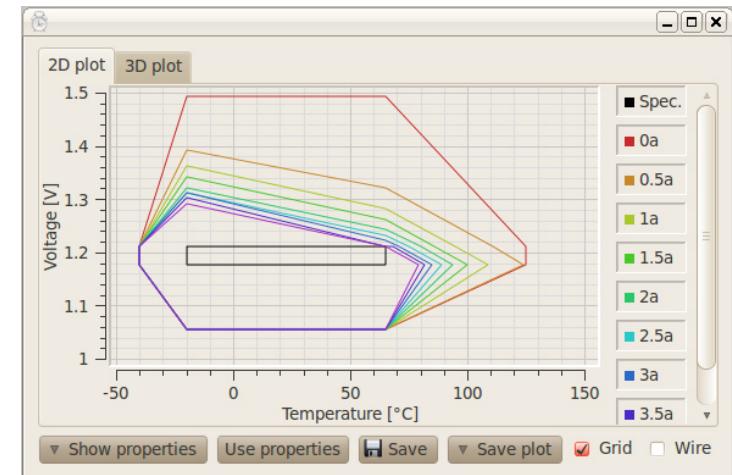
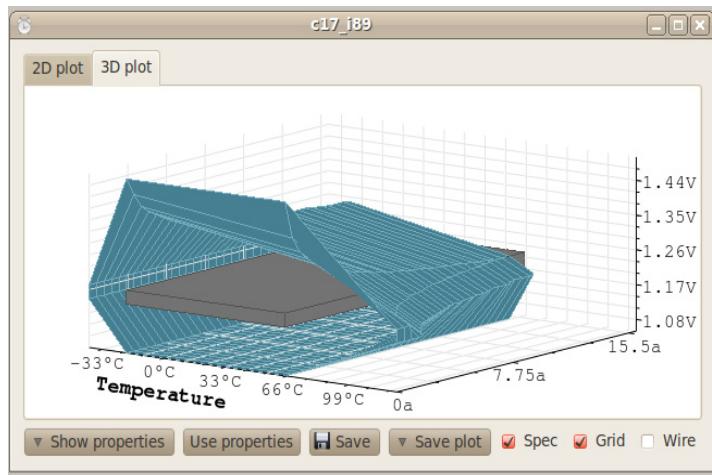


Einfluss von Alterungseffekten auf die Robustheit digitaler Schaltungen

Tutorial: Bewertung und Optimierung der Robustheit mikroelektronischer Systeme



Übersicht

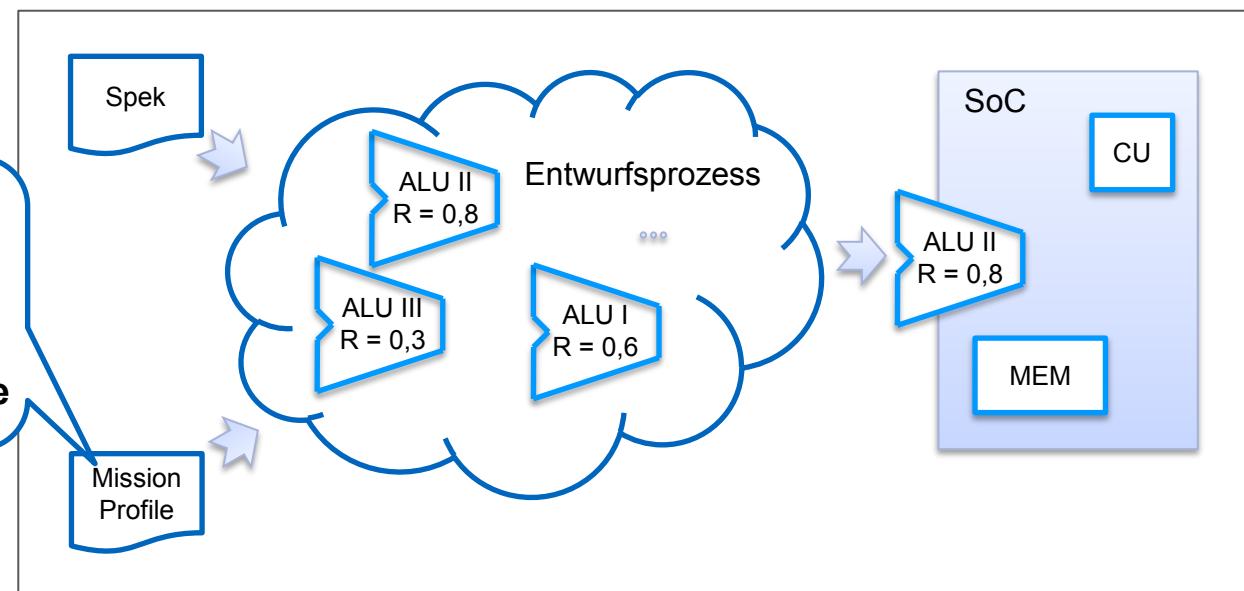
- Motivation
- Alterungsanalyse auf Gatterebene
- Alterungsmodell auf RT-Ebene
- Alterungsgewahre Robustheitsbestimmung
- Zusammenfassung

Übersicht

- Motivation
- Alterungsanalyse auf Gatterebene
- Alterungsmodell auf RT-Ebene
- Alterungsgewahre Robustheitsbestimmung
- Zusammenfassung

Warum lohnt es sich Robustheit zu messen?

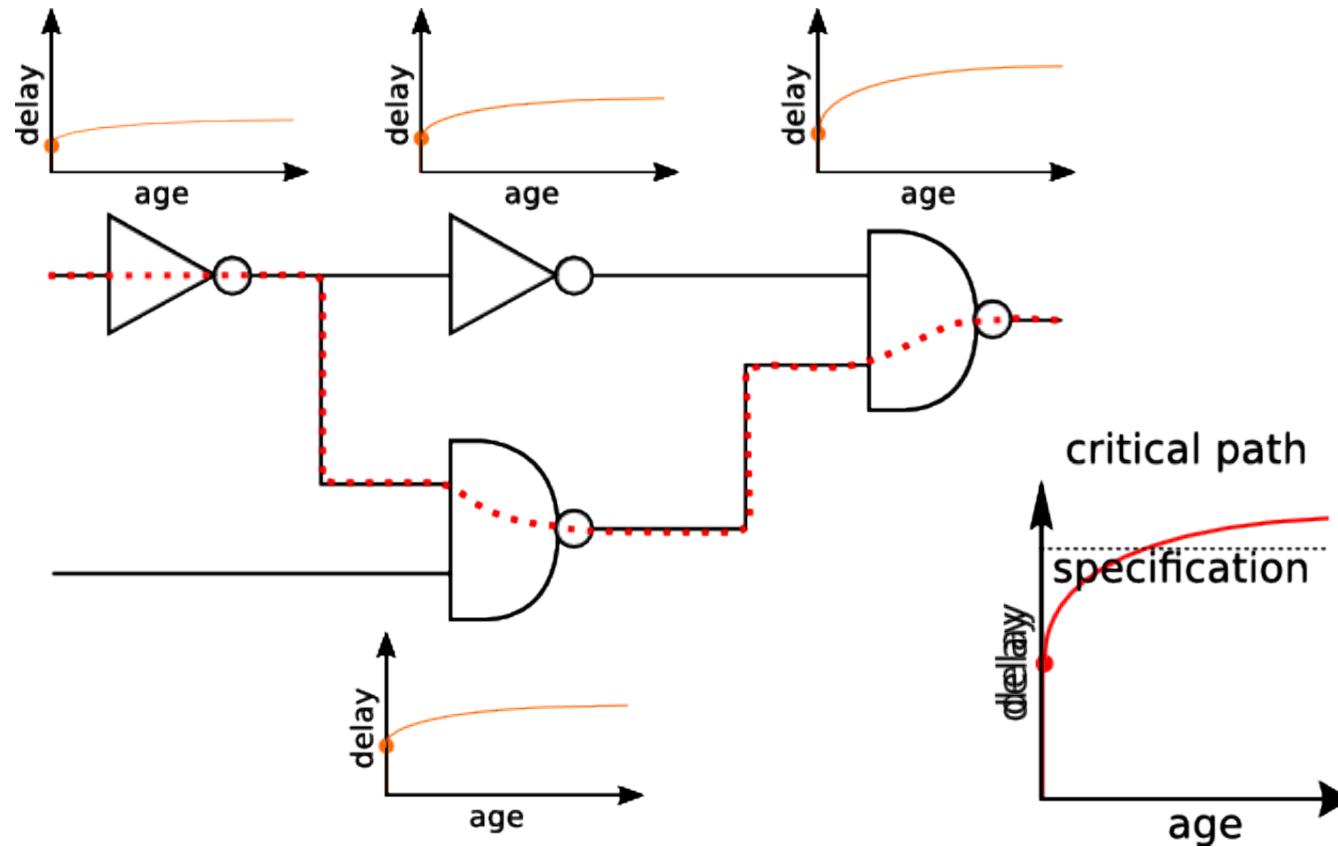
- Schaltungsdesigner haben die Möglichkeit je nach Spezifikation und Mission Profile zwischen verschiedenen Implementierungen einer Schaltung zu wählen



