definiert als:

$$T_D = \frac{\int_{t=0}^{\infty} t h(t) dt}{\int_{t=0}^{\infty} h(t) dt} \quad (3.14)$$

$h(t)$ ist die Impulsantwort des Systems, hier die Antwort am Leitungsende auf einen Diracimpuls am Leitungsanfang. Anschaulich ist T_D die t -Koordinate des Schwerpunktes der durch $h(t)$ und der Zeitachse begrenzten Fläche.

Die große Bedeutung der Elmoreverzögerung als Metrik für die Verzögerung von Leitungen rührt daher, dass sie sich für RC-Bäume mit linearer Komplexität berechnen lässt. Für einen RC-Baum gilt:

- Er besteht ausschließlich aus Kapazitäten und Widerständen.
- An jedem Schaltungsknoten i ist eine Kapazität C_i nach Masse angeschlossen.
- Kein Widerstand ist an Masse angeschlossen. Ein Widerstand sei mit R_{ij} bezeichnet, wenn er zwischen den Knoten i und j liegt.
- Es gibt keine Zyklen aus Widerständen.

Gegeben seien zwei Schaltungsknoten i und j . Der Signalfluss sei von i nach j , das heißt, die Elmore-Verzögerung von i nach j soll bestimmt werden. Die Knotenmenge J beinhaltet diejenigen Schaltungsknoten, die einen Pfad aus Widerständen zu j haben, der R_{ij} nicht enthält. Anschaulich sind dies Schaltungsknoten, die j in Signalflussrichtung folgen. Es gilt $j \in J$. Die Summe der Kapazitäten der Schaltungsknoten von J sei mit C_{jj} bezeichnet. Für die Elmore-Verzögerung T_{Dij} über einen Widerstand R_{ij} gilt:

$$T_{Dij} = R_{ij} C_{jj} \quad (3.15)$$

Die Elmore-Verzögerung zwischen zwei beliebigen, durch einen Widerstandpfad verbundenen Schaltungsknoten (im Folgenden Start- und Endknoten genannt) ist die Summe der Elmore-Verzögerungen über jeden Widerstand zwischen diesen

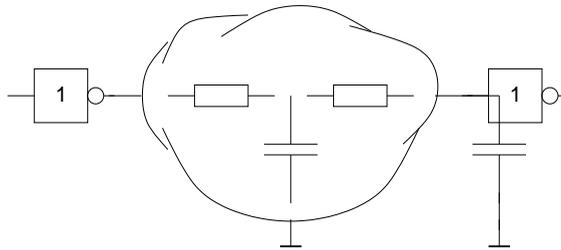
Knoten. Die Berechnung der Elmore-Verzögerung erfolgt in zwei Schritten: Erstens wird jeder Schaltungsknoten j mit der Summenkapazität in Signalflussrichtung C_{jj} und der Elmore-Verzögerung T_{Dij} über jeden Widerstand markiert. Dies erfolgt sinnvollerweise per Rekursion ausgehend vom Startknoten. Zweitens wird jeder Knoten mit der Summe der Elmore-Verzögerungen vom Startknoten zum aktuell betrachteten Knoten markiert. Bei beiden Schritten wird jeder Knoten jeweils einmal besucht.

Neben der linearen Komplexität der Berechnung spricht eine weitere Eigenschaft für die Elmore-Verzögerung als Metrik für RC-Bäume: Sie ist die Obergrenze für die 50 %-Verzögerung unter bestimmten, realistischen Annahmen über den Spannungsverlauf. Der Beweis dieser Eigenschaft erfolgte 1997 in [43].

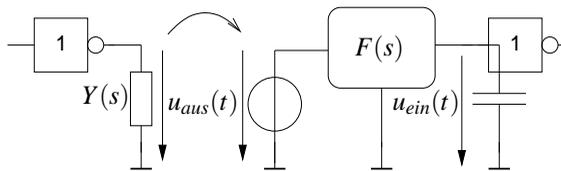
Unter Ordnungsreduktion versteht man Verfahren, die Anzahl der extrahierten Parasiten zu verringern oder eine vereinfachte Beschreibung für sie anzugeben, die für die Anwendung dominanten Eigenschaften jedoch zu erhalten. Abbildung 3.6 zeigt die prinzipielle Anwendung.

Das vom Extraktor aus dem Layout generierte Originalsystem ist im Allgemeinen diskretisiert und kann mehrere tausend Widerstands- und Kapazitätselemente pro Netz enthalten. Eine Analogsimulation eines Netzes kann bereits mehrere Stunden bis Tage dauern. Für die Validierung des Ergebnisses der statischen Timinganalyse ist diese Laufzeit unter Umständen noch akzeptabel, für die Timinganalyse selbst jedoch nicht.

Die Ordnungsreduktion liefert ein reduziertes System, das sowohl eine Verhaltensbeschreibung (als $Y(s)$ und $F(s)$ in Abbildung 3.6 dargestellt) als auch eine synthetisierte Schaltung sein kann. Für die Anwendung in der Timinganalyse sind zwei Reduktionsschritte notwendig: Eine reduzierte Ausgangsadmittanz $Y(s) = I_{aus}(s)/U_{aus}(s)$ und eine reduzierte Übertragungsfunktion $F(s) = U_{ein}(s)/U_{aus}(s)$. Die Ausgangsadmittanz wird zur Berechnung der Ausgangsspannung $u_{aus}(t)$ verwendet. Die Eingangsspannung u_{ein} wird danach aus der Ausgangsspannung und der Übertragungsfunktion $F(s)$ berechnet. In der Literatur sind zahlreiche Verfahren zur Ordnungsreduktion veröffentlicht, zum Beispiel [37][44][45][46][47]. Große Bedeutung hat insbesondere die Momentenanpassung, die in [37] und [44] behandelt wird. Zur Erläuterung des Momentenbegriffs wird auf die zitierte Literatur verwiesen. Es sei jedoch das Folgende bemerkt: 1. 1989 wurde in [37] gezeigt, dass bei der Momentenerhaltung das Zeitverhalten gut wiedergeben wird. 2. Die Elmore-Verzögerung ist das erste Moment der Impulsantwort und somit



Originalsystem



Reduziertes System

Abbildung 3.6: Anwendung der Ordnungsreduktion in der Timinganalyse

ein Spezialfall der Momentenanpassung. Jüngere Artikel konzentrieren sich auf die Erhaltung der Passivität [48].

3.5 Modellierung der Verzögerung bei kapazitiver Kopplung

3.5.1 Problemstellung

Das grundsätzliche Problem der Verzögerungsmodellierung bei vorhandener kapazitiver Kopplung liegt darin, dass auch der nicht mit dem betrachteten Gatterausgang verbundene Anschluss der Lastkapazität nicht in Ruhe ist. Die Verzöge-

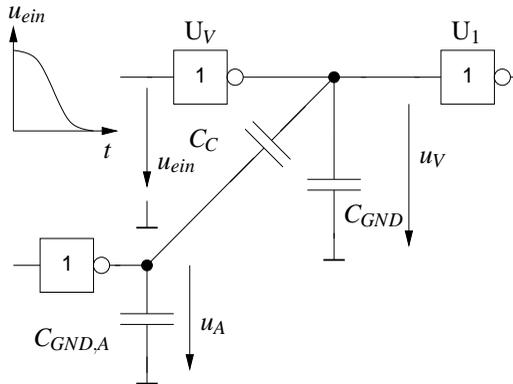


Abbildung 3.7: Zur Verzögerungsmodellierung unter Einbeziehung von Kopplung

Die bisher behandelten Verzögerungsmodelle werden passive Modelle genannt. Sie liefern ein Ausgangsverhalten, das nur von dem betrachteten Gatter selbst, seiner Eingangs- und Ausgangssituation und Konfiguration abhängt. Soll kapazitive Kopplung einbezogen werden, kommen weitere Parameter hinzu: Die Koppelkapazität C_C , die relative Ausrichtung der Ankunftszeit der Nachbarleitungen (engl. »alignment« [49]), sowie alle zur Verzögerungsberechnung relevanten Parameter (wie oben: Eingangs- und Ausgangssituation, Konfiguration) der koppelnden Gatter. Die betrachtete Leitung und das diese Leitung treibende Gatter nennt man *Opfer* (engl. »victim«). Für die einkoppelnden Nachbarleitungen und Gatter wird der Begriff *Aggressor* verwendet. Das Opfergatter ist das Gatter, das die Opferleitung treibt.

Die bisher behandelten Verzögerungsmodelle werden passive Modelle genannt. Sie liefern ein Ausgangsverhalten, das nur von dem betrachteten Gatter selbst, seiner Eingangs- und Ausgangssituation und Konfiguration abhängt. Soll kapazitive Kopplung einbezogen werden, kommen weitere Parameter hinzu: Die Koppelkapazität C_C , die relative Ausrichtung der Ankunftszeit der Nachbarleitungen (engl. »alignment« [49]), sowie alle zur Verzögerungsberechnung relevanten Parameter (wie oben: Eingangs- und Ausgangssituation, Konfiguration) der koppelnden Gatter. Die betrachtete Leitung und das diese Leitung treibende Gatter nennt man *Opfer* (engl. »victim«). Für die einkoppelnden Nachbarleitungen und Gatter wird der Begriff *Aggressor* verwendet. Das Opfergatter ist das Gatter, das die Opferleitung treibt.

Geht man von einer Verzögerungsmodellierung mit n zu charakterisierenden Parametern aus, so ist der gesamte Charakterisierungsaufwand $O(e^n)$, da alle Parameterkombinationen erfasst werden müssen. Nimmt man des Weiteren an, dass nur ein Aggressor vorhanden ist und das Verzögerungsmodell für die Kopplung nicht weiter abstrahiert werden soll, so steigt die Komplexität auf $O(e^{2n+1})$. Im

Falle des K-Faktor-Verzögerungsmodells (siehe Abschnitt 3.3.5) ist $n = 2$, bei nur einem Aggressor die Komplexität daher $O(e^5)$. Es wäre ein fünfdimensionaler Raum zu charakterisieren, was bereits bei diesem einfachen Verzögerungsmodell nahezu unmöglich ist. Eine Vereinfachung der Modellierung ist unvermeidlich.

Es gibt zwei grundsätzliche Strategien zur Behandlung von Kopplung: Passive und aktive Modellierung. Bei der passiven Modellierung werden die Koppelkapazitäten, eventuell mit verändertem Wert, als an Masse angeschlossen betrachtet. Aggressor-Gatter und -Netz sind nicht Teil des Modells. Bei der aktiven Modellierung sind der Spannungsverlauf an den Ausgängen der Aggressor-Gatter und die Koppelkapazitäten Teil des Modells.

3.5.2 Passive Modelle

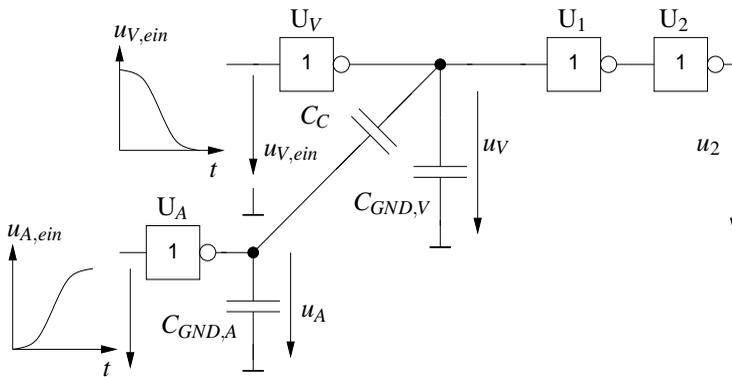


Abbildung 3.8: Schaltung zur Prüfung der Miller-Faktors

Kennzeichen der passiven Modelle ist, dass die Verzögerungsberechnung des Opfergatters mit einer an Masse angeschlossenen Kapazität als Last erfolgt. Die Koppelkapazität wird durch eine Ersatzkapazität modelliert. Der Wert der Ersatzkapazität ist größer oder gleich dem der Koppelkapazität:

$$C_{Ersatz} = kC_c, \text{ mit } k \geq 1 \quad (3.16)$$

k wird Miller-Faktor genannt. Dieser Begriff stammt ursprünglich aus der Analschaltungstechnik [6].

Häufig wird $k = 2$ gewählt. Dieses Vorgehen ist physikalisch dadurch gerechtfertigt, dass bei entgegengesetztem Schalten die Koppelkapazität mit dem Spannungshub $2U_{DD}$ umgeladen wird. Das Opfertgatter pumpt die Ladung $Q = 2U_{DD}C_C$ in die Koppelkapazität. Eine ladungserhaltende, an Masse angeschlossene Ersatzkapazität hat folglich den Kapazitätswert $2C_C$.

Obwohl dieses Modell vielfach eingesetzt wird, liefert es keine Obergrenze für die Verzögerung. Für die Beispielschaltung in Abbildung 3.8 zeigt Abbildung 3.9 die Spannungsverläufe. $u_{2,Original}$ der Originalschaltung liegt deutlich rechts von $u_{2,Ersatz}$ der Ersatzschaltung mit $C_{Ersatz} = 2C_C$ an Masse angeschlossen. Ähnliche Beobachtungen wurden in [50], [51] und [52] veröffentlicht. Das Modell ist, obwohl ladungserhaltend, für die Verzögerungsmodellierung schlecht geeignet.

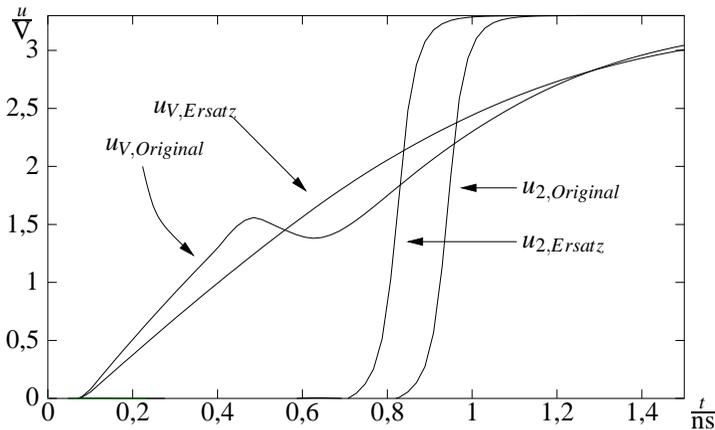


Abbildung 3.9: Plots der Spannungsverläufe der Original- und Ersatzschaltung mit $k = 2$

Diese Tatsache führte zu Vorschlägen, die Koppelkapazität zwar passiv zu modellieren, jedoch Miller-Faktoren $k > 2$ zu verwenden. Chen et. al [50] bestimmen

den Miller-Faktor anhand der bis zu einer Schwellspannung U_{th} umgesetzten Ladung:

$$\Delta Q = \left(\sum_i C_{C,i} + C_{GND} \right) \Delta u_V - \sum_i C_{C,i} \Delta u_{A,i} \quad (3.17)$$

$$= \left(\sum_i C_{C,i} \left(1 - \frac{\Delta u_{A,i}}{\Delta u_V} \right) + C_{GND} \right) \Delta u_V \quad (3.18)$$

Mit

$$\Delta u_V = U_{th} \quad (3.19)$$

folgt:

$$\Delta Q = \left(\sum_i C_{C,i} \left(1 - \frac{\Delta u_{A,i}}{U_{th}} \right) + C_{GND} \right) U_{th} \quad (3.20)$$

Der Miller-Faktor ergibt sich zu:

$$k_i = 1 - \frac{\Delta u_{A,i}}{U_{th}} \quad (3.21)$$

Der Miller-Faktor ist auch von der Anstiegszeit des Aggressorereignisses abhängig, so dass dieser Ansatz aktive Modellierung berücksichtigt. Entscheidend ist, welcher Aggressor-Spannungshub Δu_A vor Erreichen von U_{th} auf der Opferleitung erreicht wird. Im Falle von $\Delta u_A = -U_{DD}$ und $U_{th} = \frac{1}{2}U_{DD}$ ergibt sich ein Miller-Faktor von $k = 3$. Da die Zeit, die vergeht, bis die Opfer-Spannung U_{th} erreicht, selbst von Δu_A abhängt, schlagen die Autoren ein Verfahren zur schrittweisen Annäherung vor. Beginnend mit $k_i = 3$ werden die Aggressor-Spannungshübe $\Delta u_{A,i}$ und daraus neue Werte für k_i berechnet. Der Spannungsverlauf der Aggressoren wird als eingepreßt und unveränderlich behandelt. Das Modell beinhaltet zudem die Annahme, dass das Opfergatter Stromquellenverhalten hat, da der vom Opfergatter aufgebrauchte Strom nicht von der Spannung abhängt.

Das Verfahren hat den Vorteil, dass es unabhängig vom Verzögerungsmodell ist und leicht in bestehende statische Timinganalyse-Werkzeuge integriert werden kann. Für den Fall, dass der vom Betrag maximale Aggressor-Spannungshub vorausgesetzt wird, ist die vorgeschlagene Berechnung des Miller-Faktors einfach durchführbar.

Die Autoren behandeln keine widerstandsbehafteten Leitungen. Eine Erweiterung des Modells in diese Richtung bleibt unklar. Einen ausführlichen Vergleich zu Spice-Simulationen geben die Autoren nicht an.

3.5.3 Aktive Modelle

Diese Modelle sind dadurch gekennzeichnet, dass die Aggressoren und ihre Spannungsänderung Teil des Modells sind. Die Idee in [49] und [51] ist, die Auswirkungen von Kopplung auf die Opfer-Leitung durch eine Superposition zu ermitteln. Das Verfahren gliedert sich in folgende Schritte: In einem ersten Schritt werden die Admittanz- und Übertragungsfunktions-Parameter des koppelnden Systems gebildet, wozu Verfahren der Ordnungsreduktion angewendet werden (siehe Abschnitt 3.4). Die Gattermodellierung erfolgt dabei durch gesteuerte Stromquellen mit Innenwiderstand. Für jeden Eingang an der betrachteten Opferleitung erfolgt dann die Berechnung der durch das die Leitung treibende Opfer-Gatter und aller Aggressor-Gatter hervorgerufenen Spannungsverläufe. Da der Innenwiderstand der verwendeten Gattermodelle vom Arbeitspunkt abhängt, wird geprüft, ob dieser noch gültig ist und gegebenenfalls die Berechnung wiederholt. Die Verschiebung der Aggressor-Ankunftszeiten erfolgt derart, dass die Überlagerung aller Spannungskurven an den Eingängen das Erreichen des Schwellwertes von $\frac{1}{2}U_{DD}$ möglichst weit hinausschiebt (siehe auch Abschnitt 5.2).

