

**Ressourceneffiziente
Digitalschaltungen für den
Subschwellebetrieb**

Von der Fakultät für Elektrotechnik, Informatik und Mathematik
der Universität Paderborn

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

von

Dipl.-Ing. Sven Lütkemeier

Erster Gutachter: Prof. Dr.-Ing. Ulrich Rückert
Zweiter Gutachter: Prof. Dr.-Ing. Christoph Scheytt

Tag der mündlichen Prüfung: 14.05.2013

Paderborn 2013

Diss. EIM-E/231

Kurzfassung

Der Betrieb digitaler CMOS-Schaltungen im Subschwellbereich, d.h. bei Versorgungsspannung unterhalb der Schwellspannung der Transistoren, bietet eine Nische für Anwendungen, bei denen die Minimierung des Energie- oder Leistungsbedarfs das maßgebliche Entwurfskriterium darstellt, während Performanzanforderungen eher moderat und zweitrangig sind. Typischerweise kann der Energiebedarf einer Schaltung durch den Subschwellbetrieb um eine, der Leistungsbedarf um vier oder mehr Größenordnungen reduziert werden. Die Anwendungsdomäne von Subschwellschaltungen umfasst beispielsweise RFID, drahtlose Sensornetzwerke und biomedizinische Implantate. Der Entwurf einer Subschwellschaltung ist jedoch mit einer Reihe von Herausforderungen verknüpft, da der Einfluss von Prozess-, Spannungs- und Temperaturschwankungen wesentlich ausgeprägter ist als bei konventionellen Versorgungsspannungen.

Diese Arbeit befasst sich mit dem Entwurf integrierter Digitalschaltungen, die für einen ressourceneffizienten und robusten Betrieb im Subschwellbereich ausgelegt sind. Als Basis für die Implementierung beliebiger digitaler Systeme wird zunächst die Entwicklung zweier speziell für den Subschwellbetrieb optimierter Standardzellenbibliotheken in CMOS-Technologien mit Strukturgrößen von 90 nm und 65 nm beschrieben. Beide Bibliotheken gestatten eine dynamische Spannungsskalierung während des Betriebs bis hinauf zur nominellen Versorgungsspannung der jeweiligen Technologie. Basierend auf den Subschwell-Standardzellen werden zwei prototypische ASIC-Realisierungen vorgestellt, die die Vorzüge des Subschwellbetriebs belegen. Der erste integrierte Schaltkreis in der 90-nm-Technologie besteht aus vier 32-bit-ALUs. Der zweite ASIC in der 65-nm-Technologie enthält zwei Kopien eines vollständigen Subschwellprozessors mit einer 32-bit-Architektur und sechsstufiger Pipeline sowie spezielle Subschwell-SRAM-Blöcke. Dabei wird ein neuartiges System zur adaptiven Spannungs- und Frequenzskalierung unter Berücksichtigung von Prozess- und Betriebsparameterschwankungen eingesetzt. Dieser ASIC kann als Grundbaustein energieeffizienter eingebetteter Systeme, etwa für die eingangs genannten Anwendungen dienen. Im Subschwellbetrieb kann der Energiebedarf des Prozessors um den Faktor 11 gegenüber dem Betrieb bei konventioneller Versorgungsspannung reduziert werden.

Abstract

The operation of digital CMOS circuits in the subthreshold region, i.e. at supply voltage levels below the transistors' threshold voltage, offers a niche for applications where minimizing the energy or power consumption is the most important design criterion whereas performance requirements are moderate and of secondary importance. Typically, the energy consumption of a circuit can be reduced by one, the power consumption by four or more orders of magnitude during subthreshold operation. The domain of potential applications includes RFID, wireless sensor networks, and biomedical implants. The design of subthreshold circuits must, however, give rise to a number of challenges, as the influence of process, voltage, and temperature variation is much more pronounced compared to conventional supply voltages.

This thesis studies the design of integrated digital circuits aiming at a resource efficient and robust operation in the subthreshold region. In a first step, two standard cell libraries optimized for subthreshold operation in CMOS process technologies with feature sizes of 90 nm and 65 nm are presented as a basis for the implementation of arbitrary digital systems. Both libraries allow for a dynamic voltage scaling up to the nominal supply voltage of the respective technology. Based on these subthreshold standard cells, two prototypical ASIC implementations are presented, demonstrating the benefits of subthreshold operation. The first integrated circuit fabricated in the 90 nm process consists of four 32 bit ALUs. The second ASIC fabricated in the 65 nm process contains two instances of an entire subthreshold processor based on a 32 bit architecture with a six-stage pipeline as well as subthreshold SRAM blocks. A novel system providing adaptive voltage and frequency control under the influence of process and operating parameter variation is utilized. This ASIC may serve as a basic element for energy efficient embedded systems, for instance in conjunction with the applications named above. During subthreshold operation, the energy consumption of the processor can be reduced by a factor of 11 compared to a conventional supply voltage.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Verwandte Arbeiten	2
1.2	Übersicht über diese Arbeit	6

I Subschwelltechnik

2	Transistorverhalten im Subschwellbereich	11
2.1	Geometrieeffekte	19
2.2	Zusammenfassung	21
3	Digitale CMOS-Schaltungen im Subschwellbetrieb	23
3.1	Energetische Betrachtungen	23
3.2	Robustheit und Ausfallmechanismen	25
3.3	Grenzfallbedingungen	28
3.4	Zusammenfassung	29

II Standardzellen-Entwurf

4	Semi-Custom-Entwurf von Subschwellschaltungen	33
5	Robuste Subschwellbibliothek in 90 nm CMOS	35
5.1	Zellensatz	35
5.2	Dimensionierung der Logikgatter	37
5.2.1	Kombinatorische Gatter	38
5.2.2	Flipflop und Latch	41
5.3	Pegelwandler	42
5.3.1	Aufwärts-Pegelwandler	44
5.3.2	Abwärts-Pegelwandler	50
5.4	Layout	52

5.5	ALU-Prototyp	52
5.5.1	Architektur	54
5.5.2	Synthese und Layout	59
5.5.3	Messergebnisse	60
5.6	Zusammenfassung	63
6	Mehrzieloptimierte Subschwelligbibliothek in 65 nm CMOS	65
6.1	Zellendimensionierung	66
6.1.1	Mehrzieloptimierung	67
6.1.2	Ressourcenmaße	68
6.1.3	Algorithmen	75
6.1.4	Voruntersuchungen	77
6.2	Begrenzung der Transistor-Stacks	82
6.3	Kombinatorische Gatter	83
6.4	Flipflop und Latch	89
6.5	Taktbaum-Elemente	95
6.6	Pegelwandler	97
6.6.1	Aufwärts-Pegelwandler	97
6.6.2	Abwärts-Pegelwandler	105
6.7	Place-and-Route-Zellen	105
6.8	Layout	107
6.9	Zusammenfassung	108

III Adaptiver Subschwelligprozessor

7	Prototypische ASIC-Realisierung	113
7.1	CoreVA-Architektur	114
7.2	Rapid-Prototyping-System RAPTOR	117
7.2.1	ASIC-Evaluationsmodul DB-CoreVA	118
7.3	Konfiguration und Komponenten des ASICs	122
7.4	Entwurfsablauf	125
7.5	Zusammenfassung	128
8	Eigenschaften des Subschwelligprozessors	131
8.1	Simulations- und Messverfahren	131
8.2	Maximale Taktfrequenz	134
8.3	Energiebedarf	137
8.3.1	Vergleich unterschiedlicher Testprogramme	139

8.3.2	Vergleich der Bypass-Konfigurationen	140
8.4	Minimale Versorgungsspannung	141
8.5	Vergleich mit ähnlichen Implementierungen	144
8.6	Vergleich mit konventionellen Standardzellen	145
8.7	Zusammenfassung	147
9	Subschwellig-SRAM	151
9.1	Zusammenfassung	158
10	Takt- und Energieverwaltung	161
10.1	Hardware-Erweiterung PPMU	162
10.2	Adaptive Spannungs- und Frequenzsteuerung	166
10.2.1	Dynamische Eigenschaften der Regelschleife	174
10.2.2	Einfluss der Oszillator-Nichtlinearität	178
10.3	Zusammenfassung	181
11	Demonstrator	183
12	Zusammenfassung und Ausblick	187
	Abkürzungsverzeichnis	193
	Referenzen	197
	Eigene Veröffentlichungen	215
	Betreute Arbeiten	217

1 Einleitung

Der Betrieb digitaler CMOS¹-Schaltungen im Subschwelligbereich, d.h. bei Versorgungsspannung unterhalb der Schwellspannung der Transistoren, bietet eine Nische für Anwendungen, bei denen die Minimierung des Energie- oder Leistungsbedarfs das maßgebliche Entwurfskriterium darstellt, während Performanzanforderungen eher moderat und zweitrangig sind. Dabei ist es auch möglich, eine Schaltung temporär oberhalb des Subschwelligbereichs zu betreiben, wenn zeitweise eine höhere Rechenlast anfällt. Typischerweise kann der Energiebedarf einer Schaltung durch den Subschwelligbetrieb um eine, der Leistungsbedarf um vier oder mehr Größenordnungen reduziert werden. Die Anwendungsdomäne von Subschwelligschaltungen umfasst beispielsweise *Radio-Frequency Identification* (RFID) [164], drahtlose Sensornetzwerke [25, 31, 32, 101] und biomedizinische Implantate [42, 113]. Im Fall von RFID erfolgt die Energieversorgung elektromagnetisch durch die Radiowellen. Diese Art der Versorgung einer Schaltung wird als *Energy Harvesting* bezeichnet. Die dabei übertragene elektrische Leistung ist sehr gering. Andere Formen von *Energy Harvesting* nutzen etwa piezoelektrische, thermoelektrische oder photoelektrische Effekte zur Gewinnung geringer Energiemengen [123–125]. Bei Sensornetzen und Implantaten ist zwar eine Batterie- oder Akkumulator-basierte Stromversorgung möglich, allerdings kann nur eine relativ geringe Kapazität mitgeführt werden. In diesem Fall ist eine Maximierung der Zeitintervalle zwischen dem Austausch oder dem Nachladen von Batterien oder Akkumulatoren z.B. auf ein oder mehrere Jahre besonders wünschenswert. Daneben wurde der Subschwelligbetrieb als Schutzmaßnahme gegen gewisse Angriffsarten auf Kryptographie-Schaltungen vorgeschlagen [2, 4].

Die Ursprünge der Beschreibung des Transistorverhaltens im Subschwelligbereich reichen zurück bis in die 1950er Jahre [50]. Ebenso wurde bereits 1969 in [91] die Nutzung dieses Betriebsbereichs für eine integrierte Digitalschaltung beschrieben. Dennoch wurde der Subschwelligbetrieb von Digitalschaltungen bis etwa zur Jahrtausendwende in der Forschung praktisch ignoriert. Dies ist nicht zuletzt dem allgemein

¹ *Complementary Metal Oxide Semiconductor*

dominierenden Streben nach stetig steigenden Taktraten geschuldet. Mittlerweile gewinnt das Konzept zunehmend Beachtung als vielversprechender Lösungsansatz für die eingangs genannten Anwendungen.

Der Entwurf einer Subschwellschaltung ist mit einer Reihe von Herausforderungen verknüpft, da der Einfluss von Prozess-, Spannungs- und Temperaturschwankungen wesentlich ausgeprägter ist als bei konventionellen Versorgungsspannungen. Zentrale Forschungsschwerpunkte waren und sind daher insbesondere die Analyse derartiger Störeinflüsse auf das Schaltungsverhalten und die Entwicklung von Entwurfsvorschriften für eine entsprechend hohe Robustheit der Schaltungen, z.B. [11, 26–28, 56, 88, 95, 97, 107, 174], sowie die Modellierung der Abhängigkeiten des Energiebedarfs von Versorgungsspannung, Schwellspannung und Schalzhäufigkeiten, z.B. [23, 27, 93, 158]. Es wurden Schaltungen unterschiedlichster Komplexität untersucht und/oder gefertigt. Einige zu nennende Beispiele sind Volladdierer [53, 106], Mehr-Bit-Addierer [20], digitale Filter [62], Prozessoren [14, 58, 90, 136, 175] bis hin zu vollständigen SoCs¹ zur EEG²- oder EKG³-Detektion [65, 154] sowie ein Subschwellschwellschaltung⁴ [129]. Darüber hinaus gibt es Untersuchungen und Implementierungen, deren Fokus auf der Übergangsregion knapp oberhalb des Subschwellsbereichs (near-threshold) liegt, z.B. [45, 63, 172].

1.1 Verwandte Arbeiten

Moderne CMOS-Technologien erreichen für Subschwellschaltungen, u.a. je nach Versorgungsspannung und verwendeten Transistortypen, Taktfrequenzen im unteren Megahertz-Bereich, die für viele der oben genannten Anwendungen eine ausreichende Performanz bei stark reduziertem Energiebedarf bieten. Veröffentlichte Realisierungen basieren teilweise auf kommerziellen Standardzellen-Bibliotheken, teilweise jedoch auch, aus später diskutierten Gründen, auf speziell für diesen Zweck entworfenen Gattern. Nachfolgend sollen einige ausgewählte, repräsentative Realisierungen näher beschrieben werden, da sie insbesondere zur vergleichenden Bewertung des im Rahmen dieser Arbeit Erreichten herangezogen werden können. Zunächst folgt aufgrund der ähnlichen Ausrichtung dieser Arbeit eine

¹ *System on a Chip*

² *Elektroenzephalogramm*

³ *Elektrokardiogramm*

⁴ *Field Programmable Gate Array*

Übersicht über Implementierungen von Prozessoren und Mikrocontrollern für den Subschwellsbetrieb.

Die Veröffentlichungen [57, 58, 109] präsentieren einen 8-bit-RISC¹-Prozessor mit dreistufiger Pipeline in einer 130-nm-Technologie. Die beschriebene prototypische Realisierung enthält drei verschiedene Varianten des Prozessors mit jeweils unterschiedlicher Standardzellen-Dimensionierung (Minimaltransistoren, vergrößerte Kanalweiten sowie vergrößerte Kanalweiten und -längen). Die Balance zwischen NMOS- und PMOS-Treiberstärke wird durch eine Regelung der Substratspannung eingestellt. Latch- und Multiplexer-basierte Instruktions- und Datenspeicher besitzen Größen von ca. 0,19 kB und 0,125 kB. Der Prozessor ist noch bei einer Spannung von 160 mV funktionsfähig und weist eine minimale Energieaufnahme von 3,5 pJ pro Instruktion bei 350 mV und 354 kHz auf.

In [175, 176] wird ein 8-bit-CISC²-Prozessor mit zweistufiger Pipeline ebenfalls in einer 130-nm-Technologie veröffentlicht. Der Prozessor basiert auf einer kommerziellen Standardzellen-Bibliothek, wobei jedoch Zelltypen mit im Subschwellsbereich unsicherer Funktion entfernt wurden. Die Architektur mit dem Namen *Subliminal* verwendet einen optimierten Instruktionssatz sowie 0,25 kB Latch- und Multiplexer-basierten Speicher für Instruktionen und Daten. Die niedrigste Versorgungsspannung der untersuchten Exemplare beträgt im Mittel 378 mV. Das Energieminimum von 3 pJ pro Instruktion tritt bei 360 mV und ca. 1 MHz (je nach Exemplar) auf.

Ein weiterer 8-bit-Prozessor mit zweistufiger Pipeline wird in [55, 136] unter dem Namen *Phoenix* präsentiert. Der Baustein in einer 180-nm-Technologie enthält Instruktions- und Datenspeicher auf Basis einer 14T-Zelle mit Größen von 0,16 kB bzw. 0,25 kB. Die Hälfte des Instruktionsspeichers ist als Festwertspeicher (engl. *Read-Only Memory*, ROM) ausgeführt. Zur Reduktion der statischen Leistungsaufnahme während des Ruhezustands können inaktive Schaltungsteile von der Spannungsversorgung getrennt werden. Bei einer Versorgungsspannung von 500 mV beträgt die Taktfrequenz 106 kHz und der Energiebedarf 2,8 pJ pro Takt. Es werden keine Werte zur minimalen Versorgungsspannung oder Energieaufnahme angegeben.

In [65] wird ein 130-nm-SoC zur EKG-Detektion präsentiert, das neben analogen Schaltkreisen einen 8-bit-Prozessor mit zweistufiger Pipeline auf Basis der PIC16C5X-Architektur von Microchip enthält. Es kommt eine kommerzielle Standardzellen-

¹ *Reduced Instruction Set Computer*

² *Complex Instruction Set Computer*

Bibliothek mit reduziertem Zellensatz zur Verwendung. Es ist nicht bekannt, welche Speichertypen und -kapazitäten genutzt werden. Die minimale Versorgungsspannung des Prozessors beträgt 240 mV, der minimale Energiebedarf wird mit 1,51 pJ pro Takt bei 280 mV und 475 kHz angegeben.

Ein vollständiger 16-bit-Mikrocontroller auf der Grundlage der MSP430-Architektur von Texas Instruments wird in [89, 90] vorgestellt. Der 65-nm-Baustein enthält 16 kB Speicher auf Basis einer 8T-Bitzelle sowie einen Spannungswandler mit geschalteten Kapazitäten zur effizienten Erzeugung der benötigten Versorgungsspannung. Der Prozessorkern basiert auf einer speziellen Standardzellen-Bibliothek, die im Rahmen von Monte-Carlo-Simulationen auf eine festgelegte Fehlerwahrscheinlichkeit für alle Zelltypen bei einer bestimmten Versorgungsspannung ausgelegt ist. Die minimale Versorgungsspannung des Mikrocontrollers beträgt 300 mV. Die minimale Energieaufnahme von 27,2 pJ pro Takt tritt bei einer Versorgungsspannung von 500 mV und einer Taktfrequenz von 434 kHz auf.

In [14] wird ein weiterer 16-bit-Mikrocontroller auf MSP430-Basis in einer 65-nm-Technologie präsentiert. Die Schaltungsrealisierung enthält neben dem Prozessorkern 16 kB Instruktionen- und 2 kB Datenspeicher, wobei konventioneller 6T-SRAM¹ zum Einsatz kommt, der bei einer Spannung von 1 V betrieben wird. Für den Prozessor selbst wird eine kommerzielle Standardzellen-Bibliothek mit nicht-minimaler Kanallänge und reduziertem Zellensatz genutzt. Die Versorgungsspannung wird wie bei dem vorstehend beschriebenen Mikrocontroller durch einen in dem Baustein integrierten Spannungswandler mit geschalteten Kapazitäten generiert. Bei einer Spannung von 400 mV wird eine Taktfrequenz von 25 MHz bei einem Energiebedarf von 7 pJ pro Takt erreicht. Eine Aussage bezüglich minimaler Versorgungsspannung oder Lage der minimalen Energieaufnahme wird nicht getroffen.

Neben diesen Prozessor- und Mikrocontroller-Implementierungen sollen nun drei weitere relevante Schaltungen für Spezialanwendungen beschrieben werden.

In [159, 160] wird ein FFT²-Prozessor in einer 180-nm-Technologie vorgestellt, der bei Versorgungsspannungen bis hinab zu 180 mV betrieben werden kann. Es werden reellwertige FFT-Berechnungen mit Längen zwischen 128 und 1024 Punkten und Bitweiten von 8 bit und 16 bit unterstützt. Die Schaltung basiert auf einer speziellen Standardzellen-Bibliothek, die für eine minimale Versorgungsspannung unter ungünstigen Prozessparameter-Bedingungen ausgelegt ist. Neben dem FFT-

¹ *Static Random Access Memory*

² *Schnelle Fouriertransformation, engl. Fast Fourier Transform*

Datenpfad ist ein Latch- und Multiplexer-basierter RAM¹ mit einer Gesamtkapazität von 2 kB sowie ein ebenfalls Multiplexer-basierter 2 kB großer ROM zur Speicherung von Koeffizienten Teil der Implementierung. Für eine Auflösung von 16 bit und eine Länge von 1024 Punkten tritt ein Minimum der Energieaufnahme von 155 nJ pro FFT-Berechnung bei einer Spannung von 350 mV und einer Taktfrequenz von 10 kHz auf.

In [146] wird ein digitaler Basisbandprozessor für Ultrabreitband-Übertragungen vorgestellt. Die Schaltung basiert auf einer kommerziellen 90-nm-Standardzellen-Bibliothek. Die hochgradig parallele Architektur mit 620 Korrelatoren bietet einen Durchsatz von 500 Millionen Abtastwerten pro Sekunde und eine Datenrate von 100 Mbit s⁻¹ bei einer Versorgungsspannung von 400 mV und einer Taktfrequenz von 25 MHz. Die Energieaufnahme beträgt 20 pJ pro dekodiertem Bit.

Ein JPEG²-Coprozessor in einer 65-nm-Technologie wird in [120, 121] veröffentlicht. Je nach Versorgungsspannung und damit verfügbarer Performanz können verschiedene Kompressionsstandards verarbeitet werden. Die Implementierung basiert auf einer modifizierten kommerziellen Standardzellen-Bibliothek mit reduziertem Zellensatz. Des Weiteren kommt eine Regelung der Substratspannung zur Angleichung von NMOS- und PMOS-Schwelspannungen zum Einsatz. Die minimale Versorgungsspannung des Moduls zur diskreten Cosinus-Transformation und Quantisierung beträgt 400 mV. In diesem Betriebszustand wird die Kompression ruhender Bilder unterstützt, die Taktfrequenz beträgt 2,5 MHz und die Energieaufnahme 0,75 pJ pro Takt. Bei 700 mV können QXGA-Videos mit einer Auflösung von 2048 × 1536 Pixeln bei 15 Bildern pro Sekunde verarbeitet werden. Die Taktfrequenz beträgt dann 40 MHz und der Energiebedarf 2,2 pJ pro Takt.

Der in Teil III dieser Arbeit diskutierte Subschwellschwellprozessor hebt sich von den vorstehend genannten Prozessor- und Mikrocontroller-Realisierungen insbesondere durch folgende Punkte ab:

1. Die Standardzellen werden durch ein Mehrzieloptimierungsverfahren dimensioniert, das die simultane Berücksichtigung mehrerer Zielfunktionen, d.h. Eigenschaften der Zellen, während der Optimierung ermöglicht. Im Gegensatz zu bisher veröffentlichten Standardzellen-Bibliotheken für den Subschwellschwellbetrieb erlaubt dies die Berücksichtigung und Minimierung des Energiebedarfs bereits während der Entwurfsphase der Standardzellen.

¹ *Random Access Memory*

² Joint Photographic Experts Group, Verfahren zur digitalen Bildkompression

2. Die Architektur des Prozessorkerns weist gegenüber bisherigen Implementierungen eine deutlich höhere Komplexität auf, so dass die mit dem Subschwellbetrieb einhergehende Reduktion der Performanz abgemildert wird. Als Beispiele seien die Datenwortbreite von 32 bit, der sechsstufige Pipeline-Aufbau, die Multiplikations-Akkumulier- und Divisionsschritt-Einheiten sowie der SIMD-Modus genannt.
3. Der Einfluss von Prozess- und Betriebsparameter-Schwankungen wird durch eine adaptive Spannungs- und Frequenzsteuerung kompensiert. Hierdurch können notwendige Sicherheitsreserven etwa bezüglich der maximalen Taktfrequenz reduziert und dadurch die Energieeffizienz des Prozessors erhöht werden. Gegenüber einer Regelung der Substratspannung bietet die hier eingesetzte Methode den Vorteil eines geringeren Aufwands und damit niedrigerer Kosten. Zudem fällt keine zusätzliche Verlustleistung durch einen Substratspannungsregler an.

1.2 Übersicht über diese Arbeit

Die vorliegende Arbeit gliedert sich nachfolgend in drei Teile. In Teil I werden die Grundlagen der Subschwelltechnik erläutert. Dazu erfolgt in Kapitel 2 zunächst eine Beschreibung des Transistorverhaltens im Subschwellbereich. Dabei wird auch auf eine Reihe von Effekten eingegangen, die eine spezielle Rolle in diesem Transistorbetriebsbereich spielen. Zusätzlich werden grundlegende Eigenschaften der im Rahmen dieser Arbeit eingesetzten CMOS-Prozesse von STMicroelectronics mit Strukturgrößen von 90 nm und 65 nm im Subschwellbereich betrachtet. Kapitel 3 befasst sich mit den Besonderheiten des Subschwellbetriebs digitaler CMOS-Schaltungen. Dabei werden insbesondere energetische Betrachtungen und dominierende Ausfallmechanismen diskutiert.

Teil II dieser Arbeit befasst sich mit dem Entwurf zweier Standardzellen-Bibliotheken in den vorstehend genannten Technologien, die speziell für den Subschwellbetrieb ausgelegt sind. Zunächst wird in Kapitel 4 die Notwendigkeit einer solchen Standardzellen-Bibliothek zum Entwurf digitaler Subschwellschaltungen erläutert. Anschließend behandelt Kapitel 5 den Entwurf einer auf hohe Robustheit optimierten Gatterbibliothek in einer 90-nm-CMOS-Technologie. Das Kapitel diskutiert außerdem einen auf dieser Bibliothek basierenden und gefertigten Baustein mit vier 32-bit-

ALUs¹. Darauf aufbauend wird in Kapitel 6 die Entwicklung einer Subschwelligbibliothek in einer 65-nm-CMOS-Technologie durch ein Mehrziel-Optimierungsverfahren behandelt. Neben Beschreibungen des Optimierungsverfahrens sowie einzelner Entwurfsentscheidungen bezüglich des Zellsatzes wird eine neuartige, in der Bibliothek eingesetzte Pegelwandler-Schaltung vorgestellt, die einige Nachteile bisheriger Lösungen vermeidet.