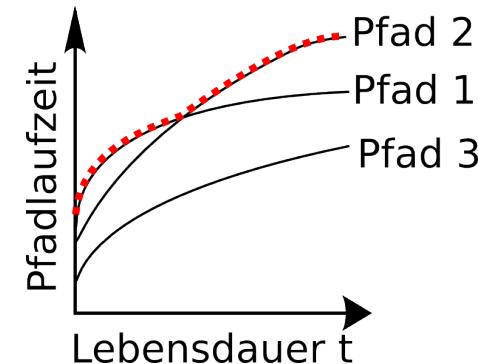
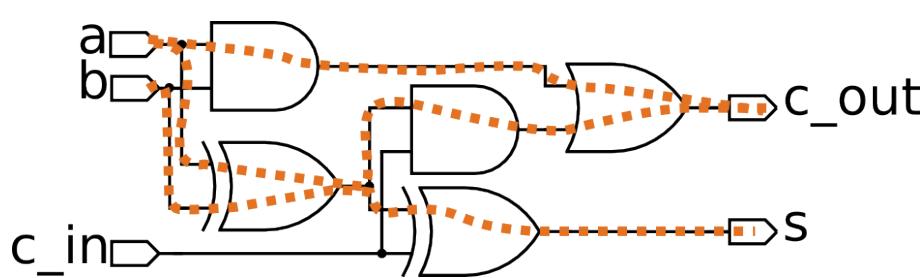
Übersicht

- Motivation
- Alterungsanalyse auf Gatterebene
- Alterungsmodell auf RT-Ebene
- Anpassung an der Robustheitskonzepte
- Zusammenfassung

Notwendigkeit einer Alterungsanalyse



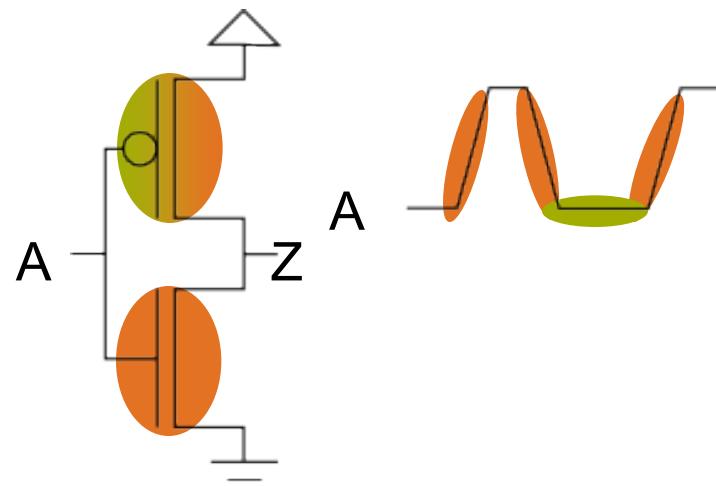
Zeitliche Verlagerung des kritischen Pfades



- Ohne Alterung: 1 kritischer Pfad
- Mit Alterung: mehrere mögliche kritische Pfade
- Abhängig von:
 - Operationsbedingungen über Lebenszeit
 - Arbeitslast (Eingangssignale über Lebenszeit)

Dominante driftbasierte Alterungseffekte

	<u>Negative Bias Temperature Instability (NBTI) *</u>	<u>Hot Carrier Injection (HCI)</u>
Transistortyp	PMOS	PMOS & NMOS
Modeliert durch	Schwellspannungsdrift ΔV_{th}	Degradation des Drain-sättigungsstroms ΔI_{on}
Stressbedingung	Transistor in Inversion	Transistor schaltet

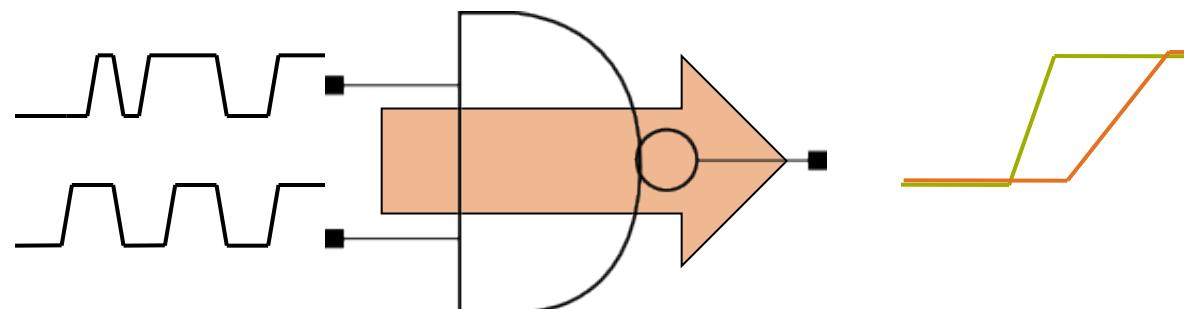
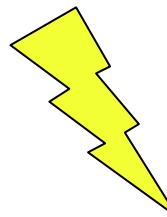


* Berücksichtigung von PBTI mit selben Ansatz leicht möglich, bei den von uns Verwendeten Technologien allerdings noch nicht relevant.

Anforderungen alterungsgewahres Timing-Modell

Use profile

- Operationsbedingungen im Laufe der Lebenszeit



Arbeitslast

- Signalwahrscheinlichkeit SP
- Übergangsdichte TD

Modellierte Effekte

- NBTI
- HCI

Gealterte

- ### Gattereigenschaften q_{aged}
- Gatterverzögerung
 - Ausgangsflankensteilheit

Alterungsgewahres Gattermodell AgeGate*

Kononisches Gattermodell

Technologiespezifische Degradationsgleichung

Strukturinformation

*“Aging Analysis of Circuit Timing Considering NBTI and HCI”, Lorenz et al., IEEE International On-Line Testing Symposium (IOLTS) 2009

Gealterte Gattereigenschaften für Driften

Kononisches Gattermodell

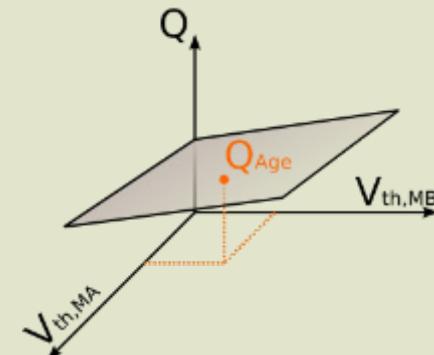
$$q_{aged} = q_{fresh} + \Delta q = q_{fresh} + \sum_{m \in G} \sum_{p \in P} \chi_{m,p}^q \cdot \Delta p_m$$

q_{fresh}, q_{aged} : fresh/aged gate performance

G, P : transistors of gate/aged parameters

χ : sensitivity coefficient

Δp : parameter drift



Technologiespezifische Degradationsgleichung

Strukturinformation

Driften durch Degradationsgleichung

Kononisches Gattermodell

Technologiespezifische Degradationsgleichung

- NBTI: $\Delta V_{th} = f(\mathbf{UP}, t_{\text{stress}}, W, L)$
- HCl: $\Delta I_{on} = f(\mathbf{UP}, t_{\text{stress}}, W, L)$

$$t_{\text{stress}} = P_{\text{stress}} \cdot t_{\text{life}}$$

UP : use profile

P_{stress} : stress probability

W, L : transistor sizes

Strukturinformation

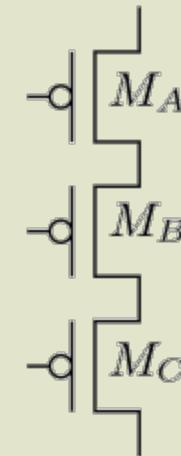
Strukturinformation für Driften benötigt

Kononisches Gattermodell

Technologiespezifische Degradationsgleichung

Strukturinformation

- Transistor-Größe
- Transistor-Stapel-Struktur



Berechnung der Stresswahrscheinlichkeit für NBTI

Stressbedingung für Transistor M : Beispiel:

- A: Logische “0” an Gate von M
- B: Logische “1” an Source oder Drain von M

$$\rightarrow M_A \text{ altert} \Leftrightarrow A \text{ auf } "0"$$

$$P_{\text{stress}} = 1 - SP_A = 0.5$$

$$\rightarrow P(B) = \prod_{t \in \text{PATH}_{NBTI}} P_{on,t}$$

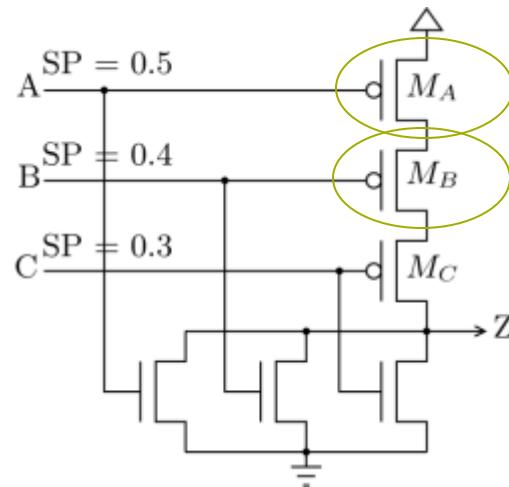
- M_B altert $\Leftrightarrow B \& A$ auf “0”

$$P_{\text{stress}} = 0.6 * 0.5 = 0.3$$

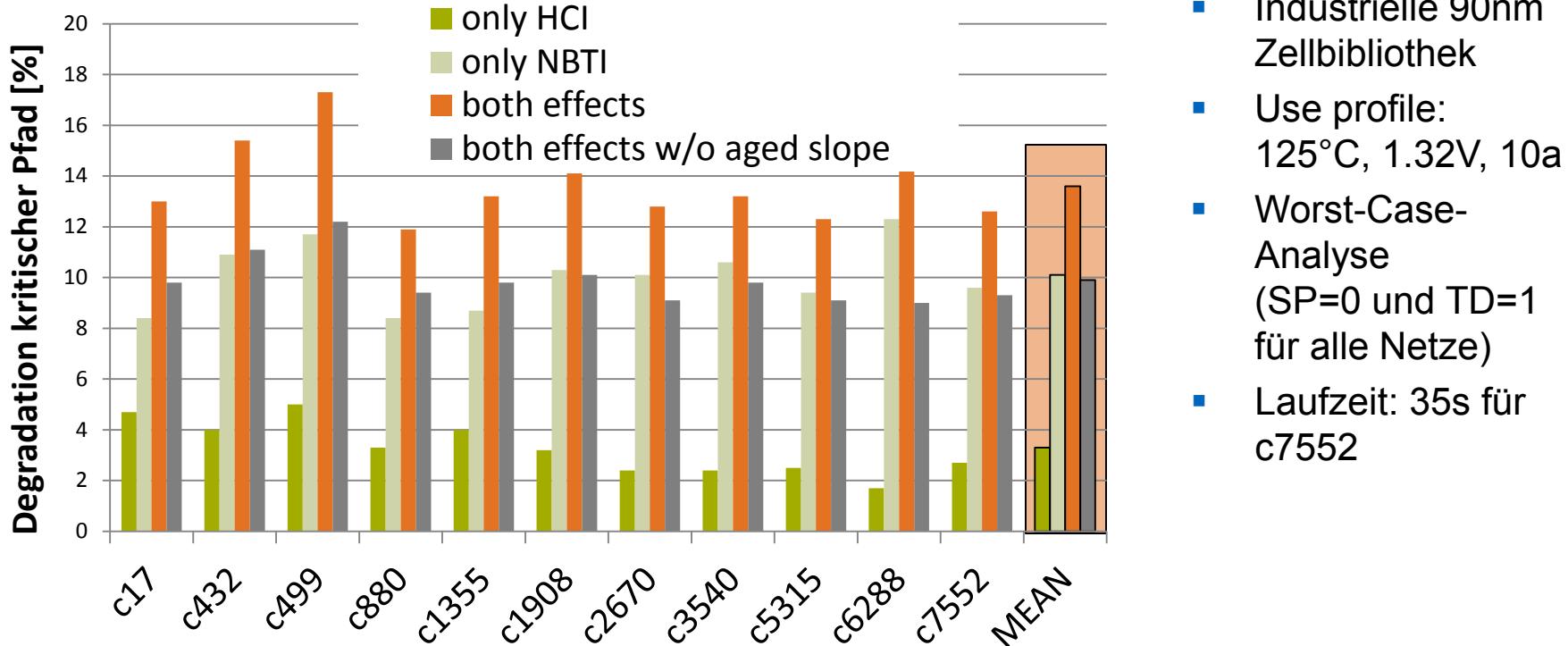
$$P_{\text{stress},NBTI} = P(A \wedge B)$$

$$= P(A) \cdot P(B)$$

NOR3



Degradation des kritischen Pfades für ISCAS'85 Benchmarkschaltungen



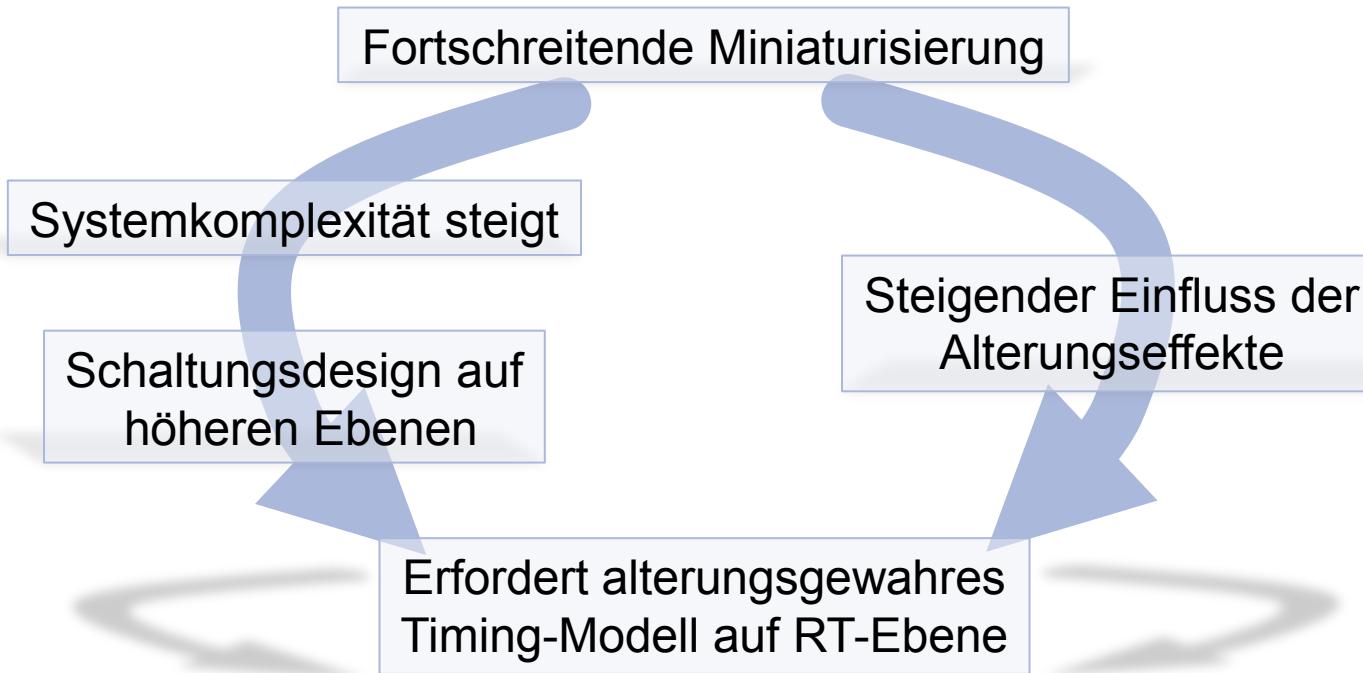
- Industrielle 90nm Zellbibliothek
- Use profile: 125°C, 1.32V, 10a
- Worst-Case-Analyse (SP=0 und TD=1 für alle Netze)
- Laufzeit: 35s für c7552

- Beide Effekte relevant
- Nichtberücksichtigung der gealterten Ausgangsflankensteilheit
→ 24% Unterschätzung

Übersicht

- Motivation
- Alterungsanalyse auf Gatterebene
- Alterungsmodell auf RT-Ebene
- Alterungsgewahre Robustheitsbestimmung
- Zusammenfassung