Vorteil dieser Methode ist die Möglichkeit zur Berücksichtigung des Leitungswiderstandes. Leider geben die Autoren keine ausführlichen Vergleichszahlen zu mittels Analogsimulationen erzielten Maximalverzögerungen an. Insbesondere wird die Verzögerung ganzer Pfade nicht validiert. So bleibt der mögliche Fehler durch die Linearisierungen und die Qualität des Modells allgemein unklar. Die Wahl des Spannungsschwellwertes von $\frac{1}{2}U_{DD}$ wird nicht gerechtfertigt.

Ebenfalls ein aktives Verzögerungsmodell mit linearisiertem Opfer-Gatter wird in [53] vorgeschlagen. Die Autoren zeigen insbesondere, dass die maximale Verzögerung nicht in jedem Fall nach der in [49] vorgestellten Methode erreicht wird. Dies ist darauf zurückzuführen, dass als Verzögerungsmaß das *erste* Erreichen eines Schwellwertes angenommen wird. Die Vorgehensweise in [49] ist korrekt, wenn der *letzte* Zeitpunkt des Schnittpunktes der Spannungskurve mit der Schwellspannung maximiert werden soll.

3.6 Kapazitive Kopplung in der statischen Timinganalyse

In Abschnitt 2.5.3 wurde dargelegt, dass die statische Timinganalyse das jeweils erste und letzte Ereignis jedes Knotens des Zeitverhaltensgraphen bestimmen kann. Diese Eigenschaft kann man sich bei der Berücksichtigung der kapazitiven Kopplung in der statischen Timinganalyse nutzbar machen. Statt in jedem Fall die ungünstigste Situation der permanenten Kopplung anzunehmen, kann im Einzelfall berücksichtigt werden, ob Aktivität auf der Nachbarleitung möglich ist. Abbildung 3.10 soll dies verdeutlichen.

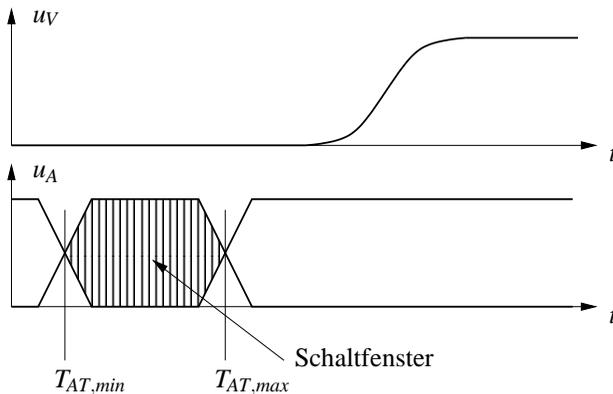


Abbildung 3.10: Schaltfenster

Für das betrachtete Ereignis auf der Opfer-Leitung ist das letzte Ereignis auf der betrachteten Aggressor-Leitung bereits verstrichen. Dies hat folgende Konsequenz: Da das Aggressorgatter in Ruhe ist, ist der Schaltungsknoten niederohmig an die Versorgungsspannung oder Masse angeschlossen. Die Koppelkapazität wirkt daher wie eine passive Lastkapazität. Es entsteht keine zusätzliche Verzögerung durch Kopplung. Die statische Timinganalyse kann dazu verwendet werden, Situationen dieser Art zu erkennen.

Schaltfenster im Zusammenhang mit Kopplung wurden zuerst in [54] und [55] beschrieben. In beiden Artikeln werden die funktionalen Auswirkungen

von Kopplung betrachtet, in [54] zusätzlich der zeitliche Einfluss auf wenige ausgewählte Netze.

In [56] und [57] wird die Berücksichtigung kapazitiver Kopplung in der statischen Timinganalyse dargestellt. In [57] wird dazu das aktive Verzögerungsmodell aus [49] verwendet. Beiden Verfahren ist gemeinsam, dass die Schaltfenster iterativ verkleinert werden. Eine detaillierte Behandlung ist in Abschnitt 5 zu finden. Das Verfahren von [57] arbeitet sowohl auf Transistorebene, wobei der in [58] und [59] präsentierte Simulatorkern verwendet wird, als auch auf Gatterebene und ist in das Programm »EinsTimer« [14] integriert worden.

Die Autoren von [60] greifen das Problem auf, dass entgegengesetztes Schalten zweier Schaltungsknoten auch funktional möglich sein muss. Durch Verwendung von Methoden der Erkennung falscher Pfade (siehe Abschnitt 2.6) wird unmögliches Gegenkoppeln ausgeschlossen. Die Autoren verwenden dazu ein einfaches Verzögerungsmodell, das Gegen-, Mit- und Nichtkopplung mit jeweils einem konstanten Wert berücksichtigt. Neben dieser Einschränkung behandeln die Autoren nur die kombinatorischen ISCAS-85-Schaltungen [61]. Nach den Erfahrungen des Verfassers kann die Kombination eines komplexen Verzögerungsmodells mit der Suche nach falschen Kopplungen zu sehr langen Programmlaufzeiten führen.

In [62] wird eine statische Timinganalyse präsentiert, die zunächst die kritischen Pfade mit einfachen Gatter- und reduzierten Leitungsmodellen ermittelt. Diese kritischen Pfade werden dann automatisiert in dem Analogsimulator »HSpice« simuliert, wobei ein BSIM3-Transistormodell [31] und unreduzierte Leitbahnparasiten verwendet werden. Diese Simulation beinhaltet auch die Koppelkapazitäten. Die Spannungskurve auf der Aggressorseite der Koppelkapazitäten wird der vorangegangenen statischen Timinganalyse entnommen. Die Autoren verwenden dazu die kleinste Anstiegs- oder Abfallzeit (siehe auch Abschnitt 5.1). Es wird behauptet, dass die Pfadverzögerung dann maximal wird, wenn Aggressor- und Opferankunftszeit gleich sind, wobei ein 50 %-Verzögerungsmodell angenommen wird. In Abschnitt 5.1 wird gezeigt, dass diese Annahme unzutreffend sein kann. Die Autoren machen bezüglich der Aggressorankunftszeiten leider keine weiteren Untersuchungen oder Vergleiche mit Spice. Ebenfalls unklar bleibt die Größe der Schaltungen und die Laufzeit der Analyse.

Die Verfasser von [63] beobachten, dass die maximale Pfadverzögerung nicht bei einer maximalen Verzögerung auf dem Opfernnetz bezüglich $\frac{1}{2}U_{DD}$ erreicht wird

(siehe auch Abschnitt 5.1). Sie schlagen eine iterative Suche vor, bei der die Ankunftszeit am Ausgang des dem Opfergatter folgenden Gatters ermittelt wird. Das Opfergatter wird als nicht-ideale Stromquelle, die Aggressor-gatter als ideale Spannungsquellen modelliert. Die Validierung des Verfahrens erfolgt derart, dass die berechneten Aggressorankunftszeiten für eine Analogsimulation der errechneten Pfade in Spice verwendet werden. Eine Validierung die Ankunftszeiten selbst findet nicht statt.

4 Kurvenformbasierte statische Timinganalyse

4.1 Einführung

Um bei der Analyse des Verzögerungseinflusses der kapazitiven Kopplung keine durch das Verzögerungsmodell gegebenen Einschränkungen hinnehmen zu müssen, wird eine statische Timinganalyse, die auf Spannungskurven $t \rightarrow u(t)$ basiert, gewählt. Alle in Kapitel 3.4 behandelten Eingangssituationen sind auf diese Weise darstellbar, so dass prinzipiell jedes Verzögerungsmodell in jeder Abstraktionsebene oberhalb und einschließlich der Transistorebene verwendet werden kann. Für die Transistorebene wird ein Analogsimulatorkern entwickelt, der für CMOS-Schaltungen hinsichtlich der Ausführungsgeschwindigkeit optimiert ist.

Die Aufgabe der Analogsimulation besteht darin, Spannungen und Ströme innerhalb einer Schaltung bei einer gegebenen Erregung numerisch zu berechnen. Man unterscheidet Gleichstrom-, Kleinsignal- und Großsignal- oder Transientenanalyse [64]. Hier ist im Wesentlichen die Transientenanalyse interessant, da sie den Verlauf von Strömen und Spannungen in Abhängigkeit von der Zeit ermittelt. Die Theorie der numerischen Schaltungssimulation ist in der Literatur hinreichend behandelt [64][65][66][67].

4.2 Modellierung der Schaltungselemente

Auf die Modellierung der einfachen Bauelemente R und C muss nicht weiter eingegangen werden. Der Kern dieses Abschnittes liegt daher in der Modellierung der MOS-Transistoren.

Die Evaluierung der Modellgleichungen in einem Analogsimulator kann bei kleinen CMOS-Schaltungen einen erheblichen Anteil der gesamten Rechenzeit erfordern [68]. Eine Möglichkeit zur Beschleunigung der Simulation ist die Verwendung einfacher MOS-Modelle. Diese Wahl kann den Nachteil haben, nicht alle Effekte hinreichend genau zu erfassen. Es wird für diese Arbeit für das Gleichstromverhalten ein Tabellenmodell gewählt.

Das Gleichstromverhalten der MOS-Transistoren wird in Abhängigkeit von Drain-, Gate- und Source-Spannung bei konstantem Substrat- beziehungsweise Wannepotential tabelliert, wobei ausschließlich der Drainstrom erfasst wird. Die Tabellierung erfolgt für die $0,5\ \mu\text{m}$ -Technologie in 50 Schritten zwischen 0 V und U_{DD} , für die 130 nm -Technologie in 44 Schritten zwischen $-0,4 \cdot U_{DD}$ und $1,4 \cdot U_{DD}$. Die Charakterisierung des Drainstromes erfolgt mit der Schaltung in Abbildung 4.1. Auf diese Weise wird der Body-Effekt (siehe Abschnitt 3.3.3), der bei der Verschiebung des Source- gegenüber dem Substratpotential auftritt, korrekt berücksichtigt.

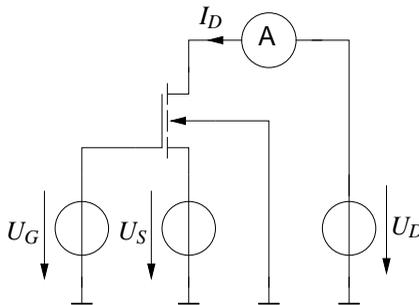


Abbildung 4.1: Charakterisierung des Gleichstrom-Ausgangsverhaltens

Die Interpolation zwischen den Tabellenwerten erfolgt durch ein Verfahren, das im zweidimensionalen Fall »Triangulierung« genannt wird. Die Übertragung auf drei Dimensionen erfordert, dass der Würfel, der durch die acht benachbarten Tabellenwerte bestimmt wird, in fünf Tetraeder zerlegt wird. Auf diese Weise wird sichergestellt, dass der Drainstrom stetig ist.

MOS-Transistoren werden als Zweipole mit den Anschlüssen Drain und Source modelliert. Gate-Potential und -Strom gehen nicht in die Simulation ein. Diese

Wahl ist motiviert durch die Tatsache, dass Gates ausschließlich Gattereingänge oder gatterinterne Eingänge folgender Stufen sind. Durch die Verwendung in einer statischen Timinganalyse werden Gattereingänge ausschließlich von Spannungsquellen gesteuert, so dass keine Rückwirkung von Drain oder Source auf das Gate berücksichtigt werden muss.

Neben dem Gleichstromverhalten spielen die Kapazitäten an Drain und Source eine Rolle für das zeitliche Verhalten eines Gatters. Die Kapazitäten teilen sich im Wesentlichen in zwei Teile, wobei Drain und Source synonym behandelt werden: Erstens in die Gate-Drain-Kapazität, zweitens in die Drain-Substrat-Kapazität (beziehungsweise Drain-Wannen-Kapazität). Die Drain-Substrat-Kapazität ist die Sperrschichtkapazität einer in Sperrrichtung geschalteten Diode. Der Kapazitätswert wird für 0 V Sperrspannung maximal [29]. Die Drain-Substrat-Kapazität ist unabhängig von einer Spannungsänderung am Gate wirksam, während die Gate-Drain-Kapazität sowohl in Ruhe als auch mit entgegengesetzter Schaltrichtung charakterisiert werden muss. Aus Gründen der Effizienz wird ebenfalls eine lineare Modellierung gewählt. Abbildung 4.2 stellt die Schaltung zur Bestimmung der Kapazität beim Schaltvorgang des Gates dar.

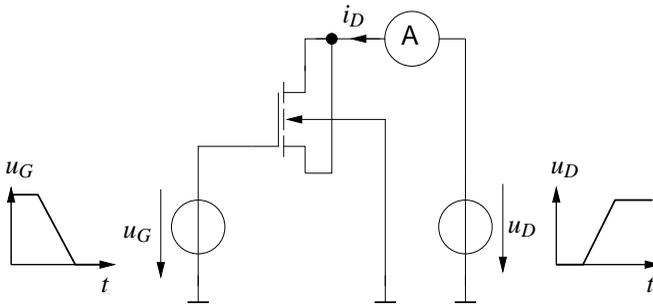


Abbildung 4.2: Charakterisierung der Drain-Kapazität

Für die Drain- und Source-Kapazität gilt:

$$C_D = C_S = \frac{1}{2U_{DD}} \int i_D(t) dt \tag{4.1}$$

Die so ermittelten Kapazitäten führen zu einer Überschätzung der Verzögerung, die darauf zurückzuführen ist, dass nicht in jedem Fall Gegenkopplung während

des Umladens der Kapazitäten vorhanden ist. Die Drain- und Source-Kapazitäten werden, sofern sie mit dem Gatterausgang verbunden sind, mit einem um 7 % verringerten Wert in der Verzögerungsberechnung verwendet.

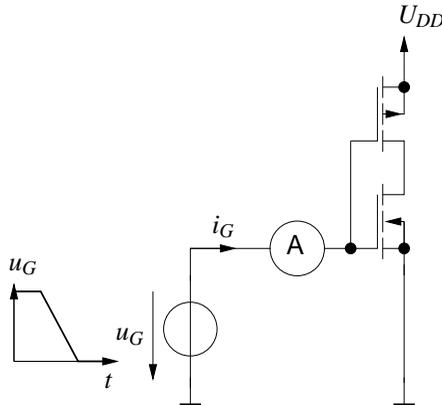


Abbildung 4.3: Charakterisierung der Gate-Kapazitäten eines Inverters

Die Gattereingänge stellen eine kapazitive Last für den Ausgang der vorgehenden Stufe dar. Diese Kapazität ist ebenfalls nichtlinear [2][3]. Die Auswirkungen der Nichtlinearität werden jedoch durch die Effekte des Schaltens des Transistors überwogen. In den verwendeten Schaltungen sind Gates von mehreren Transistoren an einen Knoten angeschlossen. Da die Summenkapazität in der Schaltungssimulation verwendet werden kann, ist eine Bestimmung der einzelnen Transistor-Gate-Kapazitäten nicht nötig. Der Effekt der Gegenkopplung während des Schaltvorgangs wird berücksichtigt. Abbildung 4.3 stellt die Schaltung zur Ermittlung der Summengatekapazität eines Inverters dar. Der Kapazitätswert ist:

$$C_{G,Summe} = \frac{1}{U_{DD}} \int i_G(t) dt \quad (4.2)$$

4.3 Verbesserte Propagierung von Spannungskurven

In Abschnitt 2.5.3 wurde als notwendige Voraussetzung für die Anwendung der Breitensuche in der statischen Timinganalyse die Erkennung des Ereignisses, das in Knoten größerer Tiefe den ungünstigsten Fall darstellt, behandelt. Im Falle des K-Faktor-Modells (siehe Abschnitt 3.3.5) ist die Auswahl einfach, da es unter keinen Umständen Spannungsverläufe, die sich überschneiden, geben kann. Das bezüglich einer beliebigen Schwelle letzte Ereignis stellt den ungünstigsten Fall dar.

Die Auswahl des zu propagierenden Ereignisses ist für beliebige, einander schneidende Spannungskurven nicht eindeutig. Wird als Auswahlkriterium ein einziger Schwellwert, typischerweise $\frac{1}{2}U_{DD}$, verwendet, so kann die Pfadverzögerung unterschätzt werden [69][70].

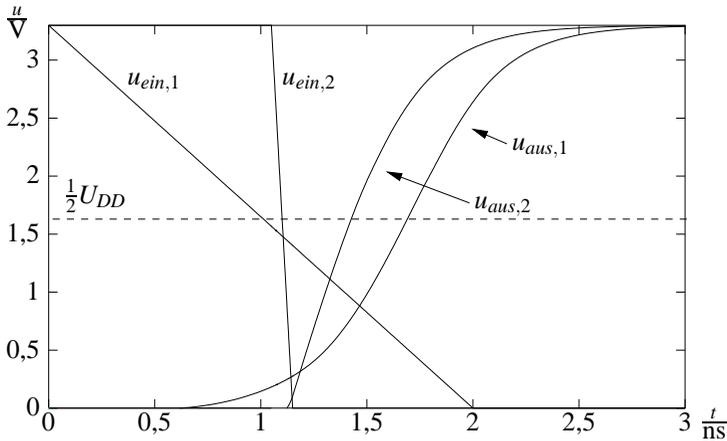


Abbildung 4.4: Propagierung von Ereignissen

Abbildung 4.4 zeigt die Ein- und Ausgangsspannungen zweier gleichartiger und mit der gleichen Kapazität belasteter Inverter. Die Eingangsspannung $u_{ein,1}$ erreicht $\frac{1}{2}U_{DD}$ 0,1 ns vor $u_{ein,2}$. Eine statische Timinganalyse mit $\frac{1}{2}U_{DD}$ als Schwellwert würde daher nur $u_{ein,2}$ propagieren. Bis auf einen kurzen Zeitraum, der ver-

nachlässigt werden kann, da die Spannung unterhalb der Transistorschwellspannung liegt, befindet sich $u_{aus,1}$ deutlich rechts von $u_{aus,2}$. Die Auswahl des verzögerungsbestimmenden Ereignisses anhand der Schwellspannung $\frac{1}{2}U_{DD}$ ist hier nicht korrekt.