Schließlich befasst sich Teil III mit der Implementierung eines vollständigen Subschwelligprozessors auf Basis der 65-nm-Subschwelligbibliothek als anwendungsspezifische integrierte Schaltung (engl. *Application Specific Integrated Circuit*, ASIC). Die prototypische Realisierung des Prozessors und der dazugehörigen Systemumgebung auf Basis des Rapid-Prototyping-Systems RAPTOR wird in Kapitel 7 vorgestellt. Die Simulations- und Messergebnisse des Prozessors werden in Kapitel 8 diskutiert. Ein Subschwellig-SRAM, der ebenfalls Teil des gefertigten ASICs ist und in Kooperation mit der Fachgruppe Nanoelektronik der Universität Oslo entwickelt wurde, wird in Kapitel 9 behandelt. Kapitel 10 befasst sich mit der Takt- und Energieverwaltung des Subschwelligprozessor-ASICs, das durch eine adaptive Spannungs- und Frequenzsteuerung die Besonderheiten des Subschwelligbetriebs berücksichtigt. Abschließend beschreibt Kapitel 11 einen Demonstrator, der die Vorteile des Subschwelligprozessors anhand einer realen Anwendung veranschaulicht. Kapitel 12 fasst die Ergebnisse dieser Arbeit zusammen.

¹ Arithmetisch-logische Einheit, engl. *Arithmetic Logic Unit*

Teil I

Subschwelltechnik

2 Transistorverhalten im Subschwelligebiet

In diesem Kapitel sollen die Grundlagen des Subschwelligebiet-Transistorbetriebs, auf dem die Erkenntnisse und Entwicklungen dieser Arbeit basieren, erläutert werden.

Abbildung 2.1 zeigt den (vereinfachten) Technologiequerschnitt eines selbstsperrenden n-Kanal-MOS¹-Transistors mit den Anschlüssen *Bulk*, *Source*, *Gate* und *Drain* [61]. Bulk ist der Substratanschluss und Gate die Steuerelektrode des Transistors. Zwischen Source (Quelle) und Drain (Senke) kommt es bei leitendem Transistor zu einem Stromfluss, dem *Drain-Strom*. Im thermodynamischen Gleichgewicht ohne angelegte Spannungen bilden sich um die n⁺-Gebiete herum Verarmungszonen aus, da an den n⁺-p-Übergängen durch Diffusion Elektronen ins p-Substrat sowie Löcher in die n⁺-Gebiete wandern und dort rekombinieren, so dass in diesen Bereichen kaum freie Ladungsträger vorhanden sind. Die zurückbleibenden Akzeptor- und Donatorionen bilden eine Raumladungszone, die auf n⁺-Seite ein positives und auf p-Seite ein negatives Vorzeichen besitzt. Durch die Raumladungszone entsteht eine Potentialdifferenz, die als Diffusionsspannung U_D bezeichnet wird, und damit ein elektrisches Feld, das die Elektronen zurück in die n⁺-Gebiete und die Löcher zurück ins p-Substrat treibt. Auf diese Weise besteht ein Gleichgewicht zwischen Diffusionsstrom und Driftstrom.

Da die Raumladungszonen an freien Ladungsträgern verarmt sind, kann beim Erhöhen der *Drain-Source-Spannung* U_{DS} nahezu kein Stromfluss zwischen Drain und Source stattfinden; der Transistor sperrt. Wird nun eine positive *Gate-Source-Spannung* U_{GS} angelegt, so wandern durch das entstehende elektrische Feld Minoritätsladungsträger, d.h. Elektronen, aus der Tiefe des p-Substrats an die Oberfläche unmittelbar unter dem Gateoxid, während gleichzeitig Majoritätsladungsträger, d.h. Löcher, durch das elektrische Feld von der Oberfläche verdrängt werden. Durch Rekombination entsteht so eine Verarmungsschicht an der Oberfläche. Ab einer

¹ Metall-Oxid-Halbleiter-Struktur, engl. *Metal Oxide Semiconductor*

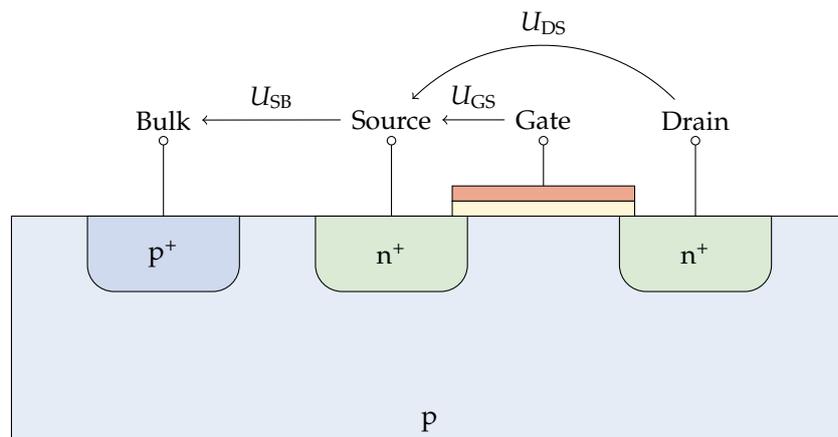


Abbildung 2.1: Technologequerschnitt eines selbst-sperrenden NMOS-Transistors

bestimmten Gate-Source-Spannung, nämlich der *Schwellspannung* U_{TH} des Transistors, befinden sich mehr Minoritätsladungsträger als Majoritätsladungsträger an der Oberfläche des Halbleiters, so dass sich ein n-leitender Kanal zwischen Drain und Source ausbildet. Man spricht in diesem Fall von *starker Inversion*. Bei angelegter Drain-Source-Spannung kann nun ein Strom fließen; der Transistor leitet.

In nicht entarteten Halbleitern gilt das Massenwirkungsgesetz [40]

$$n_0 p_0 = n_i^2 \quad (2.1)$$

wobei n_0 und p_0 die Konzentrationen der freien Elektronen bzw. Löcher sind und n_i die Eigenleitungsdichte des verwendeten Halbleiters ist. Aufgrund des multiplikativen Zusammenhangs sind also immer sowohl Elektronen als auch Löcher gleichzeitig vorhanden. Auf den NMOS-Transistor angewendet bedeutet dies, dass auch für $U_{GS} < U_{TH}$ freie Elektronen im Bereich unterhalb des Gates vorhanden sind, wenngleich kein ausgeprägter n-leitender Kanal besteht. Dieser Effekt wird bereits 1955 in [50] für die MIS¹-Struktur erläutert und als parabolischer Bereich bezeichnet. Bei Anlegen einer Drain-Source-Spannung kommt es durch Diffusionsprozesse somit zu einer Bewegung dieser freien Elektronen und dadurch zu einem geringen Stromfluss. Dieser Betriebsbereich des MOS-Transistors wird als *Subschwellbereich* bezeichnet. Man spricht auch von *schwacher Inversion* aufgrund des nicht vollständig ausgeprägten n-leitenden Kanals. Teilweise wird zur Abgrenzung des Übergangs zwischen schwacher und starker Inversion der Betrieb bei Gate-Source-Spannungen in der Nähe der Schwellspannung, also $U_{GS} \approx U_{TH}$, auch als *moderate Inversion* bezeichnet.

¹ Metall-Isolator-Halbleiter-Struktur, engl. *Metal Insulator Semiconductor*

Der Drain-Strom eines NMOS-Transistors im Subschwelligbereich kann durch

$$I_D = \frac{W}{L} I_{D0} \exp \frac{V_G}{nU_T} \left(\exp \frac{-V_S}{U_T} - \exp \frac{-V_D}{U_T} \right) \quad (2.2)$$

beschrieben werden [157]. Dabei sind V_G , V_S und V_D die Potentiale von Gate, Source und Drain bezogen auf das Substratpotential, I_{D0} ist ein charakteristischer Strom, n der sogenannte *Slope*-Faktor und $U_T = k_B T / q$ die Temperaturspannung. Für einen PMOS-Transistor sind die Vorzeichen aller Potentiale zu negieren. In dieser Form wird der Subschwelligstrom auch in Transistormodellen für Schaltungssimulationen berechnet, etwa beim BSIM4-Modell [108] oder beim EKV-Modell [17, 47]¹. Der charakteristische Strom I_{D0} ist nach dem EKV-Modell gegeben durch

$$I_{D0} = 2n\mu_c C'_{ox} U_T^2 \exp \frac{-U_{TH}}{nU_T} \quad (2.3)$$

wobei μ_c die Ladungsträgerbeweglichkeit, C'_{ox} die Gateoxid-Kapazitätsdichte und U_{TH} die Schwellspannung ist. Für den Slope-Faktor n gilt

$$n = 1 + \frac{C'_d}{C'_{ox}} \quad (2.4)$$

mit dem Kapazitätsbelag der Verarmungsschicht C'_d [157]. Anschaulich beschreibt n das Verhältnis eines kapazitiven Spannungsteilers, durch den das Oberflächenpotential des Halbleiters zwischen Drain und Source aus der Gate-Spannung hervor geht. Der Teilausdruck

$$\beta = \mu_c C'_{ox} \frac{W}{L} \quad (2.5)$$

aus Gleichungen 2.2 und 2.3 wird auch als Verstärkungsfaktor bezeichnet. Ebenso erfolgt in der Literatur gelegentlich die Ersetzung von $2n$ durch K .

Falls Source und Bulk auf dem selben Potential liegen, kann Gleichung 2.2 mit den gewohnten Spannungen U_{DS} und U_{GS} auch vereinfacht ausgedrückt werden als

$$I_D = \frac{W}{L} I_{D0} \exp \frac{U_{GS}}{nU_T} \left(1 - \exp \frac{-U_{DS}}{U_T} \right) \quad (2.6)$$

Die Gleichungen 2.2 bzw. 2.6 weisen mehrere Exponentialterme auf, die charakteristisch für den Subschwelligbereich sind. Diese exponentiellen Abhängigkeiten bewirken, dass bereits geringfügige Änderungen der Spannungen, der Temperatur oder der Prozessparameter einen starken Einfluss auf den Drain-Strom haben können. In der Folge hat dies ebenso Auswirkungen auf viele Schaltungseigenschaften wie etwa Verzögerungszeiten oder Verlustleistung.

¹ Das PSP-Modell verfolgt einen gänzlich anderen Ansatz und kommt ohne eine getrennte Modellierung von starker und schwacher Inversion aus [52].

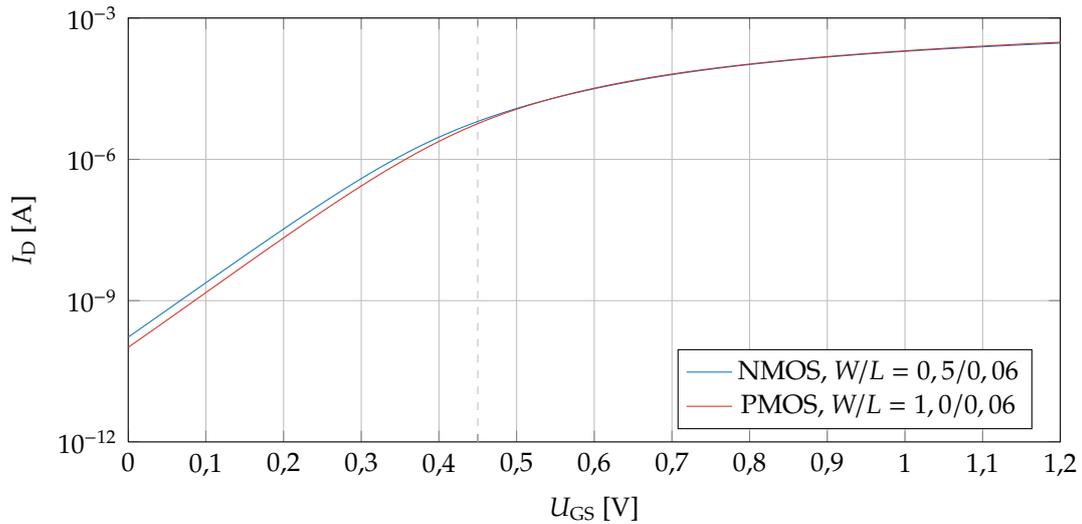


Abbildung 2.2: Transferkennlinien der 65-nm-Technologie (simuliert, $U_{DS} = 1,2\text{ V}$)

Spannungsabhängigkeit

Abbildung 2.2 zeigt die simulierten Transferkennlinien der NMOS- und PMOS-Transistoren der 65-nm-Technologie von STMicroelectronics in logarithmischem Maßstab. Für diese Simulation beträgt die Drain-Source-Spannung konstant 1,2 V und die Temperatur 300 K bzw. $U_T = 25,8\text{ mV}$. Die Transistormaße sind so gewählt, dass die Drain-Ströme von NMOS und PMOS etwa gleiche Größe besitzen. Die Schwellspannung beträgt etwa 450 mV. Unterhalb dieser Gate-Source-Spannung erkennt man deutlich den exponentiellen Charakter der Transferkennlinien. In diesem Bereich variiert der Drain-Strom über fünf Größenordnungen, während er sich oberhalb der Schwellspannung lediglich über zwei Größenordnungen ändert.

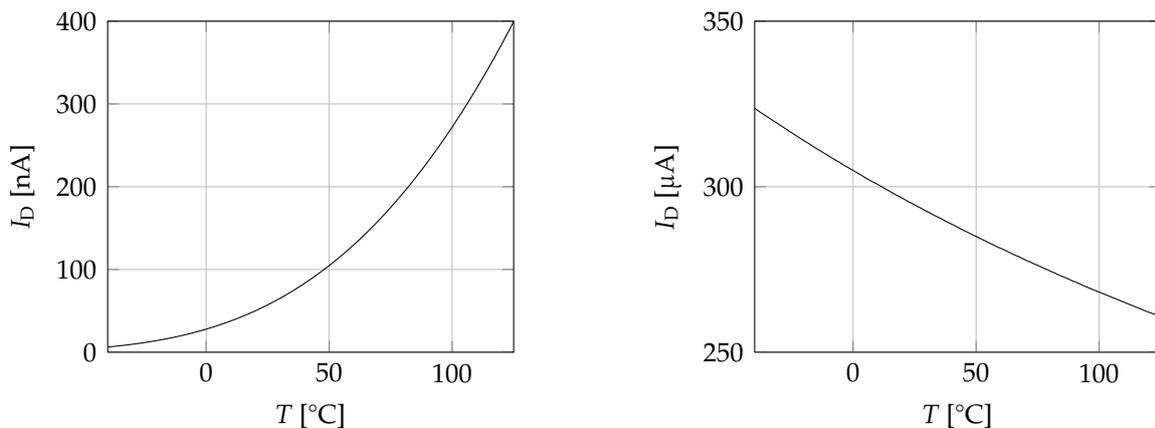
Aus den Kennlinien lassen sich die Transistorparameter I_{D0} und n aus Gleichung 2.6 durch zwei Messpunkte im Subschwelligbereich bestimmen. Seien $U_{GS1}, U_{GS2} < U_{TH}$. Es gilt:

$$\frac{I_D(U_{GS2})}{I_D(U_{GS1})} = \frac{\exp \frac{U_{GS2}}{nU_T}}{\exp \frac{U_{GS1}}{nU_T}} = \exp \frac{U_{GS2} - U_{GS1}}{nU_T} \quad (2.7)$$

$$\Rightarrow n = \frac{U_{GS2} - U_{GS1}}{U_T \ln \frac{I_D(U_{GS2})}{I_D(U_{GS1})}} \quad (2.8)$$

Technologie	Transistor	n	I_{D0} [pA]
65 nm Low Power	NMOS	1,586	20,18
	PMOS	1,540	6,14
90 nm General Purpose	NMOS	1,473	850,06
	PMOS	1,526	368,54

Tabelle 2.1: Transistorparameter n und I_{D0} der verwendeten Technologien von STMicroelectronics



(a) schwache Inversion ($U_{DS} = U_{GS} = 0,3$ V)

(b) starke Inversion ($U_{DS} = U_{GS} = 1,2$ V)

Abbildung 2.3: Temperaturabhängigkeit des Drain-Stroms (simuliert)

Nachdem n bekannt ist, kann außerdem I_{D0} bestimmt werden. Aus Gleichung 2.6 folgt:

$$I_{D0} = \frac{I_D(U_{GS2})}{\frac{W}{L} \exp \frac{U_{GS2}}{nU_T} \left(1 - \exp \frac{-U_{DS}}{U_T}\right)} \quad (2.9)$$

Mit den Messpunkten $U_{GS1} = 0$ V und $U_{GS2} = 0,4$ V ergeben sich die in Tabelle 2.1 angegebenen Werte. Zusätzlich sind die auf gleiche Weise ermittelten Transistorparameter der ebenfalls in dieser Arbeit verwendeten 90-nm-Technologie angegeben.

Temperaturabhängigkeit

Der Einfluss der Temperatur auf den Drain-Strom im Subschwelligbereich ist in Abbildung 2.3a dargestellt. Zusätzlich zeigt Abbildung 2.3b die Temperaturabhängigkeit bei starker Inversion. Im Vergleich der beiden Diagramme fallen zwei wesentliche Unterschiede auf. Zum einen ändert sich der Drain-Strom im betrachteten Tempe-

raturintervall von -40 °C bis 125 °C bei schwacher Inversion deutlich stärker als bei starker Inversion. Bezogen auf den Wert des Stroms bei Raumtemperatur von 25 °C beträgt die relative Änderung bei schwacher Inversion etwa 670 % und bei starker Inversion etwa 21 %. Zum anderen ist das Temperaturverhalten in den beiden Betriebszuständen entgegengesetzt zueinander, d.h. bei schwacher Inversion nimmt der Drain-Strom mit steigender Temperatur zu, während bei starker Inversion eine Abnahme des Stroms zu beobachten ist. Folgende Temperaturabhängigkeiten sind dabei ausschlaggebend [61]:

$$\mu_c(T) = \mu_c(300\text{ K}) \left(\frac{T}{300\text{ K}} \right)^{-a_n} \quad (2.10)$$

$$U_{\text{TH}}(T) = U_{\text{FB}} + 2\phi_F(T) + \gamma \sqrt{2\phi_F(T)} \quad (2.11)$$

$$\phi_F(T) = U_T \ln \frac{N_A}{n_i(T)} \quad (2.12)$$

$$U_T(T) = \frac{k_B T}{q} \quad (2.13)$$

$$n_i(T) = C T^{3/2} \exp \frac{-W_g}{2k_B T} \quad (2.14)$$

mit gewissen Konstanten a_n , U_{FB} , γ , N_A , k_B , q , C und W_g ¹. Es ergeben sich folgende Temperaturkoeffizienten (für μ_c und U_{TH} im Bereich der Raumtemperatur):

$$\frac{d\mu_c}{dT}(300\text{ K}) = -a_n \frac{\mu_c(300\text{ K})}{300\text{ K}} < 0 \quad (2.15)$$

$$\frac{dU_{\text{TH}}}{dT}(300\text{ K}) \approx -2\text{ mV K}^{-1} < 0 \quad (2.16)$$

$$\frac{d\phi_F}{dT} = \frac{1}{T} \left(\phi_F - \frac{W_g}{2q} \right) - \frac{3k}{2q} < 0 \quad , \text{ denn [61] } \frac{W_g}{2q} > \phi_F \quad (2.17)$$

$$\frac{dU_T}{dT} = \frac{k_B}{q} \approx 86,173\text{ }\mu\text{V K}^{-1} > 0 \quad (2.18)$$

$$\frac{dn_i}{dT} = C \frac{3k_B T + W_g}{2k_B \sqrt{T}} \exp \frac{-W_g}{2k_B T} > 0 \quad (2.19)$$

¹ a_n beträgt etwa 1,5 bis 2. U_{FB} ist die Flachbandspannung, γ der Substratsteuerfaktor, N_A die Akzeptordichte im Substrat, k_B die Boltzmann-Konstante, q die Elementarladung und W_g der Bandabstand. Es sei bemerkt, dass auch für W_g eine gewisse Temperaturabhängigkeit gilt.

Bei schwacher Inversion (Abbildung 2.3a) dominieren die Einflüsse von U_T und U_{TH} nach Gleichungen (2.6) und (2.3); der Drain-Strom nimmt mit steigender Temperatur zu. Für diesen Effekt lässt sich eine anschauliche Erklärung finden: Bei schwacher Inversion kommt der Drain-Strom überwiegend durch Diffusion zustande. Da Diffusion die Folge thermischer Bewegung ist, ist eine Zunahme des Diffusionsstroms bei steigender Temperatur zu beobachten. Zusätzlich nimmt n_i mit der Temperatur zu. Geht man von Störstellenerschöpfung aus, so gilt im Substrat $p_0 \approx N_A$ und man erkennt an Gleichung (2.1), dass die Minoritätsträgerdichte n_0 , also die Dichte der freien Elektronen, ebenfalls bei steigender Temperatur zunimmt. Der Inversionseffekt wird also bei höherer Temperatur verstärkt, was sich durch den negativen Temperaturkoeffizienten von U_{TH} zeigt und ebenfalls zum Steigen des Drain-Stroms beiträgt.

Nun wird der Fall starker Inversion betrachtet (Abbildung 2.3b). Da bei den durchgeführten Simulationen $U_{DS} = U_{GS}$ gilt, befindet sich der Transistor im Sättigungsbereich. Es gilt [61]:

$$I_D = \frac{\mu_c C'_{ox} W}{2 L} (U_{GS} - U_{TH})^2 \quad (2.20)$$

Die negativen Temperaturkoeffizienten von μ_c und U_{TH} bewirken gegenläufige Änderungen des Drain-Stroms. Bei hoher Gate-Source-Spannung dominiert die Reduktion der Ladungsträgerbeweglichkeit μ_c , was die Abnahme des Stroms erklärt. Anschaulich lässt sich dieser Effekt folgendermaßen erklären: Bei starker Inversion entsteht der Drain-Strom hauptsächlich durch Ladungsträgerdrift. Da bei höherer Temperatur die Wahrscheinlichkeit für Stöße zwischen Ladungsträgern und dem Kristallgitter des Halbleiters aufgrund der stärkeren Gitterschwingungen steigt, nimmt die mittlere freie Weglänge, damit die Beweglichkeit μ_c , und letztendlich der Drain-Strom ab.

Prozessparameterschwankungen

Durch Prozessschwankungen treten Variationen der Transistorparameter auf, die eine Streuung des Drain-Stroms unterschiedlicher Transistorexemplare bewirken. Ein Vergleich dieser Einflüsse in starker und schwacher Inversion ist in Form von Histogrammen in Abbildung 2.4 dargestellt. Hierzu wurden Monte-Carlo-Simulationen [10] mit jeweils 10 000 Durchläufen mit Cadence Analog Design Environment ausgeführt, wobei globale und lokale Parametervariationen berücksichtigt wurden. Anhand der relativen Streuung des Drain-Stroms σ/μ von 0,84 bzw. 0,08 erkennt man, dass sich Parameterschwankungen bei schwacher Inversion zehnmal stärker

2 Transistorverhalten im Subschwellbereich

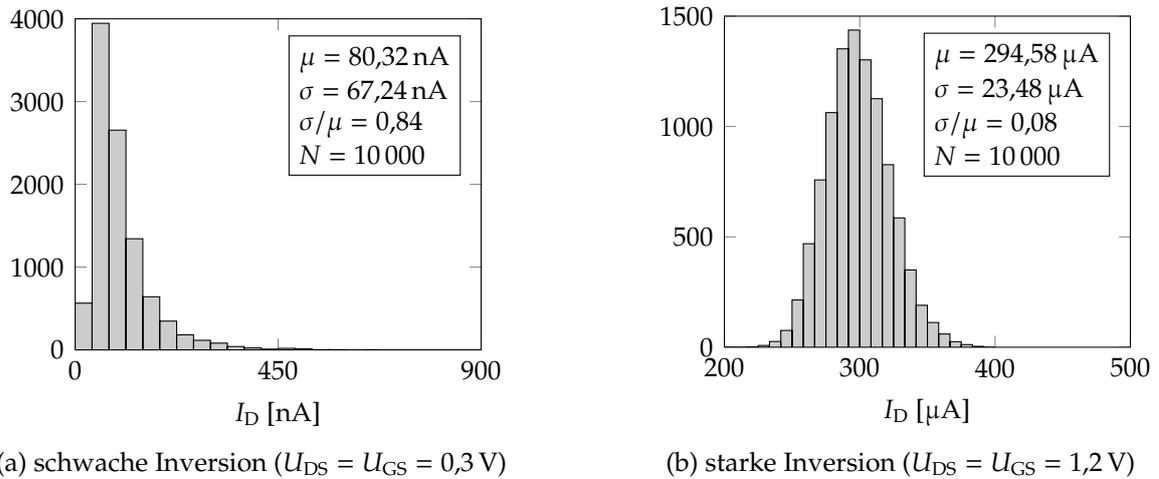


Abbildung 2.4: Streuung des Drain-Stroms durch Prozessparameterschwankungen

auf den Drain-Strom auswirken, als bei starker Inversion. Es sei an dieser Stelle bemerkt, dass der Mittelwert μ sowie die Standardabweichung σ bzw. der Variationskoeffizient σ/μ im Rahmen dieser Arbeit zur vergleichenden Bewertung verschiedener, auch nicht normalverteilter Daten herangezogen werden. Auf eine genauere Charakterisierung der jeweiligen Verteilung durch weitere Parameter wird dabei der Einfachheit halber verzichtet.