Motivation

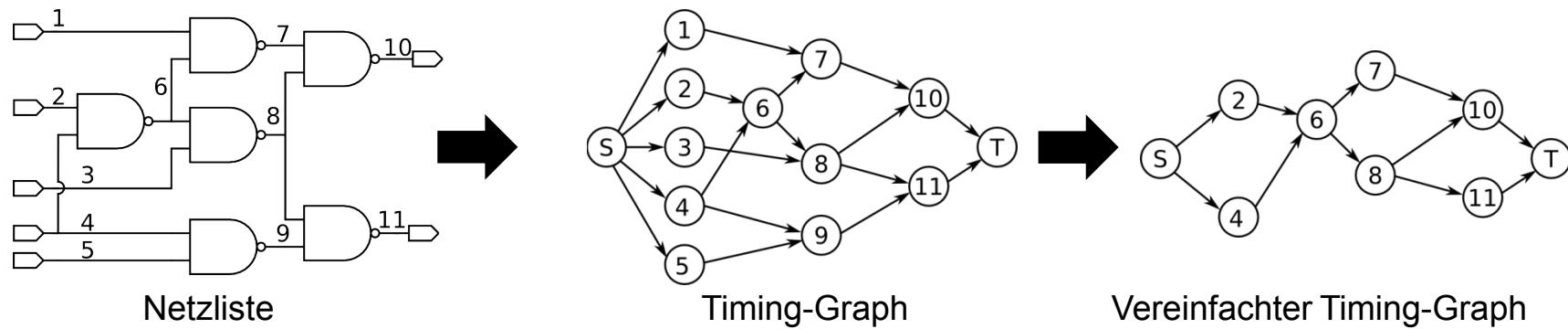


Solch ein Timing-Modell ermöglicht:

- Berücksichtigung des Einflusses von Alterung in früher Designphase
- Schnelle Performance Ermittlung auf Systemebene
- Robustheitsvalidierung

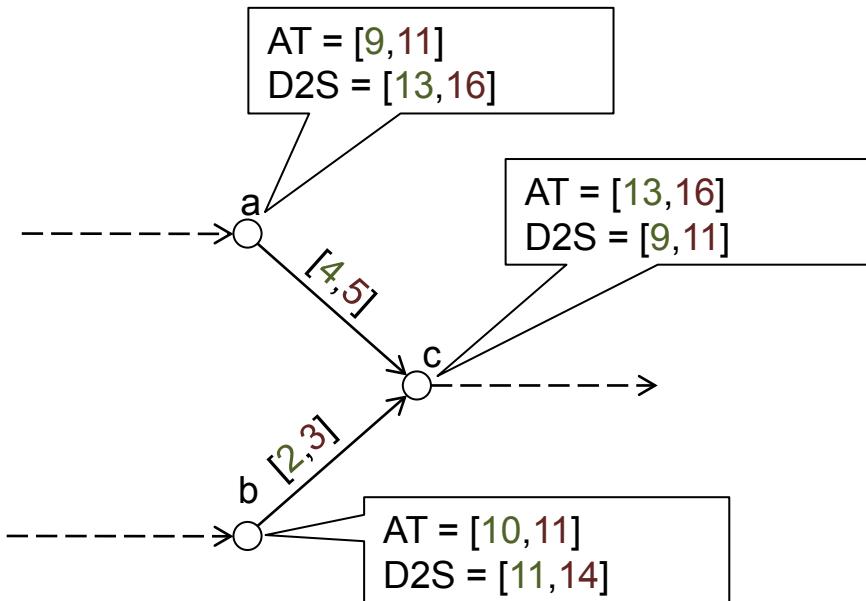
Vorgeschlagenes alterungsgewahres Timing-Modell

Idee: Vereinfachter Timing-Graph (TG), der nur noch **mögliche kritische Pfade** (MKP) enthält.



- Gleiche Genauigkeit wie Alterungsanalyse auf Gatterebene
- Speed-up durch vereinfachten TG

Reduktion des Timing-Graphen



AT = Ankunftszeit

D2S = Verzögerungszeit zum Ausgang

- Alle Timing-Eigenschaften als Intervalle $[d_0, d_{age,max}]$
- Verwendung von Intervallarithmetik
- Identifikation zu keinem MKP gehörender Elemente durch:
 1. Slackbasierte Reduktion
 2. Pfadlaufzeitbasiert Reduktion
 3. Laufzeitbasierte Reduktion einzelner Knoten
 4. Reduktion gemeinsamer Kanten
 5. Reduktion durch minimale Schaltungsalterung

Intervall-Arithmetik benötigt

- Statische Timing Analyse mit Intervallen

- Neudefinition der Summen- und Maximumsoperationen

$$\text{sum}([A_u, A_o], [B_u, B_o]) = [A_u + B_u, A_o + B_o]$$

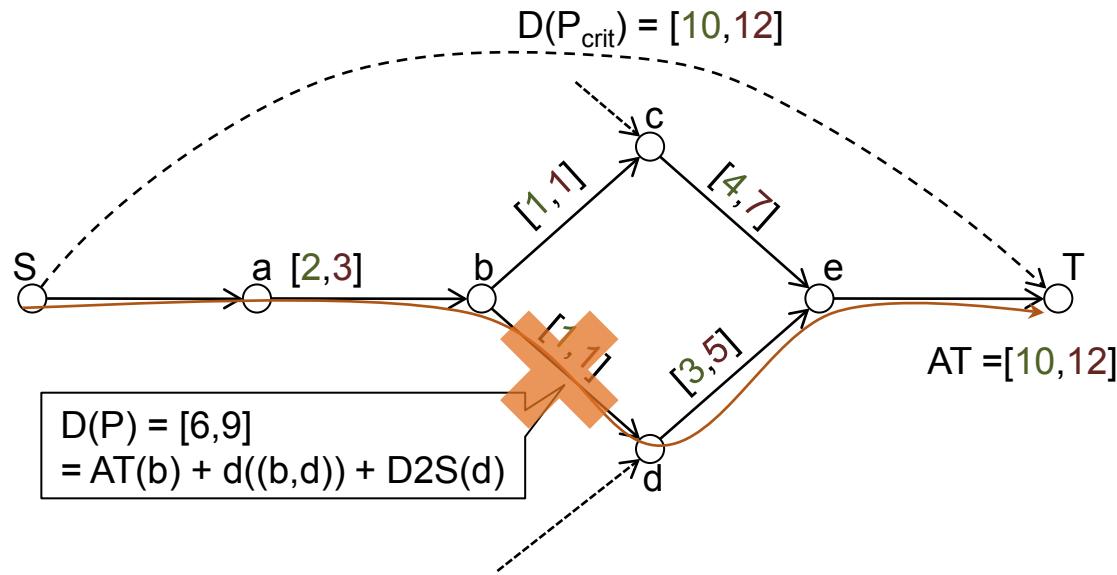
$$\text{max}([A_u, A_o], [B_u, B_o]) = [\max(A_u, B_u), \max(A_o, B_o)]$$

- Zusätzliche Vergleichsoperationen

$$[A_u, A_o] < [B_u, B_o] = A_o < B_u$$

$$[A_u, A_o] > [B_u, B_o] = B_o < A_u$$

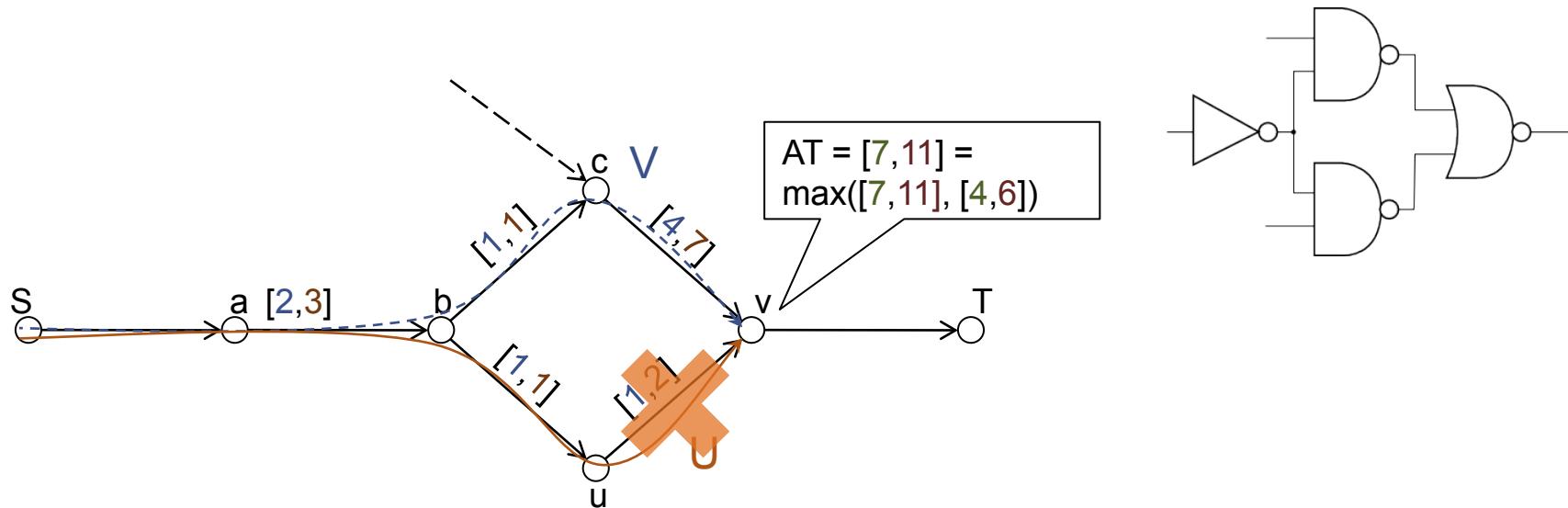
Pfadlaufzeitbasierte Reduktion



- Gealterte Laufzeit eines MKPs muss größer als die ungealterte Laufzeit des kritischen Pfades sein
- Pfadlaufzeit:

$$D(P) = AT(u) + d(u,v) + D2S(v)$$
- Wenn $D(P) < D(P_{crit})$: entferne Kante (u,v)

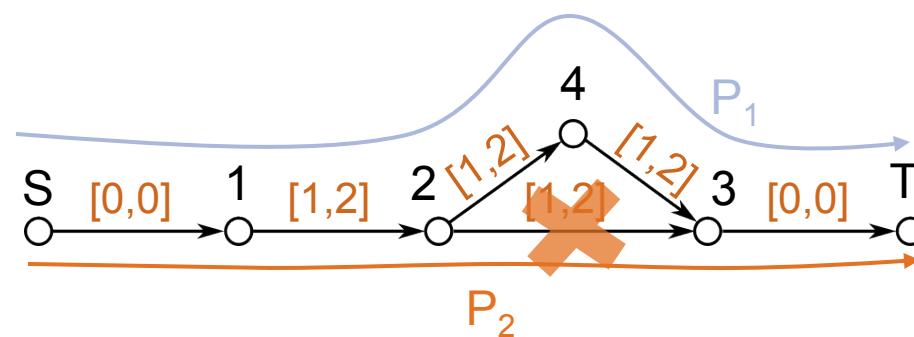
Ankunftszeitbasierte Reduktion einzelner Knoten



- Statische Timing Analyse (STA) mit Intervallen

Reduktion gemeinsamer Kanten

Idee: Überprüfung, ob der schnellere zweier rekonvergenter Pfade für alle Belastungsfälle schneller als der andere ist.



Ganzer Pfad

$$D(P_1) = [3, 6]$$

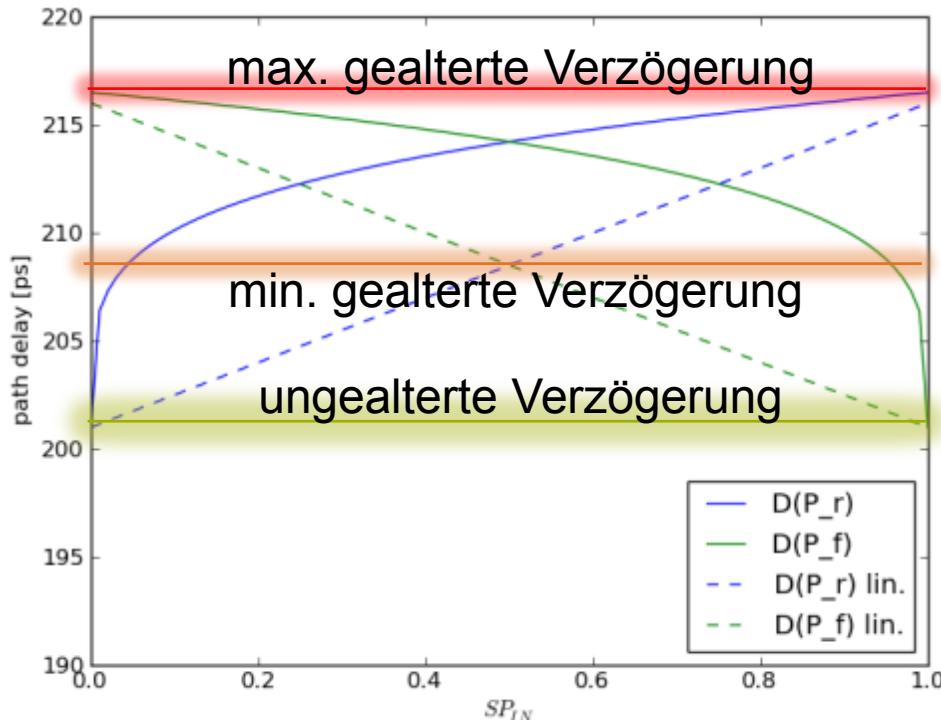
$$D(P_2) = [2, 4]$$

Nicht gemeinsame Teilpfade

$$D(P_1 \setminus P_2) = [2, 4]$$

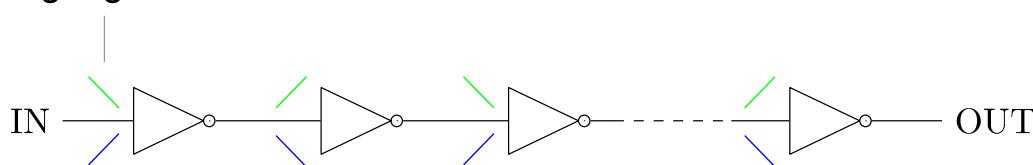
$$D(P_2 \setminus P_1) = [1, 2]$$

Reduktion durch minimale Schaltungsalterung



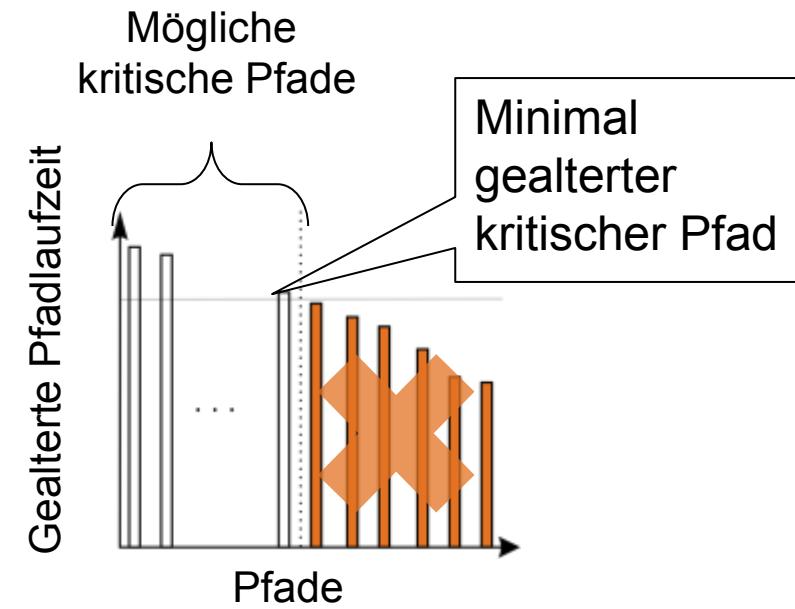
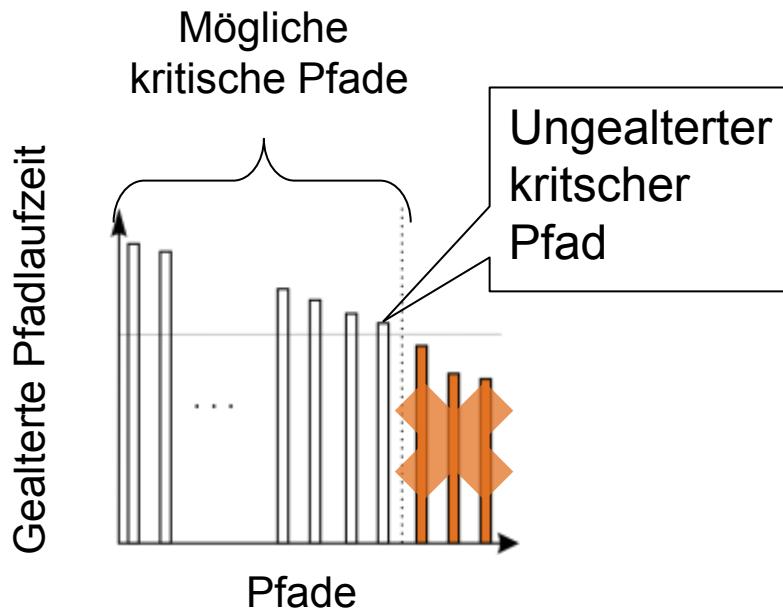
- Bei fast allen Schaltungen tritt unabhängig von der tatsächlichen Belastung eine Mindestalterung ein
- Inverter-Kette als einfaches Beispiel
- Lineares Optimierungsproblem zur Ermittlung der minimalen Schaltungsalterung

Eingangsflanke



Pfadbasierte Reduktion mit minimaler Schaltungsalterung

Idee: Verwendung der Verzögerungszeit des minimal gealterten kritischen Pfades.