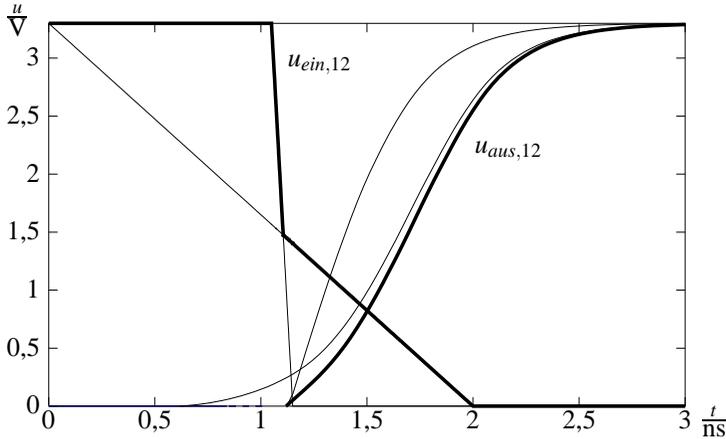


Abbildung 4.5: Neues Verfahren zur Propagierung von Ereignissen

Zu einer korrekten Auswahl wird folgender neuer Ansatz vorgeschlagen: Es wird das Ereignis propagiert, das bezüglich *jeder* Spannungsschwelle zwischen 0 V und U_{DD} das zeitlich letzte Ereignis darstellt. Im Falle einander schneidender Kurven wird ein virtuelles Ereignis propagiert, das die rechtsseitig einhüllende aller dieser Kurven ist. Abbildung 4.5 zeigt die neue Spannungscurve $u_{ein,12}$, die $u_{ein,1}$ und $u_{ein,2}$ ersetzt. Die Ausgangsspannung $u_{aus,12}$ liegt rechts der Ausgangsspannungen $u_{aus,1}$ und $u_{aus,2}$. Obwohl dieser Ansatz naheliegend erscheint, ist dem Verfasser keine Veröffentlichung bekannt.

In [69][70] wird vorgeschlagen, alle Ereignisse mit einander schneidenden Spannungscurven zu propagieren. Begonnen wird mit dem spätesten Ereignis bezüglich einer beliebigen Spannungsschwelle zwischen 0 V und U_{DD} . Ereignisse, die komplett links von bereits propagierten Ereignissen liegen, werden nicht propagiert. Da bei diesem Ansatz die grundsätzliche Gefahr besteht, exponentiell viele Ereignisse propagieren zu müssen, schlagen die Autoren ebenfalls ein Verfahren vor, das ein einziges virtuelles Ereignis propagiert. Das virtuelle Ereignis wird

gebildet aus der spätesten 90 %-Ankunftszeit für steigende oder der spätesten 10 %-Ankunftszeit für fallende Ereignisse und der kleinsten Anstiegszeit aller Ereignisse. Dieser Ansatz ist ebenfalls korrekt unter der oben gemachten Annahme, beschränkt sich jedoch auf Ereignisse, deren Spannungskurven Strecken zwischen 0 V und U_{DD} sind.

5 Verzögerung bei kapazitiver Kopplung

5.1 Theoretische Betrachtungen

Zur Berücksichtigung der kapazitiven Kopplung in der statischen Timinganalyse ist es notwendig, den Einfluss der Kopplung auf die Verzögerung zu beschreiben. Es ist ein Verfahren anzugeben, das die Pfadverzögerung bei einer gegebenen Kopplungssituation maximiert.

Im Gegensatz zur bekannten Gatterverzögerung, bei der im Wesentlichen nur zwei Parameter, der Verlauf der Eingangsspannung und die Ausgangslast (siehe Abschnitt 3.3), eingehen, ist die Verzögerung bei vorhandener Kopplung von weiteren Größen abhängig:

- den Koppelkapazitäten,
- den Spannungsverläufen an den Eingängen der Aggressorgatter,
- den Ausgangslasten der Aggressorgatter.

Die Komplexität wird zusätzlich dadurch vergrößert, dass es nicht nur einen sondern eine Vielzahl von Aggressoren geben kann.

Im Folgenden soll angenommen werden, dass der Spannungsverlauf der Aggressoren frei wählbar, also nicht von der Art des Gatters abhängig, ist. Es soll untersucht werden, welche Art von Spannungskurve auf der Aggressorleitung zu einer maximalen Verzögerung führt.

Betrachtet wird zuerst das Opfergatter, das heißt, es wird zunächst ignoriert, welchen Einfluss der Spannungsverlauf auf der Opferleitung auf die folgenden CMOS-Gatter hat. Es stellt sich das Problem, eine Verzögerungsmetrik für die Verzögerung einer beliebigen, nicht notwendigerweise monotonen Kurvenform

festzulegen. Als Verzögerungsmaß wird das Erreichen einer Spannungsschwelle U_{th} angenommen. Der Wert dieser Schwelle wird zunächst nicht festgelegt.

Da das Ausgangsverhalten von MOS-Transistoren einen Stromquellen- wie auch einen Widerstandsbereich beinhaltet, werden diese Verhalten getrennt untersucht. Betrachtet wird zunächst die Stromquelleneigenschaft (Abbildung 5.1). Das Verhalten des Opfertgatters wird durch eine Konstantstromquelle, die bei Erreichen von U_{DD} abschaltet, beschrieben. Die Spannungskurve der Aggressorleitung ist eine Strecke zwischen U_{DD} und 0 V.

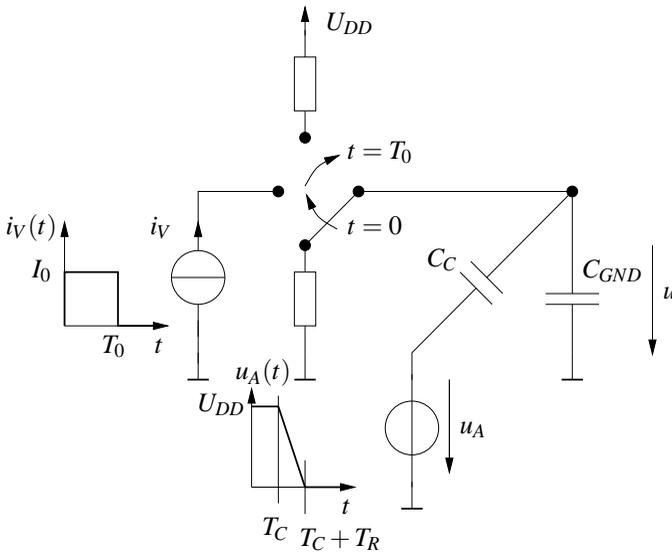


Abbildung 5.1: Modellierung des Opfertgatters durch eine Stromquelle

Für den Quellenstrom gilt:

$$i_V(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ I_0 & \text{für } 0 < t \leq T_0 \\ 0 & \text{für } T_0 < t \end{cases} \quad (5.1)$$

$$u(t \geq T_0) = U_{DD} \quad (5.2)$$

T_0 ist der Zeitpunkt, bei dem $u(t)$ die Versorgungsspannung U_{DD} erreicht. Der Zeitpunkt ist last- und kopplungsabhängig und wird später bestimmt. Das Aggressorverhalten wird durch eine stückweise lineare Spannungsquelle modelliert:

$$u_A(t) = \begin{cases} U_{DD} & \text{für } t \leq T_C \\ U_{DD} - \frac{t-T_C}{T_R} U_{DD} & \text{für } T_C < t \leq T_C + T_R \\ 0 & \text{für } T_C + T_R < t \end{cases} \quad (5.3)$$

mit

$$0 \leq T_C \leq (T_C + T_R) \leq T_{Max} \leq T_0$$

Die Spannung $u(t)$ der Opferleitung errechnet sich zu:

$$u(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ \frac{I_0 \cdot t}{C_{GND} + C_C} & \text{für } 0 < t \leq T_C \\ \frac{I_0 \cdot t}{C_{GND} + C_C} - \frac{U_{DD} \cdot C_C \cdot (t - T_C)}{(C_{GND} + C_C) \cdot T_R} & \text{für } T_C < t \leq T_C + T_R \\ \frac{I_0 \cdot t}{C_{GND} + C_C} - \frac{U_{DD} \cdot C_C}{C_{GND} + C_C} & \text{für } T_C + T_R < t \leq T_0 \\ U_{DD} & \text{für } T_0 < t \end{cases} \quad (5.4)$$

mit

$$T_0 = \frac{U_{DD} \cdot (C_{GND} + 2C_C)}{I_0} \quad (5.5)$$

Setzt man voraus, dass das Aggressorereignis vor Erreichen der Spannungsschwelle U_{th} abgeschlossen ist, errechnet sich die Verzögerung T_D zu:

$$T_D = \frac{1}{I_0} (U_{th} \cdot (C_{GND} + C_C) + U_{DD} \cdot C_C) \quad (5.6)$$

Die zeitliche Differenz der Spannungsverläufe mit und ohne Kopplung ist:

$$\Delta T_D = \frac{U_{DD} \cdot C_C}{I_0} \quad (5.7)$$

Eine Plausibilitätskontrolle zeigt, dass die Verzögerungsdifferenz für eine größere Treiberstärke (I_0 vergrößert) verringert sowie für größere C_C vergrößert wird. Erwartungsgemäß sind T_D und ΔT_D völlig unabhängig von T_C und T_R . Das bedeutet,

5 Verzögerung bei kapazitiver Kopplung

dass der Verlauf der Aggressorspannung keinen Einfluss auf die Verzögerung hat, wenn das Aggressorereignis vor Erreichen von U_{th} abgeschlossen ist. Es muss jedoch die zusätzliche Ladung $U_{DD} \cdot C_C$ vom Opfertaster aufgebracht werden. Die herkömmliche Modellierung der Kopplung durch eine geerdete Kapazität mit verdoppeltem Wert ist durch das Stromquellenmodell nur im Fall $U_{th} = U_{DD}$ gerechtfertigt. Mit $U_{th} = \frac{1}{2}U_{DD}$ ergibt sich, wie in Abschnitt 3.5.2 gezeigt, ein Miller-Faktor von drei. Das in [50] beschriebene Verfahren impliziert folglich eine Modellierung des Opfertasters durch eine Stromquelle.

Im Folgenden soll ein Widerstandsverhalten des Opfertasters betrachtet werden (Abbildung 5.2). Das Gatter wird durch eine ideale, gesteuerte Spannungsquelle mit einem konstanten Innenwiderstand R_I modelliert. Die Spannungsquelle führt zur Zeit $t = 0$ einen Sprung von 0 auf U_{DD} aus. Die Aggressorspannungsquelle hat das in Gleichung 5.3 dargelegte Verhalten.

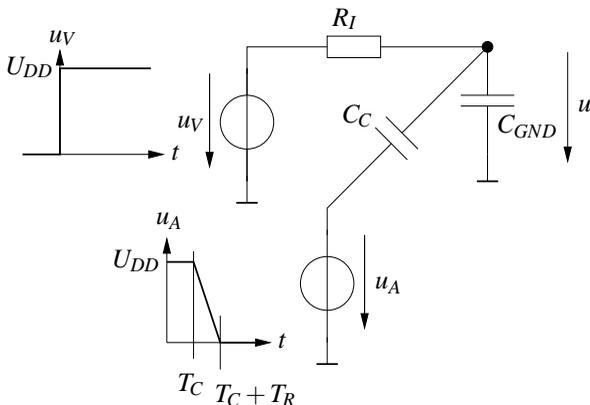


Abbildung 5.2: Modellierung des Opfertasters durch eine Spannungsquelle mit Innenwiderstand

Zur einfacheren Darstellung wird $R_I(C_{GND} + C_C)$ durch τ substituiert. Die Spannung $u(t)$ errechnet sich zu:

$$u(t) = \begin{cases} U_{DD} \left(1 - e^{-\frac{t}{\tau}}\right) & \text{für } 0 < t \leq T_C \\ U_{DD} \left(1 - e^{-\frac{t}{\tau}} - \frac{R_I C_C}{T_R} \left(1 - e^{-\frac{t-T_C}{\tau}}\right)\right) & \text{für } T_C < t \leq T_C + T_R \\ U_{DD} \left(1 - e^{-\frac{t-T_0}{\tau}}\right) & \text{für } T_C + T_R < t \end{cases} \quad (5.8)$$

Für $t > T_C + T_R$ erhält man die ursprüngliche Exponentialfunktion, die um T_0 verschoben ist. Für T_0 gilt:

$$T_0 = \tau \ln \left(1 + C_C R_I \cdot e^{\frac{T_C}{\tau}} \frac{e^{\frac{T_R}{\tau}} - 1}{T_R}\right) \quad (5.9)$$

Setzt man, wie bereits für das Stromquellenmodell geschehen, voraus, dass das Aggressorereignis vor Erreichen der Spannungsschwelle beendet ist, so ist T_0 bereits die zeitliche Differenz zwischen dem Spannungsverlauf ohne und dem Spannungsverlauf mit Kopplung, also $\Delta T_D = T_0$. Es stellt sich die Frage, unter welchen Umständen T_0 maximal wird. Die Ableitung nach T_C

$$\frac{\partial T_0}{\partial T_C} = \frac{C_C R_I \left(e^{\frac{T_R}{\tau}} - 1\right) e^{\frac{T_C}{\tau}}}{T_R + C_C R_I \cdot e^{\frac{T_C}{\tau}} \left(e^{\frac{T_R}{\tau}} - 1\right)} \quad (5.10)$$

hat keine Nullstellen in T_C und somit keine Minima. Es müssen daher die Randwerte betrachtet werden. Da es sich um ein zweidimensionales Problem handelt, wird zunächst T_R betrachtet. Die Ableitung von T_0 nach T_R

$$\frac{\partial T_0}{\partial T_R} = \frac{C_C R_I \cdot e^{\frac{T_C}{\tau}} \left(\frac{T_R}{\tau} e^{\frac{T_R}{\tau}} - e^{\frac{T_R}{\tau}} + 1\right)}{T_R \left(T_R + C_C R_I \cdot e^{\frac{T_C}{\tau}} \left(e^{\frac{T_R}{\tau}} - 1\right)\right)} \quad (5.11)$$

hat die Zählernullstelle $T_R = 0$, die ebenfalls eine Nennernullstelle ist. Die zweifache Anwendung der de L'Hospital-Regel zeigt, dass $T_R = 0$ tatsächlich keine Nullstelle von $\frac{\partial T_0}{\partial T_R}$ ist. Es ist also auch im Fall T_R eine Randwertbetrachtung erforderlich.

Die folgenden Gleichungen geben die Randwerte an, wobei bei $T_R \rightarrow 0$ die de L'Hospital-Regel angewendet wird.

$$T_0(T_R \rightarrow 0, T_C \rightarrow 0) = \tau \ln \left(1 + \frac{C_C R_I}{\tau} \right) \quad (5.12)$$

$$T_0(T_R \rightarrow 0, T_C \rightarrow T_{Max}) = \tau \ln \left(1 + \frac{C_C R_I}{\tau} e^{\frac{T_{max}}{\tau}} \right) \quad (5.13)$$

$$T_0(T_R \rightarrow T_{Max}, T_C \rightarrow 0) = \tau \ln \left(1 + \frac{C_C R_I}{T_{Max}} \left(e^{\frac{T_{max}}{\tau}} - 1 \right) \right) \quad (5.14)$$

Der Vergleich der Gleichungen 5.13 und 5.12 ergibt:

$$T_0(T_R \rightarrow 0, T_C \rightarrow T_{Max}) > T_0(T_R \rightarrow 0, T_C \rightarrow 0) \quad (5.15)$$

$$\tau \ln \left(1 + \frac{C_C R_I}{\tau} e^{\frac{T_{max}}{\tau}} \right) > \tau \ln \left(1 + \frac{C_C R_I}{\tau} \right) \quad (5.16)$$

$$e^{\frac{T_{Max}}{\tau}} > 1 \quad \text{q.e.d.} \quad (5.17)$$

Der Vergleich der Gleichungen 5.13 und 5.14 ergibt:

$$T_0(T_R \rightarrow 0, T_C \rightarrow T_{Max}) > T_0(T_R \rightarrow T_{Max}, T_C \rightarrow 0) \quad (5.18)$$

$$\tau \ln \left(1 + \frac{C_C R_I}{\tau} e^{\frac{T_{max}}{\tau}} \right) > \tau \ln \left(1 + \frac{C_C R_I}{T_{Max}} \left(e^{\frac{T_{max}}{\tau}} - 1 \right) \right) \quad (5.19)$$

$$\frac{1}{\tau} e^{\frac{T_{max}}{\tau}} > \frac{1}{T_{Max}} \left(e^{\frac{T_{max}}{\tau}} - 1 \right) \quad (5.20)$$

$$\frac{T_{max}}{\tau} > \frac{e^{\frac{T_{Max}}{\tau}} - 1}{e^{\frac{T_{Max}}{\tau}}} \quad (5.21)$$

Mit der Substitution

$$x := \frac{T_{Max}}{\tau} \quad (5.22)$$

folgt

$$x > \frac{e^x - 1}{e^x} \quad (5.23)$$

$$e^x(x-1) > -1 \quad (5.24)$$

$$e^x(x-1)+1 > 0 \quad (5.25)$$

Es ist erkennbar, dass die Ungleichung im Bereich $x > 1$ erfüllt ist. Für den Bereich $0 < x \leq 1$ wird die Funktion

$$f(x) := e^x(x-1) + 1 \quad (5.26)$$

gebildet, deren Ableitung nach x

$$\frac{df(x)}{dx} = xe^x \quad (5.27)$$

eine Nullstelle bei $x = 0$ hat. Der Funktionswert ergibt sich zu $f(x = 0) = 0$. Aus $x = 0$ folgt $T_{Max} = 0$ und damit $T_C = T_{Max} = 0$, womit die Gleichungen 5.13 und 5.14 denselben Fall beschreiben. Da $f(x = 1) = 1$ ist, folgt $f(x) > 0$ für $0 < x \leq 1$.

Eine Veranschaulichung der Überlegungen zeigt Abbildung 5.3. Die mit » T_C verkleinert« bezeichnete Spannungskurve liegt nach der Kopplung links der Kurve » $T_R = 0$ «, die Kurve » $T_R > 0$ « rechts. Im letzten Fall muss jedoch ein höherer Schwellwert zugelassen werden.

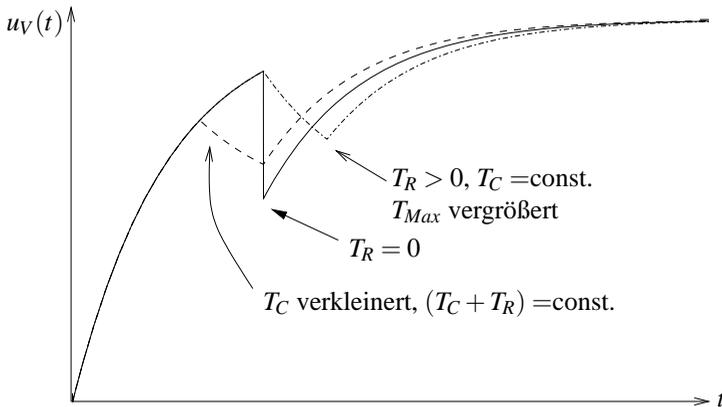


Abbildung 5.3: Variation von T_C und T_R

Hier erfolgt eine Zusammenfassung der bisherigen Überlegungen: Unter der Annahme, dass CMOS-Gatter ein Widerstandsverhalten haben, ist die Verzögerung durch Kopplung maximal, wenn das Kopplungsereignis zum spätestmöglichen Zeitpunkt erfolgt und die Aggressorleitung einen Spannungssprung der Höhe

U_{DD} und der Dauer 0 s ausführt. Dieses Ergebnis unterliegt der Annahme, dass als Verzögerungsmetrik das letztmalige Erreichen einer Spannungsschwelle der Opferleitung maßgeblich ist. Unter den genannten Bedingungen kann die Spannungskurve auf der Aggressorleitung nicht-monoton sein.