Die Streuung der Transistorparameter U_{TH} , n und β durch globale Prozessparameterschwankungen ist abhängig von den Transistormaßen W und L . Transistoren mit größerer Kanalfläche weisen geringere Parameterschwankungen auf als Transistoren mit kleinerer Kanalfläche. Dies wird durch folgende Zusammenhänge beschrieben [116]:

$$\sigma(U_{TH}) = \frac{A_{UTH}}{\sqrt{WL}} \quad (2.21)$$

$$\sigma(n) = \frac{A_n}{\sqrt{WL}} \quad (2.22)$$

$$\frac{\sigma(\beta)}{\beta} = \frac{A_\beta}{\sqrt{WL}} \quad (2.23)$$

Hierdurch ergibt sich ein Ansatzpunkt zur Reduktion des Einflusses von Prozessschwankungen auf das Transistorverhalten: Vergrößert man Transistorweite und -länge jeweils um einen Faktor $s > 1$, so vergrößert sich die Kanalfläche $W \cdot L$ um den Faktor s^2 . Dadurch reduzieren sich $\sigma(U_{TH})$, $\sigma(n)$ und $\sigma(\beta)/\beta$ um den Faktor $1/s$. Gleichzeitig bleibt dabei der Verstärkungsfaktor β erhalten und es ergibt

sich unter Vernachlässigung der im nachfolgenden Abschnitt 2.1 beschriebenen Geometrieeffekte der gleiche Drain-Strom.

Bei der Schaltungsanalyse unter Grenzfallbedingungen der Prozessparameter betrachtet man häufig Konstellationen, in denen die Parameter der Transistormodelle um ein bestimmtes Vielfaches der Standardabweichung σ höher oder niedriger als der jeweilige Mittelwert μ bei Nominalbedingungen liegen. Sind die Parameter derart verschoben, dass sich dadurch eine Erhöhung des Drain-Stroms ergibt, so spricht man von einem schnellen Transistor. Bei einer Parametermodifikation in umgekehrter Richtung spricht man von einem langsamen Transistor.

2.1 Geometrieeffekte

Die Schwellspannung U_{TH} hängt nicht nur, wie im vorangegangenen Abschnitt gezeigt, von der Temperatur ab, sondern auch von Weite und Länge des jeweiligen Transistors. Nachfolgend werden die hierfür verantwortlichen Ursachen erläutert.

Kurzkanaleffekt

Bei Transistoren mit geringer Kanallänge tritt eine Überlappung der von Drain und Source verursachten Verarmungszonen auf [61, 161]. Dadurch weisen Transistoren mit kurzen Kanälen eine geringere Schwellspannung auf als Transistoren mit langen Kanälen. Dieser Effekt wird als *Kurzkanaleffekt* (engl. *Short Channel Effect, SCE*) bezeichnet. Technologisch ist der SCE von der Substratdotierung abhängig, da hierdurch die Ausdehnung der Verarmungszonen beeinflusst wird. Zusätzlich wirkt sich auch die Drain-Source-Spannung auf die Größe der Verarmungszonen aus, wobei eine Erhöhung zu einer Reduktion der Schwellspannung führt. Dieser ebenfalls bei kurzen Transistorkanälen stärker ausgeprägte Effekt wird als *Drain Induce Barrier Lowering (DIBL)* bezeichnet.

Inverser Kurzkanaleffekt

Aufgrund verschiedener Effekte kommt es im Kanalbereich durch unerwünschte Diffusionsprozesse zu einem inhomogenen, von der Transistorlänge abhängigen

Profil der Substratdotierung [98, 102, 114, 122, 142]. Einige dieser Effekte sind beschleunigte Diffusion durch Ionenimplantation (*Transient Enhanced Diffusion*), durch Oxidation (*Oxidation Enhanced Diffusion*) und durch Silizidbildung (*Silicidation Enhanced Diffusion*). Die Auswirkung ist eine Erhöhung der Schwellspannung bei sehr kurzen Transistorkanälen, was als *inverser Kurzkanaleffekt* (engl. *Reverse Short Channel Effect*, RSCE) bezeichnet wird. Eine Erhöhung der Substratdotierung in Source- und Drain-Nähe wird bei Prozessen mit Halo-Dotierung durch spezielle Prozessschritte zur Unterdrückung von SCE und DIBL bewusst herbeigeführt. Diese Prozesse weisen einen ausgeprägten RSCE auf [30]. Die in dieser Arbeit eingesetzten Technologien setzen allerdings keine Halo-Dotierungen ein.

Schmalkanaleffekt und inverser Schmalkanaleffekt

Die Aktivgebiete¹ unterschiedlicher Transistoren werden umlaufend durch Dickoxid gegeneinander isoliert. Bei Technologien größer 0,25 μm kommt dabei *Local Oxidation of Silicon* (LOCOS), bei neueren Technologien *Surface Trench Isolation* (STI) zum Einsatz [60]. Durch das bis ins Substrat hineinreichende Oxid kommt es in Richtung der Kanalweite an den Seiten des Transistorkanals zu Randeffecten durch den geänderten Verlauf der elektrischen Feldlinien [61]. Diese Randeffecte wirken sich bei geringen Kanalweiten stärker auf das Transistorverhalten aus als bei größeren Kanalweiten. Bei LOCOS-Prozessen bewirken die Randeffecte eine Erhöhung der Schwellspannung bei schmalen Transistoren, was als *Schmalkanaleffekt* (engl. *Narrow Channel Effect*, NCE) bezeichnet wird. In STI-Prozessen tritt hingegen eine Absenkung der Schwellspannung auf, dies wird *inverser Schmalkanaleffekt* (engl. *Reverse Narrow Channel Effect*, RNCE) genannt. Bei den im Rahmen dieser Arbeit eingesetzten Technologien handelt es sich um STI-Prozesse.

Die Auswirkungen der oben genannten Effekte beim Betrieb eines Transistors in schwacher und starker Inversion werden in Abbildung 2.5 deutlich. Dargestellt ist der Drain-Strom I_D eines NMOS-Transistors in der 65-nm-Technologie in Abhängigkeit von Kanallänge und Kanalweite. Bei starker Inversion erkennt man, wie aus klassischer Sichtweise zu erwarten, eine monotone Zunahme des Drain-Stroms bei größerer Kanalweite und kleinerer Kanallänge. Hier überwiegt also der Einfluss des Faktors W/L nach Gleichung 2.20 gegenüber Geometrie-bedingten Änderungen der Schwellspannung U_{TH} . Im Subschwelligbereich hingegen tritt bei kleinen Längen unterhalb etwa 400 nm eine Abnahme von I_D auf, die durch die Erhöhung der

¹ Zusammenfassung von Source-, Drain- und Kanalbereich

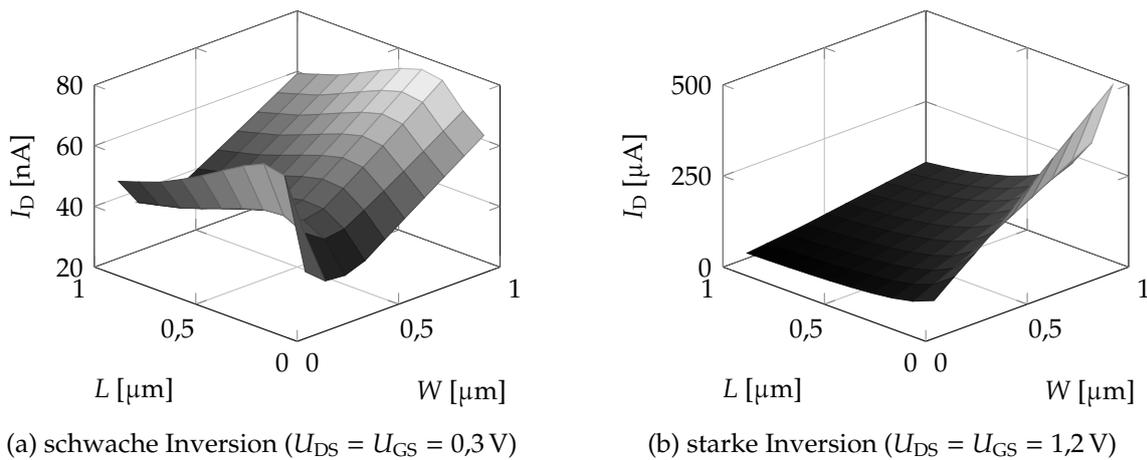


Abbildung 2.5: Abhängigkeit des Drain-Stroms von den Transistormäßen

Schwellschwung aufgrund des RSCE zu erklären ist. Außerdem nimmt I_D bei Kanalweiten unter ca. 250 nm durch den RNCE bedingt zu.

Aufgrund dieser Zusammenhänge ist eine besondere Aufmerksamkeit bei der Dimensionierung von Schaltungen für den Subschwellschwung erforderlich. Es ist bei CMOS-Gattern beispielsweise ohne weiteres nicht ausreichend, den PMOS-Transistor etwa mit der dreifachen Weite des NMOS-Transistors auszulegen, um den entsprechenden Unterschied bezüglich der Beweglichkeiten von Löchern und Elektronen auszugleichen. Stattdessen sind Schaltungssimulationen auf Basis von Transistormodellen mit akkurater Modellierung des Subschwellschwungsbereichs erforderlich. Bei einem derartigen Vorgehen können RSCE und RNCE sogar zu einem Vorteil bei der Auslegung von Subschwellschwungen genutzt werden [7, 77, 78, 83].

2.2 Zusammenfassung

In diesem Kapitel wurden das Verhalten von MOS-Transistoren im Subschwellschwungsbereich und die damit verbundenen Besonderheiten erläutert. Unter Beachtung dieser Besonderheiten lässt sich der Subschwellschwungsbereich für den Betrieb integrierter Schaltungen nutzen. Nachfolgend wird untersucht, welche Konsequenzen sich für den Subschwellschwungsbetrieb digitaler CMOS-Schaltungen ergeben.

3 Digitale CMOS-Schaltungen im Subschwellebetrieb

Der Betrieb einer digitalen integrierten CMOS-Schaltung im Subschwellebereich hat einen Einfluss auf wichtige Eigenschaften wie Energiebedarf, Taktfrequenz und Robustheit. Dieses Kapitel betrachtet die Vorteile und Herausforderungen eines solchen Schaltungsentwurfs.

3.1 Energetische Betrachtungen

Eine Möglichkeit die Verlustleistung von CMOS-Schaltungen zu reduzieren besteht darin, die Versorgungsspannung zu senken. Unter Vernachlässigung von statischen Anteilen lässt sich die Verlustleistung einer CMOS-Schaltung näherungsweise durch

$$P_{\text{dyn}} = \frac{1}{2} f C_{\text{eff}} V_{\text{DD}}^2 \quad (3.1)$$

ausdrücken [6]. Dabei ist f die Taktfrequenz, V_{DD} die Versorgungsspannung und C_{eff} eine Proportionalitätskonstante, die die effektiv pro Takt umgeladene Kapazität der Schaltung angibt. Es ist zu beachten, dass f selbst von der Versorgungsspannung abhängt. Geht man davon aus, dass die durchschnittliche Lastkapazität eines Gatters C_g während einer Taktperiode durch den konstanten Strom I_{on} eines Transistors gerade umgeladen wird, so gilt [161]:

$$f = \frac{I_{\text{on}}}{C_g V_{\text{DD}}} \quad (3.2)$$

Der Strom I_{on} hängt vom Betriebsbereich der Transistoren ab. Bei starker Inversion gilt etwa die Näherung [161]

$$f = \frac{(V_{\text{DD}} - U_{\text{TH}})^\alpha}{L_{\text{DP}} K_1 C_g V_{\text{DD}}} \quad (3.3)$$

und im Subschwellebereich [161]

$$f = \frac{I_{D0} \exp \frac{V_{DD}}{nU_T}}{L_{DP} K_2 C_g V_{DD}} \quad (3.4)$$

$K_{1,2}$ und α sind Konstanten und L_{DP} gibt die Logiktiefe des kritischen Pfades an. α liegt zwischen 1 und 2, so dass sich bei starker Inversion eine linear bis quadratisch mit der Versorgungsspannung wachsende Taktfrequenz ergibt. Im Subschwellebereich ist der Zusammenhang zwischen Versorgungsspannung und Taktfrequenz exponentiell. Betrachtet man anstelle der Verlustleistung die pro Takt umgesetzte Energiemenge, so erhält man den von der Taktfrequenz unabhängigen Ausdruck

$$E_{\text{dyn}} = \frac{P}{f} = \frac{1}{2} C_{\text{eff}} V_{DD}^2 \quad (3.5)$$

Aus den obigen Gleichungen wird deutlich, dass die Versorgungsspannung einen starken Einfluss sowohl auf die Verlustleistung als auch auf den Energiebedarf hat. Wie bereits erwähnt, vernachlässigen diese Überlegungen jedoch die statischen Verlustleistungsanteile. Die statische Verlustleistung eines CMOS-Inverters wird näherungsweise durch den Subschwellestrom des jeweils sperrenden Transistors bestimmt. Geht man von idealen Logikpegeln an Ein- und Ausgängen aus, so gilt $U_{GS} = 0$ und $|U_{DS}| = V_{DD}$ und es ergibt sich aus Gleichung 2.6 für die gesamte Schaltung der Leckstrom

$$I_{\text{leak}} = K_3 I_{D0} \left(1 - \exp \frac{-V_{DD}}{U_T} \right) \quad (3.6)$$

Die Konstante K_3 berücksichtigt dabei die Gatteranzahl und das mittlere Verhältnis W/L der Gatter. Sofern gilt $V_{DD} \gg U_T$ lässt sich dieser Ausdruck vereinfachen zu

$$I_{\text{leak}} = K_3 I_{D0} \quad (3.7)$$

Für eine Versorgungsspannung im Subschwellebereich beträgt der statische Energieumsatz damit

$$E_{\text{stat}} = \frac{I_{\text{leak}} V_{DD}}{f} = K_2 K_3 C_g L_{DP} V_{DD}^2 \exp \frac{-V_{DD}}{nU_T} \quad (3.8)$$

Für $V_{DD} \gg nU_T$ überwiegt der exponentielle Einfluss der Versorgungsspannung, so dass der statische Energieumsatz zu kleineren Versorgungsspannungen hin ansteigt. Dadurch besitzt die insgesamt pro Takt benötigte Energie $E_{\text{tot}} = E_{\text{dyn}} + E_{\text{stat}}$ bei einer bestimmten optimalen Versorgungsspannung ein Minimum. Dies ist in Abbildung 3.1 illustriert. In vielen Fällen kann die Energieaufnahme einer Schaltung in diesem optimalen Betriebspunkt um einen Faktor 10 bis 20 gegenüber dem Betrieb

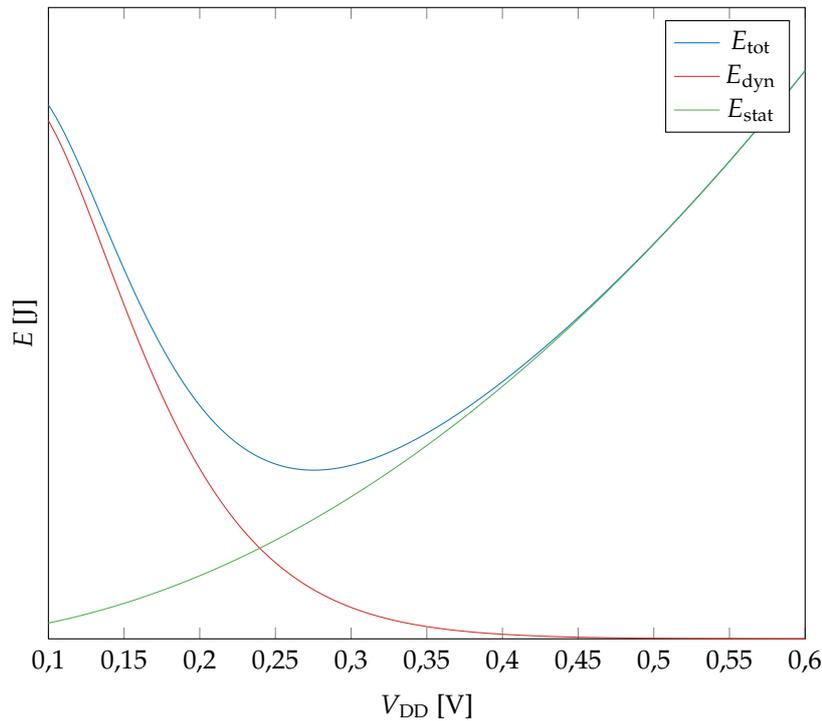


Abbildung 3.1: Energieminimum einer CMOS-Schaltung im Subschwelligbereich

bei nomineller Versorgungsspannung reduziert werden [134]. Die genaue Lage dieses Punktes hängt insbesondere vom Verhältnis zwischen statischen und dynamischen Energieanteilen und letztere wiederum von den durchschnittlichen Schalthäufigkeiten ab, die einen direkten Einfluss auf C_{eff} haben. Eine hohe Schalthäufigkeit führt zu einem hohen dynamischen Energieumsatz, so dass sich bei konstanter statischer Energie die Lage des Energieminimums zu kleineren Spannungen hin verschiebt. Umgekehrt resultiert eine niedrige Schalthäufigkeit in einer höheren optimalen Versorgungsspannung. Eine eingehende Analyse und Modellierung verschiedener Abhängigkeiten der Lage des Energieminimums findet sich etwa in [161].

3.2 Robustheit und Ausfallmechanismen

Die energetisch vorteilhaften Eigenschaften des Subschwelligbetriebs werden durch eine Reihe von Herausforderungen erkauft, die sich in einer reduzierten Robustheit der Schaltungen gegenüber verschiedenen Störeinflüssen wie etwa Rauschen, kapazitive Kopplungen unterschiedlicher Netze oder Temperatur- und Spannungsschwankun-

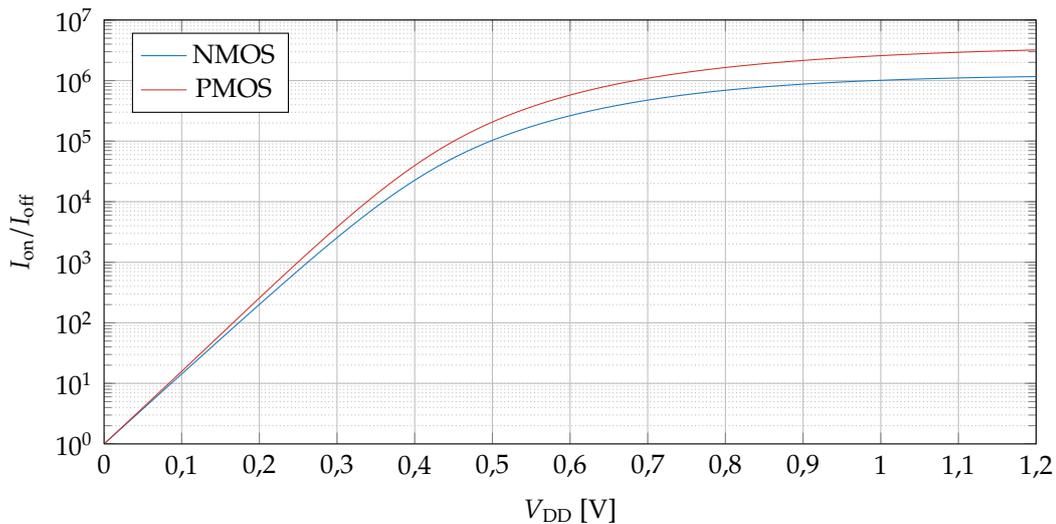


Abbildung 3.2: I_{on}/I_{off} -Verhältnis in Abhängigkeit von der Versorgungsspannung

gen darstellen. Ein wesentlicher Grund für die reduzierte Robustheit ist die Abnahme des Verhältnisses I_{on}/I_{off} mit sinkender Spannung. Als *On-Strom* I_{on} eines Transistors wird derjenige Strom bezeichnet, der fließt, wenn U_{DS} und U_{GS} (betragsmäßig) der Versorgungsspannung entsprechen. Dies ist der maximale Strom, der während des Umschaltens eines Gatterausgangs fließt. Der *Off-Strom* I_{off} ist dadurch bestimmt, dass $U_{GS} = 0$ ist, während U_{DS} weiterhin der Versorgungsspannung entspricht. I_{off} ist also der Leckstrom, der durch den jeweils sperrenden Transistorzweig eines Gatters im stationären Zustand fließt. I_{on}/I_{off} beschreibt anschaulich das Verhältnis der Leitfähigkeit zwischen einem leitenden und einem sperrenden Transistor. Dieser Wert hat einen direkten Einfluss darauf, wie ausgeprägt die Logikpegel eines Gatters sind und damit auch auf die Größe der Störabstände. Für die im Rahmen dieser Arbeit betrachtete 65-nm-Technologie ist das I_{on}/I_{off} -Verhältnis einzelner NMOS- und PMOS-Transistoren in Abhängigkeit von der Versorgungsspannung in Abbildung 3.2 dargestellt. Während bei voller Versorgungsspannung das Verhältnis bei über 10^6 liegt, sinkt dieser Wert im Subschwellebereich auf unter 1000 und nimmt dort weiter exponentiell ab, wie der Abbildung zu entnehmen ist. Bei sinkender Versorgungsspannung nehmen daher die relativen Störabstände ab und die Robustheit der Gatter sinkt.

Wie in Kapitel 2 gezeigt wurde, steigt außerdem die Empfindlichkeit der Drainströme unter anderem gegenüber Prozessschwankungen. Durch entgegengesetzte Prozessbedingungen für NMOS und PMOS, also z.B. schneller NMOS, langsamer PMOS, kann es damit im Subschwellebereich zu einem stark unsymmetrischen Span-

Schaltung, Technologie, Besonderheiten	Spannung	Referenz
Ringoszillator	100 mV	[145]
Inverter und NAND2, 180 nm, Regelung der Substratvorspannung	100 mV	[16]
8-bit-Multiplizier-Akkumulier-Einheit, 140 nm, Regelung der Substratvorspannung, sehr niedrige Schwellspannung (50 mV)	175 mV	[72]
Ringoszillator, 180 nm, Regelung der Substratvorspannung	80 mV	[64]
16-bit-FFT-Prozessor, 1024 Punkt, 180 nm	180 mV	[160]
8-bit-FIR-Filter 8. Ordnung, 130 nm, Regelung der Substratvorspannung	84 mV	[62]
8-bit-RISC-Prozessor, 130 nm, Regelung der Substratvorspannung	160 mV	[57]
8-bit-Multiplizierer, 130 nm, Schmitt-Trigger-Logik	62 mV	[96]

Tabelle 3.1: Implementierungen mit sehr niedriger Versorgungsspannung

nungsübertragungsverhalten der Gatter und damit zu einer weiteren signifikanten Abnahme der Störabstände kommen. Aus diesem Grund kann die Versorgungsspannung einer CMOS-Schaltung nicht beliebig weit abgesenkt werden, ohne dass es zu einer Fehlfunktion kommt. Verschiedene Veröffentlichungen untersuchen die theoretische untere Grenze der Versorgungsspannung für CMOS-Schaltungen, z.B. [16, 74, 75, 144], wobei unterschiedliche Autoren Werte zwischen $1U_T$ und $8U_T$ als Minimalwert vorschlagen. Dies entspricht bei Raumtemperatur Versorgungsspannungen zwischen etwa 25 mV und 200 mV. Eine Übersicht über praktische Implementierungen von Schaltungen, die in diesem Spannungsbereich realisiert wurden, ist in Tabelle 3.1 angegeben.

Eine weitere Fehlerursache, die bei reduzierter Versorgungsspannung zum Tragen kommt, entsteht durch die zeitlichen Anforderungen sequenzieller Elemente, also insbesondere Flipflops, an die eingehenden Signale. Dies sind die *Setup*- und *Hold*-Zeiten. Durch den starken Einfluss der Prozessschwankungen, hier insbesondere innerhalb eines Schaltungsexemplars (engl. *On-Chip Variation*, OCV) auf die Gatterverzögerungen im Subschwelligbereich kommt es hier beim Senken der Versorgungsspannung zu einer zeitlichen Verschiebung unterschiedlicher Logikpfade zueinander. Dies kann Verletzungen der Setup- und Hold-Anforderungen zur Folge haben. Während sich Setup-Verletzung durch eine Reduktion der Taktfrequenz beheben lassen, füh-

ren Hold-Verletzungen unabhängig von der verwendeten Taktfrequenz zu einer Schaltungsfehlfunktion. Es ist daher sehr wichtig, bei der Implementierung von Subschwellschaltungen ein besonderes Augenmerk auf Hold-Anforderungen unter dem Einfluss von Prozessschwankungen zu legen. Eine detaillierte Beschreibung dieses Effektes und der zur Vermeidung erforderlichen Schritte ist in Abschnitt 7.4 gegeben.

Insgesamt lässt sich festhalten, dass die wichtigsten Fehlerursachen für den Schaltungsbetrieb im Subschwellbereich unzureichende Störabstände bzw. Logikpegel und Hold-Verletzungen sind. Zu diesem Schluss kommt beispielsweise auch [15].