Ergebnisse im Vergleich zum Stand der Technik*

Schaltung	#Gatter	#Pfade	#MKPs Stand der Technik*	#MKPs vorgestellter Ansatz	Reduktion
c499	534	452608	1487	375	4.0×
c1355	589	522368	3376	2224	1.5×
c2670	708	31286	21	21	1.0×
c3540	905	4248254	15276	1345	11.4×
c5315	1484	738816	1568	899	1.7×
c6288	2601	5.1e+16	6.8e+12	4.1e+12	1.6×
c7552	2242	448564	3173	522	6.1×

- #Gatter > 500
- ~25x speed-up gegenüber einer Alterungsanalyse auf Gatterebene
- #MKPs können um einen Faktor 3x reduziert werden

Setup: 90 nm Zellbibliothek; 125°C; 1.32V; 10a; nur NBTI

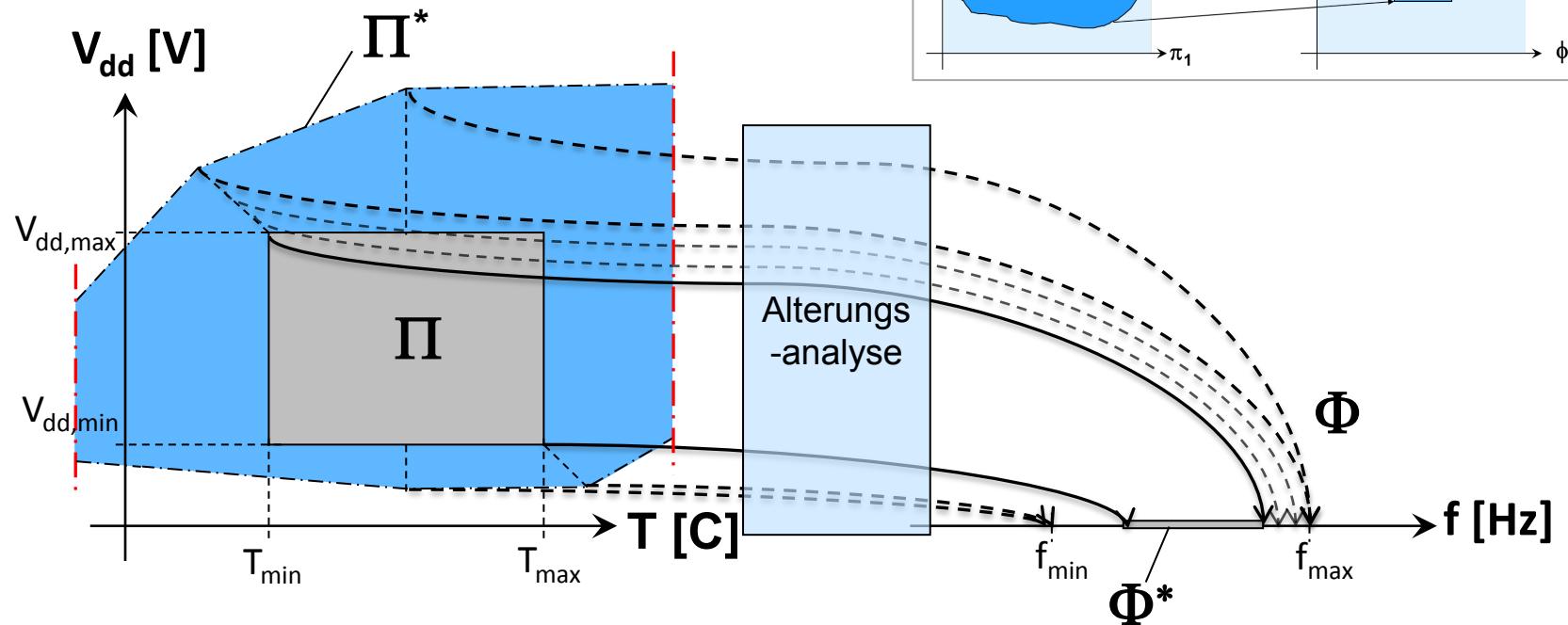
* “Testing for transistor aging”, Baba, VTS’09

Übersicht

- Motivation
- Alterungsanalyse (Gatterebene)
- Alterungsmodell auf RT-Ebene
- Alterungsgewahre Robustheitsbestimmung
- Zusammenfassung

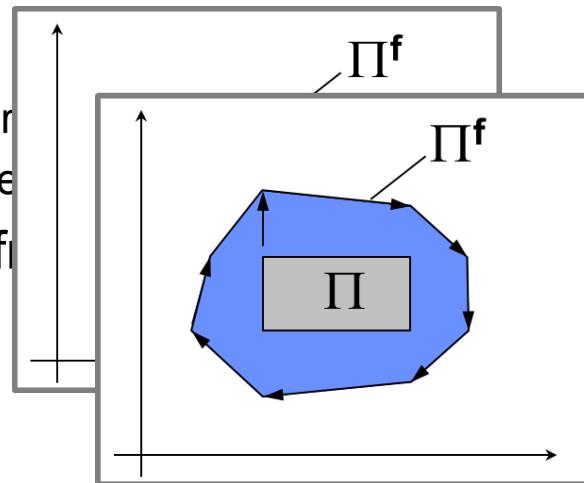
Robustheitsvalidierung unter Berücksichtigung von Alterungseffekten

- Eigenschaften Φ : Frequenz f
- Operationsbedingungen Π :
 - Temperatur T , Betriebsspannung V_{dd}

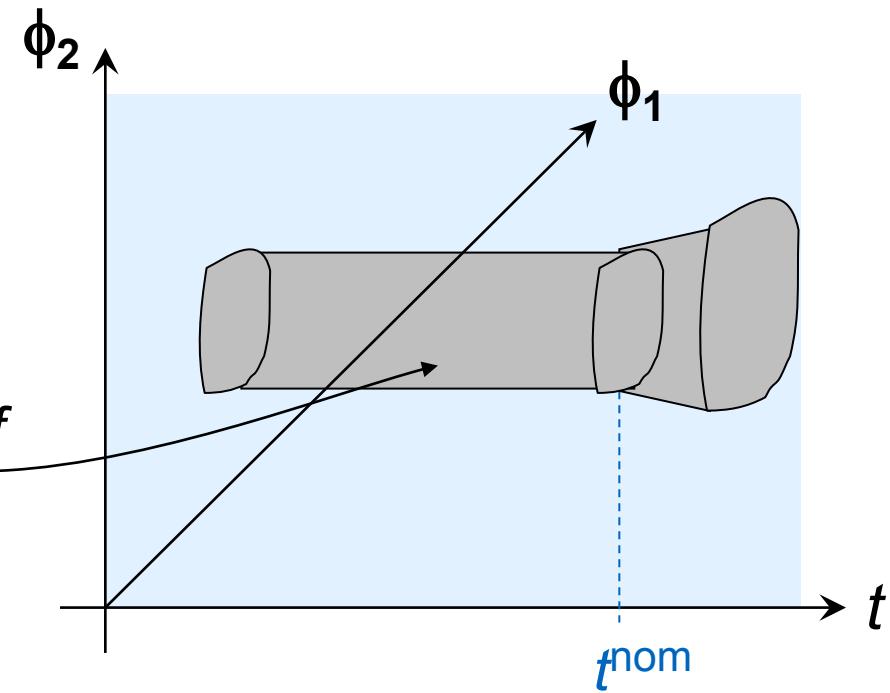
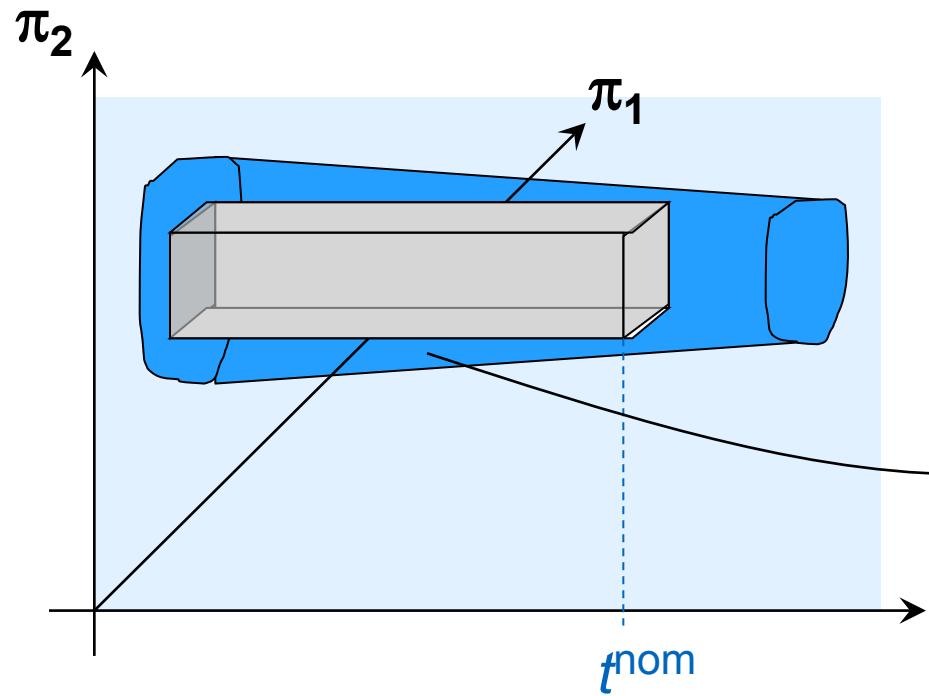