Es ist die Frage zu klären, welche Spannungsschwelle sinnvollerweise gewählt wird. Folgende Möglichkeiten kommen in Betracht:

- Die jeweiligen Transistor-Schwellspannungen $U_{DD} + U_{th,p}$ und $U_{th,n}$. Die Definition ist vom gewählten Transistormodell abhängig. Für die Betrachtungen wird als Transistor-Schwellspannung diejenige Gate-Source-Spannung angesehen, bei der durch den Transistor ein Tausendstel des maximalen Drain-Stromes fließt. Der Vorteil dieses Schwellwertes ist, dass bereits die folgende Stufe keine durch Kopplung hervorgerufenen Monotoniewechsel aufweise. Nachteil ist die möglich Unterschätzung der Verzögerung.
- Die klassische CMOS-Spannungsschwelle $\frac{1}{2}U_{DD}$. Für dieses Modell spricht der höhere Spannungswert. Da die vom Betrag höchsten differentiellen Spannungsverstärkungen erreicht werden, kann ein durch Kopplung hervorgerufener Monotoniewechsel durch mehrere Stufen propagiert werden.
- Eine Spannungsschwelle zwischen diesen beiden Werten.

Abbildung 5.4 zeigt eine Beispielschaltung, anhand derer verschiedene Schwellwerte veranschaulicht werden sollen. Die Abbildungen 5.5 und 5.6 zeigen die Spannungsverläufe für u_V und $u_{aus,1}$ bei verschiedenen Schwellspannungen. In der 0,5 μm -Technologie ist $U_{DD} = 3,3 \text{ V}$, die Transistorschwellspannungen betragen:

$$U_{th,n} = 0,65 \text{ V} \quad (5.28)$$

$$U_{th,p} = -0,86 \text{ V} \quad (5.29)$$

$$U_{DD} + U_{th,p} = 2,44 \text{ V} \quad (5.30)$$

Eine Schwellspannung von $\frac{1}{2}U_{DD}$ ist ungeeignet, da die Verzögerung unterschätzt werden kann. In Abbildung 5.6 ist sichtbar, dass die anderen Spannungskurven im relevanten Bereich zwischen den Transistorschwellspannungen rechts der Kurve

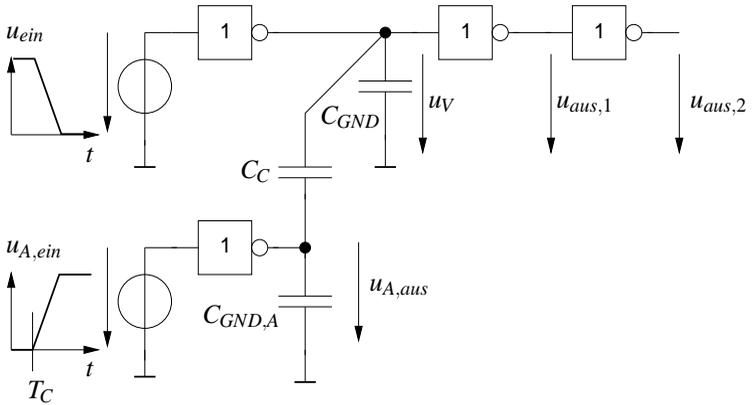


Abbildung 5.4: Schaltung zur Herleitung von Spannungsschwellen bei Kopplung

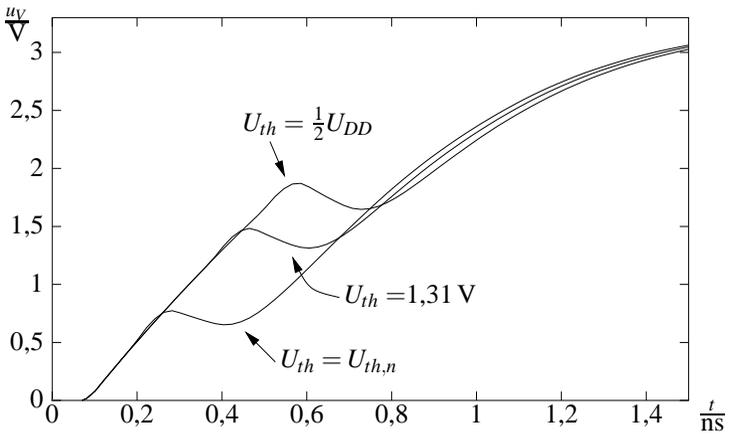


Abbildung 5.5: u_V für verschiedene Spannungsschwellen

für $U_{th} = \frac{1}{2}U_{DD}$ liegen. Das späteste Ereignis wird in diesem Beispiel nach Unterschreiten von $u_{aus,1} = U_{DD} + U_{th,p}$ für $U_{th} = 1,31$ V erreicht.

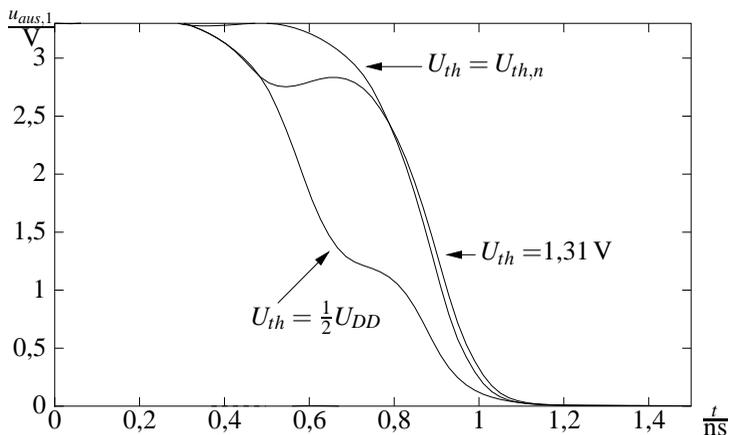


Abbildung 5.6: $u_{aus,1}$ für verschiedene Spannungsschwellen

Es ist sinnvoll, Schwellwerte zwischen den Transistorschwellspannungen und $\frac{1}{2}U_{DD}$ zu wählen. Es werden die Werte gewählt, bei denen die Kleinsignal-Gleichspannungsverstärkung eines Inverters eins ist. Auf diese Weise ist sichergestellt, dass sich der Betrag der Differenz der durch Kopplung hervorgerufenen lokalen Spannungsminima und -maxima verringert. Die in der $0,5\ \mu\text{m}$ -Technologie verwendeten Spannungsschwellen sind $1,31\ \text{V}$ für steigende und $1,83\ \text{V}$ fallende Ereignisse. Nebenbei bemerkt ist der geringe Abstand der Spannungsschwellen von nur $0,52\ \text{V}$ die Ursache der starken Angleichung verschiedener Spannungsverläufe nach nur wenigen Gattern. Die resultierenden Ausgangsspannungen sind $3,01\ \text{V}$ und $0,28\ \text{V}$. Sie liegen damit unterhalb beziehungsweise oberhalb der jeweiligen Transistor-Schwellspannungen $U_{th,n}$ und $U_{DD} + U_{th,p}$. Die Spannungsschwellen für Kopplung in der $130\ \text{nm}$ -Technologie betragen $0,52\ \text{V}$ für steigende und $0,80\ \text{V}$ für fallende Ereignisse bei einer Versorgungsspannung von $1,4\ \text{V}$.

5.2 Verzögerungsmodelle

Als Ableitung der in Abschnitt 5.1 durchgeführten Überlegungen werden im Folgenden zwei Verzögerungsmodelle vorgeschlagen. Das erste modelliert die maximale Verzögerung ohne, das zweite unter Berücksichtigung der Aggressorsituation.

Für die Integration eines Verzögerungsmodells in eine statische Timinganalyse ist zu beachten, dass die Verzögerung nach Möglichkeit unmittelbar berechnet und nicht durch ein iteratives Suchverfahren ermittelt wird. Insbesondere sollte die Schlüsseleigenschaft der statischen Timinganalyse, die lineare Komplexität (siehe Abschnitt 2.5.2), nicht verloren gehen.

Das erste Modell ignoriert die Aggressorsituation. Das Modell hat diese Eigenschaften:

- Als Aggressorverhalten wird eine Sprungfunktion des Betrages U_{DD} angenommen.
- Der Sprung wird zeitlich so ausgeführt, dass nach seiner Beendigung die Spannung der Opferleitung $U_{th,u}$ für steigende Ereignisse beziehungsweise $U_{th,o}$ für fallende Ereignisse beträgt.
- Eventuell auftretende Monotoniewechsel werden ignoriert.

Die Motivation für die erste Eigenschaft ergibt sich unmittelbar aus der theoretischen Herleitung in Abschnitt 5.1. Die Motivation für Punkt zwei ergibt sich aus der Forderung, den Zeitpunkt des Erreichens der Schwellspannung möglichst weit hinauszuschieben. Punkt drei ist eine Vereinfachung, die das Ziel hat, bereits die Opferspannung monoton zu halten. Dieses Vorgehen könnte die Verzögerung vergrößern und ist somit für die Berechnung der größten Verzögerung korrekt.

Abbildung 5.7 stellt den Spannungsverlauf des ersten Verzögerungsmodells für Kopplung schematisch dar.

Dieses Modell ist für die statische Timinganalyse gut geeignet. Die tatsächliche Spannungskurve der Aggressorleitungen muss nicht berechnet werden. Der durch Kopplung ausgelöste Spannungshub kann zu

$$\Delta U = U_{DD} \frac{C_C}{C_{GND} + C_C} \quad (5.31)$$

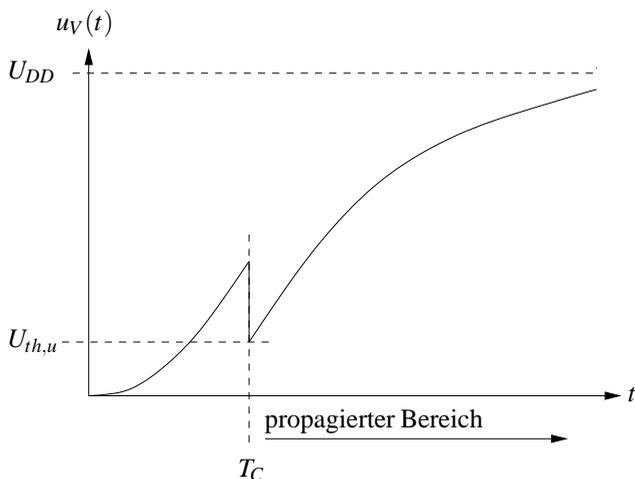


Abbildung 5.7: Schematische Darstellung der Opferspannungskurve für das erste Verzögerungsmodell

berechnet werden. Daraus folgt, dass sich auch T_C einfach ermitteln lässt. Die Berechnung der Opferspannungskurve erfolgt ohne Kopplung bis

$$u_V(T_C) = U_{th,u} + \Delta U \quad \text{beziehungsweise} \quad (5.32)$$

$$u_V(T_C) = U_{th,o} - \Delta U \quad (5.33)$$

gilt. Die dann folgende Berechnung erfolgt ab $t = T_C$ mit $u_V = U_{th,u}$ oder $u_V = U_{th,o}$. Nur dieser Teil der Kurve wird zur Verzögerungsberechnung folgender Stufen verwendet. Der Nachteil dieses Modells liegt im Ignorieren der Aggressorkurvenformen. Dies kann zu einer Überschätzung der Verzögerung führen.

Um diesem Nachteil zu begegnen, wird ein zweites Verzögerungsmodell vorgeschlagen. Dieses lehnt sich an das in [49] und [51] präsentierte Vorgehen an. Die Berechnung des Modells erfolgt in zwei Schritten: Im ersten Schritt werden die Spannungskurven aller Knoten, die in diesem Fall als Aggressoren betrachtet werden, berechnet. Ziel hierbei ist einerseits, Spannungskurven zu erhalten, die zu einer maximalen Verzögerung des Opferpfades führen, aber andererseits möglichst realistisch sind. Die Berechnung dieser Verläufe erfolgt mit einem Sprung

als Eingangsspannung und der Massekapazität des Netzes als Ausgangslast. Die Koppelkapazitäten werden ignoriert, wodurch eine Mitkopplung abgebildet werden soll.

Nachdem alle Aggressorspannungsverläufe berechnet sind, erfolgt mit der eigentlichen Verzögerungsberechnung der zweite Schritt. Die Problemstellung besteht auch hier darin, zu ermitteln, zu welcher Zeit die Kopplung die größte Auswirkung auf die Verzögerung hat. Zur Lösung wird der Überlagerungssatz verwendet. Damit dies möglich ist, ist es erforderlich, den Innenwiderstand des Opfergatters in den jeweiligen Arbeitspunkten für steigende und fallende Ereignisse von u_V zu bestimmen. Da das Ziel ist, den Zeitpunkt des Erreichens der Schwellspannung zu verschieben, ergeben sich die Arbeitspunkte $u_V \approx U_{th,u}$ für steigende und $u_V \approx U_{th,o}$ für fallende Ereignisse.

Abbildung 5.8 zeigt die Charakterisierung für einen Inverter. ΔU ist hier eine kleine Spannungsdifferenz und nicht der durch Kopplung ausgelöste Spannungshub. Der Innenwiderstand errechnet sich zu:

$$R_I = \frac{2\Delta U}{i_{aus,2} - i_{aus,1}} \tag{5.34}$$

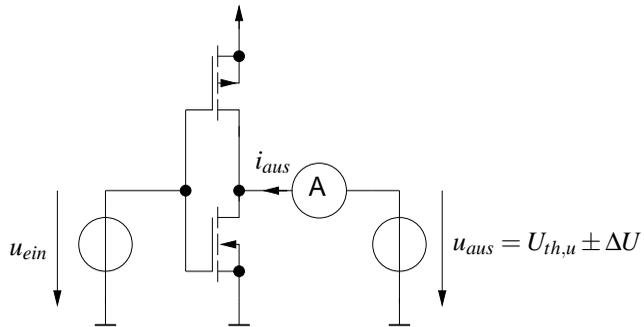


Abbildung 5.8: Charakterisierung des Innenwiderstandes

Die Eingangsspannung u_{ein} beeinflusst den zu bestimmenden Innenwiderstand. Je nach Aggressorereignis kann sich das Opfergatter im Stromquellen- oder im Widerstandsbereich befinden. Es wäre möglich, diesen Widerstand für jede Kopplung neu zu bestimmen. Es zeigt sich jedoch, dass relativ gute Abschätzungen

auch für einen konstanten Widerstandswert erreicht werden, wenn dieser für den Widerstandsbereich des Gatters bestimmt wurde. Es wird $u_{ein} = U_{DD}$ für fallende und $u_{ein} = 0\text{ V}$ für steigende Ausgangsereignisse gewählt.

Der nächste Schritt der Verzögerungsberechnung besteht in der Ermittlung der durch die Aggressoren hervorgerufenen Spannungskurven auf der Opferleitung. Das Verhalten der Aggressorleitung wird durch eine Spannungsquelle, die die vorher errechnete Kurve einprägt, modelliert. Abbildung 5.9 zeigt die Vorgehensweise für einen Aggressor bei steigendem Opferereignis.

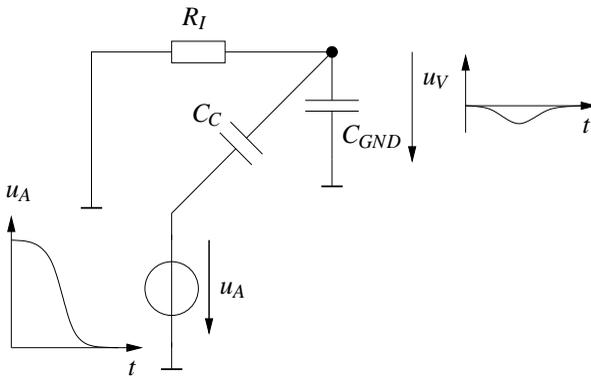


Abbildung 5.9: Berechnung der durch Kopplung verursachten Spannungskurven auf der Opferleitung

Danach sind diese Kurven derart zu der Opferspannungskurve ohne Kopplung zu überlagern, dass das Erreichen von $u_V = U_{th,u}$ beziehungsweise $u_V = U_{th,o}$ möglichst weit verschoben wird. Die Berechnung ist analytisch nicht durchzuführen und numerisch sehr aufwändig. In [51] wird eine sehr einfache Lösung aufgezeigt: Gibt es mehrere Aggressoren, werden die durch Kopplung hervorgerufenen Opferkurven relativ zueinander verschoben, so dass ihre Spitzenwerte $U_{P,i}$ gleichzeitig erreicht werden. Die Addition der Kurven ergibt einen resultierenden Spitzenwert $U_P = \sum_i U_{P,i}$. Für den Zeitpunkt T_C , an dem die Kopplung bezüglich der Verzögerung am wirksamsten ist, gilt $u_V(T_C) = U_{th,u} + U_P$.

Im letzten Schritt der Verzögerungsberechnung werden die Aggressorcurven um die berechnete Zeit verschoben und eine Simulation mit Verzögerungsmodell auf Transistorebene wie in Kapitel 4 beschrieben durchgeführt.

Der beschriebene Ansatz ist an [51] angelehnt. Die Unterschiede sind im Wesentlichen die folgenden:

- In [51] wird die Linearisierung nicht nur für den Kopplungsfall, sondern auch für die Verzögerungsberechnung selbst durchgeführt. Im hier vorgeschlagenen Modell erfolgt die Verzögerungsberechnung mit einem nichtlinearen Tabellenmodell.
- In [51] wird nur eine Stufe betrachtet und als Verzögerungsmaß das Erreichen des 50%-Versorgungsspannungswertes verwendet. Wie gezeigt kann dies zu einer Unterschätzung der Verzögerung führen. Das vorgeschlagene Modell zielt auf eine möglichst große Pfadverzögerung und nimmt daher die oben behandelten Spannungsschwellen an.
- Das vorgeschlagene Modell betrachtet alle Aggressorkapazitäten als an Masse angeschlossen. In [51] werden die Aggressorgatter auch für den Ruhe-Arbeitspunkt linearisiert. Dieser Innenwiderstand geht in ein reduziertes π -Modell (siehe Abschnitt 3.4), das wiederum unter Berücksichtigung des Opfergatters zu einer effektiven Kapazität reduziert wird [39]. Dieses Vorgehen ist prinzipiell genauer, erhöht allerdings den Aufwand. Eine Vernachlässigung des Ruhe-Innenwiderstandes der Aggressorgatter führt zu größeren Verzögerungen und ist somit nicht falsch. Die Bedeutung des Ruhe-Innenwiderstandes wird als gering erachtet.