3.3 Grenzfallbedingungen

In Kapitel 2 wurde gezeigt, dass der Drain-Strom bei starker und schwacher Inversion ein entgegengesetztes Temperaturverhalten zeigt. Diese Besonderheit muss bei der Definition der bezüglich der Gatterlaufzeiten besten und schlechtesten PVT¹-Betriebsbedingungen (engl. *Best Case* und *Worst Case*) beachtet werden. Bei gewöhnlichen CMOS-Digitalschaltungen wird der ungünstigste Fall in der Regel durch die Kombination langsamer NMOS, langsamer PMOS, eine um 10 % reduzierte Versorgungsspannung und eine Temperatur von 125 °C festgelegt. Der beste Fall lautet dementsprechend schneller NMOS, schneller PMOS, eine um 10 % erhöhte Versorgungsspannung und eine Temperatur von -40 °C. Für eine Subschwellschaltung sind die Temperaturwerte aus obigem Grund zu vertauschen, es gilt also als ungünstigster Fall die Kombination langsamer NMOS, langsamer PMOS, eine um 10 % reduzierte Versorgungsspannung und eine Temperatur von -40 °C und als bester Fall die Kombination schneller NMOS, schneller PMOS, eine um 10 % erhöhte Versorgungsspannung und eine Temperatur von 125 °C. Dies gilt jedoch nur für reine CMOS-Logik und beispielsweise nicht für Pegelwandler. Die für Pegelwandler geltenden besten und schlechtesten PVT-Bedingungen werden im entsprechenden Abschnitt 6.6.1 erläutert.

¹ *Process, Voltage, Temperature*

3.4 Zusammenfassung

In diesem Kapitel wurden die Vorteile und Herausforderungen eines digitalen Schaltungsentwurfs für den Subschwellsbetrieb betrachtet. Während die Leistungsaufnahme und die maximale Taktfrequenz bei sinkender Versorgungsspannung im Subschwellsbereich näherungsweise exponentiell abnehmen, ergibt sich für den Energiebedarf bei einer bestimmten Versorgungsspannung ein Minimum. Die Lage dieses Energieoptimums hängt unter anderem von den Schalthäufigkeiten der Signale in der betrachteten Schaltung ab. Gegenüber dem Betrieb bei maximaler Versorgungsspannung lassen sich Energieeinsparungen um den Faktor 10 bis 20 erreichen. Dieser energetische Vorteil wird durch eine reduzierte Robustheit erkauft, so dass mit sinkender Versorgungsspannung die Ausfallwahrscheinlichkeit einer Subschwellschaltung steigt. Die beiden hierfür im Wesentlichen verantwortlichen Effekte sind die Verschlechterung der Störabstände bzw. Logikpegel sowie Verletzungen der Hold-Zeit-Anforderungen der Flipflops. Eine weitere Besonderheit ergibt sich aus der Umkehr der Temperaturabhängigkeit des Drain-Stroms im Subschwellsbereich. Die schlechtesten Betriebsbedingungen treten daher bei niedriger Temperatur, die besten Betriebsbedingungen bei hoher Temperatur auf. Bei einem konventionellen Schaltungsentwurf ergeben sich entgegengesetzte Verhältnisse. Diese Besonderheiten müssen bei einem Schaltungsentwurf für den Subschwellsbetrieb beachtet werden.

Die Optimierung der Robustheit im Sinne der Störabstände erfordert einen angepassten Entwurf der Logikgatter. Der nachfolgende Teil dieser Arbeit befasst sich daher mit dem Entwurf zweier Standardzellen-Bibliotheken für den Subschwellsbetrieb in Technologien mit Strukturgrößen von 90 nm und 65 nm.

Teil II

Standardzellen-Entwurf

4 Semi-Custom-Entwurf von Subschwellschaltungen

Die Implementierung ressourceneffizienter Digitalschaltungen für den Subschwellbetrieb erfordert zunächst den Entwurf einer Standardzellen-Bibliothek, die den speziellen Anforderungen dieses Einsatzgebiets gerecht wird. Einfache Logikschaltungen, vor allem mit regelmäßigen logischen und physikalischen Strukturen, lassen sich in einigen Fällen noch mit akzeptablem Aufwand von der Schaltplan-Eingabe über die Dimensionierung bzw. Optimierung bis zur Layout-Erstellung als Handentwurf (*Full-Custom-Entwurf*) entwickeln. Solche Schaltungen können beispielsweise Decoder oder Addierer unterschiedlicher Bitbreite und Implementierungsart sein. Bei komplexen Schaltungen wie etwa Prozessoren, Digitalfiltern oder Modulen zur drahtgebundenen oder drahtlosen Datenübertragung lässt sich ein derartiges Entwurfsverfahren jedoch nicht mehr mit vertretbarem Aufwand realisieren. In diesen Fällen ist ein *Semi-Custom-Entwurf* vorteilhaft, bei dem die gewünschte Funktionalität der Schaltung in einer Hardware-Beschreibungssprache wie VHDL¹ oder Verilog auf Register-Transfer-Ebene (engl. *Register Transfer Level*, RTL) formuliert wird und anschließend die Übertragung in eine physikalische Halbleiterschaltung durch Software-Werkzeuge vorgenommen wird. Von Vorteil ist bei dieser Vorgehensweise auch, dass Modifikationen an der Schaltungsfunktionalität in der Regel einfacher möglich sind, als dies bei einem Handentwurf der Fall ist. Zudem kann der Entwurfsprozess durch Wiederverwendung und ggf. Weiterentwicklung von bestehendem Code in Form von kommerziellen, frei verfügbaren oder eigen-entwickelten IP²-Modulen verkürzt werden. Ein solcher Entwurfsablauf ist industriell als Standard anzusehen und daher auch im Bereich digitaler Subschwellschaltungen anzustreben.

Ein Semi-Custom-ASIC-Entwurf erfordert stets das Vorhandensein einer Standardzellen-Bibliothek für die gewählte Halbleiter-Technologie. Die Entwurfswerkzeuge

¹ *Very High Speed Integrated Circuit Hardware Description Language*

² *Intellectual Property*

bilden die vorgegebene Verhaltensbeschreibung in den groben Schritten Synthese, Platzierung und Verdrahtung auf die Gatter der Standardzellen-Bibliothek ab. Konventionelle, kommerzielle Standardzellen-Bibliotheken sind jedoch in der Regel nur eingeschränkt zum Entwurf von Subschwellschaltungen geeignet, da diese für die nominelle Versorgungsspannung der entsprechenden Technologie ausgelegt sind. Da sich bei geringerer Spannung, insbesondere beim Erreichen des Subschwellschwellbetriebs, das Transistorverhalten grundlegend verändert, kann nicht davon ausgegangen werden, dass solche Gatter bei deutlich geringeren Versorgungsspannungen noch funktionsfähig sind. So kann das optimale Weitenverhältnis der NMOS- und PMOS-Transistoren im Subschwellschwellbetrieb deutlich vom Betrieb mit nomineller Spannung abweichen. Ferner ergibt sich im Subschwellschwellbereich eine stark erhöhte Empfindlichkeit gegenüber Schwankungen der Prozessparameter, die ebenfalls beim Entwurf konventioneller Standardzellen keine Berücksichtigung findet. In Abschnitt 6.1.4 wird zudem anhand einer Optimalitätsbetrachtung gezeigt, dass konventionelle Standardzellen im Subschwellschwellbetrieb nicht ressourceneffizient sind.

Im Rahmen dieser Arbeit wurden zwei Standardzellen-Bibliotheken entwickelt, mit denen beliebige Digitalschaltungen in einem Semi-Custom-Entwurfsverfahren implementiert werden können. Kapitel 5 behandelt die Entwicklung einer Subschwellschwellbibliothek in einer 90-nm-CMOS-Technologie, wobei die Gatter für maximale Robustheit bei einer Versorgungsspannung von 200 mV ausgelegt sind. Außerdem werden Messergebnisse eines auf der Bibliothek basierenden Bausteins mit 32-bit-ALUs präsentiert. In Kapitel 6 wird eine weiterentwickelte Subschwellschwellbibliothek für eine 65-nm-Low-Power-CMOS-Technologie vorgestellt. Hier wurden durch den Einsatz eines Mehrzieloptimierungsverfahrens die Gatter unter Berücksichtigung weiterer Zielfunktionen optimal ausgelegt. Der Entwurf eines vielseitig verwendbaren Prozessors basierend auf dieser Subschwellschwellbibliothek sowie Messergebnisse gefertigter Prozessorexemplare werden in Teil III dieser Arbeit behandelt.

5 Robuste Subschwelligbibliothek in 90 nm CMOS

In diesem Kapitel wird die Entwicklung einer Standardzellen-Bibliothek für den Subschwelligbetrieb in einer 90-nm-CMOS-Technologie sowie ein darauf basierender Baustein (engl. *Chip*) mit 32-bit-ALUs beschrieben. Einige Aspekte dieses Kapitels wurden auch in der Dissertation von Kaulmann [73] und der gemeinsamen Veröffentlichung [182] behandelt. Während in der vorliegenden Arbeit das Augenmerk vor allem auf dem Entwurfsprozess der Bibliothek, der Entwicklung des Testbausteins sowie den zugehörigen Messergebnissen liegt, wird in [73] die Subschwelligbibliothek insbesondere als Grundlage zur digitalen Implementierung pulscodierter neuronaler Netze eingesetzt.

Alle nachfolgend in dieser Arbeit vorgestellten analogen Schaltungssimulationen wurden mittels Cadence Spectre innerhalb von Cadence Analog Design Environment durchgeführt. Zur automatisierten Charakterisierung der Standardzellen kam Cadence Encounter Library Characterizer zum Einsatz.

5.1 Zellensatz

Die Synthese beliebiger synchroner digitaler Schaltungen ist prinzipiell bereits mit einer Bibliothek lediglich bestehend aus einem NAND- oder NOR-Gatter mit zwei Eingängen [137] sowie einem Speicherelement möglich. Derartige Implementierungen wären jedoch äußerst ineffizient in Bezug auf Geschwindigkeit, Verlustleistung und Fläche, da eine große Anzahl an Gattern zur Realisierung einer beliebig vorgegebenen Logikfunktion erforderlich ist. Ein angemessener Zellensatz umfasst daher wesentlich mehr Gattertypen. Zur Kommunikation mit Schaltungsteilen, die nicht im Subschwelligbereich arbeiten, sind außerdem Pegelwandler erforderlich, um die

Spannungspegel der beteiligten Signale anzupassen. Für die entwickelte 90-nm-Subschwelligbibliothek wurden daher folgende Gatter ausgewählt:

- Inverter (INV)
- Puffer (BUF)
- Tristate-Inverter (TRI_INV)
- NAND-Gatter mit 2 Eingängen (NAND2)
- NAND-Gatter mit 3 Eingängen (NAND3)
- NOR-Gatter mit 2 Eingängen (NOR2)
- AND-OR-INVERT-Gatter, Logikfunktion $\overline{(A \wedge B) \vee (C \wedge D)}$ (AOI)
- D-Flipflop (DFF)
- D-Latch (DL)
- Aufwärts-Pegelwandler (LVL_SHIFT_UP)
- Abwärts-Pegelwandler (LVL_SHIFT_DOWN)

Auf diese Weise steht dem Synthese-Werkzeug bereits eine Vielzahl von Möglichkeiten zur Realisierung einer gegebenen Logikfunktion zur Verfügung. Insbesondere die Bereitstellung des AND-OR-INVERT-Gatters ist vorteilhaft, da Boole'sche Funktionen stets in disjunktiver Normalform

$$f(X) = \bigvee_i \bigwedge_j (\neg)x_{ij} \quad (5.1)$$

dargestellt werden können [112]. Das AND-OR-INVERT-Gatter bietet unmittelbar die Möglichkeit zur Berechnung einer solchen Funktion mit zwei Konjunktionstermen zu je zwei Variablen.

Von allen Gattern wurde zunächst jeweils ein Typ mit geringer Treiberstärke realisiert. Inverter und Puffer wurden in je drei unterschiedlichen Treiberstärken implementiert, um Netze mit hoher kapazitiver Last besser treiben zu können. Eine weitere Möglichkeit besteht in solchen Fällen für das Synthese-Werkzeug darin, betroffene

Gatter zu vervielfältigen und somit die Last aufzuteilen. Die Puffer und Inverter unterschiedlicher Treiberstärken werden darüber hinaus für die Taktbaum-Synthese benötigt.

Alle Gatter wurden in gewöhnlicher CMOS-Schaltungstechnik realisiert, da bekannt ist, dass CMOS-Gatter auch im Subschwellebetrieb gute Eigenschaften (Robustheit, Verlustleistung, usw.) besitzen. Funktionsfähige Subschwelle-CMOS-Schaltungen wurden beispielsweise in [24, 58, 62, 72, 90, 121, 136, 148, 160, 175, 176] veröffentlicht. Bei einigen dieser Implementierungen kommt dabei sogenanntes *Body Biasing* zum Einsatz, also die Verwendung von positiver oder negativer Substrat- und Wannenvorspannung. Dennoch sind diese Schaltungen in CMOS-Technik aufgebaut. Alternative Schaltungstechniken weisen oftmals im Subschwellebetrieb Defizite in Bezug auf die Robustheit auf [161] und werden daher für die weiteren Betrachtungen ausgeschlossen.

5.2 Dimensionierung der Logikgatter

Die Schwellspannung der Transistoren in der verwendeten 90-nm-Technologie beträgt je nach Transistormaßen etwa 250 mV bis 300 mV. Für den Subschwellebetrieb wurde daher der Nennwert der Versorgungsspannung auf 200 mV festgelegt. Aufgrund der in Kapitel 3 aufgeführten Problematik bei der Implementierung von Subschwelle-Schaltungen wird die Dimensionierung der kombinatorischen Gatter im Interesse der bestmöglichen Robustheit, d.h. Störsicherheit gegenüber Pegelverfälschungen, bei dieser Nennversorgungsspannung durchgeführt. Hierzu wird zunächst der Umschaltpunkt U_{sp} des Gatters definiert durch

$$U_{sp} = U_{out}(U_{in} = U_{sp}) \quad (5.2)$$

Um eine große Toleranz gegenüber Pegelverfälschungen zu erreichen, ist eine Übertragungskennlinie mit symmetrischem Verlauf anzustreben, was erreicht wird, wenn gilt:

$$U_{sp} = \frac{V_{DD}}{2} \quad (5.3)$$

bzw.

$$U_{out}(U_{in} = \frac{V_{DD}}{2}) = \frac{V_{DD}}{2} \quad (5.4)$$

Dabei ist U_{out} die Ausgangsspannung und U_{in} die Spannung an allen Eingängen des Gatters. Da an allen Eingängen für diese Betrachtung also die gleiche Spannung anliegt, kann das jeweils betrachtete Gatter durch Zusammenfassen aller NMOS-Transistoren und aller PMOS-Transistoren durch einen äquivalenten Inverter beschrieben werden.

Um obige Forderung zu erfüllen, kann zunächst eine Abschätzung anhand der Stromgleichung 2.6 vorgenommen werden. Es muss gelten:

$$I_{D,n}|_{U_{\text{in}}=V_{\text{DD}}/2=U_{\text{out}}} = I_{D,p}|_{U_{\text{in}}=V_{\text{DD}}/2=U_{\text{out}}} \quad (5.5)$$

$$\Rightarrow \frac{W_n}{L_n} I_{D0,n} \exp \frac{V_{\text{DD}}}{2n_n U_T} \left(1 - \exp \frac{-V_{\text{DD}}}{2U_T} \right) = \frac{W_p}{L_p} I_{D0,p} \exp \frac{V_{\text{DD}}}{2n_p U_T} \left(1 - \exp \frac{-V_{\text{DD}}}{2U_T} \right) \quad (5.6)$$

$$\Rightarrow r := \frac{W_p/L_p}{W_n/L_n} = \frac{I_{D0,n}}{I_{D0,p}} \exp \left(\frac{V_{\text{DD}}}{2n_n U_T} - \frac{V_{\text{DD}}}{2n_p U_T} \right) \quad (5.7)$$

Dabei sind $I_{D,n/p}$ die Drainströme, $W_{n/p}$ die Kanalweiten und $L_{n/p}$ die Kanallängen der (äquivalenten) NMOS- bzw. PMOS-Transistoren. Die Versorgungsspannung V_{DD} beträgt 200 mV. Die Werte der Transistorparameter I_{D0} und n können Tabelle 2.1 entnommen werden. Durch Einsetzen erhält man das Ergebnis $r = 2,53$.

Eine genauere Bestimmung von r ist möglich, indem der Umschaltzeitpunkt U_{sp} in Abhängigkeit von r bei einer Versorgungsspannung von 200 mV direkt ermittelt wird. Das Simulationsergebnis ist in Abbildung 5.1 dargestellt. Man erkennt, dass für $r = 2,3$ der optimale Umschaltzeitpunkt von $U_{\text{sp}} = V_{\text{DD}}/2 = 100$ mV erreicht wird. Dieser Wert wird im Folgenden zur Dimensionierung der Subschwelliggatter verwendet.

5.2.1 Kombinatorische Gatter

Die minimalen Transistormaße betragen für den verwendeten Prozess: $W_{\text{min}} = 120$ nm, $L_{\text{min}} = 100$ nm. Um Parameterstreuungen zu halbieren und damit die Robustheit der Gatter zu erhöhen, wurden diese Grenzen verdoppelt. Damit steht die Dimensionierung des Minimal-Inverters, also des Inverters mit der kleinsten Treiberstärke, bereits fest: $W_n = 240$ nm, $W_p = 550$ nm (gerundet auf 5-nm-Raster), $L = 200$ nm. Für die übrigen kombinatorischen Gatter mit mehr als einem Eingang sind folgende Schritte zu berücksichtigen:

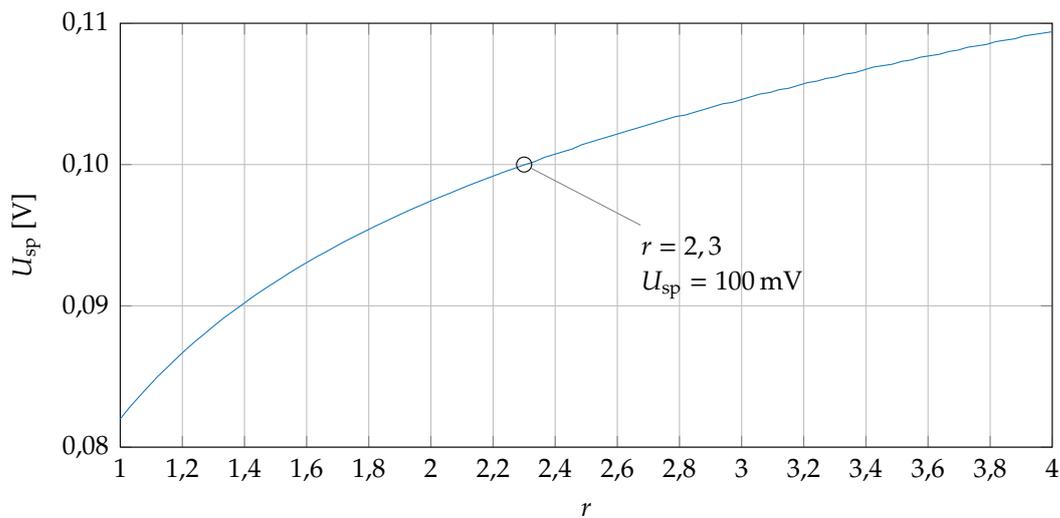


Abbildung 5.1: Umschalt­punkt U_{sp} in Ab­hän­gig­keit vom Ver­hät­nis r

- Zusammenfassen von Serien- und Parallelschaltungen von Transistoren durch entsprechendes Vervielfältigen der Weiten bzw. Längen, so dass ein äquivalenter Inverter betrachtet werden kann
- Dimensionierung der Weite des äquivalenten PMOS-Transistors um den Faktor r größer als die Weite des äquivalenten NMOS-Transistors
- Wahl der Weiten so, dass im Fall des ungünstigsten Umschaltens die Treiberstärke des Minimal-Inverters erreicht wird.

Um Inverter mit zweifacher und dreifacher Treiberstärke zu realisieren, werden die Weiten des Minimal-Inverters entsprechend vervielfacht. Für die Auslegung der Puffer wird zusätzlich ein Weiten-Verhältnis einer Stufe zur vorangegangenen Stufe von 3 festgelegt. Auf diese Weise werden Puffer mit dreifacher und neunfacher Treiberstärke mit zweistufigem Aufbau sowie ein Puffer mit 27-facher Treiberstärke in vierstufigem Aufbau realisiert. Die sich aus den obigen Regeln ergebenden Dimensionierungen der kombinatorischen Gatter sind in Tabelle 5.1 gegeben.

Die Robustheit der Gatter unter dem Einfluss von Prozessschwankungen wurde durch Monte-Carlo-Simulationen untersucht. Detaillierte Ergebnisse entsprechender Simulationen werden in Abschnitt 6.3 in Bezug auf den Entwurf der mehrzieloptimierten Standardzellen-Bibliothek diskutiert. Prinzipiell ergaben sich für die hier betrachtete Subschwelligbibliothek vergleichbare Ergebnisse und es wurde festgestellt,

Gatter	W_n	W_p	
INVX1	240	550	nm
INVX2	480	1100	nm
INVX3	720	1650	nm
BUFX3			
1. Stufe	240	550	nm
2. Stufe	720	1650	nm
BUFX9			
1. Stufe	720	1650	nm
2. Stufe	2160	4950	nm
BUFX27			
1. Stufe	240	550	nm
2. Stufe	720	1650	nm
3. Stufe	2160	4950	nm
4. Stufe	6480	14850	nm
TRI_INV			
Inverter	240	550	nm
Tristate-Stufe	480	1100	nm
NAND2	960	550	nm
NAND3	2160	550	nm
NOR2	240	2200	nm
AOI	480	1100	nm

Tabelle 5.1: Dimensionierung der kombinatorischen Gatter (Gatelänge: 200 nm)

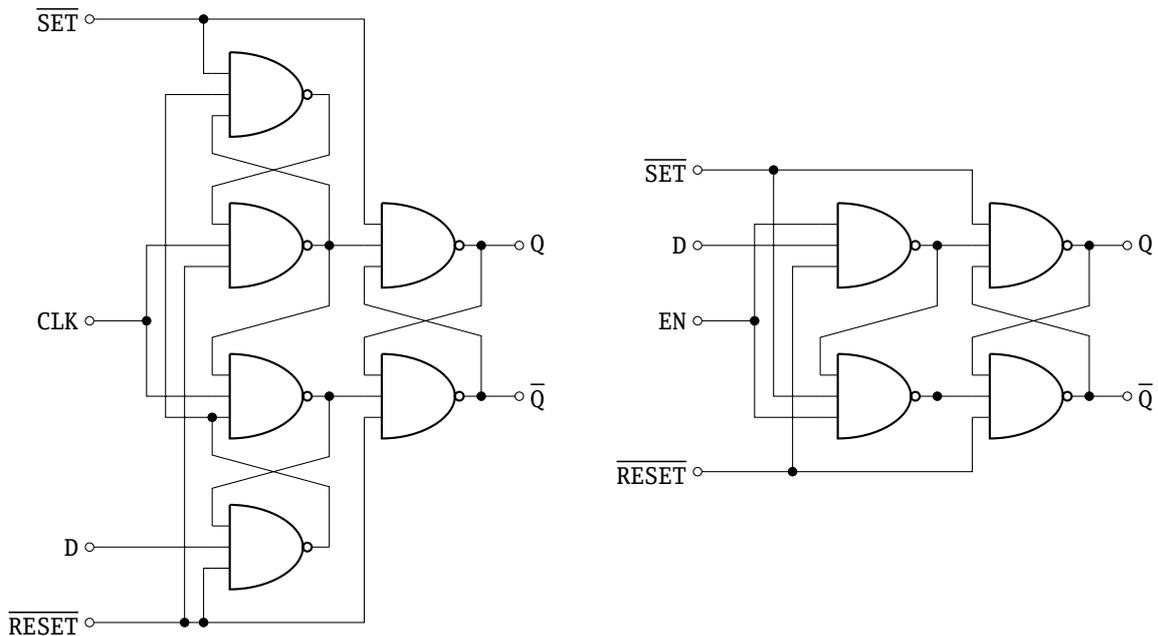


Abbildung 5.2: NAND3-basiertes Flipflop [147] und Latch

dass alle Gatter eine ausreichend hohe Robustheit gegenüber Prozessschwankungen besitzen.

5.2.2 Flipflop und Latch

Im Rahmen dieser Arbeit werden taktflankengesteuerte Speicherelemente als *Flipflops* und taktzustandgesteuerte Speicherelemente als *Latches* bezeichnet. Für die Realisierung von Flipflop und Latch wurden konventionelle, ausschließlich auf NAND3-Gattern, also rein kombinatorischen Zellen basierende Implementierungen gewählt [147]. Die Speicherfunktion wird durch geeignete Rückkopplungspfade erreicht. Die Implementierungen bieten zusätzlich zur Grundfunktionalität Eingänge zum asynchronen Setzen und Rücksetzen des Speicherinhalts sowie komplementäre Ausgänge. Die Gatterschaltpläne sind in Abbildung 5.2 dargestellt.