Robustheitsberechnung

- “StopSign” Ansatz (4 Ecken der Spezifikation \rightarrow 8 Punkte)
 - Feste Schrittänge
 - Bisektion/Dichotomy
- Gradientenbasiert
- Größte
 - Inner
 - Auße
- Paretof
- ...

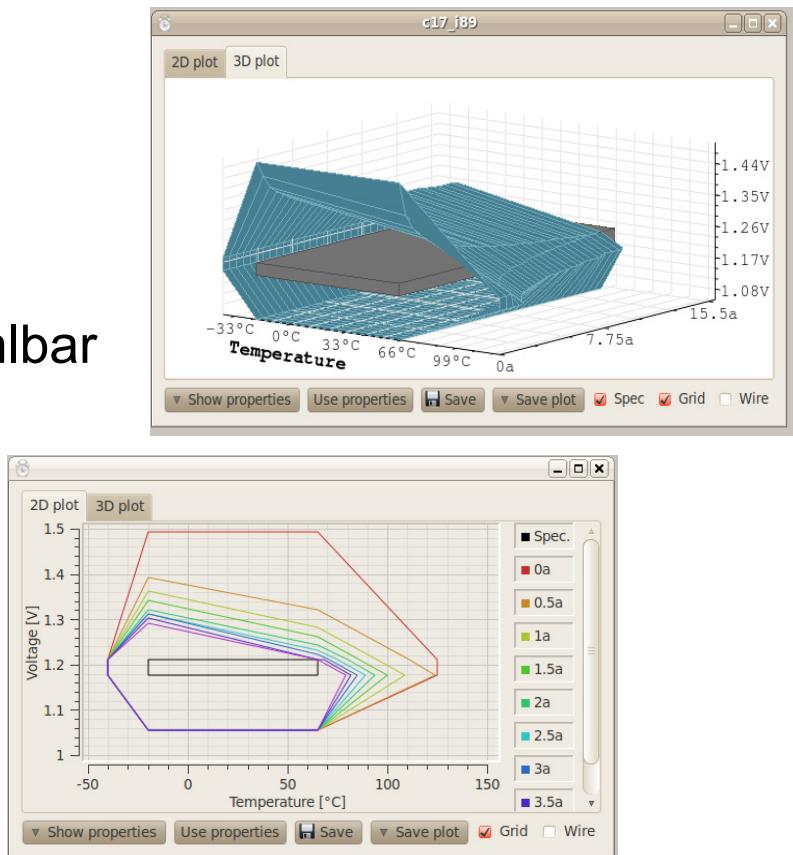


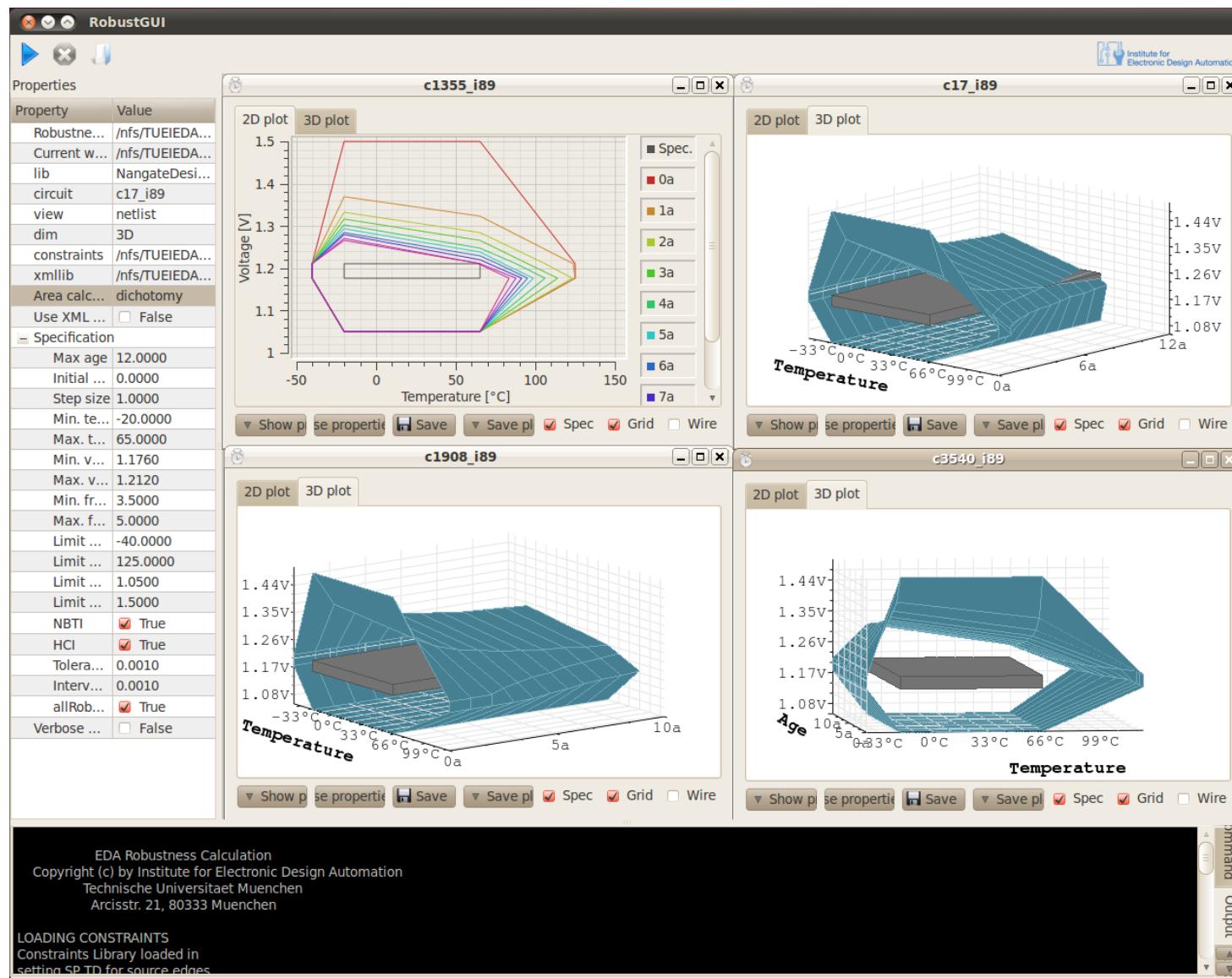
Zeitabhängigkeiten



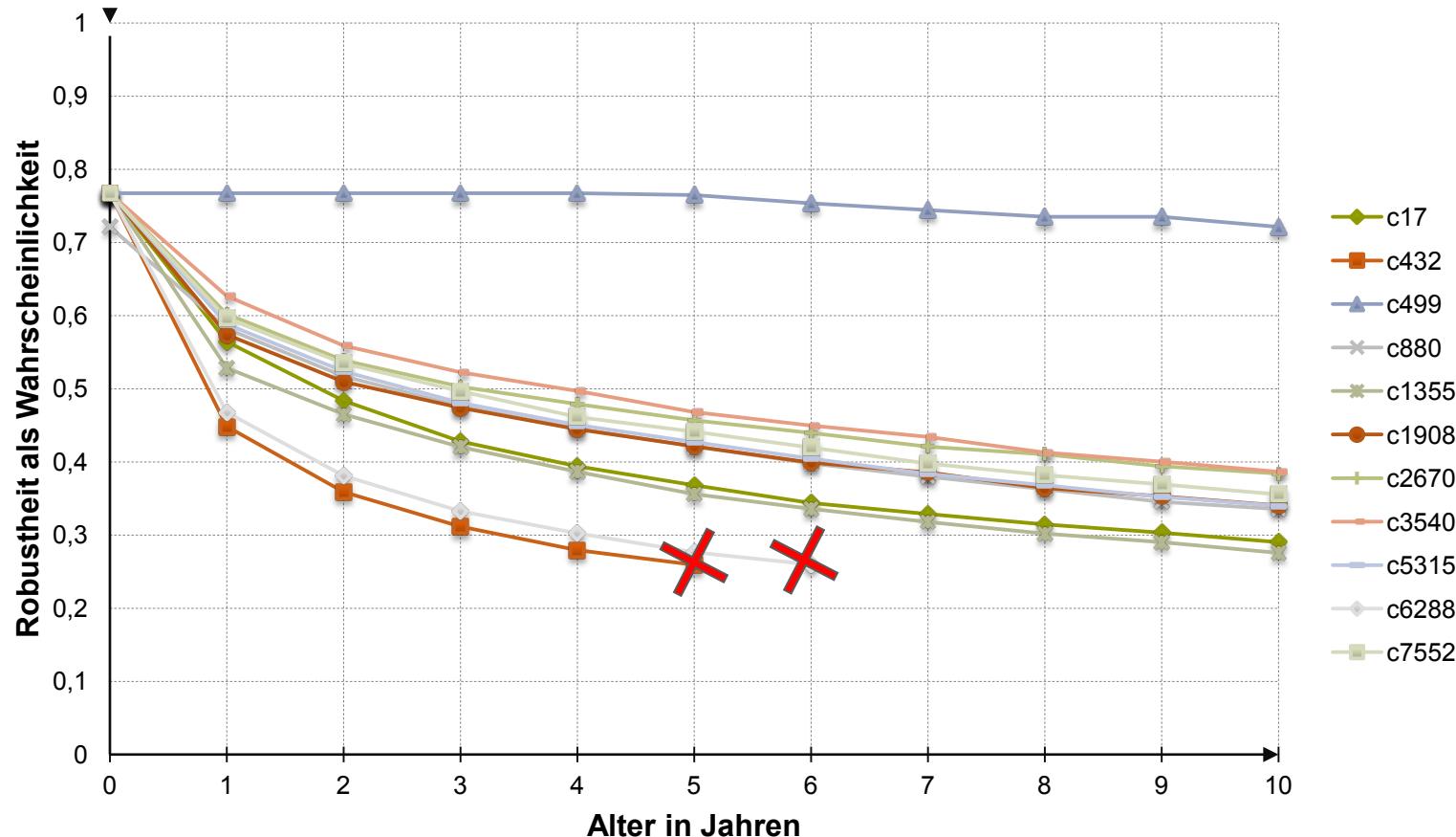
Robustheitsvalidierungs-GUI

- Berücksichtigung von NBTI und HCI (Erweiterung leicht möglich)
- Komandozeilenbasiert
- Benutzerfreundliche GUI
- Spezifikation/Mission Profiles frei wählbar
- ISCAS'85 Benchmarksschaltungen
- 2D/3D Option (Integration der Zeit)





Robustheitsdegradation durch Alterung



Setup: 90 nm Zellbibliothek; $-20^{\circ}\text{C} < T < 65^{\circ}\text{C}$; $1,17\text{V} < V_{\text{DD}} < 1,22\text{V}$; $-4\% < f < +5\%$

Übersicht

- Motivation
- Alterungsanalyse (Gatterebene)
- Alterungsmodell auf RT-Ebene
- Alterungsgewahre Robustheitsbestimmung
- Zusammenfassung

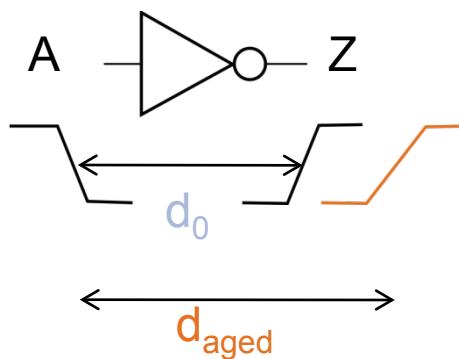
Zusammenfassung

- Alterungsanalyse durch alterungsgewahres Gattermodell
- Reduzierung der Sicherheitspuffer
- Vereinfachter TG mit allen möglichen kritischen Pfaden
- Gleiche Genauigkeit wie Timing Analyse auf Gatterebene
- Speed-up von bis zu 56x
- Robustheitsmessung macht Implementierungen vergleichbar
- Robustheitsmodell basierend auf Φ und Π Raum und einer Abbildung durch System F bietet eine sehr hohe Flexibilität