5.3 Empirische Validierung

5.3.1 Erstes Verzögerungsmodell

Die in den vorangegangenen Abschnitten durchgeführten Untersuchungen sollen durch Analogsimulationen validiert werden. Abbildung 5.10 zeigt die verwendete Schaltung. Die Aggressorcurvenform wird durch eine Rampe mit dem Spannungshub U_{DD} modelliert. Während wiederholter Simulationen werden Aggressorstartzeit T_C und -anstiegszeit T_R variiert. Das zu maximierende Verzöge-

5 Verzögerung bei kapazitiver Kopplung

rungsmaß T_D ist das Erreichen der Schwelle $u_2(T_D) = \frac{1}{2}U_{DD}$. Die Anstiegszeit der Eingangsspannung u_{ein} sowie die Kapazitäten C_{GND} und C_C sind für jeweils eine Versuchsreihe konstant.

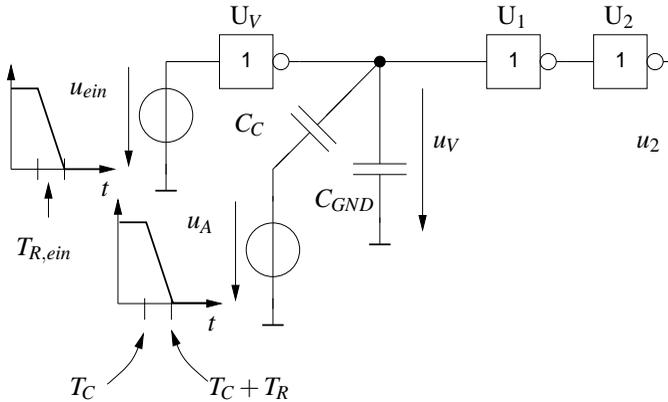


Abbildung 5.10: Schaltung zur empirischen Bestimmung maximaler Verzögerung nach dem ersten Verzögerungsmodell

Die Versuchsreihe wird für einen Inverter sowie für ein NAND- und NOR-Gatter für beide Eingänge durchgeführt. Die NAND- und NOR-Gatter sollen die Reihenschaltung von Transistoren erfassen. Andere Gatter unterscheiden sich prinzipiell nicht.

Die Eingangs-Anstiegszeit $T_{R,in}$, die Last-Massekapazität C_{GND} und die Koppelkapazität C_C werden in mehreren Stufen variiert. Die obere Grenze der Gesamtlast ist so gewählt, dass bei ihrer Verwendung eine Anstiegszeit von zirka 2 ns erreicht wird. Die Koppelkapazität wird nach oben derart begrenzt, dass sie höchstens 50 % der Gesamtlast ausmacht. Diese Annahme ist vernünftig, da bei größeren Koppelanteilen funktionale Fehler auftreten können. Das Verzögerungsmodell ist in »Spice3« implementiert. Es werden zwei Analogsimulationen durchgeführt: Bei der ersten wird Kopplung ignoriert und der Zeitpunkt T_C des Erreichens der für die jeweilige Kopplung notwendigen Schwellwerte von u_V festgehalten. Die zweite Simulation ist weitgehend identisch mit der ersten mit der Erweiterung, dass u_V bis zur Zeit $t = T_C$ auf U_{th} geklemmt wird, und zur Zeit $t = T_C$ die

Klemmung beendet wird. Die Tabellen 5.1 und 5.2 listen die verwendeten Werte auf.

Gatter/Eingangspin	INV/1, NAND2/1, NAND2/2, NOR2/1, NOR2/2
$T_{R, \text{ein}}/\text{ns}$	0,1 0,2 0,5 1,0 1,5 2,0
C_{GND}/fF	1 2 3 5 7 10 20 30 40 50
C_C/fF mit $C_C \leq C_{GND}$	1 2 3 5 7 10 20 30 40 50
Anzahl der Kombinationen	2220

Tabelle 5.1: Variierung der Parameter zur Validierung des ersten Verzögerungsmodells, 0,5 μm -Technologie

Gatter/Eingangspin	INV/1, NAND2/1, NAND2/2, NOR2/1, NOR2/2
$T_{R, \text{ein}}/\text{ns}$	0,05 0,1 0,2 0,5 1,0 1,5 2,0
C_{GND}/fF	2 5 10 20 30 50 70 100 150 200
C_C/fF mit $C_C \leq C_{GND}$	2 5 10 20 30 50 70 100 150 200
Anzahl der Kombinationen	2340

Tabelle 5.2: Variierung der Parameter zur Validierung des ersten Verzögerungsmodells, 130 nm-Technologie

Die Tabellen 5.3 und 5.4 zeigen die Ergebnisse des ersten Verzögerungsmodells im Vergleich zur empirisch ermittelten Maximalverzögerung.

Minimaler Fehler	-0,32 %
Maximaler Fehler	3,34 %
Mittelwert des Fehlers	0,29 %
Empirische Standardabweichung des Fehlers	0,60 %

Tabelle 5.3: Validierung des ersten Verzögerungsmodells, 0,5 μm -Technologie

Minimaler Fehler	-0,20 %
Maximaler Fehler	4,32 %
Mittelwert des Fehlers	0,47 %
Empirische Standardabweichung des Fehlers	0,72 %

Tabelle 5.4: Validierung des ersten Verzögerungsmodells, 130 nm-Technologie

5.3.2 Zweites Verzögerungsmodell

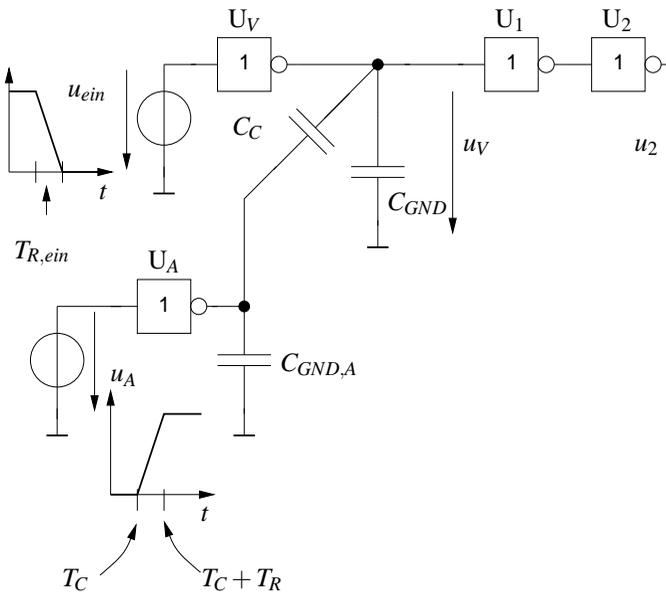


Abbildung 5.11: Schaltung zur empirischen Bestimmung maximaler Verzögerung nach dem zweiten Verzögerungsmodell

Da das zweite Modell auch die Situation der Aggressorleitung berücksichtigt, werden als weitere Parameter der Typ des Aggressorgatters und die Anstiegszeit an dessen Eingang variiert. Abbildung 5.11 zeigt die verwendete Schaltung, die

Tabellen 5.5 und 5.6 die Variierung der Parameter für die Validierung des zweiten Verzögerungsmodells.

Opfergatter/Eingangspin	INV/1, NAND2/1, NAND2/2, NOR2/1, NOR2/2
$T_{R, \text{ein}}/\text{ns}$	0,1 0,2 0,5 1,0 2,0
Aggressorgatter	INV in vier Treiberstärken
$T_{R, A}/\text{ns}$	0,1 0,2 0,5 1,0 2,0
C_{GND}/fF	1, 2, 5, 10, 20, 50
C_C/fF	1, 2, 5, 10, 20
mit $C_C \leq \frac{2}{5}C_{GND}$	
Anzahl der Parameterkombinationen	48550

Tabelle 5.5: Variierung der Parameter des zweiten Verzögerungsmodells, 0,5 μm -Technologie

Opfergatter/Eingangspin	INV/1, NAND2/1, NAND2/2, NOR2/1, NOR2/2
$T_{R, \text{ein}}/\text{ns}$	0,05 0,1 0,2 0,5 1,0 2,0
Aggressorgatter	INV in vier Treiberstärken
$T_{R, A}/\text{ns}$	0,05 0,1 0,2 0,5 1,0 2,0
C_{GND}/fF	5, 10, 20, 50, 100, 200, 500
C_C/fF	5, 10, 20, 50, 100, 200
mit $C_C \leq \frac{2}{5}C_{GND}$	
Anzahl der Parameterkombinationen	71750

Tabelle 5.6: Variierung der Parameter des zweiten Verzögerungsmodells, 130 nm-Technologie

Die Tabellen 5.7 und 5.8 zeigen die Ergebnisse der Validierung des zweiten Verzögerungsmodells.

Minimaler Fehler	-1,38 %
Maximaler Fehler	12,90 %
Mittelwert des Fehlers	0,39 %
Empirische Standardabweichung des Fehlers	0,97 %

Tabelle 5.7: Validierung des zweiten Verzögerungsmodells, 0,5 μm -Technologie

Minimaler Fehler	-2,34 %
Maximaler Fehler	17,99 %
Mittelwert des Fehlers	0,77 %
Empirische Standardabweichung des Fehlers	1,75 %

Tabelle 5.8: Validierung des zweiten Verzögerungsmodells, 130 nm-Technologie

5.3.3 Bewertung

Die Simulationsergebnisse des ersten Verzögerungsmodells werden wie folgt bewertet:

- Die maximale Verzögerung wird in jedem Fall für $T_R = \min$ erreicht. Dies ist eine gute Bestätigung der theoretischen Betrachtungen aus Abschnitt 5.1, des verwendeten Verzögerungsmodells und der gewählten Spannungsschwellen. T_R ist auf 1 ps begrenzt.
- Die mittels einer gerichteten Suche durch Variieren der Parameter gefundene Maximalverzögerung ist nur in Einzelfällen größer als die durch das Modell vorhergesagte Verzögerung. Sie ist vom Absolutbetrag nicht größer als 1 ps.
- Die größte Abweichung (26 ps oder 4,32 % in der 130 nm-Technologie) wird für die maximale Koppelkapazität C_C erreicht. Hierfür kommen zwei Erklärungen in Betracht: Zum einen ist die Anstiegszeit der Aggressorleitung 1 ps, das Modell nimmt null an, zum anderen egalisiert das Modell Monotoniewechsel.
- Der mittlere relative Fehler ist mit 0,47 % gering.

Zusammenfassend kann festgestellt werden, dass das erste Modell mit hoher Sicherheit eine obere Grenze für die Verzögerung voraussagt. Im Unterschied zu

beispielsweise [50] und [51] zieht dieses Modell das Verhalten der folgenden Stufen in Betracht.

Die Simulationsergebnisse des zweiten Verzögerungsmodells werden wie folgt bewertet:

- Die maximale Verzögerung wird um bis zu 2,34 % oder absolut 11 ps in der 130 nm-Technologie unterschätzt. Dieser Fall tritt bei dem stärksten Aggressorgatter auf. Der Fehler kann akzeptiert werden.
- Der maximale Fehler tritt bei dem schwächsten Aggressorgatter und der maximalen Koppelkapazität auf. Dies ist darauf zurück zu führen, dass der Aggressorspannungsverlauf im Modell eingeprengt wird, während in der Validierung der Aggressor selbst durch das Opfer belastet wird. Die Untersuchungen wurden auf $C_C \leq \frac{2}{5}C_{GND}$ beschränkt, weil größere Verhältnisse im Allgemeinen nicht erreicht werden. Die Überschätzung würde weiter zunehmen.
- Der mittlere Fehler und die empirische Standardabweichung sind gering.

Auch das zweite Modell ermöglicht eine gute Vorhersage der Verzögerung.

5.4 Berücksichtigung kapazitiver Kopplung in der statischen Timinganalyse

In Abschnitt 3.6 wurde das Konzept der Aktivitätsfenster bereits dargelegt. Dazu werden im Folgenden zwei Algorithmen beschrieben. Sie erfordern, dass Verzögerungsberechnung und statische Timinganalyse integriert sind.

Der erste Algorithmus ist ein Einschrittverfahren. Dieses erhöht den Rechenaufwand der herkömmlichen statischen Timinganalyse nicht. Im Folgenden wird die Berechnung eines Ereignisses beschrieben: Im ersten Teil des Verfahrens wird eine Spannungskurve unter der Annahme, dass keine Kopplung auftritt, berechnet. Diese stellt die im Late-Mode günstigste Kurvenform dar. In Abbildung 5.12 ist die Situation für ein steigendes Ereignis auf der Opferleitung dargestellt. Der Zeitpunkt T_V , zu dem die Spannung der Opferleitung dauerhaft oberhalb von U_{th} liegt, wird als Anfangs-Zeitpunkt des Ereignisses gewählt. Als Schwellwerte werden die Transistorschwellspannungen $U_{th,n}$ und $U_{DD} + U_{th,p}$ verwendet.

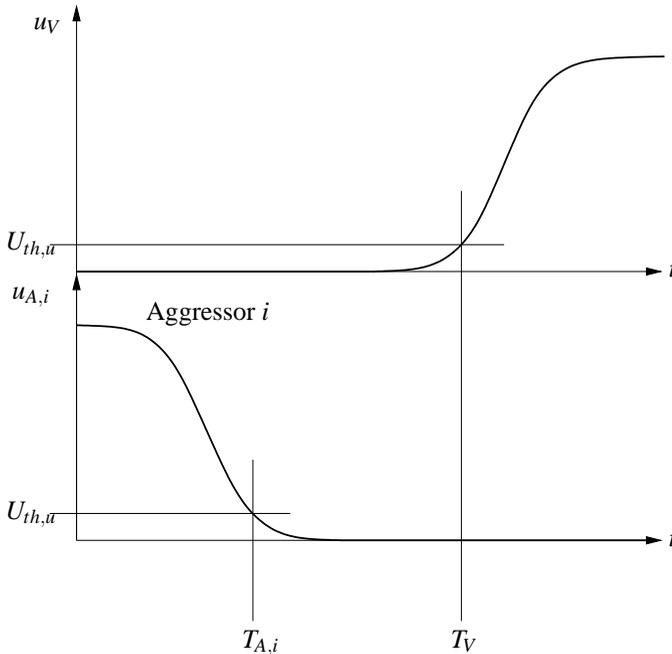


Abbildung 5.12: Zur Beurteilung von Kopplung

Im zweiten Schritt werden die End-Zeitpunkte aller Aggressorereignisse ermittelt. Da hier die fallenden Ereignisse von Interesse sind, wird als Schwellwert ebenfalls U_{th} verwendet. Durch Vergleich der Zeitpunkte $T_{A,i}$ mit T_V , wird entschieden, ob Gegenkopplung zeitlich möglich ist. Es kann den Fall geben, dass die Nachbarleitung noch nicht berechnet wurde. Hier ist als ungünstigster Fall das Vorhandensein von Kopplung anzunehmen. Als letzter Schritt wird die Spannungskurve der Opferleitung derart berechnet, dass die im zweiten Schritt ermittelte mögliche Kopplung berücksichtigt wird.

Dieses Verfahren stellt einen Fortschritt gegenüber der Annahme, dass in jedem Fall Kopplung vorliegt, dar. Der konstante Term der Rechenzeit ist erhöht, da die Spannungskurve der Opferleitung zweifach berechnet werden muss. Nachteil des Verfahrens ist, dass es benachbarte Leitungen, die noch nicht berechnet wurden,

5.4 Berücksichtigung kapazitiver Kopplung in der statischen Timinganalyse

geben kann. Für diesen Fall muss man das Vorhandensein von Kopplung annehmen.

Der Nachteil kann durch ein Mehrschrittverfahren überwunden werden, bei dem die vollständige statische Timinganalyse wiederholt durchgeführt wird. Nach der initialen Iteration sind die letzten Ereignisse aller Leitungen bekannt, so dass keine Annahme über das Auftreten von Kopplung getroffen werden muss. Die vollständige Durchführung der statischen Timinganalyse kann wiederholt werden, bis keine Verbesserung mehr erreicht wird.

6 Ergebnisse

6.1 Voraussetzungen

Als Schaltungsbeispiele zur Demonstration der Qualität der implementierten statischen Timinganalyse und der Effekte kapazitiver Kopplung werden die Schaltungen aus der ISCAS-89-Benchmark-Sammlung [71] gewählt. Es handelt sich im Gegensatz zu den ISCAS-85-Designs um sequentielle Schaltungen. Es werden die drei größten Schaltungen, »s35932«, »s38417« und »s38584«, verwendet.

Diese Schaltungen werden auf die Gatter einer 0,5 μm - und 130 nm-Technologie übertragen. Diese Netzlisten werden dem physikalischen Entwurf unterzogen, das heißt, platziert, gegebenenfalls optimiert, mit einem Taktverteilungsbaum versehen, verdrahtet und extrahiert. Tabelle 6.1 vergleicht Prozessdaten und Entwurfsschritte der Technologien. Zwei wesentliche Unterschiede zwischen den Entwurfsverfahren in den beiden Technologien sind hervorzuheben: Erstens ist die Anzahl der Eingangspins pro Netz in der 130 nm-Technologie im physikalischen Entwurf durch das Einfügen von Inverterbäumen auf 15 begrenzt. Der Entwurfsprozess in der 0,5 μm -Technologie sieht dieses nicht vor. Zweitens wurden die Netze in der 130 nm-Technologie mit diskretisierten Leitungswiderständen extrahiert. Die Extraktion in der 0,5 μm -Technologie liefert eine diskrete Kapazität pro Netz ohne Leitungswiderstände.

Das Ergebnis der statischen Timinganalyse ist, neben dem minimalen Slack, der Pfad, der diesen Slack hervorruft. Dieser Pfad kann in der extrahierten Spice-Netzliste wiedererkannt und einschließlich der Parasiten isoliert werden. Der Pfad muss mit einer Stimulations-Spannungsquelle am Startpunkt und den Transistornetzlisten der Gatter versehen werden, um ihn transient mit dem Analogsimulator »Spice3« simulieren zu können. Das verwendete Transistormodell in Spice3 ist »BSIM3«.