Die Dimensionierung der NAND3-Teilgatter mit den Werten aus dem vorangegangenen Abschnitt führte jedoch zu einem ungünstigen Layout mit großer Fläche. Um ein effizienteres Layout insbesondere des für sequenzielle Logik wichtigen

Flipflops zu erreichen, wurden daher die Transistormaße gegenüber den Werten des eigenständigen NAND3-Gatters um den Faktor 0,6 verkleinert:

$$W_n = 1300 \text{ nm} \quad , \quad W_p = 330 \text{ nm} \quad , \quad L = 120 \text{ nm}$$

Mit diesen Werten wird zwar die zuvor vereinbarte minimale Transistorlänge unterschritten, das Flipflop ist jedoch auch mit diesen Maßen ausreichend robust: Der Proportionalitätsfaktor für die Streuung der Transistorschwellspannung nach Gleichung 2.21 beträgt für den vereinbarten Minimaltransistor

$$\frac{1}{\sqrt{0,24 \mu\text{m} \cdot 0,20 \mu\text{m}}} \approx 4,56 \mu\text{m}^{-1}$$

bzw. für einen Transistor mit den Minimalmaßen der Technologie

$$\frac{1}{\sqrt{0,12 \mu\text{m} \cdot 0,10 \mu\text{m}}} \approx 9,13 \mu\text{m}^{-1}$$

Für die Transistoren des Flipflops betragen die Faktoren

$$\frac{1}{\sqrt{1,3 \mu\text{m} \cdot 0,12 \mu\text{m}}} \approx 2,53 \mu\text{m}^{-1} \quad (\text{NMOS})$$

$$\frac{1}{\sqrt{0,33 \mu\text{m} \cdot 0,12 \mu\text{m}}} \approx 5,02 \mu\text{m}^{-1} \quad (\text{PMOS})$$

Damit ist die Streuung der NMOS-Schwelspannung immer noch geringer als die des vereinbarten Minimaltransistors. Lediglich die Streuung der PMOS-Schwelspannung ist gegenüber diesem um ca. 10% größer, aber immer noch 45% kleiner als die eines Transistors mit den Minimalmaßen der Technologie. Die robuste Funktion der Schaltung, d.h. die korrekte Datenspeicherung mit dieser Dimensionierung wurde durch Monte-Carlo-Simulationen überprüft.

5.3 Pegelwandler

Die durch den Subschwelligbetrieb von digitalen CMOS-Schaltungen mögliche Reduktion der Leistungsaufnahme ist mit einer Reduktion der maximalen Taktfrequenz in ähnlicher Größenordnung verbunden. In einem komplexen digitalen System kann es daher wünschenswert sein, Module nach der erforderlichen Rechenleistung aufzuteilen. Schaltungsteile mit strikten Performanzvorgaben werden mit

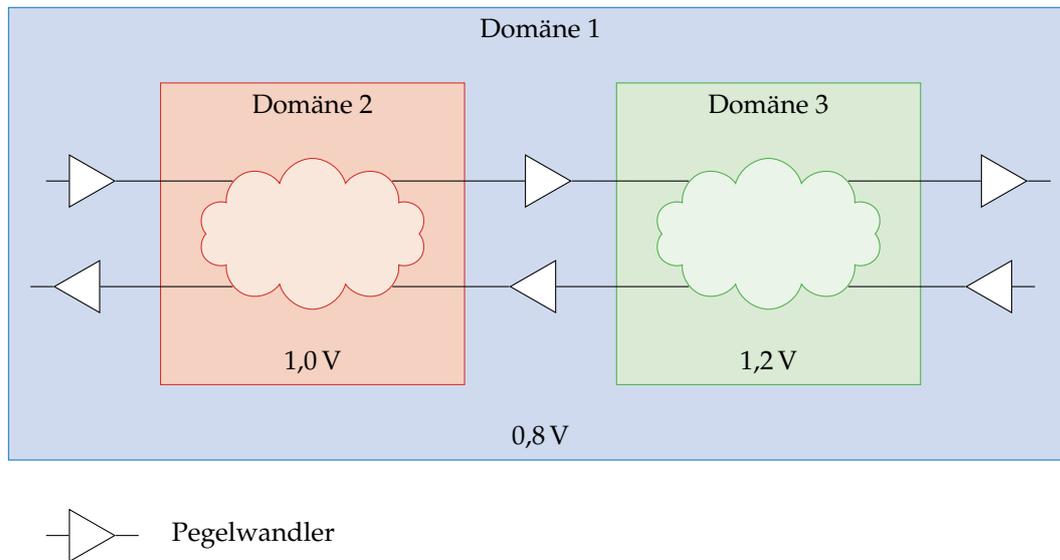


Abbildung 5.3: Ein System mit mehreren Spannungsdomänen

hoher Versorgungsspannung betrieben, während Schaltungsteile, bei denen die Verarbeitungsgeschwindigkeit zweitrangig ist, im Subschwellebetrieb bei niedriger Versorgungsspannung betrieben werden. Dadurch entstehen Systeme mit mehreren Spannungsdomänen (engl. *Multiple Supply Voltage*, MSV). Selbst wenn die gesamte in einem Bauteil implementierte Logik für den Subschwellebetrieb ausgelegt werden soll, werden noch immer mit hoher Versorgungsspannung arbeitende Anschlusszellen (engl. *Pad-Zellen*) für die Kommunikation mit externen Schaltungen benötigt.

Sofern die verwendeten Logikgatter dies zulassen, kann die Versorgungsspannung der Module auch im laufenden Betrieb variiert werden. Zusätzlich muss die Taktfrequenz bei reduzierter Versorgungsspannung herab gesetzt werden, um den erhöhten Gatterlaufzeiten Rechnung zu tragen. Man spricht dann von dynamischer Spannungs- und Frequenzskalierung (engl. *Dynamic Voltage and Frequency Scaling*, DVFS). DVFS erlaubt es, in Phasen mit hoher Rechenlast Module mit hoher Taktfrequenz und Versorgungsspannung zu betreiben, um der Verarbeitung der Daten nachzukommen, während in weniger rechenintensiven Zeitabschnitten Taktfrequenz und Versorgungsspannung abgesenkt werden, so dass im Subschwellebetrieb weniger Leistung bzw. Energie benötigt wird. Diese Technik bietet sich daher besonders für Systeme mit wechselndem Bedarf an Rechenleistung an.

An der Schnittstelle zwischen Subschwellemodulen und konventionellen Schaltungsteilen sind Pegelwandler erforderlich, um zwischen den unterschiedlichen Spannungspegeln der Signale zu vermitteln. Bei DVFS-Systemen müssen die Pe-

gelwandler in der Lage sein, über einen weiten Spannungsbereich für eine oder sogar beide Versorgungsspannungen fehlerfrei zu arbeiten. Da ein Pegelwandler keinen Beitrag zur logischen Funktionalität der Schaltung beiträgt, muss er darüber hinaus ressourceneffizient implementiert werden, um die Erhöhung von Verzögerungszeiten, Verlustleistung und Fläche klein zu halten.

Im Rahmen dieser Arbeit wird davon ausgegangen, dass die höhere Versorgungsspannung am Pegelwandler stets der vollen bzw. nominellen Versorgungsspannung der verwendeten Technologie entspricht.

5.3.1 Aufwärts-Pegelwandler

Ein Aufwärts-Pegelwandler dient dazu, Signale von einem niedrigen Spannungspegel auf einen hohen Spannungspegel zu transformieren. Dabei kann auch der Grenzfall zugelassen sein, dass beide Spannungen übereinstimmen. Dies ist etwa dann der Fall, wenn in einem DVFS-System die Spannung eines Moduls im Betrieb zur beschleunigten Verarbeitung von Daten auf den nominellen Wert angehoben wird.

Im Folgenden wird der prinzipielle Aufbau eines Aufwärts-Pegelwandlers nach Abbildung 5.4 zugrunde gelegt. Da alle betrachteten Pegelwandler neben dem eigentlichen Eingangssignal A auch das logisch Inverse des Eingangssignals AN benötigen, ist der Inverter M1, M2 vorgesehen. Da insbesondere bei Subschwellanwendungen die zentrale Pegelwandler-Stufe nur eine begrenzte Fähigkeit besitzt, angeschlossene Lasten zu treiben, wird außerdem von einer Pufferung des Ausgangssignals durch M3 – M6 ausgegangen. In den nachfolgenden Abschnitten werden unterschiedliche Implementierungen für die zentrale Pegelwandler-Stufe sowie deren Vor- und Nachteile vorgestellt.

5.3.1.1 Konventioneller Aufwärts-Pegelwandler

Ein konventioneller Pegelwandler, der auf einem sogenannten *Halb-Latch* basiert, ist in Abbildung 5.5a dargestellt [6]. Die Schaltung ist auch unter dem Namen *Differential Cascode Voltage Switch* (DCVS) bekannt. Die eingangsseitigen Signale A und AN sind mit den Gates der NMOS-Transistoren M1 bzw. M3 verbunden. Die PMOS-Transistoren M2, M4 stellen einen Halb-Latch dar. Bei jeder eintreffenden

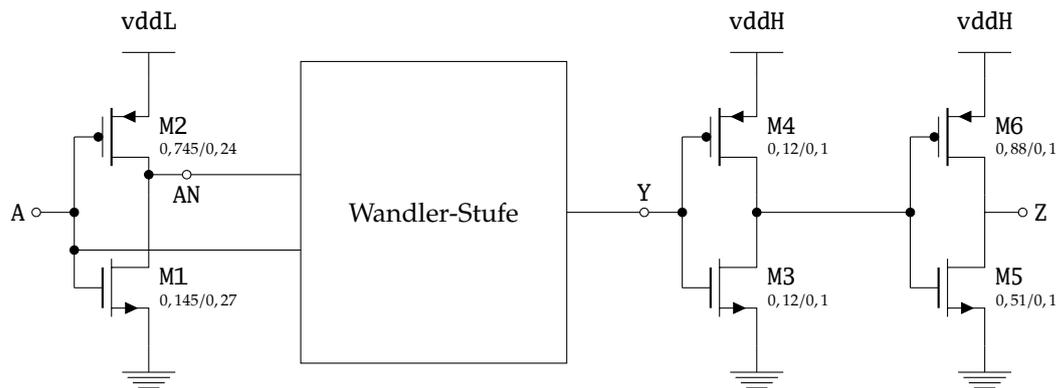


Abbildung 5.4: Prinzipieller Aufbau der betrachteten Aufwärts-Pegelwandler

Signalflanke muss entweder M1 oder M3 die Treiberstärke des korrespondierenden PMOS-Transistors überwinden, um den Halb-Latch zu kippen. Dies erreicht man durch eine geeignete Dimensionierung, d.h. indem die NMOS-Transistoren mit einer hinreichend großen Weite ausgelegt werden. In [39] wird gezeigt, dass das erforderliche Verhältnis von NMOS- zu PMOS-Weite exponentiell ansteigt, wenn die niedrige Versorgungsspannung des Pegelwandlers in den Subschwelligbereich abgesenkt wird, da der *Pull-Down*-Transistor nur einen On-Strom auf Subschwelligniveau liefern kann, während der *Pull-Up*-Zweig des Halb-Latches in starker Inversion arbeitet. In der betrachteten 90-nm-Technologie würde ein Weitenverhältnis von ≈ 2400 benötigt, um den konventionellen Pegelwandler bei einer Versorgungsspannung von 200 mV funktionsfähig zu dimensionieren. Aufgrund der daraus resultierenden großen Weite des NMOS-Transistors ist diese Schaltung ungeeignet als Aufwärts-Pegelwandler für Subschwelligschaltungen.

5.3.1.2 Existierende Implementierungen

Zur Lösung dieser Problematik werden verschiedene Ansätze in der Literatur vorgeschlagen. In der Regel handelt es sich dabei um Abwandlungen der DCVS-Schaltung. In den im Folgenden zitierten Veröffentlichungen erfolgt der Abgriff des Ausgangssignals der eigentlichen Pegelwandler-Stufe für die anschließende Pufferung auf unterschiedlichen Seiten der DCVS-Struktur. Im Interesse einer einheitlichen Darstellung zeigt Abbildung 5.5 alle Pegelwandler ohne Pufferung des Ausgangs und mit Abgriff des Zwischensignals Y auf der rechten, nicht invertierenden Seite der Schaltung.

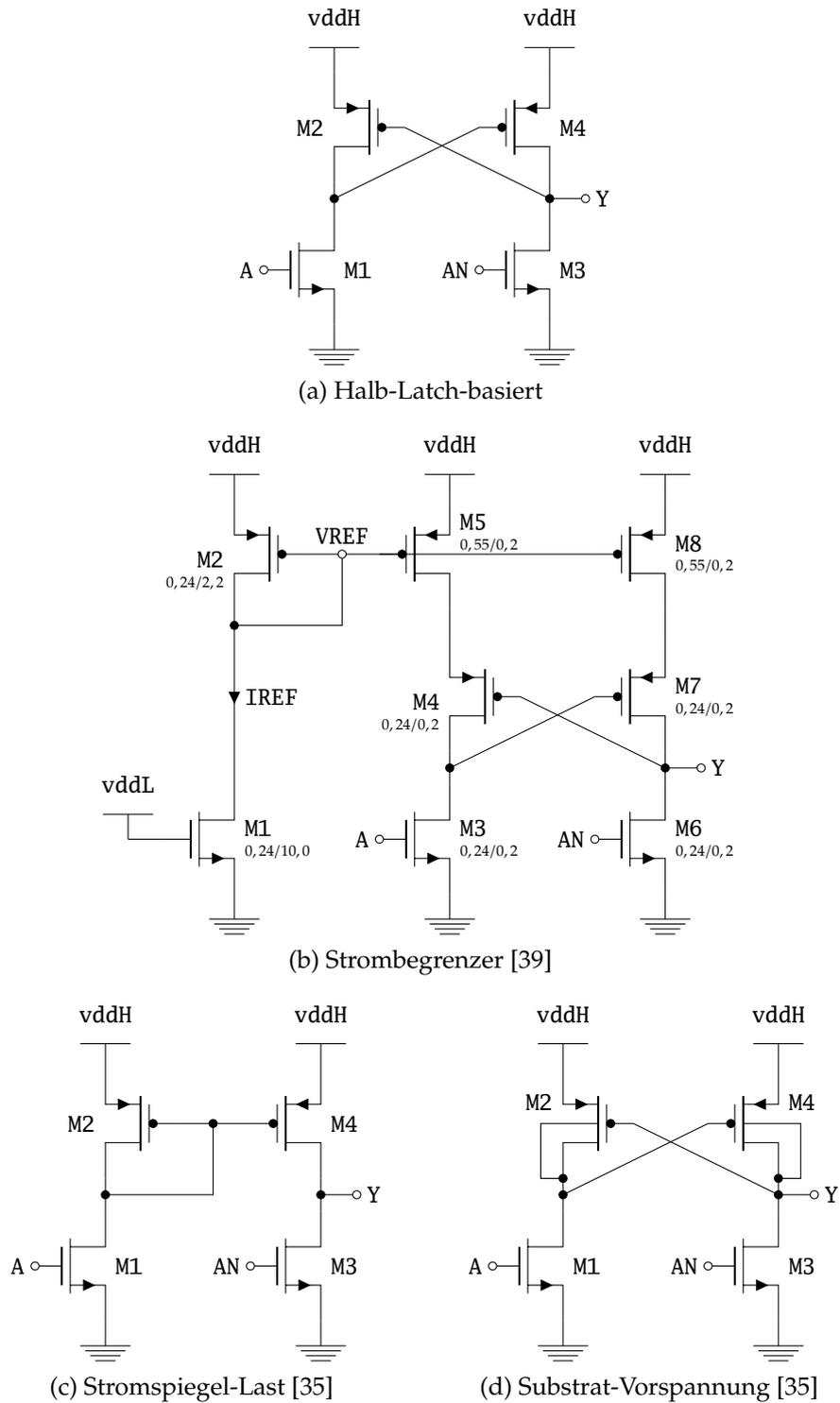
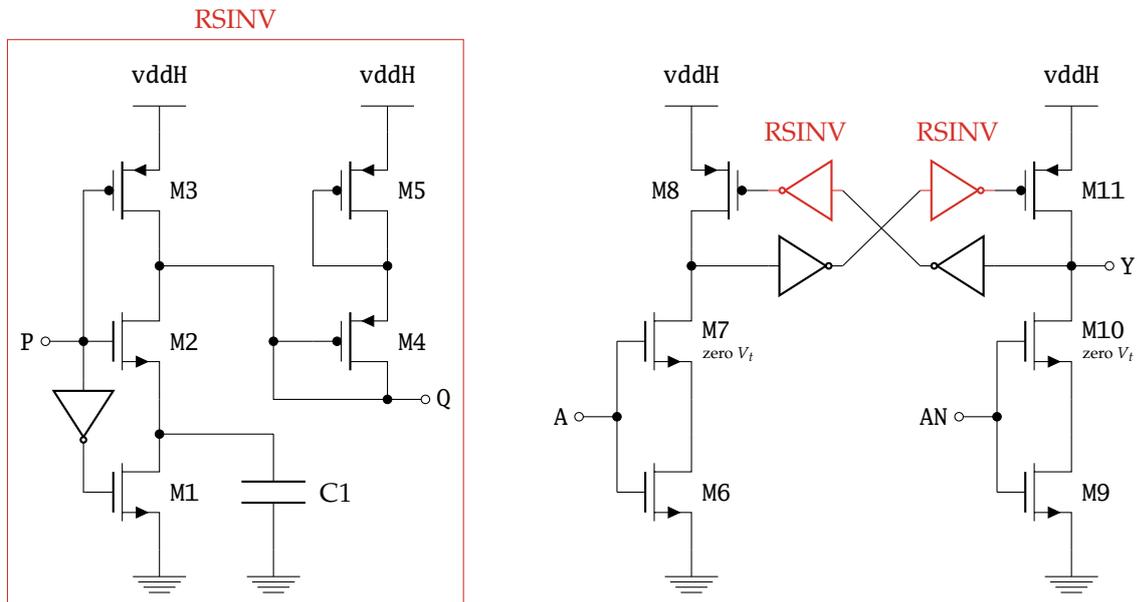
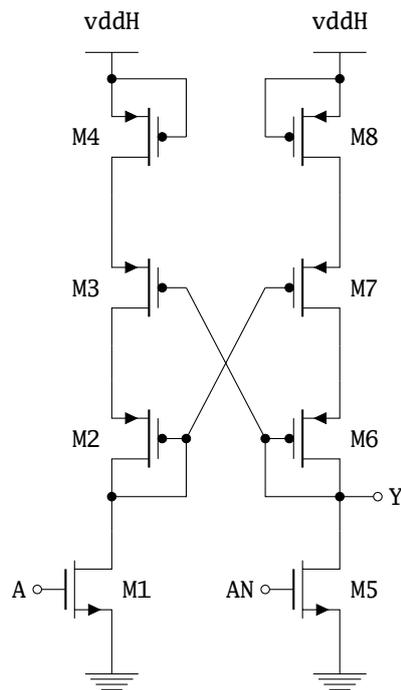


Abbildung 5.5: Unterschiedliche Aufwärts-Pegelwandler



(e) Inverter mit reduziertem Spannungshub [33]



(f) PMOS im Sperrbetrieb [59]

Abbildung 5.5: Unterschiedliche Aufwärts-Pegelwandler (Fortsetzung)

In [39] wird der konventionelle Pegelwandler um PMOS-Strombegrenzer bzw. -Stromspiegel erweitert, die die Treiberstärke des Pull-Up-Zweigs auf Subschwelligniveau reduzieren (siehe Abbildung 5.5b). Diese Schaltung ist in einem weiten Spannungsbereich auf der Eingangsseite funktionsfähig und ist robust gegenüber Schwankungen der Prozessparameter, der Versorgungsspannung und der Temperatur. Ein Nachteil der Schaltung ist die Notwendigkeit des Referenzpfades für die Strombegrenzer, der einen statischen On-Strom IREF benötigt, wodurch sich die Verlustleistung insbesondere für hohe eingangsseitige Versorgungsspannungen erhöht. Eine Möglichkeit diesen Nachteil abzuschwächen besteht darin, einen Referenzpfad für mehrere Pegelwandler beispielsweise in einer lokaler Umgebung einer integrierten Schaltung zu nutzen.

Bei der in [35] vorgestellten Lösung wird der PMOS-Halb-Latch durch eine Stromspiegel-Last ersetzt, wodurch der Referenzpfad der zuvor beschriebenen Lösung entfallen kann (siehe Abbildung 5.5c). Diese Schaltung weist jedoch ebenfalls einen statischen On-Strom durch M1 und M2 auf, wenn am Eingang ein hoher Logikpegel anliegt, woraus ebenfalls eine hohe statische Verlustleistung resultiert.

In [35] werden darüber hinaus drei Pegelwandler präsentiert, bei denen die Treiberstärke der Transistoren durch geeignete Verbindungen der Substrat-Anschlüsse angepasst wird. Eine dieser Schaltungs-Varianten ist in Abbildung 5.5d dargestellt. Diese Technik ist jedoch vor allem für SOI¹-Technologien geeignet, da für jeden Transistor separate Substratanschlüsse benötigt werden. In herkömmlichen CMOS-Technologien wären isolierte p-Wannen erforderlich, die neben zusätzlichen Herstellungsschritten vor allem einen deutlich erhöhten Flächenbedarf aufgrund von einzuhaltenden Abstandsregeln erfordern. Ein weiterer Nachteil dieser Art von Pegelwandler-Schaltungen ist, dass auch hier statische Strompfade durch im Vorwärtsbetrieb befindliche Source-Bulk-Dioden bestehen, die wiederum zur statischen Verlustleistung beitragen.

Die Treiberstärke der Pull-Up-Transistoren kann auch durch das Einfügen von Invertern mit reduziertem Spannungshub in die Rückkopplungspfade des Halb-Latches erreicht werden [33]. In Abbildung 5.5e ist der entsprechende Pegelwandler abgebildet. Durch eine geeignete Dimensionierung kann die Gate-Spannung an den Pull-Up-Transistoren derart begrenzt werden, dass diese lediglich einen On-Strom auf Subschwelligniveau liefern. Die vorgestellte Schaltung benötigt jedoch mit insgesamt 24 Transistoren mehr als doppelt so viele Transistoren wie die zuvor genannten Pegelwandler, was in einem erhöhten Flächenbedarf resultiert. Darüber hinaus besitzt

¹ *Silicon on Insulator*

der Inverter mit reduziertem Spannungshub nach Aussage der Autoren eine hohe Empfindlichkeit gegenüber Prozessparameter-Variationen. Im Subschwellobetrieb ist daher mit starken Exemplarstreuungen sowie ggf. einer hohen Ausfallrate zu rechnen.

Eine weitere Möglichkeit zur Reduktion der On-Ströme des Halb-Latches wird in [59] vorgestellt. Hier kommen als Diode geschaltete sowie im Sperrbetrieb arbeitende PMOS-Transistoren zum Einsatz, um die Treiberstärke des Pull-Up-Netzwerks einzustellen. Durch die Verwendung von Transistoren mit mehreren verschiedenen Schwellspannungen (engl. *Multithreshold CMOS*, MTCMOS) wird ein optimaler Kompromiss zwischen Schaltgeschwindigkeit und Verlustleistung realisiert. Ein Schaltplan des Pegelwandlers ist in Abbildung 5.5f dargestellt. Der Einfachheit halber wurde die in der Original-Schaltung vorhandene Logik zur Deaktivierung des Pegelwandlers hier ausgelassen. Die Schaltung arbeitet in einem weiten Spannungsbereich von 180 mV bis 1 V bei Temperaturen zwischen -40°C und 150°C auch unter ungünstigen Prozessparameter-Bedingungen robust. Ein Nachteil der Schaltung ist, dass die Verzögerungszeit aufgrund der im Sperrbereich betriebenen Transistoren schlecht skaliert, wenn die Versorgungsspannung oberhalb der Schwellspannung liegt. Daher ist diese Schaltung für DVFS-Systeme mit zeitweise höheren Performanzanforderungen weniger geeignet.

5.3.1.3 Optimierung

Für die Implementierung eines Aufwärts-Pegelwandlers für die 90-nm-Subschwellobibliothek wurde die Schaltung nach [39] ausgewählt, bei der die Treiberstärke der PMOS-Transistoren durch Strombegrenzer auf Subschwellobniveau eingestellt wird. Die Dimensionierung des Pegelwandlers wurde mithilfe der Software Cadence Virtuoso Analog Circuit Optimization mit dem Ziel einer Minimierung des *Energy Delay Product* (EDP) [141] durchgeführt, um simultan die Energie pro Schaltvorgang und die Verzögerungszeit zu optimieren. Zusätzlich wurde ein möglichst geringer statischer Querstrom des Referenzpfades bestehend aus M1 und M2 angestrebt. Es wurde eine Spannung von 200 mV auf der Eingangsseite und 1 V auf der Ausgangsseite des Pegelwandlers vorgegeben. Aufgrund der wechselseitig schaltenden Pfade besitzt der Pegelwandler von sich aus eine gute Schaltsymmetrie, die im Anschluss an die Optimierung durch Simulationen überprüft wurde.

Die im Rahmen der Optimierung gewonnenen Transistormaße der zentralen Wandler-Stufe sind in Abbildung 5.5b, die Transistormaße des zusätzlichen eingangsseitigen Inverters und des ausgangsseitigen Puffers in Abbildung 5.4 in der Form W/L in der Einheit μm angegeben. Die Forderung nach möglichst geringem statischem Querstrom führte zur Wahl einer großen Länge von $10\ \mu\text{m}$ für M1 in Verbindung mit einer entsprechenden Gewichtung der Stromspiegel M2, M5, M8. Dieser n-Kanal-Transistor wurde daraufhin in eine Reihenschaltung aus fünf Transistoren mit jeweils $2\ \mu\text{m}$ Kanallänge aufgeteilt, um ein günstigeres mäanderförmiges Layout zu ermöglichen.