- Konkrete Eigenschaften ϕ_i und Operationsbedingungen π_i müssen für den jeweiligen Anwendungsfall definiert werden (Mission Profile)
 - Frequenz f
 - Temperatur T , Versorgungsspannung V_{dd}

BACKUP

Alterungsgewahres Gattermodell wird für Erzeugung/Auswertung des TM benötigt

Operationsbedingungen & Arbeitslast



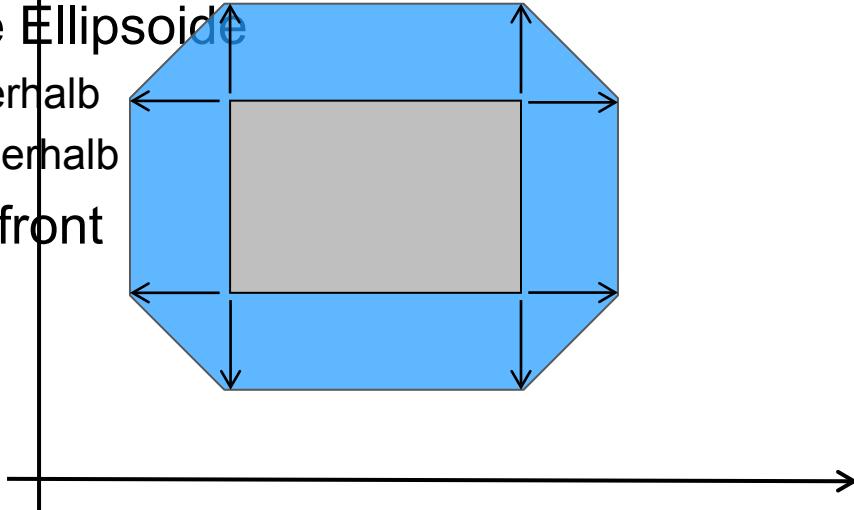
d_0 : ungealterte Gatterlaufzeit
 d_{aged} : gealterte Gatterlaufzeit



- Ansatz unabhängig von konkretem Gattermodell
- Intervall für Gatterlaufzeit, da tatsächliche Gatterlaufzeit unbekannt
- Gültigkeitsbereich:
 - Max. Lebensdauer
 - Max. Temperatur
 - Max. Versorgungsspannung

Robustheitsberechnung

- “StopSign” Ansatz
 - Feste Schrittänge
 - Bisektion/Dichotomy
- Gradientenbasiert
- Größte Ellipsoide
 - Innerhalb
 - Außerhalb
- Paretofront
- ...



Zusammenfassung

- Alterungsanalyse durch alterungsgewahres Gattermodell
- Vereinfachter TG mit allen **möglichen kritischen Pfaden**
- Gleiche Genauigkeit wie Timing Analyse auf Gatterebene
- Speed-up von bis zu 56x (25x im Mittel)
- Anzahl MKPs um weitere 3x reduziert gegenüber Stand der Technik
- Robustheitsmessung macht Implementierungen vergleichbar
- Robustheitsmodell basierend auf Φ und Π Raum und einer Abbildung durch System F bietet eine sehr hohe Flexibilität
- Konkrete Eigenschaften ϕ_j und Operationsbedingungen π_i müssen für den jeweiligen Anwendungsfall definiert werden (Mission Profile)
 - Frequenz f
 - Temperatur T , Versorgungsspannung V_{dd}