	0,5 μm - Technologie	130 nm- Technologie
Anzahl der verwendeten Metalllagen	3	3
Versorgungsspannung	3,3 V	1,4 V
Optimierung vor dem phys. Entwurf	ja	nein
Optimierung im physikalischen Entwurf	nein	ja
Inverterbäume zur Fanoutbegrenzung	nein	ja
Taktverteilungsbaum	ja	ja
maximale Anstiegs- bzw. Abfallzeit	ca. 2 ns	ca. 0,2 ns
Leitungswiderstände extrahiert	nein	ja

Tabelle 6.1: Vergleich der Technologien und des physikalischen Entwurfs

Um die Fehlerquelle der eventuell falschen Modellierung der Eingangskapazitäten der Gatter auszuschließen, sind die ersten Stufen aller den interessierenden Pfad verlassender Nebenpfade in der zu simulierenden Netzliste enthalten. Abbildung 6.1 stellt diesen Vorgang dar: Gatter, die mittels eines Eingangs mit dem zu untersuchenden Pfad verbunden sind, werden mit ihrer Ausgangslast berücksichtigt. Alle Seiteneingänge werden auf den nicht-kontrollierenden Wert gelegt.

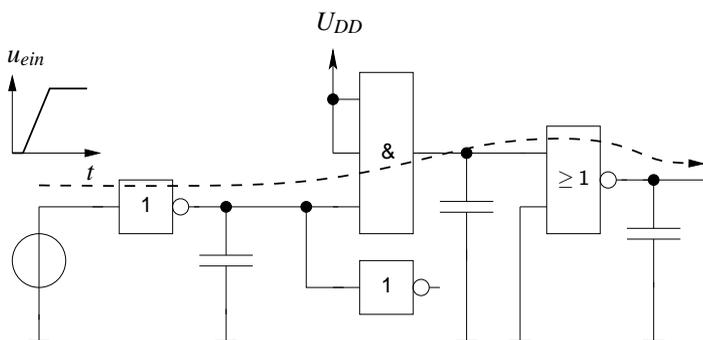


Abbildung 6.1: Zur Analogsimulation eines Pfades

Zum Vergleich der statischen Timinganalyse mit der Analogsimulation des ermittelten Pfades wird die Verzögerung zwischen dem Durchschreiten des 50 %-Versorgungsspannungswertes am Startpunkt und am Endpunkt herangezogen.

gen. Dieses Vorgehen ist sinnvoll, da die Anstiegszeiten deutlich kleiner als die Verzögerung des gesamten Pfades sind. Der Verzögerungsbegriff wird in diesem Fall also zu Vergleichszwecken gebraucht.

Die statische Timinganalyse und die Analogsimulation werden unter dem Betriebssystem »Linux« auf einem Rechner mit einem AMD-XP2600+-Prozessor mit 2,083 GHz Taktfrequenz durchgeführt. Die angegebenen Laufzeiten beziehen sich auf diesen Rechner.

Dieses Kapitel ist im Folgenden in zwei Teile gegliedert: Im ersten Teil werden von der Timinganalyse ermittelte Pfade mit »Spice3« simuliert und die Verzögerungswerte der Pfade verglichen. Hiermit sollen die Abweichungen der statischen Timinganalyse grundsätzlich beurteilt werden. Im zweiten Teil werden Ergebnisse unter Verwendung der in den Kapiteln 4 und 5 vorgestellten Methoden behandelt.

6.2 Validierung der statischen Timinganalyse und der Verzögerungsmodelle

6.2.1 Passive Verzögerungsmodellierung

Die statische Timinganalyse wird hier im herkömmlichen Modus, also mit ausschließlich geerdeten Kapazitäten ohne Berücksichtigung der Kopplung, verwendet. Die 100 kritischen Pfade der Schaltung werden validiert, wobei jeweils nur ein Pfad mit steigendem oder fallendem Ereignis pro Endpunkt untersucht wird. Es werden, sofern diese Pfade existieren, 100 steigende und fallende Ereignisse an jedem Knoten berechnet. Auf diese Weise wird die Komplexität einer einhundertmal so großen Schaltung simuliert. Im Falle der Schaltung »s35932« werden 1,5 Millionen Simulationen durchgeführt. Die Koppelkapazität wird mit einfachem, verdoppeltem und verdreifachtem Wert berücksichtigt. Bei drei Schaltungen ergeben sich 900 zu untersuchende Pfade pro Technologie und somit ebensoviele Analogsimulationen. Die Tabellen 6.2 und 6.3 zeigen die relativen Abweichungen, die Tabellen 6.4 und 6.5 die Rechenzeit und den Speicherbedarf für die Berechnung von jeweils 100 Pfaden pro Endpunkt.

Minimum des relativen Fehlers	-0,90 %
Maximum des relativen Fehlers	2,55 %
Mittlerer relativer Fehler	0,61 %
Empirische Standardabweichung	0,64 %

Tabelle 6.2: Vergleich zur Analogsimulation bei passiver Kopplungsmodellierung, 0,5 μm -Technologie

Minimum des relativen Fehlers	-0,19 %
Maximum des relativen Fehlers	3,08 %
Mittlerer relativer Fehler	1,58 %
Empirische Standardabweichung	0,70 %

Tabelle 6.3: Vergleich zur Analogsimulation bei passiver Kopplungsmodellierung, 130 nm-Technologie

Schaltung	Laufzeit/min:s	Maximaler Speicherbedarf/MByte
s35932	5:41	357
s38417	5:00	350
s38584	4:34	205

Tabelle 6.4: Laufzeit und Speicherbedarf bei passiver Kopplungsmodellierung, 0,5 μm -Technologie

Schaltung	Laufzeit/min:s	Maximaler Speicherbedarf/MByte
s35932	35:02	984
s38417	21:03	898
s38584	16:58	606

Tabelle 6.5: Laufzeit und Speicherbedarf bei passiver Kopplungsmodellierung, 130 nm-Technologie

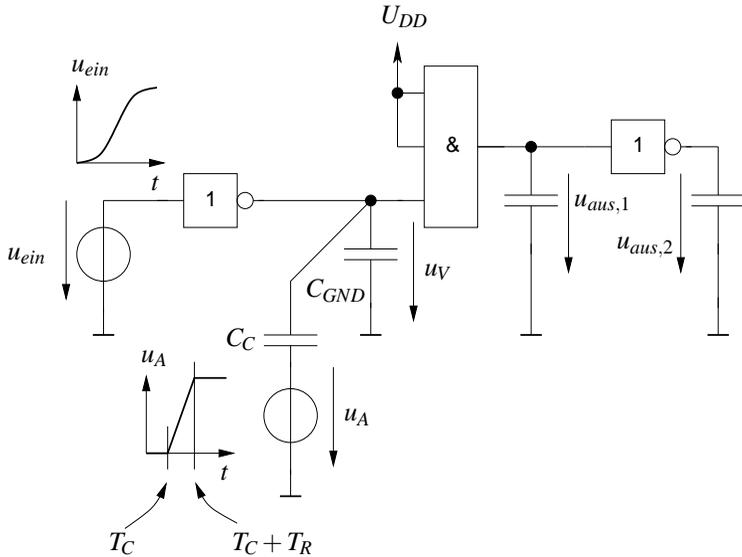


Abbildung 6.2: Analogsimulation einer Stufe eines Pfades, Kopplung nach Modell 1

6.2.2 Erstes Verzögerungsmodell

Im Unterschied zur passiven Modellierung sind zur Validierung der Kopplung nach dem ersten Modell eine wesentliche Modifikation erforderlich: Für jeden zu untersuchenden Pfad sind mehrere Analogsimulationen erforderlich. Dies hat seine Ursache darin, dass die Spannungsquelle, die an die Koppelkapazität eines jeden Knotens angeschlossen ist, in Ankunftszeit T_C und Anstiegszeit T_R variiert werden muss, um die maximale Verzögerung zu ermitteln.

Die Simulation des gesamten Pfades einschließlich aller Kopplungsknoten kann bereits bei diesem einfachen Modell mehrere Minuten dauern. Um das Verfahren zu beschleunigen, wird jede Stufe des Pfades separat simuliert (siehe Abbildung 6.2). Die Eingangsspannung u_{ein} ist die Opferspannung u_V der vorgehenden Stufe. Neben dem Gatter der betrachteten Stufe werden die Gatter der folgenden Stufen mit ihren Massekapazitäten berücksichtigt. T_C und T_R werden derart

variiert, dass das Erreichen der Schwellspannung $u_{aus,2}(T_D) = \frac{1}{2}U_{DD}$ maximiert wird.

Ist das Maximum von T_D gefunden, ist die Simulation der betrachteten Stufe beendet. Die folgende Stufe wird mit einer Eingangsspannung $u_{ein} := u_V$, die wiederum von einer Spannungsquelle eingepreßt wird, bearbeitet.

Die Simulationsergebnisse sind in den Tabellen 6.6 und 6.7 und dargestellt.

Minimum des relativen Fehlers	-0,51 %
Maximum des relativen Fehlers	2,64 %
Mittlerer relativer Fehler	0,84 %
Empirische Standardabweichung	0,65 %

Tabelle 6.6: Validierung des ersten Verzögerungsmodells, 0,5 µm-Technologie

Minimum des relativen Fehlers	-0,39 %
Maximum des relativen Fehlers	2,70 %
Mittlerer relativer Fehler	1,32 %
Empirische Standardabweichung	0,60 %

Tabelle 6.7: Validierung des ersten Verzögerungsmodells, 130 nm-Technologie

6.2.3 Zweites Verzögerungsmodell

Im Unterschied zum ersten Verzögerungsmodell bleiben bei diesem Modell die Aggressorgatter erhalten. Abbildung 6.3 zeigt die Simulation einer Stufe mit einem Aggressor. Die Spannungskurve an den Eingängen der Aggressorgatter wird durch eine Rampe, deren Anstiegszeit aus der statischen Timinganalyse entnommen ist, modelliert. Der Einfluss der Anstiegszeit an den Aggessoreingängen auf die Verzögerung des Opferpfades wird als gering erachtet. Die so getroffene Vereinfachung ermöglicht, nur einen Parameter, die Ankunftszeit T_C , statt derer zwei zu variieren.

Die der Simulation zugrunde liegende extrahierte Netzliste beinhaltet im Allgemeinen weitere Koppelkapazitäten des betrachteten Aggessornetzes zu anderen

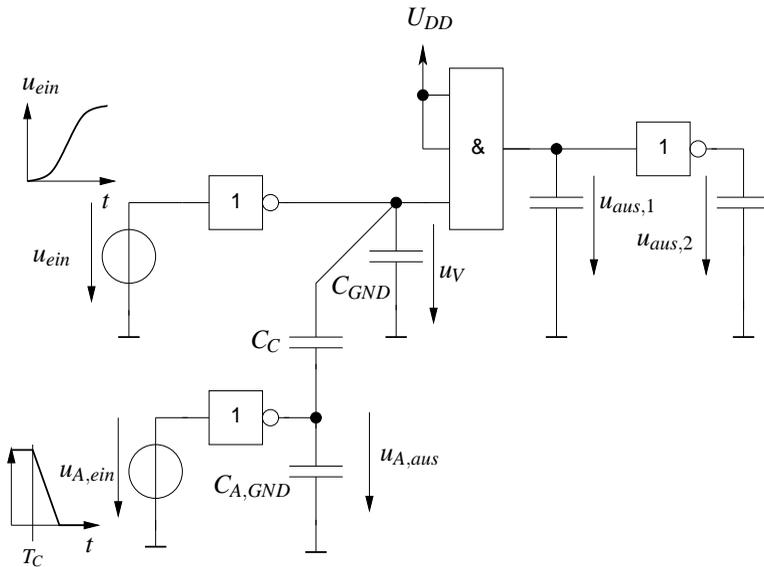


Abbildung 6.3: Analogsimulation einer Stufe eines Pfades, Kopplung nach Modell 2

Netzen. Um den ungünstigsten Kopplungsfall zu erfassen, müssten diese ebenfalls zum Opfernnetz gegenkoppeln, das heißt zu anderen Aggressornetzen mitkoppeln. Da in diesem Fall die Anzahl der zu betrachtenden Gatter und Netze leicht die gesamte Schaltung erreicht, was eine Simulation unmöglich macht, werden nur Koppelkapazitäten, die direkt mit dem Opfernnetz verbunden sind, einbezogen. Alle anderen Koppelkapazitäten werden ignoriert und sind somit keine Last für die Aggressoren.

Ziel der Simulationen einer Stufe ist, wie in Abschnitt 6.2.2, die Zeit T_D mit $u_{aus,2}(T_D) = \frac{1}{2}U_{DD}$ zu maximieren. Dies erfolgt durch ein gerichtetes Suchverfahren, das im Folgenden durch Pseudocode beschrieben wird. Es wird ein fallendes Opferereignis angenommen. In einem ersten Schritt werden die Ankunftszeiten auf den Opfer- und Aggressornetzen bestimmt.

setze jeden Aggressor in Ruhe ($u_{A, \text{ein}} = U_{DD}$)
 berechne die Opferankunftszeit T_V mit $u_V(T_V) = \frac{1}{2}U_{DD}$
 setze das Opfernnetz in Ruhe ($u_{\text{ein}} = 0$)
 für jeden Aggressor i
 setze andere Aggressoren in Ruhe
 berechne Aggressorankunftszeit $T_{A,i}$ mit $u_{A, \text{in}, i}(T_{A,i})$ so,
 dass $u_{A, \text{aus}, i}(T_V) = \frac{1}{2}U_{DD}$

Auf diese Weise wird ein Startwert für jede Aggressorankunftszeit $T_{A,i}$ bestimmt. Dieser Wert wird durch das folgende gerichtete Suchverfahren variiert.

berechne T_D
 $\Delta t = \Delta t_{\text{max}}$
 während $\Delta t > \Delta t_{\text{min}}$
 für jeden Aggressor i , beginnend mit $C_{C,i} = \text{max}$
 L1 : für jedes $\tau := (-2\Delta t, -\Delta t, 0, \Delta t, 2\Delta t)$
 $T_{A,i, \text{tmp}} := T_{A,i} + \tau$
 berechne $T_{D, \text{tmp}}$
 falls $T_{D, \text{tmp}} > T_D$
 $T_D := T_{D, \text{tmp}}$
 $T_{A,i} := T_{A,i, \text{tmp}}$
 gehe zu L1
 $\Delta t := \frac{1}{2}\Delta t$

Bei diesem Verfahren wird davon ausgegangen, dass eine größere Koppelkapazität die Pfadverzögerung stärker als eine kleinere beeinflusst. Es erscheint daher sinnvoll, die Ankunftszeiten dieser Aggressoren zuerst zu bestimmen. Im dargestellten Pseudocode wird die Verzögerung für eine Konfiguration der Aggressorankunftszeiten gegebenenfalls mehrfach berechnet. Diese Ineffizienz kann durch einen einfachen Cache-Mechanismus überwunden werden.

Die Komplexität des vorgeschlagenen Suchverfahrens ist $O(n)$ mit n als Anzahl aller Aggressoren. Da diese Anzahl häufig dreistellig ist, kann die Berechnung einer Stufe bereits mehrere Tage in Anspruch nehmen. Es wird die folgende Vereinfachung verwendet: Maximal zehn koppelnde Aggressornetze werden berücksichtigt, beginnend mit dem größten Summenkapazitätswert. Es werden zwei Simulationen durchgeführt: In der ersten wird ein Verhalten gemäß des ersten Kopplungsmodells, das den ungünstigsten Fall darstellt, angenommen. Die ermit-

6.2 Validierung der statischen Timinganalyse und der Verzögerungsmodelle

telte Pfadverzögerung ist die obere Schranke. In der zweiten Simulation werden die oben genannten, kleinwertigen Koppelkapazitäten geerdet. Diese Pfadverzögerung ist die untere Schranke.

Tabelle 6.8 zeigt, dass die untere Schranke in drei Fällen unterschritten wird. Das heißt, die Pfadverzögerung in der statischen Timinganalyse ist kleiner als die durch Analogsimulation ermittelte. Die obere Schranke wird in der 0,5 μm -Technologie von etwa 26 % aller Pfade überschritten, in der 130 nm-Technologie von 94 %. Tabelle 6.9 zeigt die vom Betrag maximalen Fehler. In den Tabellen 6.10 und 6.11 sind die relativen Differenzen zwischen oberer und unterer Schranke der Pfadverzögerung bezüglich der oberen Schranke dargestellt.

Untersuchte Pfade	600
Unterschreitungen der unteren Schranke	3
Überschreitungen der oberen Schranke, 0,5 μm	78
Überschreitungen der oberen Schranke, 130 nm	282

Tabelle 6.8: Validierung des zweiten Verzögerungsmodells, Anzahl der Pfade, 0,5 μm - und 130 nm-Technologie

Minimale Unterschreitung der unteren Schranke	-0,46 %
Maximale Überschreitung der oberen Schranke	2,73 %

Tabelle 6.9: Validierung des zweiten Verzögerungsmodells, relative Fehler, 0,5 μm - und 130 nm-Technologie

Minimale relative Fenstergröße	5,41 %
Maximale relative Fenstergröße	8,79 %
Mittelwert der relative Fenstergröße	7,21 %

Tabelle 6.10: Validierung des zweiten Verzögerungsmodells, Schranken, 0,5 μm -Technologie

Minimale relative Fenstergröße	0,00 %
Maximale relative Fenstergröße	1,64 %
Mittelwert der relative Fenstergröße	0,49 %

Tabelle 6.11: Validierung des zweiten Verzögerungsmodells, Schranken, 130 nm-Technologie

6.2.4 Verzögerungsmodell nach Chen, Kirkpatrick und Keutzer

Da das in [50] beschriebene Modell in dem Programm »EinsTimer« [14] unterstützt wird, wird es zum Vergleich herangezogen. Das Modell wird dazu in das behandelte Werkzeug zur statischen Timinganalyse integriert. Es ist das verbesserte Modell, bei dem der Miller-Faktor beginnend mit $k = 3$ iterativ verkleinert wird, implementiert. Das Verfahren der Validierung ist identisch mit dem in Abschnitt 6.2.3 beschriebenen. Es werden ebenfalls 100 Pfade für jede Schaltung berechnet. Die Ergebnisse sind in den Tabellen 6.12 und 6.13 dargestellt. Da das Modell die Verzögerung überschätzt, wird der Vergleich nur mit der oberen Schranke gezogen.