Zur Reduktion von Flächenbedarf und statischer Verlustleistung ist es, wie bereits erwähnt, möglich eine gemeinsame Referenzspannungsquelle für mehrere Pegelwandler, die sich beispielsweise in einer gewissen Umgebung innerhalb einer integrierten Schaltung befinden, zu nutzen. Hierbei ergibt sich allerdings der zusätzliche Verdrahtungsaufwand für die Referenzspannung V_{REF} , die als Analog-Signal zu behandeln ist, so dass ggf. vergrößerte Abstände zu anderen Signalen eingehalten werden müssen oder sogar eine Schirmung des Signals erforderlich ist. Andernfalls kann es durch kapazitive Kopplungen mit anderen Signalen in Verbindung mit der hohen Ausgangsimpedanz der Referenzspannungsquelle zu Übersprech-Störungen (Crosstalk) und damit einem veränderten Verhalten oder gar einer Fehlfunktion des Pegelwandlers kommen. Trotz vergrößerten Signalabständen und Schirmung kann jedoch eine Verschlechterung der Spannungsstabilität mit steigender Anzahl an gemeinsam an einem Referenzpfad betriebenen Pegelwandler nicht gänzlich verhindert werden. Aus diesem Grund und um die Handhabung des Pegelwandlers durch Synthese- und *Place-and-Route*¹-Werkzeuge zu vereinfachen, wird im Rahmen dieser Arbeit auf die Möglichkeit einer gemeinsamen Nutzung der Referenzspannungsquelle verzichtet und der Pegelwandler als vollständige, abgeschlossene Komponente behandelt.

5.3.2 Abwärts-Pegelwandler

Abbildung 5.6 zeigt den Schaltplan des aus zwei aufeinander folgenden Inverterstufen aufgebauten Abwärts-Pegelwandlers.

Die Kanallänge in der mit der höheren Versorgungsspannung betriebenen ersten Stufe beträgt $100\ \text{nm}$ und in der mit der niedrigen Versorgungsspannung betriebe-

¹ Platzieren und Verdrahten

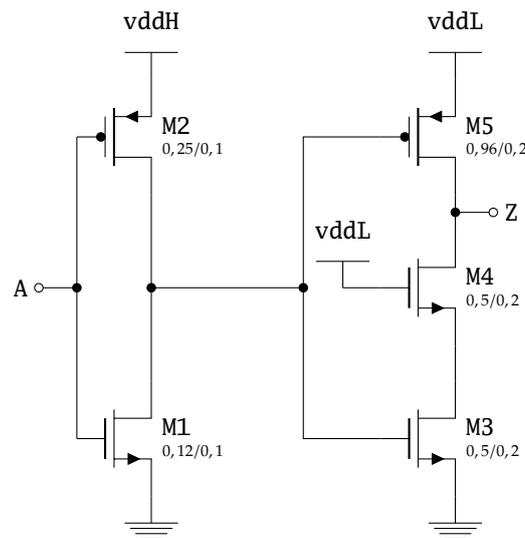


Abbildung 5.6: Schaltplan des Abwärts-Pegelwandlers

nen zweiten Stufe 200 nm. Durch die doppelte Kanallänge im empfindlichen Teil des Pegelwandlers wird wie bei allen übrigen Zellen der Bibliothek eine höhere Robustheit erreicht.

Der zusätzliche Transistor M4 im ausgangsseitigen Inverter wurde im Interesse eines symmetrischen Schaltverhaltens in die Schaltung eingefügt, wie nachfolgend erläutert werden soll. Ist der PMOS-Transistor M5 durchgeschaltet, so beträgt seine Gate-Source-Spannung $-V_{DDL}$, nämlich dann, wenn ein *High*-Pegel (logisch 1) an A anliegt. M5 kann also einen Strom auf Subschwelniveau liefern. Ist dagegen M3 durchgeschaltet, so beträgt hier die Gate-Source-Spannung V_{DDH} – dabei liegt ein *Low*-Pegel (logisch 0) an A an. Damit kann durch M3 ein Strom fließen, der weit über Subschwelniveau liegt. Diese stark asymmetrische Stromergiebigkeit kann zu nicht optimalen, asymmetrischen Störabständen und damit zu schlechter Robustheit führen. Außerdem führt die schnelle Abfallzeit des Pegelwandlers zu steilen Signalflanken, wodurch verstärkt Übersprechen zu anderen Subschwellsignalen auftritt. Dies kann letztlich zu einer Fehlfunktion der Schaltung führen. Daher begrenzt der zusätzlich eingefügte Transistor M4 den Stromfluss im Pulldown-Pfad ebenfalls auf Subschwelniveau.

Bei der Dimensionierung der Transistorweiten wurde als Optimierungswerkzeug wie auch beim Aufwärts-Pegelwandler die Software Cadence Virtuoso Analog Circuit Optimization genutzt. Ziel waren maximale Störabstände bei minimalen

Verzögerungszeiten. Diese Optimierungen wurden bei 200 mV durchgeführt und anschließend im Bereich bis 1 V überprüft.

5.4 Layout

Abbildung 5.7 zeigt beispielhaft die Layouts des Inverters und des Flipflops. Es kommt der gleiche Standardzellenrahmen wie bei den konventionellen Standardzellen der Technologie zum Einsatz.

5.5 ALU-Prototyp

Zur Überprüfung der entwickelten 90-nm-Subschwelligbibliothek sowie der Werkzeugkette zur Zellen-Charakterisierung, Synthese und Place-and-Route wurde ein prototypischer integrierter Schaltkreis entworfen, der eine auf dieser Bibliothek basierende Digitalschaltung enthält. Neben einer rein funktionalen Prüfung der Testschaltung sollten insbesondere Messergebnisse zur minimal möglichen Versorgungsspannung ohne Fehlfunktion sowie die Abhängigkeiten der Schaltungseigenschaften – maximale Taktfrequenz und Verlustleistung bzw. Energieaufnahme – von der Versorgungsspannung gewonnen werden. Außerdem sollte untersucht werden, ob die durch den Technologie-Anbieter verfügbaren Transistormodelle im Subschwelligbereich zuverlässige Daten liefern. Da der Subschwelligbereich der Transistormodelle in Zusammenhang mit konventionellen Digital- oder Analog-Schaltungen im Wesentlichen zur Abschätzung der statischen Verlustleistung, nicht jedoch zur exakten Beschreibung bzw. Simulation des funktionalen Schaltungsverhaltens benötigt wird, besteht die Möglichkeit, dass bei der hier vorgesehenen Verwendung Abweichungen zwischen Simulation und Realität auftreten. Aufgrund der exponentiellen Abhängigkeiten im Subschwelligbereich können diese ggf. ein signifikantes Ausmaß annehmen.

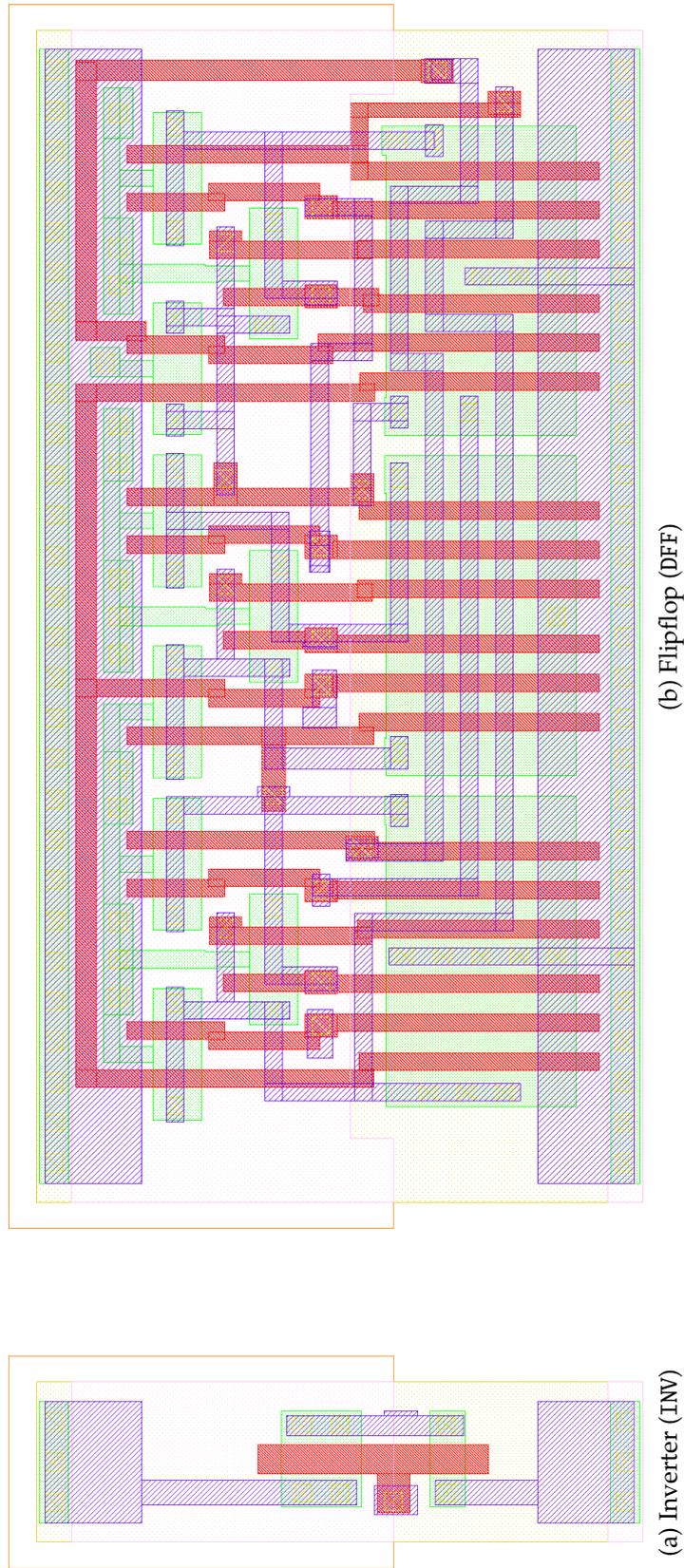


Abbildung 5.7: Layouts des Inverters und des Flipflops der 90-nm-Subschwelligbibliothek

5.5.1 Architektur

Die Gesamtarchitektur des Testbausteins ist in Abbildung 5.8 dargestellt. Der größte Teil des Schaltkreises besteht aus vier identischen ALUCORE-Modulen mit den Instanznamen CORE_0 bis CORE_3, die jeweils aus einer 32-bit-ALU und zwei Operandenregistern bestehen. Das BASIC-Modul enthält je eine Instanz der Zellen NAND2, NOR2 und AOI, sowie zusätzliche Eingangsregister. Ein Ringoszillator bildet den OSC-Block. Alle Moduleingänge sind über Abwärts-Pegelwandler an die Eingangspads angeschlossen, auf der anderen Seite sind zwischen alle Modulausgänge und die Ausgangspads Aufwärts-Pegelwandler geschaltet. Für jede ALUCORE-Instanz sind eigene Pegelwandler vorgesehen, um für den Fall, dass ein Pegelwandler nicht funktionsfähig ist, die übrigen, nicht betroffenen Module trotzdem verwenden und testen zu können. Die Operanden werden bitweise über Schieberegister in den Baustein geladen, da für das parallele Anlegen der Daten zu wenig Pads auf der zur Verfügung stehenden Fläche untergebracht werden können (siehe Abbildung 5.11). Jede ALUCORE-Instanz besitzt aus den selben Redundanzgründen wie zuvor ein eigenes Schieberegister, wobei jedoch alle Schieberegister mit gleichen Eingangsdaten über das SCAN_IN-Pad beschrieben werden.

Abbildung 5.9a zeigt den Aufbau der ALUCORE-Module. OPA_REG und OPB_REG sind 32-bit-Register, deren Flipflops zu einer *Scan Chain*, d.h. als Schieberegister zusammen geschaltet werden können (siehe Abbildung 5.9b). Über den SCAN_EN-Eingang kann zwischen normalem Rechenbetrieb und seriellem Schiebetrieb umgeschaltet werden. Die 32 bit breite ALU unterstützt die in Tabelle 5.2 aufgelisteten Instruktionen. Dabei sind die logischen Verknüpfungen NAND, NOR und XOR bitweise zu verstehen. Im Schiebetrieb (High-Pegel an SCAN_EN) werden mit der steigenden Taktflanke die an SCAN_IN angelegten Daten bitweise durch die beiden Operandenregister geschoben. Die dabei heraus geschobenen Daten sind über den SCAN_OUT-Ausgang jeder ALUCORE-Instanz zugänglich. Im normalen Rechenbetrieb (Low-Pegel an SCAN_EN) wird das von der ALU berechnete Ergebnis in das als Akkumulator dienende Operandenregister OPA_REG zurück geschrieben, während OPB_REG den zuvor gespeicherten Wert beibehält. Hierdurch wird erreicht, dass nach dem Initialisieren der Operanden mehrere Takte hintereinander mit maximaler Frequenz Instruktionen ausgeführt werden können, wodurch sich beispielsweise ein 32-bit-Zähler realisieren lässt. Anschließend kann das Ergebnis über die Schieberegister ausgelesen und überprüft werden. Aufgrund der verfügbaren Chipfläche wurde das ALUCORE-Modul viermal instanziiert, um die Anzahl funktionierender

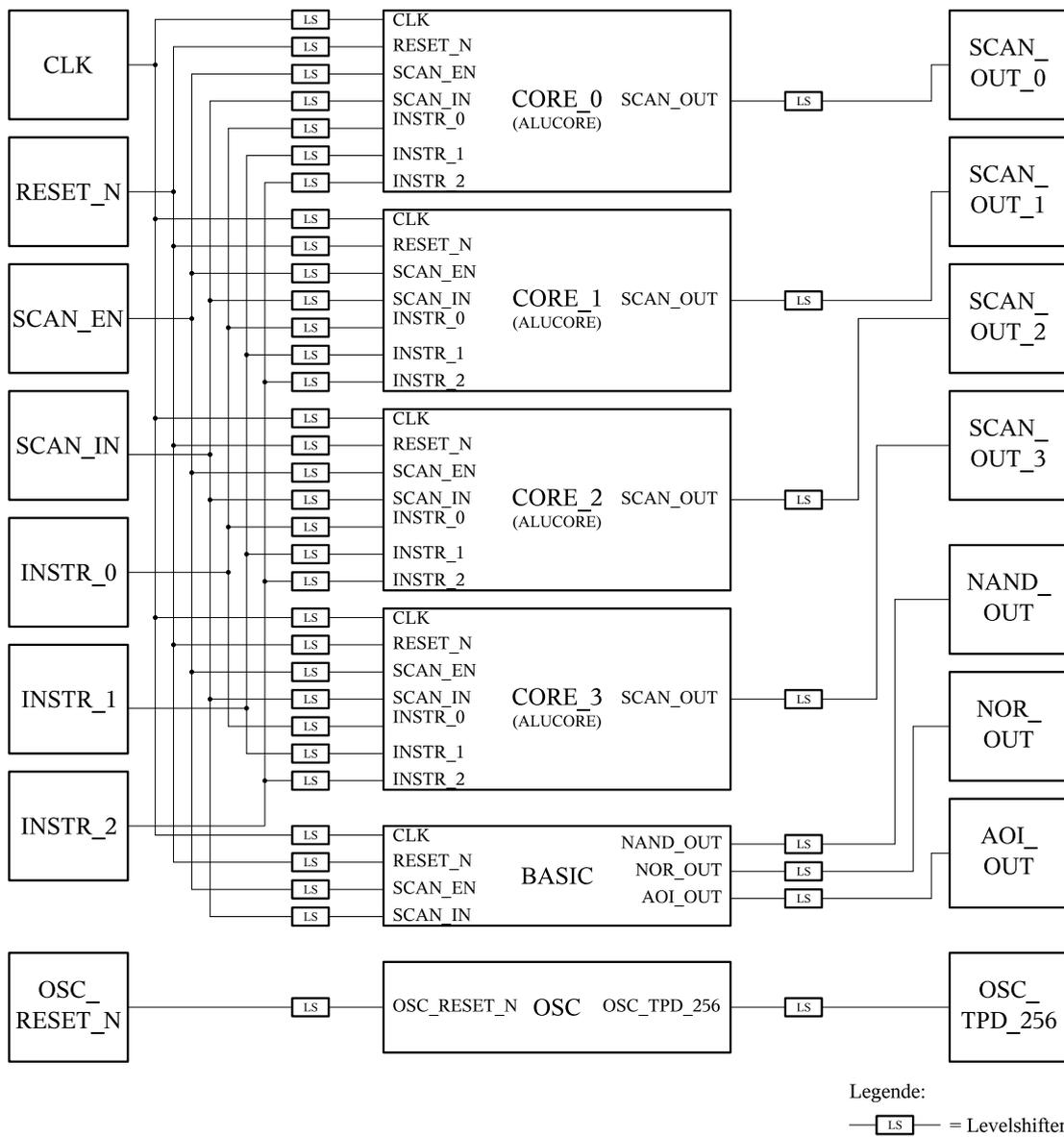


Abbildung 5.8: Architektur des ALU-Prototypen

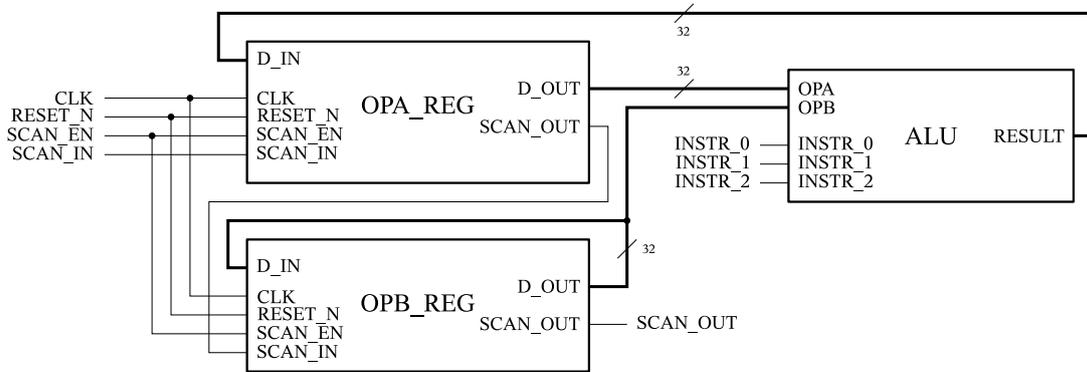
Instruktion	Codierung	RESULT
NOP	000	OPA
NAND	001	$\overline{OPA \wedge OPB}$
NOR	010	$\overline{OPA \vee OPB}$
XOR	011	$(OPA \wedge \overline{OPB}) \vee (\overline{OPA} \wedge OPB)$
ADD	100	OPA + OPB
SUB	101	OPA – OPB
SHL	110	OPA(30...0), "0"
ROL	111	OPA(30...0), OPA(31)

Tabelle 5.2: Instruktionen der ALU

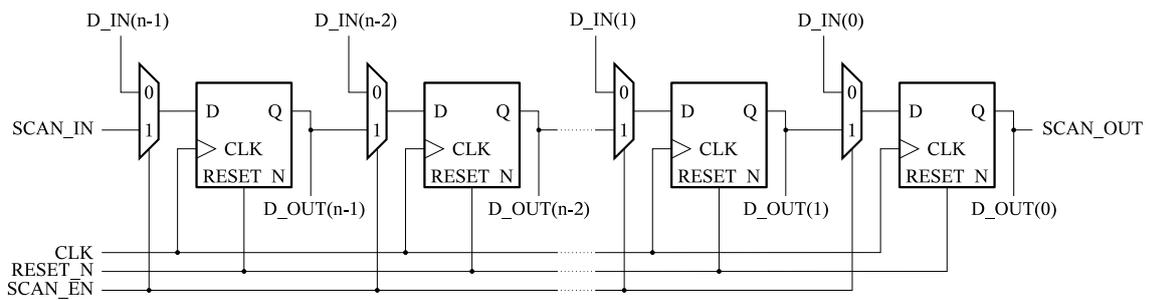
Module zu vergrößern und möglichst viele verschiedene Exemplare vergleichen zu können.

Während die vier ALUCORE-Module zum Testen größerer Zusammenschaltungen der Subschwellig-Standardzellen dienen, ist das BASIC-Modul zum Testen kleinerer Logikschaltungen gedacht. Abbildung 5.9c zeigt deren Aufbau. Die Eingangsdaten für die drei Grundgatter werden über ein 4-bit-Schieberegister zur Verfügung gestellt. Ein SCAN_OUT-Ausgang ist nicht vorhanden, da die Daten innerhalb des Schieberegisters nicht verändert werden. NAND2-, NOR2- und AOI-Gatter wurden im VHDL-Code hart instanziiert, um zu vermeiden, dass das Synthese-Werkzeug die zugehörigen Logikfunktionen auf andere Weise realisiert. Die Gatterausgänge sind über Pegelwandler direkt auf Ausgangspads geführt, so dass die Funktion der Einzelgatter z.B. betriebsspannungsabhängig überprüft werden kann.

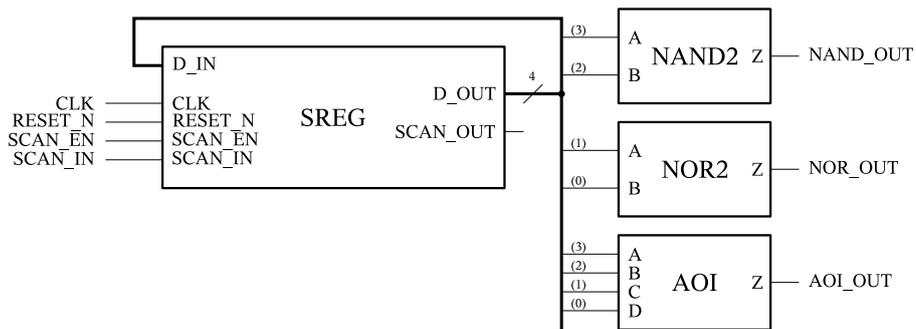
Eine Untersuchung, bei welcher Frequenz die Subschwellig-Zellen in Abhängigkeit von der Versorgungsspannung verwendet werden können, ist mit den ALUCORE- und BASIC-Modulen nur durch „Probieren“ möglich, d.h. es kann nur geprüft werden, ob die Schaltung mit der angelegten Taktfrequenz noch korrekt arbeitet. Eine direkte Messung der Gattergeschwindigkeit ist durch den OSC-Block möglich. Hierbei handelt es sich um einen Ringoszillator, der aus 33 Inverterstufen und einem 8-bit-Takteiler besteht. Diese Schaltung wurde als *Hard-Macro*, d.h. als vollständig platzierter und verdrahteter Block realisiert, da das Layout handoptimiert und zur Vermeidung von Störungen anderer Schaltungsteile mit einem umlaufend dotierten Schutzring (engl. *Guard Ring*) versehen wurde. Der Schaltplan des OSC-Blocks ist in Abbildung 5.10 dargestellt. Eine der Inverterstufen ist durch ein NAND2-Gatter realisiert. Der dadurch vorhandene zusätzliche Eingang $\overline{OSC_RESET}$ kann



(a) ALUCORE-Module



(b) Operandenregister (für Wortbreite n)



(c) BASIC-Block

Abbildung 5.9: Implementierungsdetails einzelner Module

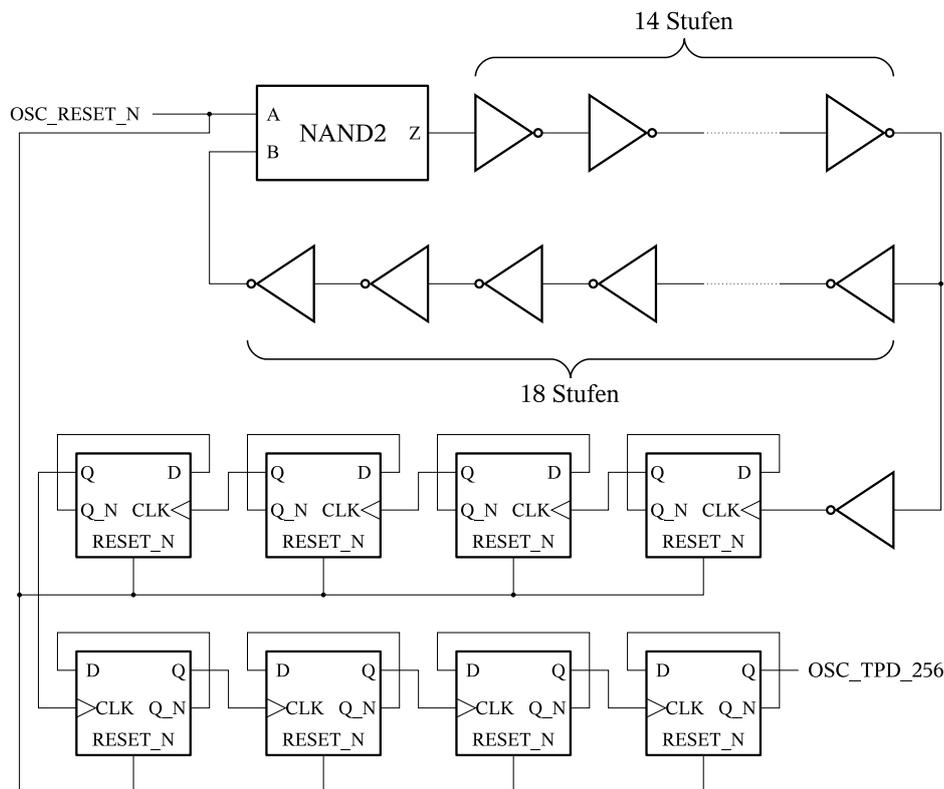


Abbildung 5.10: Aufbau des Ringoszillators

zum Anhalten der Oszillation verwendet werden, während andere Schaltungsteile getestet werden, so dass Störungen durch den Ringoszillator sicher auszuschließen sind. Vom NAND2-Gatter aus gesehen nach 14 Inverterstufen wird das Oszillatorsignal abgegriffen und durch einen weiteren Inverter gepuffert. Die gewählte Position hat den Vorteil, dass zum einen die beiden „Störungen“ des Rings (NAND2-Gatter und Abgriff des Oszillationssignals) relativ weit auseinander liegen und dadurch Asymmetrien vermindert werden, und zum anderen ein Layout mit besonders günstiger Verdrahtung erreicht wird. Die vom Inverter-Ring erzeugte Taktfrequenz wird anschließend durch acht Flipflops (durch Rückkopplung von \bar{Q} auf D als T-Flipflop geschaltet) auf $1/256$ herunter geteilt. Das Ausgangssignal des Taktteilers wird dann über Pegelwandler und Pad extern zur Verfügung gestellt. Das Herunterteilen des Taktsignals ist notwendig, da beim Betrieb mit 1 V Versorgungsspannung Oszillationsfrequenzen auftreten, die außerhalb des Spezifikationsbereichs der verfügbaren Padzellen liegen.

5.5.2 Synthese und Layout

Die Implementierung des Testbausteins erfolgte basierend auf einer Hardware-Beschreibung in einem Semi-Custom-Entwurfsablauf [179]. Dabei kam die Software Synopsys Design Compiler als Synthese-Werkzeug sowie Cadence SoC Encounter als Layout-Werkzeug zum Einsatz. Da der Entwurf Teilkomponenten mit unterschiedlichen Versorgungsspannungen enthält – die ALUCORE-Module arbeiten bei einer Spannung von 200 mV, während die IO-Logik sowie die Padzellen mit 1 V betrieben werden – ist ein MSV-gerechter Ablauf von Synthese und Layout erforderlich, wozu entsprechende Skripte entwickelt wurden. Die Synthese ist dabei derart eingerichtet, dass Pegelwandler automatisch an allen Stellen eingefügt werden, an denen Signale von einer Spannungsdomäne in eine andere übertreten. Die Auswahl der korrekten Pegelwandler-Richtung erfolgt durch das Synthese-Werkzeug. Zielvorgabe für die Synthese war eine Taktfrequenz von 1,5 MHz unter typischen Betriebsbedingungen (Prozess, Spannung und Temperatur). Diese Frequenz liegt einige Prozentpunkte unter der durch Versuchs-Synthesen ermittelten maximal möglichen Taktfrequenz für den gegebenen Entwurf, wodurch ein übermäßiges starkes Anwachsen von Gatteranzahl bzw. Fläche und Verlustleistung vermieden wird. Insbesondere die Vermeidung einer unnötig hohen Verlustleistung bei nur geringfügig reduzierter Performanz ist eine wichtige Entwurfsentscheidung für die Implementierung energieeffizienter Systeme.

Das Layout des ALU-Prototypen ist in Abbildung 5.11 als Chipfotografie dargestellt. Es enthält zwei Spannungsdomänen, in denen die Standardzellen für 200 mV und 1 V Versorgungsspannung getrennt angeordnet sind. Dabei umschließt die schmale, auf dem Foto kaum sichtbare 1-V-Domäne die innere 200-mV-Domäne ringförmig. Sie enthält insbesondere alle Pegelwandler-Zellen, die auf diese Weise bestmöglich zwischen Ein- oder Ausgangspad und zugehörige Kernlogik platziert werden können. Die innere Spannungsdomäne ist in mehrere Bereiche aufgeteilt, um die Anordnung der einzelnen Module zu definieren. Im oberen Teil sind vier gleich große Bereiche für die ALUCORE-Instanzen ALU1 bis ALU4 vorgesehen. Die Fläche darunter wird für das Standardzellen-Testmodul BASIC und den Ringoszillator OSC verwendet. Die Anordnung der Pads wurde so ausgelegt, dass möglichst kurze Verbindungsleitungen zur Kernlogik entstehen und eine gleichmäßige Verteilung der Versorgungspads gewährleistet ist.

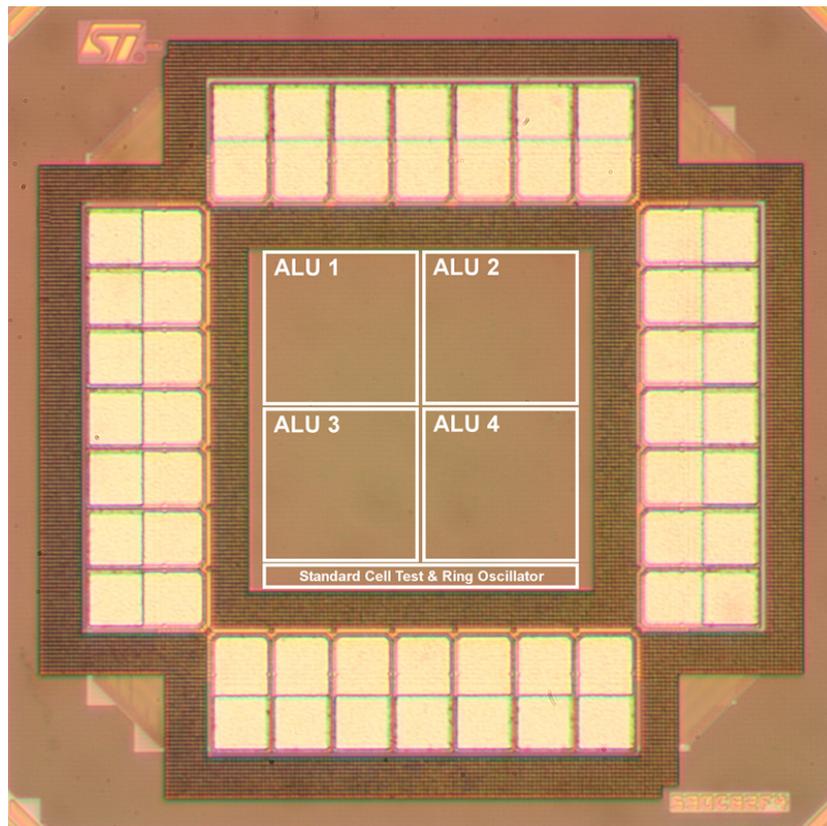


Abbildung 5.11: Foto des ALU-Prototypen

5.5.3 Messergebnisse

Im Rahmen einer Prototypen-Fertigung wurden 15 Exemplare des Testbausteins hergestellt und in Gehäuse integriert. Somit standen insgesamt 60 gefertigte ALUs für Messungen zur Verfügung.

Für die nachfolgend beschriebenen Messergebnisse wurde folgender Testablauf verwendet: Bei jeder Messung wird zunächst eine zu testende Versorgungsspannung festgelegt und an das Testexemplar angeschlossen. Anschließend werden die vorgegebenen Eingangsdaten mit Hilfe der Schieberegister als Operanden der ALUs geladen, wobei alle ALUs mit den selben Daten initialisiert werden. Hierfür werden 64 Taktzyklen benötigt. Durch eine niedrige Taktfrequenz wird sichergestellt, dass bei dieser Initialisierung möglichst keine Fehler auftreten. Anschließend werden die ALUs in den Rechenbetrieb umgeschaltet und es wird eine ebenfalls für diese Messung vorgegebene Arbeits-Taktfrequenz angelegt. Insgesamt werden in dieser

Berechnungsphase durch die ALUs jeweils 1 000 000 Additionen ausgeführt, wobei alle Bits des Zwischenergebnis-Registers bei korrekter ALU-Funktion mehrfach ihren Logikpegel wechseln. Die Addition wurde als auszuführende Operation ausgewählt, weil der kritische Pfad der ALU durch den Addierer verläuft. Das Endergebnis wird im Anschluss bei wiederum reduzierter Taktfrequenz über die Schieberegister ausgelesen und auf Korrektheit überprüft. Mit Hilfe dieses Testverfahrens lässt sich durch „Ausprobieren“ bestimmen, bis zu welcher maximalen Taktfrequenz die einzelnen Exemplare in Abhängigkeit von der Versorgungsspannung fehlerfrei betrieben werden können.

Das Ergebnis dieser Messungen ist in Abbildung 5.12a dargestellt. Alle Messungen wurden bei einer Temperatur von 25 °C durchgeführt. Bei der Synthese bzw. der Layout-Erzeugung wurde unter typischen Prozess- und Betriebsbedingungen eine Taktfrequenz von 1,5 MHz bei einer Versorgungsspannung von 200 mV vorgegeben. ALU-Exemplare, die diese Synthese-Vorgabe nicht erfüllen, sind in der Abbildung rot dargestellt. Ferner sind drei Exemplare, die knapp unterhalb von 200 mV im Vergleich zu den übrigen Exemplaren eine stärker abfallende Taktfrequenz aufweisen, ebenfalls rot gezeichnet.

Bei einer Versorgungsspannung von 200 mV erfüllen 85 % der getesteten ALUs die obige Synthese-Vorgabe. Dies deutet darauf hin, dass die Simulationsmodelle der Transistoren eine hinreichende Genauigkeit für die Beschreibung des Verhaltens im Subschwelligbereich besitzen. Es sind 97 % aller ALUs bei 200 mV unabhängig von der Taktfrequenz voll funktionsfähig. Die übrigen zwei Exemplare liefern auch bei langsamster Taktfrequenz ein falsches Rechenergebnis. Beim weiteren Absenken der Versorgungsspannung geht die Ausbeute korrekt arbeitender ALUs weiter zurück und erreicht einen Wert von 35 % bei 120 mV. Knapp darunter, bei einer Spannung von 115 mV, berechnen noch fünf ALU-Exemplare das korrekte Additionsergebnis. Die statistische Verteilung der niedrigsten möglichen Versorgungsspannung ist in Abbildung 5.12b in Form eines Histogramms dargestellt. Der Stichprobenmittelwert der minimalen Versorgungsspannung beträgt etwa 143 mV. Im Vergleich mit den in Tabelle 3.1 angegebenen Implementierungen ist dieses Ergebnis für eine Schaltung entsprechender Komplexität als sehr gut zu bewerten, insbesondere vor dem Hintergrund, dass für den Testbaustein keine Regelung der Substratvorspannung zum Abgleich der Leitfähigkeiten von NMOS- und PMOS-Transistoren verwendet wird.

Abbildung 5.12c zeigt die Energieaufnahme der ALUs pro Takt bzw. pro Instruktion. Hierbei ist zu beachten, dass aufgrund der limitierten Pad-Anzahl die Energieaufnahme

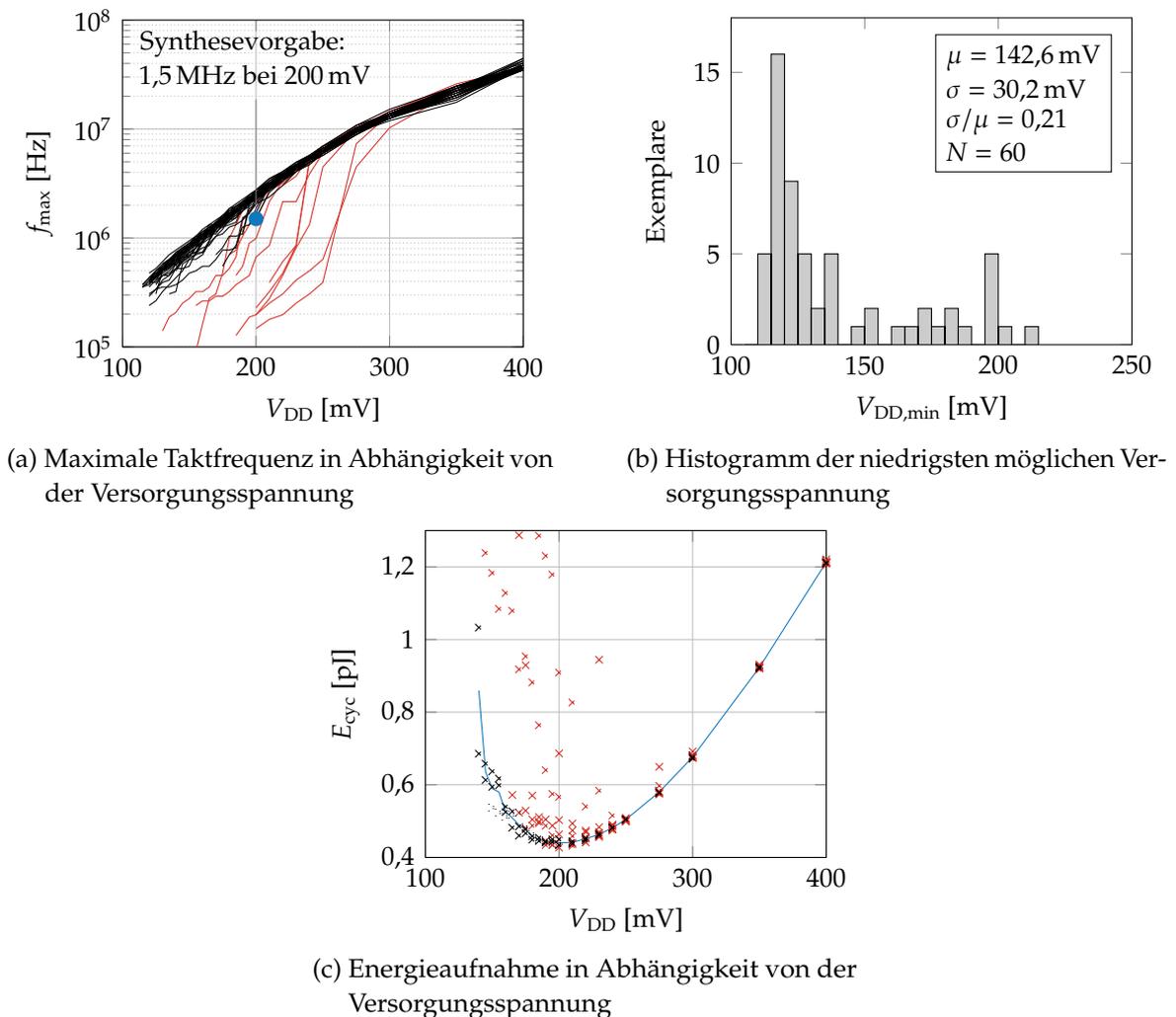


Abbildung 5.12: Messergebnisse der ALU-Prototypen

me nur für den jeweils gesamten aktuell untersuchten Testbaustein gemessen werden konnte und nicht separat für jede ALU-Instanz. Dabei wurden die Messungen jeweils bei der höchsten Taktfrequenz durchgeführt, bei der alle ALUs des Testexemplars im aktuellen Spannungsschritt jeweils noch voll funktionsfähig waren. Die als Kreuze dargestellten Einzel-Messwerte in der Abbildung sind somit als Mittelwerte der vier ALUs bei dieser Taktfrequenz zu verstehen. Schwarze Kreuze kennzeichnen in der Abbildung Schaltungsexemplare, bei denen alle ALUs die Synthese-Vorgabe erfüllen. Für diese Exemplare ist zusätzlich eine Ausgleichskurve durch Mittelwert-Bildung eingefügt. Alle übrigen Messwerte sind rot dargestellt. Da ein Einbrechen der Taktfrequenz mit einer entsprechenden Zunahme der Periodendauer verbunden ist, steigt in diesem Fall der Anteil der statischen Energieaufnahme an der Gesamtenergie

pro Takt im gleichen Verhältnis. Dies erklärt die zahlreichen Abweichungen der Messwerte oberhalb des Energie-Durchschnitts, da bereits die Frequenzabnahme einer einzelnen ALU auf einem Exemplar genügt, um diesen Effekt zu bewirken. Schaltkreise, bei denen alle ALUs die Synthese-Vorgabe erfüllen, zeigen dagegen ein homogenes Verhalten bezüglich der Energieaufnahme, wie sich anhand der schwarz markierten Messwerte und der Ausgleichskurve erkennen lässt. Die minimale Energieaufnahme der ALUs beträgt 0,45 pJ und tritt bei einer Versorgungsspannung von 210 mV sowie einer Taktfrequenz von 3 MHz auf. Bei dieser Spannung ist nur ein einzelnes ALU-Exemplar nicht funktionsfähig. Das Energieminimum liegt nahe am Entwurfsziel der Subschwelligbibliothek von 200 mV.

Zur Bewertung der oben aufgeführten Eigenschaften der ALU folgt nun ein Vergleich mit einer ähnlichen Implementierung. In [20] wird ein 32-bit-Kogge-Stone-Addierer [84] mit einer dem ALU-Prototypen ähnlichen Akkumulator-Struktur ebenfalls in einer 90-nm-Technologie vorgestellt. Da der Addierer des ALU-Prototypen den größten Anteil der Logik sowie die Verzögerung des kritischen Pfades bestimmt, sind beide Architekturen in etwa vergleichbar. Der Addierer aus [20] weist eine minimale Energieaufnahme von 0,1 pJ pro Takt bei einer Spannung von 330 mV und einer Taktfrequenz von 50 kHz auf. Der Energiebedarf ist damit um einen Faktor 4,5 geringer, die Taktfrequenz jedoch um einen Faktor 60 geringer als die entsprechenden Werte einer ALU im optimalen Betriebspunkt. Damit ist das EDP des Addierers aus [20] im Energieminimum mit 2 aJs um einen Faktor 13,3 größer (schlechter) als das EDP der ALU von 0,15 aJs. Zudem tritt das Energieminimum des Addierers bei einer höheren Spannung auf als das Energieminimum der ALU. Diese Beobachtungen deuten darauf hin, dass für die Implementierung aus [20] eine sogenannte Low-Power-Technologie mit größerer Gateoxid-Dicke und höherer Schwellspannung eingesetzt wurde. Die hier verwendete Technologie ist dagegen eine General-Purpose-Technologie mit geringerer Gateoxid-Dicke und Schwellspannung. Es ist ferner zu bemerken, dass die minimale Versorgungsspannung des Addierers aus [20] mit 200 mV deutlich höher liegt als die minimale Versorgungsspannung der ALU von 143 mV.

5.6 Zusammenfassung

In diesem Kapitel wurde der Entwurf einer Standardzellen-Bibliothek für den Subschwelligbetrieb in einer 90-nm-Technologie von STMicroelectronics betrachtet. Die

Bibliothek enthält 15 Logikgatter, deren hohe Robustheit durch eine Optimierung der Schaltsymmetrie bei 200 mV erreicht wurde. Basierend auf diesen Grundelementen wurde ein prototypischer integrierter Schaltkreis mit vier 32-bit-ALUs implementiert, gefertigt und charakterisiert. Die ALUs besitzen eine minimale Energieaufnahme von 0,45 pJ bei 210 mV und 3 MHz. Es wird bei dieser Spannung eine Ausbeute funktionsfähiger ALU-Exemplare von 98,3 % erreicht. Der Stichprobenmittelwert der minimalen Versorgungsspannung beträgt etwa 143 mV. Bei einer Spannung von 115 mV sind noch fünf Exemplare voll funktionsfähig. Dieses Ergebnis ist für eine Schaltung entsprechender Komplexität ohne eine Regelung der Substratspannung zur Kompensation von Prozessschwankungen beachtlich.

Das nachfolgende Kapitel behandelt den Entwurf einer weiteren Standardzellen-Bibliothek für den Subschwelligbetrieb. Dabei kommt eine 65-nm-Technologie sowie ein Mehrzieloptimierungsverfahren für die Zellendimensionierung zum Einsatz. Dieses Verfahren gestattet die Berücksichtigung weiterer Zielfunktionen wie die Verzögerungszeiten oder den Energiebedarf, so dass ein günstigerer Kompromiss im Entwurfsraum gefunden werden kann.

6 Mehrzieloptimierte Subschwelligbibliothek in 65 nm CMOS

Im vorangegangenen Kapitel wurde eine Standardzellen-Bibliothek vorgestellt, deren Dimensionierung mit dem Fokus maximaler Robustheit im Form eines symmetrischen Schaltverhaltens im Subschwelligbereich durchgeführt wurde. Nachteilig an diesem Vorgehen ist, dass weitere Kriterien wie etwa die Schaltgeschwindigkeit oder der Energiebedarf der Gatter beim Entwurf nicht berücksichtigt werden. Dieses Kapitel befasst sich mit der Entwicklung einer mehrzieloptimierten Gatterbibliothek. Es wurden mehrere Maßnahmen ergriffen, um bestmögliche Eigenschaften der daraus später aufgebauten Digitalschaltungen zu erzielen.

Als Grundlage dieser Standardzellen-Bibliothek wurde eine moderne 65-nm-CMOS-Technologie von STMicroelectronics ausgewählt, die eine Auswahl an insgesamt sechs unterschiedlichen Transistortypen (jeweils NMOS und PMOS) für digitale oder analoge Schaltungen bietet¹. Dabei stehen zwei unterschiedliche Gateoxid-Dicken für niedrige Verlustleistung (*Low Power*, LP) oder hohe Performanz (*General Purpose*, GP) sowie jeweils drei unterschiedliche Schwellspannungen (*Low Vt* (LVT), *Standard Vt* (SVT), *High Vt* (HVT)) zur Verfügung. Die Transistorauswahl fiel hier zunächst im Interesse einer niedrigen Verlustleistung auf den LP-Transistortyp und innerhalb dieser Gruppe auf die mittlere Schwellspannung SVT, um einen Kompromiss bezüglich der erreichbaren Taktfrequenzen zu erzielen. Es können jedoch später bei Bedarf ohne Geometrie-Änderungen Gatter der anderen Transistorarten durch Hinzufügen von Markierungen in entsprechenden CAD²-Ebenen realisiert werden. Die nachfolgenden Betrachtungen beziehen sich jedoch stets auf LPSVT-Transistoren.

¹ Darüber hinaus existieren weitere Transistortypen für I/O-Zwecke und SRAM-Zellen.

² *Computer-Aided Design*

Eine ausschließliche Betrachtung der Robustheit als Kriterium zur Dimensionierung von Gattern führt zu einer unnötigen Einschränkung der Sichtweise, da eine solche Einzielloptimierung keine Information über weitere Schaltungseigenschaften in der Umgebung eines gefundenen Optimums liefert. Es ist beispielsweise denkbar, dass es Dimensionierungen gibt, die deutlich geringere Verlustleistung und/oder Verzögerungszeit auf Kosten nur geringfügig niedrigerer Störabstände des betrachteten Gatters aufweisen. Solche Dimensionierungen wären evtl. denen maximaler Störabstände vorzuziehen. Um derartige Konstellationen sinnvoll bewerten zu können, wurde bei der Entwicklung der in diesem Kapitel vorgestellten Standardzellen-Bibliothek eine Mehrzieloptimierungsstrategie angewendet. Dabei wurde auch keine Vereinfachung komplexer Gatter zu einem Inverter vorgenommen, sondern jeder Gattertyp individuell optimiert, um das Optimierungspotential bestmöglich auszuschöpfen. Details zu den eingesetzten Verfahren werden im nachfolgenden Abschnitt 6.1 erläutert.

6.1 Zellendimensionierung

Die Dimensionierung einer integrierten Schaltung ist gekennzeichnet durch die Suche nach den Transistormaßen, für die bestimmte Eigenschaften der Schaltung die bestmöglichen Werte annehmen. Oftmals ist man dabei daran interessiert, mehrere Schaltungseigenschaften simultan zu betrachten. Da in der Regel unterschiedliche Eigenschaften in Konkurrenz zueinander bezüglich der Dimensionierung stehen, müssen optimale Kompromisse gefunden werden. Mathematisch lässt sich dieses Vorgehen als das Lösen eines Mehrzieloptimierungsproblems auffassen.

Teile dieses Kapitels sind auch in der Dissertation von Blesken [12] sowie in der gemeinsamen Publikation [178] veröffentlicht. Während [12] sich schwerpunktmäßig mit der Lösung des algorithmischen Teilproblems der Aufgabenstellung befasst, beschäftigt sich diese Arbeit insbesondere mit den schaltungstechnischen Aspekten sowie den speziellen Anforderungen und Entwurfsentscheidungen bei der Anwendung der Optimierungsverfahren auf die Dimensionierung von Standardzellen für den Subschwelligbetrieb.