Minimale Unterschreitung der unteren Grenze	0,00 %
Minimale Überschreitung der oberen Grenze	2,75 %
Maximale Überschreitung der oberen Grenze	8,82 %
Mittelwert der Überschreitung der oberen Grenze	5,90 %

Tabelle 6.12: Validierung des Verzögerungsmodells nach Chen et. al., relative Fehler, 0,5 μm -Technologie

Minimale Unterschreitung der unteren Grenze	0,00 %
Minimale Überschreitung der oberen Grenze	1,11 %
Maximale Überschreitung der oberen Grenze	4,07 %
Mittelwert der Überschreitung der oberen Grenze	2,21 %

Tabelle 6.13: Validierung des Verzögerungsmodells nach Chen et. al., relative Fehler, 130 nm-Technologie

6.2.5 Bewertung

Der Vergleich der passiven Modellierung mit der Analogsimulation zeigt Fehlergrenzen im Bereich von -0,90 % bis 3,08 %. Trotz starker Vereinfachung der Modellierung, hier ist insbesondere die Linearisierung aller gatterinternen parasitären Kapazitäten zu nennen, ist der Fehler als gering zu bezeichnen. Die Laufzeiten für die ohne Widerstände extrahierte 0,5 μm -Technologie sind sehr gering. Im Falle der 130 nm-Technologie wird ein Preis für die fehlende Ordnungsreduktion bezahlt. Die Rechenzeit der Schaltung »s35932« ist etwa um den Faktor sieben größer. Die Validierung des ersten Verzögerungsmodells für Kopplung liefert ähnliche Fehlergrenzen wie die der passiven Modellierung.

Bei der Validierung des zweiten Modells gibt es einen Unterschied zu den bisher besprochenen Verfahren: Zur Reduzierung der Rechenzeit mussten die teilweise mehr als 100 Koppelkapazitäten pro Schaltungsknoten auf weniger als zehn reduziert werden. Statt eines singulären Verzögerungswertes liefert die Validierung eine obere und untere Schranke. Obwohl die Unterschiede zwischen oberer und unterer Schranke nur in der Behandlung kleinwertiger Koppelkapazitäten liegen, ist die mittlere Differenz dieser Größen mit 7,21 % im Falle der 0,5 μm -Technologie groß, im Falle der 130 nm-Technologie praktisch nicht vorhanden (0,49 %). Es deutet sich an, dass der Einfluss von Kopplung auf die Verzögerung in der 130 nm-Technologie sehr gering ist.

Die statische Timinganalyse verletzt die durch Analogsimulation bestimmten Grenzen der Verzögerung nur gering, maximal mit 2,73 %. Das zum Vergleich herangezogene Modell von Chen et. al. [50] überschreitet die obere Grenze in jedem Fall, im Mittel sogar um 5,90 % und maximal um 8,82 % in der

0,5 μm -Technologie. Das Chen-Modell kann daher als pessimistischer als das vorgeschlagene zweite Verzögerungsmodell für Kopplung angesehen werden.

Zusammenfassend wird festgestellt, dass es eine gute Übereinstimmung der durch die statische Timinganalyse mit denen durch Analogsimulation ermittelten Verzögerungen gibt. Das zweite Verzögerungsmodell hat geringere Fehler als das von Chen et. al. vorgeschlagene Modell.

6.3 Vergleich der Verzögerungsmodelle

6.3.1 Kopplung ohne Berücksichtigung der Aggressorsituation

Die Tabellen 6.14 bis 6.19 zeigen die Verzögerungen der kritischen Pfade. Das erste Modell wird als Norm angesehen, da es gute Übereinstimmung mit der Analogsimulation zeigt.

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/s
passiv, $k = 1$	11,378	-14,9	16
passiv, $k = 2$	12,557	-6,1	16
passiv, $k = 3$	13,709	2,5	16
Modell 1	13,369	0,0	16

Tabelle 6.14: Ergebnisse Schaltung s35932, 0,5 μm -Technologie

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/s
passiv, $k = 1$	21,482	-16,2	18
passiv, $k = 2$	23,869	-6,9	18
passiv, $k = 3$	26,336	2,8	18
Modell 1	25,627	0,0	18

Tabelle 6.15: Ergebnisse Schaltung s38417, 0,5 μm -Technologie

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/s
passiv, $k = 1$	20,293	-15,0	27
passiv, $k = 2$	22,456	-5,9	27
passiv, $k = 3$	24,579	3,0	28
Modell 1	23,863	0,0	27

Tabelle 6.16: Ergebnisse Schaltung s38584, 0,5 μm -Technologie

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/min:s
passiv, $k = 1$	1,253	-7,0	2:10
passiv, $k = 2$	1,315	-2,4	2:17
passiv, $k = 3$	1,378	2,2	2:22
Modell 1	1,348	0,0	2:25

Tabelle 6.17: Ergebnisse Schaltung s35932, 130 nm-Technologie

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/min:s
passiv, $k = 1$	2,098	-1,6	1:37
passiv, $k = 2$	2,126	-0,3	1:38
passiv, $k = 3$	2,157	1,1	1:38
Modell 1	2,133	0,0	1:42

Tabelle 6.18: Ergebnisse Schaltung s38417, 130 nm-Technologie

Modell	Verzögerung/ns	rel. Differenz/%	Laufzeit/s
passiv, $k = 1$	2,071	-3,7	2:10
passiv, $k = 2$	2,124	-1,2	2:13
passiv, $k = 3$	2,176	1,2	2:15
Modell 1	2,150	0,0	2:17

Tabelle 6.19: Ergebnisse Schaltung s38584, 130 nm-Technologie

6.3.2 Kopplung unter Berücksichtigung der Aggressorsituation

Die Tabellen 6.20 bis 6.25 zeigen die Verzögerungen der kritischen Pfade der untersuchten Schaltungen. Die nach dem ersten Verzögerungsmodell berechne-

te Verzögerung dient hier zu Vergleichszwecken, die Aggressorsituation ist in diesem Modell unberücksichtigt. Die Zeile »Modell 2, Schaltfenster« ist die Umsetzung des Ein-Schritt-Verfahrens, das in Abschnitt 5.4 beschrieben ist, »Modell 2, iterativ« die Erweiterung um den wiederholten Aufruf. Das zweite Verzögerungsmodell wird als Norm betrachtet.

Modell	Verzög./ns	rel. Diff./%	Laufzeit/h:min:s
Modell 1	13,369	6,5	16
Modell nach Chen et. al	13,445	7,1	31
Modell 2	12,556	0,0	53
Modell 2, Schaltfenster	12,126	-3,4	50
Modell 2, iterativ	11,910	-5,1	1:13

Tabelle 6.20: Ergebnisse Schaltung s35932, 0,5 µm-Technologie

Modell	Verzög./ns	rel. Diff./%	Laufzeit/min:s
Modell 1	25,627	5,0	18
Modell nach Chen et. al	26,161	7,2	33
Modell 2	24,397	0,0	59
Modell 2, Schaltfenster	23,521	-3,6	55
Modell 2, iterativ	22,271	-8,7	1:00

Tabelle 6.21: Ergebnisse Schaltung s38417, 0,5 µm-Technologie

Modell	Verzög./ns	rel. Diff./%	Laufzeit/min:s
Modell 1	23,863	4,7	27
Modell nach Chen et. al	24,365	6,9	49
Modell 2	22,794	0,0	1:25
Modell 2, Schaltfenster	22,127	-2,9	1:20
Modell 2, iterativ	21,673	-4,9	1:36

Tabelle 6.22: Ergebnisse Schaltung s38584, 0,5 µm-Technologie

Modell	Verzög./ns	rel. Diff./%	Laufzeit/min:s
Modell 1	1,348	2,0	2:25
Modell nach Chen et. al	1,358	2,7	6:34
Modell 2	1,322	0,0	11:05
Modell 2, Schaltfenster	1,320	-0,2	10:06
Modell 2, iterativ	1,309	-1,0	19:34

Tabelle 6.23: Ergebnisse Schaltung s35932, 130 nm-Technologie

Modell	Verzög./ns	rel. Diff./%	Laufzeit/min:s
Modell 1	2,133	0,5	1:42
Modell nach Chen et. al	2,130	0,3	3:23
Modell 2	2,123	0,0	5:24
Modell 2, Schaltfenster	2,104	-0,9	4:56
Modell 2, iterativ	2,104	-0,9	8:37

Tabelle 6.24: Ergebnisse Schaltung s38417, 130 nm-Technologie

Modell	Verzög./ns	rel. Diff./%	Laufzeit/min:s
Modell 1	2,150	1,2	2:17
Modell nach Chen et. al	2,136	0,5	5:14
Modell 2	2,125	0,0	10:46
Modell 2, Schaltfenster	2,095	-1,4	9:32
Modell 2, iterativ	2,088	-1,7	25:57

Tabelle 6.25: Ergebnisse Schaltung s38584, 130 nm-Technologie

6.3.3 Bewertung

Die Modelle ohne Berücksichtigung der Aggressorsituation werden wie folgt bewertet: Wie erwartet unterschätzen die passiven Modelle mit $k = \{1, 2\}$ die Pfadverzögerung. Der in [50] hergeleitete maximale Miller-Faktor von drei führt zu einer Überschätzung der Verzögerung. Ein Ignorieren der Kopplung in der Verzögerungsberechnung ($k = 1$) kann zu einer beträchtlichen Unterschätzung, hier bis zu 16,2 %, der Pfadverzögerung führen. Die Überschätzung mit einem Miller-Faktor von $k = 3$ ist gering (maximal 3 %).

Die Bewertung der Modelle mit Berücksichtigung der Aggressorsituation ist diese: Das Modell nach Chen et. al. überschätzt die Verzögerung in der 0,5 µm-Technologie deutlich. Obwohl dieses Modell die Aggressorsituation berücksichtigt, liegen die Verzögerungen teilweise über denen des die Aggressorsituation nicht berücksichtigenden ersten Verzögerungsmodells. Zieht man in Betracht, dass die Verringerung der Verzögerung des kritischen Pfades eine der Hauptaufgaben des Chipentwurfs ist und, dass das Modell von Chen et. al. teilweise als Fertigungskriterium Anwendung findet, so kann durch den Einsatz eines besseren Verzögerungsmodells für die Kopplung erheblich Aufwand gespart werden.

Das zweite Verzögerungsmodell hat, verglichen mit dem ersten und dem Modell nach Chen et. al. eine deutlich höhere Rechenzeit. Die Verbesserung durch die Beachtung der Schaltfenster ist insbesondere beim iterativen Verfahren erheblich (bis zu 8,7 %). Das Modell nach Chen et. al. kann prinzipiell auch mit der Beachtung von Schaltfenstern kombiniert werden. Dieses Verfahren wurden jedoch nicht untersucht.

Es fällt auf, dass der Effekt der Kopplung auf die Verzögerung in der 0,5 µm-Technologie wesentlich größer als auf die der 130 nm-Technologie ist. Die Geometrien und Parasiten der Technologien liefern keine Erklärung für dieses Phänomen. Der wesentliche Unterschied liegt in der Behandlung der Schaltungen im physikalischen Entwurf, insbesondere in den Möglichkeiten im Entwurfsverfahren der 130 nm-Technologie, Inverterbäume einzubauen und die Anstiegs- und Abfallzeiten zu limitieren. Zur Untermauerung dieser These wurde in der Schaltung »s35932« in der 130 nm-Technologie die Treiberstärke zweier Gatter derart vermindert, dass ihre Ausgangsanstiegszeiten im Bereich 1 ns lagen. Tabelle 6.26 zeigt die Ergebnisse der Timinganalyse.

Modell	Verzög./ns	rel. Diff./%
Modell 1	2,067	1,3
Modell nach Chen et. al	2,157	5,7
Modell 2	2,039	0,0

Tabelle 6.26: Ergebnisse der modifizierten Schaltung s35932, 130 nm-Technologie

Die Ergebnisse stimmen mit den Beobachtung in der 0,5 μm -Technologie überein. Das Modell nach Chen et. al. überschätzt die Verzögerung und ist noch pessimistischer als das erste Verzögerungsmodell für Kopplung.

6.4 Verbesserte Spannungskurvenpropagierung

In Abschnitt 4.3 wurde ein Verfahren zur verbesserten Spannungskurvenpropagierung vorgeschlagen. Die kritischen Pfade der oben genannten Schaltungen wurden mit und ohne das neue Verfahren ermittelt. Es wurden 18 Konfiguration untersucht. Die maximalen Verzögerungen wurden in zwei Fällen durch das neue Auswahlverfahren verringert (um bis zu 9 ps), in zehn Fällen vergrößert (um bis zu 20 ps).

Da eine Verringerung der Verzögerung durch das neue Verfahren in der Theorie ausgeschlossen ist und die Verringerung beziehungsweise Vergrößerung im Promillebereich (-0,07 % und 0,076 %) liegt, kann vermutet werden, dass der Effekt kleiner als die Berechnungsungenauigkeit des Simulators ist.

6.5 Grenzen der vorgeschlagenen Verfahren

Für die Gattermodellierung wurden drei Vereinfachungen zur Verringerung der Rechenzeiten gemacht:

1. Das Gleichstrom-Ausgangsverhalten ist tabelliert.
2. Alle gatterinternen Kapazitäten sind linear.
3. Alle gatterinternen Kapazitäten sind an den Masseknoten angeschlossen.

Während die erste Änderung vermutlich zu nur geringen Fehlern führt, dürften die zweite und dritte Vereinfachung die Hauptursache für die Differenzen zwischen der statischen Timinganalyse und der Spice-Simulationen der kritischen Pfade sein. Ein Neuimplementierung sollte nichtlineare Kapazitäten vorsehen.

Das erste Verzögerungsmodell ignoriert die Aggressorsituation. Von dieser Voraussetzung abgesehen werden jedoch keine weiteren nachteiligen Annahmen getroffen.

Bei dem zweiten Verzögerungsmodell wird die Aggressorsituation berücksichtigt. Die zeitlichen Auswirkungen der Kopplung sind, wie in Abschnitt 5.1 gezeigt, bei kleineren Aggressoranstiegszeiten größer. In dem implementierten zweiten Verzögerungsmodell wie auch in der Validierung wird eine Mitkopplung in den Aggressorknoten durch Weglassen dieser Koppelkapazitäten berücksichtigt. Der Miller-Faktor für Mitkopplung wird also zu null angenommen. In [50] wird ein minimaler Miller-Faktor von $k = -1$ für Mitkopplung angegeben. Um das vorgeschlagene zweite Verzögerungsmodell zu verbessern, muss die Mitkopplung ebenfalls genauer untersucht werden.

Eine andere Einschränkung haben das zweite Verzögerungsmodell und das Modell nach Chen et. al gemeinsam: Die Aggressor-Spannungskurven werden als eingepreßt angesehen. Bei großen Koppelkapazitäten und starken Opfergattern führt das zu einer Überschätzung der Verzögerung, die auch schon in Abschnitt 5.3.3 beobachtet wurde.

7 Zusammenfassung

In dieser Arbeit wurde der Einfluss kapazitiver Kopplung auf die Verzögerung integrierter digitaler CMOS-Schaltungen untersucht.

Um den Einfluss kapazitiver Kopplung in einer statischen Timinganalyse verwenden zu können, ist es sinnvoll, Spannungsschwellen zu definieren, anhand derer das verzögerungsbestimmende Ereignis bestimmt wird. Theoretische Betrachtungen zeigen, dass die Verzögerung für größere Schwellenspannungen bei ausschließlicher Betrachtung einer Stufe größer wird. Werden folgende Stufen in Betracht gezogen, nimmt die Pfadverzögerung bei größeren Schwellenspannungen wieder ab. Dies ist bereits bei $\frac{1}{2}U_{DD}$ der Fall. Die Schwellenspannungen werden daher empirisch ermittelt. Sie liegen zwischen den Transistor-Schwellenspannungen und $\frac{1}{2}U_{DD}$.

Um bezüglich des verwendeten Verzögerungsmodells keinen Einschränkungen zu unterliegen, wurde eine transistorbasierte statische Timinganalyse implementiert. Die Verzögerungsberechnung erfolgt durch einen integrierten Analogsimulator. Ereignisse sind Spannungskurven der Form $t \rightarrow u(t)$. Gegenüber klassischen Analogsimulatoren sind Verfahren zur Beschleunigung notwendig. Dazu zählen unter anderen eine Beschränkung auf Zweipole, die Verwendung der einfachen Knotenanalyse und Tabellenmodelle zur Beschreibung des Gleichstromverhaltens der Transistoren.

Zur Modellierung der Verzögerung bei Kopplung wurden zwei Modelle vorgeschlagen: Eines ohne und eines unter Beachtung der Aggressorsituation. Beide Modelle wurden durch Analogsimulation validiert und in die statische Timinganalyse integriert. Das von Chen et. al. vorgeschlagene Verzögerungsmodell [50] hat bereits industrielle Anwendung gefunden und ist ebenfalls implementiert.

Zur Validierung der Methoden wurden die drei größten Schaltungen der ISCAS-89-Benchmark in einer 0,5 μm - und 130 nm-Technologie platziert, verdrahtet, extrahiert und der implementierten statischen Timinganalyse unterzogen. Die er-

rechneten kritischen Pfade wurden danach durch Analogsimulationen mit dem Programm »Spice3« validiert. Es zeigte sich eine gute Übereinstimmung sowohl ohne als auch mit Kopplung. Das Modell von Chen et. al. überschätzt die Verzögerung deutlich. Es zeigte sich außerdem, dass der Effekt von Kopplung auf das Zeitverhalten der Schaltungen in der 130 nm-Technologie geringer als in der 0,5 μm -Technologie ist. Dies ist jedoch nicht auf Eigenschaften der Technologie, sondern des Entwurfsverfahrens zurückzuführen.

Eine Schlüsseleigenschaft der statischen Timinganalyse ist, dass sie das letzte Ereignis jedes Knotens innerhalb eines Taktzyklusses berechnet. Diese Eigenschaft kann bei der Betrachtung der Kopplung genutzt werden: Erfolgt das Opferereignis nach dem spätesten Aggressorereignis, so kann das Opferereignis durch Kopplung nicht zusätzlich verzögert werden. Dazu wurden zwei Verfahren umgesetzt.

Literaturverzeichnis

- [1] D. Gajski und R. H. Kuhn, »New VLSI Tools«. IEEE Computer, Bd. 16, Nr. 12, S. 11–14, 1983.
- [2] N. H. E. Weste und K. Eshraghian, »Principles of CMOS VLSI Design«. Addison-Wesley Publishing, 1994.
- [3] P. Pirsch, »Entwurf integrierter digitaler Schaltungen«. Vorlesungsskript, Universität Hannover, Institut für theoretische Nachrichtentechnik und Informationsverarbeitung, 1994.
- [4] E. Barke und R. Brück, »Layout integrierter Schaltungen«. Vorlesungsskript, Universität Hannover, Institut für Mikroelektronische Systeme, 1995.
- [5] H. Engesser, Hrsgb., »Duden Informatik«. Bibliographisches Institut & F. A. Brockhaus AG, Mannheim, 1993.
- [6] U. Tietze und C. Schenk, »Halbleiter-Schaltungstechnik«. Springer-Verlag, 1993.
- [7] D. H. C. Du, S. H. C. Yen und S. Ghanta, »On the General False Path Problem in Timing Analysis«. Proc. of the 26th ACM/IEEE DAC, S. 555–560, 1989.
- [8] P. C. McGeer und R. K. Brayton, »Efficient Algorithms for Computing the Longest Viable Path in a Combinational Network«. Proc. of the 26th ACM/IEEE DAC, S. 561–567, 1989.
- [9] B. Korte und J. Vygen, »Combinatorial Optimization«. Springer-Verlag, 2001.
- [10] J. W. Freeman, »Improvements to Propositional Satisfiability Search Algorithms«. Dissertation an der University of Pennsylvania, 1995.

- [11] P. Pirsch, »Digitalschaltungen der Elektronik«. Vorlesungsskript, Universität Hannover, Institut für theoretische Nachrichtentechnik und Informationsverarbeitung, 1993.
- [12] D. E. Wallace und C. H. Séquin, »ATV: An Abstract Timing Verifier«. Proc. of the 25th ACM/IEEE DAC, S. 154–159, 1988.
- [13] J. Schietke, »Timing Optimierung beim physikalischen Layout von nicht-hierarchischen Designs«. Dissertation an der Universität Bonn, 1999.
- [14] »EinsTimer User's Guide«. IBM Corp., September 2002.
- [15] N. P. Jouppi, »Timing Analysis for nMOS VLSI«. Proc. of the 20th ACM/IEEE DAC, S. 411–418, 1983.
- [16] N. P. Jouppi, »Timing Analysis and Performance Improvement of MOS VLSI Designs«. IEEE Transactions on Computer-Aided Design, Bd. 6, Nr. 4, S. 650–665, Juli 1987.
- [17] R. B. Hitchcock, G. L. Smith und D. D. Cheng, »Timing Analysis of Computer Hardware«. IBM Journal of Research and Development, Bd. 26, Nr. 1, S. 100–105, Januar 1982.
- [18] J. K. Ousterhout, »A Switch-Level Timing Verifier for Digital MOS VLSI«. IEEE Transactions on Computer-Aided Design, Bd. 4, Nr. 3, S. 336–349, Juli 1985.
- [19] J. Benkoski, E. V. Meersch, L. Claesen und H. D. Man, »Efficient Algorithms for Solving the False Path Problem in Timing Verification«. Proc. of the ICCAD, S. 44–47, 1987.
- [20] J. Benkoski, E. V. Meersch, L. Claesen und H. D. Man, »Timing Verification Using Statically Sensitizable Paths«. IEEE Transactions on Computer-Aided Design, Bd. 9, Nr. 10, S. 1073–1084, Oktober 1990.
- [21] P. C. McGeer und R. K. Brayton, »Integrating Functional and Temporal Domains in Logic Design«. Kluwer Academic Publishers, 1991.
- [22] H. Chang und J. A. Abraham, »VIPER: An Efficient Vigorously Sensitizable Path Extractor«. Proc. of the 30th ACM/IEEE DAC, S. 112–117, 1993.
- [23] M. Ringe, T. Lindenkreuz und E. Barke, »Path Verification using Boolean Satisfiability«. Proc. of the DATE '98, S. 965–966, 1998.

- [24] K. Sakallah, T. Mudge und O. Olukotun, »Analysis and Design of Latch-Controlled Synchronous Digital Circuits«. IEEE Transactions on Computer-Aided Design, Bd. 11, Nr. 3, S. 322–333, März 1992.
- [25] J.-F. Lee, D. T. Tang und C. K. Wong, »A Timing Analysis Algorithm for Circuits with Level-Sensitive Latches«. IEEE Transactions on Computer-Aided Design, Bd. 15, Nr. 5, S. 535–543, Mai 1996.
- [26] B. R. Chawla, H. K. Gummel und P. Kozak, »MOTIS-A MOS Timing Simulator«. IEEE Transactions on Circuits and Systems, Bd. 22, Nr. 12, S. 901–910, Dezember 1975.
- [27] V. Chandramouli und K. A. Sakallah, »Selection of Voltage Thresholds for Delay Measurement«. Analog Integrated Circuits and Signal Processing, Bd. 14, Nr. 9, S. 9–28, Dezember 1993.
- [28] »Library Compiler Handbuch«. Synopsys Inc., 1998.
- [29] S. M. Sze, »Physics of Semiconductor Devices«. John Wiley & Sons, Inc., New York, 1981.
- [30] T. Sakurai und A. R. Newton, »Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas«. IEEE Journal of Solid-State Circuits, Bd. 25, Nr. 4, S. 584–594, April 1990.
- [31] W. Liu, X. Jin, J. Chen, M.-C. Jeng, Z. Liu, Y. Cheng, K. Chen, M. Chan, K. Hui, J. Huang, R. Tu, P. K. Ko und C. Ho, »BSIM3v3.2.2 MOSFET Model User’s Manual«. Entwicklungsbericht, University of California, 1999.
- [32] X. Xi, M. Dunga, J. He, W. Liu, K. M. Cao, X. Jin, J. J. Ou, M. Chan, A. M. Niknejad und C. Hu, »BSIM4.4.0 MOSFET Model User’s Manual«. Entwicklungsbericht, University of California, 2004.
- [33] R. E. Bryant, »A Switch-Level Model and Simulator for MOS Digital Systems«. IEEE Transactions on Computers, Bd. 33, S. 336–349, Februar 1984.
- [34] T. Sakurai und A. R. Newton, »Delay Analysis of Series Connected MOSFETs«. IEEE Journal of Solid-State Circuits, Bd. 26, Nr. 2, S. 122–133, Februar 1991.
- [35] S. Embaki und R. Damodaran, »Delay Models for CMOS, BiCMOS and BiNMOS Circuits and Their Applications for Timing Simulations«. IEEE

Transactions on Computer-Aided Design, Bd. 13, Nr. 9, S. 1132–1142, Februar 1993.

- [36] V. Raghavan und R. A. Rohrer, »A New Nonlinear Driver Model for Interconnect Analysis«. Proc. of the 28th ACM/IEEE DAC, S. 561–566, 1991.
- [37] L. T. Pillage, X. Huang und R. A. Rohrer, »Asymptotic Waveform Evaluation for Timing Analysis«. Proc. of the 26th ACM/IEEE DAC, S. 634–637, 1989.
- [38] M. Misheloff, »Improved Modeling and Characterization System for Logic Simulation«. Proc. of the IEEE Asic Intl. Conference, 1992.
- [39] J. Qian, S. Pullela und L. Pillage, »Modeling the “Effective Capacitance” for the RC Interconnect of CMOS Gates«. IEEE Transactions on Computer-Aided Design, Bd. 13, Nr. 12, S. 1526–1535, Dezember 1994.
- [40] F. Dartu, N. Menezes, J. Qian und L. T. Pillage, »A Gate-Delay Model for High-Speed CMOS Circuits«. Proc. of the 31st ACM/IEEE DAC, S. 576–580, 1994.
- [41] F. Dartu, N. Menezes und L. T. Pileggi, »Performance Computations for Precharacterized CMOS Gates with RC Loads«. IEEE Transactions on Computer-Aided Design, Bd. 15, Nr. 5, S. 544–553, Mai 1996.
- [42] W. C. Elmore, »The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers«. Journal of Applied Physics, Bd. 19, Nr. 1, S. 11–12, Januar 1948.
- [43] R. Gupta, B. Tutuianu und L. Pileggi, »The Elmore Delay as a Bound for RC Trees with Generalized Input Signals«. IEEE Transactions on Computer-Aided Design, Bd. 16, Nr. 1, S. 95–104, Januar 1997.
- [44] P. Feldmann und R. W. Freund, »Efficient Linear Circuit Analysis by Padé Approximation via the Lanczos Process«. IEEE Transactions on Computer-Aided Design, Bd. 14, Nr. 5, S. 639–649, Mai 1995.
- [45] K. J. Kerns und A. T. Yang, »Stable and Efficient Reduction of Large, Multiport RC Networks by Pole Analysis via Congruence Transformations«. Proc. of the 33rd ACM/IEEE DAC, S. 285–290, 1996.

- [46] E. Chiprout und M. S. Nakhla, »Analysis of Interconnect Networks Using Complex Frequency Hopping (CFH)«. IEEE Transactions on Computer-Aided Design, Bd. 14, Nr. 2, S. 186–200, Februar 1995.
- [47] A. Odabasioglu, M. Celik und L. T. Pileggi, »PRIMA: passive reduced-order interconnect macromodeling algorithm«. Proc. of the ICCAD, S. 58–65, November 1997.
- [48] R. W. Freund, »SPRIM: Structure-Preserving Reduced-Order Interconnect Macromodeling«. Proc. of the ICCAD, S. 80–87, November 2004.
- [49] P. D. Gross, R. Arunachalam, K. Rajagopal und L. T. Pileggi, »Determination of Worst-Case Aggressor Alignment for Delay Calculation«. Proc. of the ICCAD, S. 212–219, November 1998.
- [50] P. Chen, D. A. Kirkpatrick und K. Keutzer, »Miller Factor for Gate-Level Coupling Delay Calculation«. Proc. of the ICCAD, S. 68–75, November 2000.
- [51] F. Dartu und L. T. Pileggi, »Calculating Worst-Case Gate Delays due to Dominant Capacitance Coupling«. Proc. of the 34th ACM/IEEE DAC, S. 46–51, 1997.
- [52] G. Yee, R. Chandra, V. Ganesan und C. Sechen, »Wire Delay in the Presence of Crosstalk«. Proc. of the IEEE/ACM International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems, S. 170–175, 1997.
- [53] S. Sirichotiyakul, D. Blaauw, C. Oh, R. Levy, V. Zolotov und J. Zuo, »Driver Modeling and Alignment for Worst-Case Delay Noise«. Proc. of the 38th ACM/IEEE DAC, 2001.
- [54] K. L. Shepard, V. Narayanan, P. C. Elmendorf und G. Zheng, »Global Harmony: Coupled Noise Analysis for Full-Chip RC Interconnect Networks«. Proc. of the ICCAD, S. 139–146, 1997.
- [55] T. Stoehr, M. Alt, A. Hetzel und J. Koehl, »Analysis, Reduction and Avoidance of Crosstalk on VLSI Chips«. Proc. of the International Symposium on Physical Design 1998, S. 211–218, 1998.
- [56] M. Ringe, T. Lindenkreuz und E. Barke, »Static Timing Analysis Taking Crosstalk into Account«. Proc. of the DATE 2000, S. 451–455, 2000.

- [57] R. Arunachalam, K. Rajagopal und L. T. Pileggi, »TACO: Timing Analysis With COupling«. Proc. of the 37th ACM/IEEE DAC, S. 266–269, 2000.
- [58] F. Dartu und L. T. Pileggi, »TETA: Transistor-Level Engine for Timing Analysis«. Proc. of the 35th ACM/IEEE DAC, S. 595–598, 1998.
- [59] E. Acar, F. Dartu und L. T. Pileggi, »TETA: Transistor-Level Waveform Evaluation for Timing Analysis«. IEEE Transactions on Computer-Aided Design, Bd. 21, Nr. 5, S. 605–616, Mai 2002.
- [60] R. Arunachalam, R. D. Blanton und L. T. Pileggi, »False Coupling Interactions in Static Timing Analysis«. Proc. of the 38th ACM/IEEE DAC, S. 726–731, 2001.
- [61] D. Bryan, »The ISCAS '85 Benchmark Circuits and Netlist Format«. Entwicklungsbericht 919-248-1432, MCNC, September 1988.
- [62] Y. Li, R. Murgai, T. Miyoshi und A. Verma, »XTalkDelay: A Crosstalk-Aware Timing Analysis Tool for Chip-Level Designs«. Proc. of the ICCD, S. 208–215, Oktober 2004.
- [63] I. Keller, K. Tseng und N. Verghese, »A robust cell-level crosstalk delay change analysis«. Proc. of the ICCAD, S. 147–154, November 2004.
- [64] J. Vlach und K. Singhal, »Computer Methods for Circuit Analysis and Design«. Van Nostrand Reinhold, 1983.
- [65] L. W. Nagel, »SPICE2: A Computer Program to Simulate Semiconductor Circuits«. Entwicklungsbericht ERL-M520, University of California, Mai 1975.
- [66] E. H. Horneber, »Simulation elektrischer Schaltungen auf dem Rechner«. Springer-Verlag, 1985.
- [67] H. Wupper, »Rechnerunterstützter Schaltungsentwurf für die Großintegration«. Vorlesungsskript, Universität Duisburg, 1986.
- [68] A. Devgan, »Accurate Device Modeling Techniques for Efficient Timing Simulation of Integrated Circuits«. Proc. of the ICCD, S. 138–143, 1995.
- [69] D. Blaauw, V. Zolotov, S. Sundareswaran, C. Oh und R. Panda, »Slope Propagation in Static Timing Analysis«. Proc. of the ICCAD, S. 338–343, November 2000.

- [70] D. Blaauw, V. Zolotov, S. Sundareswaran und C. Oh, »Slope Propagation in Static Timing Analysis«. IEEE Transactions on Computer-Aided Design, Bd. 21, Nr. 10, S. 1180–1195, Oktober 2002.
- [71] F. Brglez, D. Bryan und K. Kozminski, »The ISCAS '89 Benchmark Circuits and Netlist Format«. Entwicklungsbericht, MCNC, Oktober 1989.

Index

- π -Modell, 31
- Abschaltbereich
 - eines Transistors, 32
- Adjust, 16
- Admittanz, 31
- Aggressor, 42
- Algorithmus
 - nichtdeterministischer, 7
- Alignment, 42
- Ankunftszeit, 15
 - notwendige, 16
- AT, 15
- Ausgangssituation, 28, 30
- BFS, 16
- Body-Effekt, 33, 36, 52
- Boolesche Algebra, 5
- Breitensuche, 16
 - Voraussetzungen, 19
- Charakterisierung, 29
- Chen-Verzögerungsmodell, 45
 - Pfadvalidierung, 92
- Datenpfad, 12
- Disjunktion, 5
- Early-Mode, 15
- effektive Kapazität, 38
- Eingangssituation, 28, 30
- Einschrittverfahren, 79
- Elmore-Verzögerung, 39
- Endpunkt, 4
- Ereignis, 8
- Erfüllbarkeit, 5, 7
- Erstes Verzögerungsmodell, 69
 - Pfadvalidierung, 87
 - Validierung, 73
- Extraktion, 11
- falscher Pfad, 22
- Flipflop, 9
- Gatter, 8
- Grad, 4
- Graph, 4
 - azyklisch, 4
 - gerichtet, 4
 - gewichtet, 4
- Holdzeit, 10
- Innenwiderstand, 71
- ISCAS-89, 83
- K-Faktor-Modell, 30, 37
- Kante, 4
- Knoten, 4
- Konjunktion, 5

- kritischer Pfad, 15
- Latch, 9
 - transparent, 10
- Late-Mode, 15
- Literal, 5
- Mehrschrittverfahren, 81
- Miller-Faktor, 44, 62
- Momentenanpassung, 40
- MOS-Transistor, 32
- Netz, 11
- Netzliste, 8
- nicht-kontrollierende Funktion, 6
- Normalform
 - disjunktive, 5
 - konjunktive, 5
- np-hartes Problem, 7
- NP-Problem, 7
- NP-Vollständigkeit, 7
- Opfer, 42
- Ordnung, 6
- Ordnungsreduktion, 40
- P-Problem, 6
- Parasiten, 1, 11
 - eines Transistors, 33
- Pfad, 4
 - falscher, 22
 - kritischer, 15
 - Länge, 5
- Primärausgang, 11
- Primäreingang, 11
- Produktform, 5
- Propagierung, 55
- Pulsweite, 10
- RAT, 16
- RC-Baum, 39
- Register, 9
- Rekonvergenz, 5
- Schaltermodell, 35
- Schaltfenster, 15, 47
- Schaltungsknoten, 8
- Schlinge, 4
- Schwellspannung
 - eines Transistors, 32, 66
 - für Kopplung, 66
- Setupzeit, 10
- Slack, 16
- Speicher, 9
- Startpunkt, 4
- Stromquellenbereich eines Transistors, 32
- Stromquellenverhalten
 - eines Transistors, 60
- Summenform, 5
- Switch-Level-Modell, 35
- Takt, 20
- Taktpfad, 12
- Tiefe, 4, 16
- Transistor
 - Abschaltbereich, 32
 - Gleichstromverhalten, 32, 52
 - Parasiten, 33
 - Stromquellenbereich, 32
 - Widerstandsbereich, 32, 62
- Transistormodell
 - α -Power-Law-, 36
 - BSIM3-, 33
 - BSIM4-, 33
 - Schaltermodell, 35

- Shockley-, 32
- Transparenz, 23
- Treiberstärke, 29

- Validierung, 1
 - erstes Verzögerungsmodell, 73
 - Pfadvalidierung des ersten Verzögerungsmodells, 87
 - Pfadvalidierung des zweiten Verzögerungsmodells, 88
 - zweites Verzögerungsmodell, 76
- Verifikation, 1
- Verzögerung, 3, 25
 - Elmore-, 39
- Verzögerungskante, 11
- Verzögerungsmodell
 - erstes, 69
 - Pfadvalidierung, 87
 - Validierung, 73
 - zweites, 70
 - Pfadvalidierung, 88
 - Validierung, 76
- Victim, 42

- Widerstandsbereich
 - eines Transistors, 32
- Widerstandsverhalten
 - eines Transistors, 62

- Zeitverhaltensgraph, 14
- Zweites Verzögerungsmodell, 70
 - Pfadvalidierung, 88
 - Validierung, 76

Lebenslauf

Matthias Ringe

16. April 1970 geboren in Minden in Nordrhein-Westfalen
- 1976 - 1980 Besuch der Grundschule Eisbergen
in Porta Westfalica
- 1980 - 1989 Besuch des Städtischen Gymnasiums
in Porta Westfalica
- 1989 Abitur
- 1989 - 1990 Zivildienst
- 1990 - 1995 Studium der Elektrotechnik an der Universität Hannover,
Studienrichtung Nachrichtentechnik
- 1995 Abschluss des Studiums als Diplom-Ingenieur
- 1995 - 1999 selbständig tätig
- seit 1999 Entwicklungsingenieur bei der
IBM Deutschland Entwicklung GmbH
in Böblingen