6.1.1 Mehrzieloptimierung

Zur Definition eines Mehrzieloptimierungsproblems (MOP) sind zunächst einige zugrunde liegende Begriffe zu erklären: Sei $S \subset \mathbb{R}^n$ der Suchraum und $F : S \rightarrow \mathbb{R}^k, F(s) = (f_1(s), \dots, f_k(s))$ mit $k \geq 2$ die mehrkriterielle Bewertungsfunktion des MOP. Die einzelnen Vektorkomponenten der Bewertungsfunktion $f_j : S \rightarrow \mathbb{R}$ werden als Zielfunktionen bezeichnet. Jedes Element $s = (s_1, \dots, s_n) \in S$ stellt in dem hier betrachteten Zusammenhang eine mögliche Implementierungsalternative der integrierten Schaltung dar, wobei s_i beispielsweise die frei wählbaren Transistormaße sein können. Die Werte s_i werden auch als freie Entwurfparameter bezeichnet. Die quantifizierten Eigenschaften einer möglichen Schaltungsimplementierung s werden durch die Bildpunkte $F(s) = b = (b_1, \dots, b_k)$ beschrieben. Die Gesamtheit aller Bildpunkte ist der Bildraum $F(S)$. Im hier betrachteten Anwendungsfall ist die Auswertung eines Funktionswertes $b = F(s)$ mit dem Ausführen einer analogen Schaltungssimulation unter Vorgabe der Entwurfparameter s_i und der anschließenden Bestimmung der Schaltungseigenschaften b_j gleichbedeutend. Mit obigen Festlegungen lässt sich das MOP als

$$\min_{s \in S} F(s) \quad (6.1)$$

formulieren. Da $F(s)$ vektorwertig ist, muss zur Bestimmung des Minimums ein geeigneter Vergleichsbegriff spezifiziert werden. Gegeben seien zwei Bildpunkte $a, b \in \mathbb{R}^k$. a dominiert b genau dann, wenn gilt $a_j \leq b_j \forall j = 1, \dots, k$ und $a \neq b$. Dieser Zusammenhang wird durch die Schreibweise $a < b$ dargestellt. Die Lösung des MOP ist mit dieser Festlegung also gegeben durch

$$P = \{s \in S \mid \nexists t \in S, F(t) < F(s)\} \quad (6.2)$$

Die Lösungsmenge $P \subset S$ wird als *Pareto-Menge*, das Bild von P unter F als *Pareto-Front* $F(P)$ bezeichnet. In Abbildung 6.1 ist beispielhaft ein zweidimensionaler Bildraum mit Pareto-Front dargestellt.

Zu beachten ist, dass die Pareto-Menge in der Regel eine Vielzahl möglicher Lösungen eines Optimierungsproblems liefert. Ist man an einer konkreten Einzellösung interessiert, so ist ein nachfolgendes Auswahlverfahren erforderlich, um die gewünschte Lösung aus der Pareto-Menge zu bestimmen. Das für die Dimensionierung der Standardzellen eingesetzte Auswahlverfahren wird in Abschnitt 6.1.4 erläutert.

Um die Dimensionierung einer integrierten Schaltung als Mehrzieloptimierungsproblem zu formulieren, müssen zunächst die freien Entwurfparameter, die Zielfunktionen sowie Rahmenbedingungen festgelegt werden. Beim konventionellen

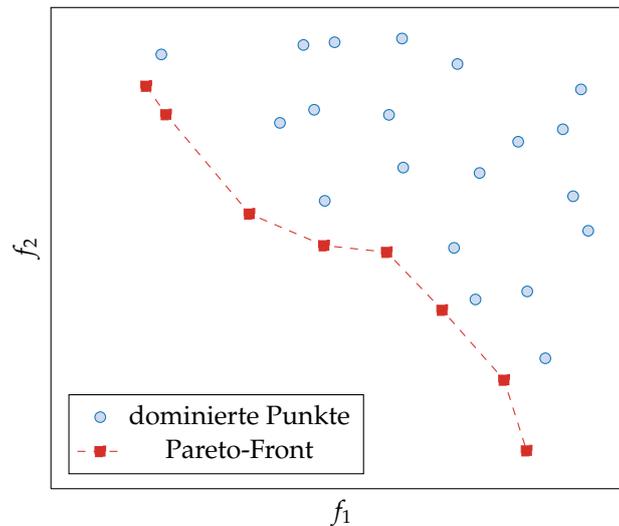


Abbildung 6.1: Zweidimensionaler Bildraum mit Pareto-Front

Entwurf von Standardzellen werden üblicherweise die Kanalweiten der Transistoren als freie Parameter herangezogen, während für die Kanallängen die in der jeweiligen Technologie minimal zulässigen Werte verwendet werden. Diese Wahl bietet den kleinsten Flächenbedarf, die geringsten (parasitären) Gatekapazitäten sowie beim Betrieb mit nomineller Spannung die kleinsten Verzögerungszeiten. Im Subschwelligbereich ist diese Festlegung der Kanallängen jedoch aus zwei Gründen nicht optimal. Zum einen führt dies wegen Gleichung 2.21 zu relativ großen Streuungen der Schwellspannung und anderer Transistor-Parameter und damit auch zu starken Streuungen oder sogar Fehlfunktionen des Schaltungsverhaltens. Zum anderen kann sich im Subschwelligbetrieb eine nicht-minimale Kanallänge durch den RSCE vorteilhaft auf die Schalteigenschaften auswirken, wie in Abschnitt 2.1 erläutert wurde. Damit kommen für den Entwurf von Subschwelligattern als freie Entwurfparameter prinzipiell die Weiten und Längen sämtlicher Transistoren in Betracht.

6.1.2 Ressourcenmaße

Zur Charakterisierung von Logikgattern bieten sich verschiedene Ressourcenmaße an, die einen direkten quantitativen Vergleich unterschiedlicher Implementierungsalternativen erlauben. Diese Eigenschaften eines Gatters entsprechen dem Verbrauch bestimmter Ressourcen, die im Rahmen dieser Arbeit nach Zeit, Energie, Fläche und Robustheit klassifiziert werden. Es sei an dieser Stelle angemerkt, dass es auf

den ersten Blick ungewohnt erscheint, den Begriff der Robustheit als Ressource zu bezeichnen. Hinsichtlich des Abwägens und der Konkurrenz untereinander gibt es jedoch keinerlei Unterschied in der Behandlungsweise der vier genannten Begriffe. Im Interesse einer konsistenten Bezeichnungsweise soll hier daher auch die Robustheit als Ressource aufgefasst werden.

Grundsätzlich sind viele der nachfolgend beschriebenen Ressourcenmaße von dem betrachteten Eingangs-Ausgangs-Pfad des Gatters abhängig. Dementsprechend müssten für alle Pfade die Ressourcenmaße separat bei der Optimierung berücksichtigt werden. Sollen zu verschiedenen Gattereingängen gehörige Transistoren jedoch mit gleichen Weiten und Längen dimensioniert werden und ist der Schaltungsaufbau des Gatters für alle Eingänge symmetrisch, so besteht zwischen den Ressourcenmaßen unterschiedlicher Pfade kein Optimierungskonflikt und es genügt zur Vereinfachung, die Ressourcenmaße nur für einen Eingangs-Ausgangs-Pfad zur Optimierung heranzuziehen. Generell liegen nicht schaltende Eingänge auf konstanten Pegeln, die ein Schalten des Ausgangs erlauben¹.

Schaltzeiten

Die Dauer, die ein Logikgatter benötigt, um aus den Eingangssignalen entsprechend der Boole'schen Funktion ein Ausgangssignal zu bilden, wird durch die Schaltzeiten beschrieben. Dabei unterscheidet man zwischen *Anstiegs-* und *Abfallzeiten* sowie *Verzögerungszeiten*. Die Anstiegszeit beschreibt die Dauer, die der Gatterausgang braucht, um von einem Low-Pegel auf einen High-Pegel zu wechseln. Analog beschreibt die Abfallzeit die Dauer eines umgekehrten Pegelwechsels. Es werden dabei die Punkte bei Erreichen von 10% und 90% der Versorgungsspannung zur Bestimmung des Zeitintervalls verwendet. Die Verzögerungszeiten hingegen beschreiben das Verzögerungsverhalten zwischen Eingang und Ausgang des Gatters, wobei auch hier zwischen steigender und fallender Flanke am Gatterausgang unterschieden wird. Die Messpunkte liegen dabei eingangs- und ausgangsseitig bei 50% der Versorgungsspannung. Die Anstiegszeit wird mit t_{LH} , die Abfallzeit

¹ Beispiel: Beim NAND2-Gatter ist ein Eingang konstant auf High-Pegel zu legen, damit eine Pegeländerung am anderen Eingang eine Pegeländerung am Ausgang hervorruft.

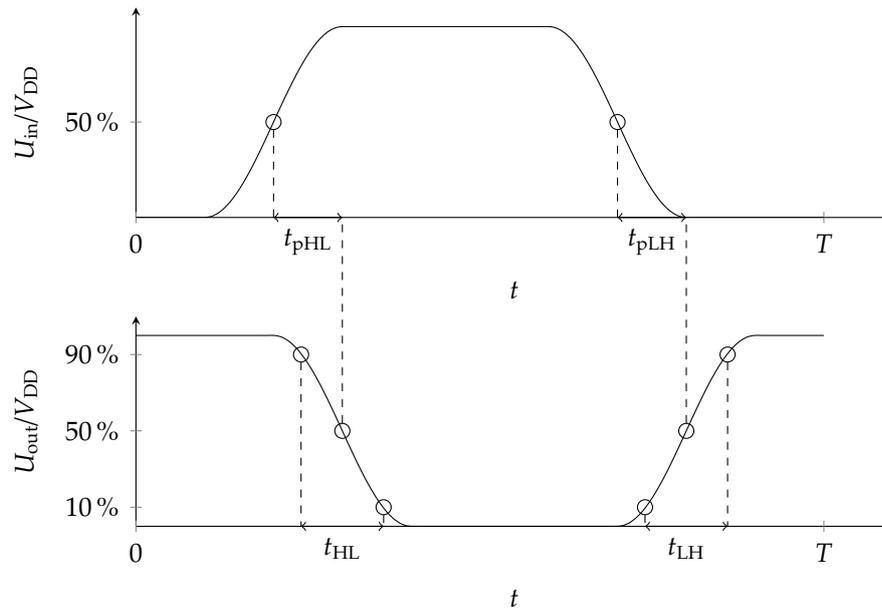


Abbildung 6.2: Schaltzeiten

mit t_{HL} und die Verzögerungszeiten mit t_{pLH} bzw. t_{pHL} bezeichnet. Abbildung 6.2 veranschaulicht die Definition der Schaltzeiten. Es gelten folgende Festlegungen:

$$t_{LH} = t \Big|_{\substack{U_{out}=0.9V_{DD} \\ U_{out}=0.1V_{DD}}} \quad (6.3)$$

$$t_{HL} = t \Big|_{\substack{U_{out}=0.1V_{DD} \\ U_{out}=0.9V_{DD}}} \quad (6.4)$$

$$t_{pLH/pHL} = t \Big|_{\substack{U_{out}=0.5V_{DD} \\ U_{in}=0.5V_{DD}}} \quad (6.5)$$

Im Rahmen der Optimierung ist man in vielen Fällen daran interessiert, die Schaltzeiten einer Flankenrichtung nicht zulasten der Schaltzeiten für die entgegengesetzte Flankenrichtung zu verbessern, da dies zu unsymmetrischem Schaltverhalten führt. Diese Situation kann dadurch vermieden werden, indem bei der Optimierung das Maximum der Schaltzeiten beider Flankenrichtungen betrachtet wird. Es wird daher definiert:

$$t_{rf} = \max(t_{LH}, t_{HL}) \quad (6.6)$$

$$t_{pd} = \max(t_{pLH}, t_{pHL}) \quad (6.7)$$

Verlustleistung bzw. -energie

Die Informationsverarbeitung in Logikgattern geht mit einer gewissen elektrischen Leistungsaufnahme einher. Dabei existieren mehrere Anteile, die nachfolgend erläutert werden. Grundsätzlich lässt sich unterscheiden zwischen *statischer Verlustleistung* P_{stat} , die permanent, auch bei Inaktivität von dem Gatter aufgenommen wird, und *dynamischer Verlustleistung* P_{dyn} , die nur bei Pegelwechseln von Signalen anfällt.

$$P_{\text{tot}} = P_{\text{stat}} + P_{\text{dyn}} \quad (6.8)$$

In CMOS-Gattern sind Leckströme die Ursache der statischen Verlustleistung. Diese treten als Subschwelligströme der nicht ideal sperrenden Transistoren, als Sperrströme der Source-Bulk- und Drain-Bulk-Dioden, sowie als Tunnelströme durch das Gateoxid auf. Fasst man alle Stromanteile zu einem Gesamtleckstrom I_{leak} zusammen, so lässt sich die statische Verlustleistung ausdrücken als

$$P_{\text{stat}} = V_{\text{DD}} \cdot I_{\text{leak}} \quad (6.9)$$

Da der Leckstrom in der Regel von der Pegelkonstellation an den Gattereingängen abhängt, repräsentiert I_{leak} dabei den Mittelwert über alle Möglichkeiten.

Die dynamische Verlustleistung lässt sich unterteilen in Verluste, die durch das Umladen von Lastkapazitäten entstehen (P_{sw}), und Verluste, die während des Umschaltens durch Querströme und Umladevorgänge innerhalb des Gatters entstehen (P_{int}). Beide Anteile sind proportional zur *Schalzhäufigkeit* (engl. *Toggle Rate*), also der mittleren Anzahl an Pegelwechseln des Gatterausgangs pro Sekunde. Dieser Wert ist von der Beschaltung des Gatters und von der konkreten Anwendung abhängig und daher im Zuge der Gatteroptimierung unbekannt. Sinnvoller ist an dieser Stelle die Betrachtung der entsprechenden Energien E_{dyn} , E_{sw} und E_{int} , die von der Schalzhäufigkeit unabhängig sind. Im Rahmen dieser Arbeit beziehen sich diese Energiewerte auf eine vollständige Periode des Ausgangssignals, also auf steigende und fallende Flanke.

$$E_{\text{dyn}} = E_{\text{sw}} + E_{\text{int}} \quad (6.10)$$

$$E_{\text{sw}} = \frac{1}{2} C_L V_{\text{DD}}^2 \quad (6.11)$$

$$E_{\text{int}} = V_{\text{DD}} \int_{\perp\perp} I_{\text{int}}(t) dt \quad (6.12)$$

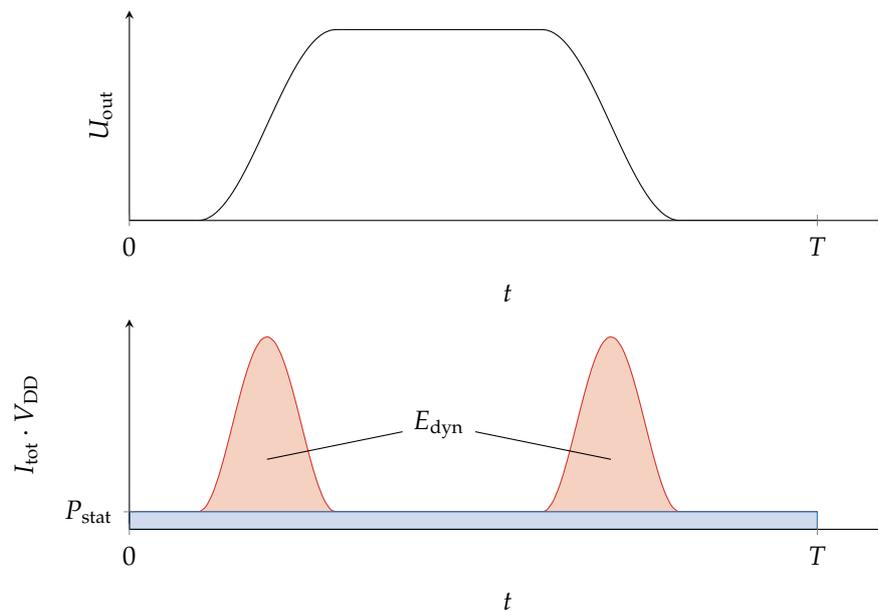


Abbildung 6.3: Bestimmung der dynamischen Energie

Dabei ist $I_{\text{int}}(t)$ der dynamische Anteil des Stroms innerhalb des Gatters. E_{sw} hängt nur von der Betriebsspannung und der Lastkapazität ab und ist daher im Rahmen der Gatteroptimierung konstant, wenn diese Werte als Rahmenbedingungen der Optimierung einmal festgelegt wurden. Es ist daher unerheblich, ob zur Optimierung E_{int} oder E_{dyn} betrachtet wird. Da die Messung von I_{int} nicht direkt möglich ist, ist es einfacher, E_{dyn} zu bestimmen. Dazu muss zunächst die Gesamtenergie während einer Periode über die Gesamtstromaufnahme I_{tot} bestimmt und anschließend der statische Anteil subtrahiert werden:

$$E_{\text{dyn}} = V_{\text{DD}} \int_{\square} I_{\text{tot}}(t) dt - P_{\text{stat}} \cdot T \quad (6.13)$$

Abbildung 6.3 illustriert das Vorgehen zur Bestimmung von E_{dyn} .

Ein weiterer Anteil der Verlustenergie, der einem Gatter zugeordnet werden kann, ist die Energie, die durch das Umladen der Eingangskapazitäten, also im Wesentlichen der Gatekapazitäten des Gatters umgesetzt wird.

$$E_{\text{in}} = \frac{1}{2} C_{\text{in}} V_{\text{DD}}^2 \quad (6.14)$$

E_{in} ist direkt proportional zur Eingangskapazität C_{in} . Bei der Optimierung kann also wahlweise E_{in} oder C_{in} als Kriterium verwendet werden. E_{in} lässt sich u.a. gemeinsam mit E_{dyn} in einem Simulationslauf direkt ermitteln:

$$E_{\text{in}} = V_{\text{DD}} \int_{\text{J}} I_{\text{in}}(t) dt \quad (6.15)$$

In diesem Fall erstreckt sich das Integrationsintervall lediglich über den Aufladevorgang der Eingangskapazität, also auf die steigende Flanke des Eingangssignals, da andernfalls der Wert des Integrals 0 wäre. Es sei bemerkt, dass E_{in} Teil der Verlustenergie $E_{\text{sw,prev}}$ des vorangegangenen Gatters ist.

Zusammengefasst kann auch die Gesamtenergie, die ein Gatter pro Umschaltvorgang umsetzt, als Zielfunktion der Optimierung verwendet werden:

$$E_{\text{gate}} = E_{\text{dyn}} + E_{\text{in}} \quad (6.16)$$

Störabstände

Die *Störabstände* sind ein Maß für die Unempfindlichkeit eines Gatters gegenüber verfälschten Signalpegeln und Rauschen. Dazu wird die Spannungsübertragungskennlinie des Gatters zwischen Ein- und Ausgang betrachtet, siehe Abbildung 6.4. Die äußeren Bereiche stabiler bzw. sicherer Pegelwerte mit flachem Kennlinienverlauf sind durch einen Unsicherheitsbereich mit größerer Kennliniensteigung getrennt. Diese Bereiche lassen sich durch die folgendermaßen definierten Spannungspegel trennen:

$$U_{\text{IL}} = \min \left\{ U_{\text{in}} \left| \frac{dU_{\text{out}}}{dU_{\text{in}}} = -1 \right. \right\} \quad (6.17)$$

$$U_{\text{IH}} = \max \left\{ U_{\text{in}} \left| \frac{dU_{\text{out}}}{dU_{\text{in}}} = -1 \right. \right\} \quad (6.18)$$

$$U_{\text{OL}} = \min \left\{ U_{\text{out}} \left| \frac{dU_{\text{out}}}{dU_{\text{in}}} = -1 \right. \right\} \quad (6.19)$$

$$U_{\text{OH}} = \max \left\{ U_{\text{out}} \left| \frac{dU_{\text{out}}}{dU_{\text{in}}} = -1 \right. \right\} \quad (6.20)$$

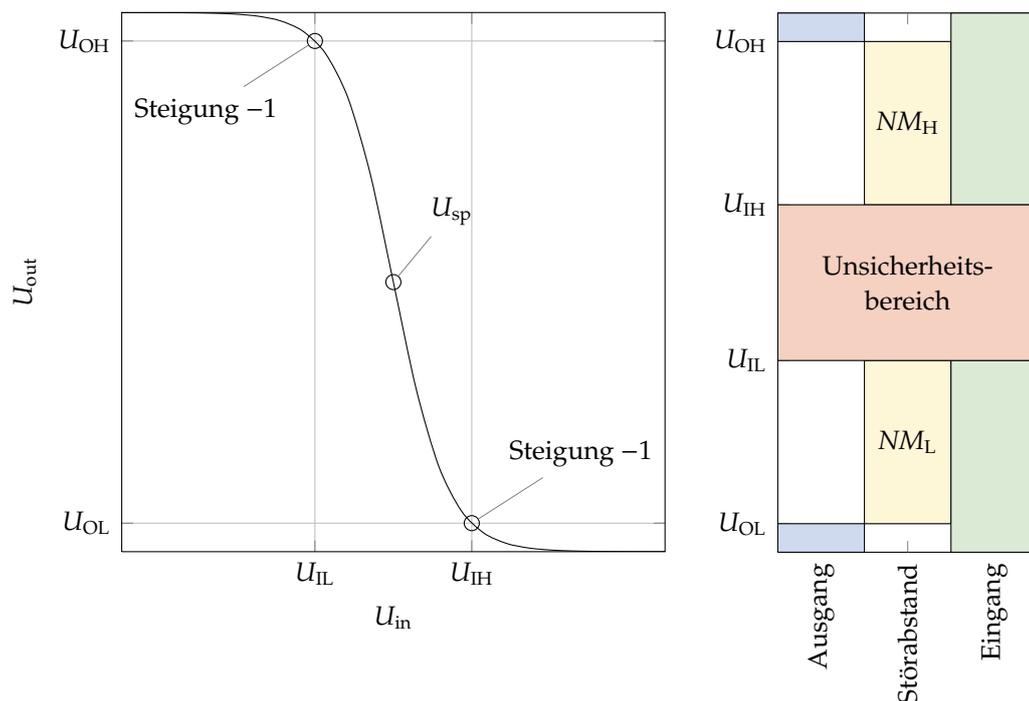


Abbildung 6.4: Spannungsübertragungskennlinie und Störabstände

Für invertierende Gatter gilt $U_{OL} = U_{out}(U_{IH})$ und $U_{OH} = U_{out}(U_{IL})$. Mithilfe der Spannungspegel lassen sich nun die Störabstände definieren als

$$NM_H = U_{OH} - U_{IH} \quad (6.21)$$

$$NM_L = U_{IL} - U_{OL} \quad (6.22)$$

$$NM = \min(NM_H, NM_L) \quad (6.23)$$

NM gibt den schlechteren der beiden separat für High- und Low-Pegel bestimmten Störabstände an. Der Wert der Störabstände beschreibt, wie groß die Verfälschung einer Signalspannung bei mehreren hintereinander geschalteten Gattern des gleichen Typs im ungünstigsten Fall sein muss, um einen Signalpegel im Unsicherheitsbereich zu erreichen. Spannungsverfälschungen mit kleinerer Amplitude können die logische Funktionalität des Gatters nicht beeinträchtigen. Man beachte, dass im Gegensatz zu den übrigen Ressourcenmaßen *größere* Störabstände eine bessere Bewertung bedeuten. Bei der Formulierung des Mehrzieloptimierungsproblems in Form von Gleichung 6.1 ist daher der Störabstand mit negativem Vorzeichen ($-NM$) als Zielfunktion anzusetzen.

Fläche

Als Ressourcenmaße für den Flächenbedarf eines Gatters können einfache Abschätzungen oder tatsächliche Layout-Maße herangezogen werden. Abschätzungen können etwa die Transistoranzahl oder die akkumulierte Fläche der Transistorkanäle (also die Summe des Produktes aus Kanalweite und Kanallänge aller Transistoren) sein. Im Rahmen der Optimierung eines bestimmten Gattertyps ist jedoch die Transistoranzahl konstant und kommt daher als Flächenmaß nicht in Frage. Andererseits ist es nicht praktikabel, für jede mögliche Implementierungsalternative ein vollständiges Layout zu erstellen, um den Flächenbedarf zu ermitteln. Daher ist auch dieses Maß als Zielfunktion der Optimierung ungeeignet. Prinzipiell kommt lediglich die akkumulierte Kanalfläche

$$A_{\text{gate}} = \sum_i W_i L_i \quad (6.24)$$

als Optimierungskriterium in Betracht. Dieses Ressourcenmaß lässt sich ohne Simulation direkt aus den vorgegebenen Entwurfparametern ermitteln. Allerdings ist insbesondere beim Entwurf von Standardzellen zu beachten, dass der tatsächliche Flächenbedarf eines Gatterlayouts neben der Kanalfläche auch von der Verdrahtungskomplexität, dem Format des Standardzellenrahmens, dem Verdrahtungsraster (engl. *Routing Grid*) und anderen Faktoren abhängt. Daher ist A_{gate} nur unter Vorbehalt als Zielfunktion einer Optimierung geeignet. Alternativ besteht auch die Möglichkeit, die Fläche als Ressourcenmaß erst bei der Auswahl geeigneter Implementierungen aus der Pareto-Menge nach abgeschlossener Mehrzieloptimierung zu berücksichtigen. Damit kann auch eine Verzerrung der Pareto-Menge durch ein ungenaues, geschätztes Flächenmaß vermieden werden.

6.1.3 Algorithmen

Die Lösung des oben beschriebenen MOPs erfordert den Einsatz geeigneter Algorithmen zur Approximation der Pareto-Menge. Im Rahmen dieser Arbeit wurden zwei unterschiedliche Verfahren für diese Aufgabe eingesetzt, die unter den Namen *Global Analysis of Invariant Objects* (GAIO) [43, 130–132] und *Strength Pareto Evolutionary Algorithm* (SPEA) [177] veröffentlicht und in [12, 13] auf die Dimensionierung ressourceneffizienter integrierter Schaltungen angewendet wurden. Die Algorithmen sind auf Basis von Mathworks Matlab implementiert und sind an Cadence Spectre für

analoge Schaltungssimulationen angebunden. Nachfolgend werden die Grundzüge dieser Algorithmen in Anlehnung an die genannten Quellen erläutert.

GAIO

Bei Verwendung von GAIO wird die Pareto-Menge eines MOPs durch „Boxen“ approximiert. Dabei kommen im Wesentlichen zwei Teilalgorithmen zum Einsatz. Durch den Unterteilungsalgorithmus wird in jedem Schritt zunächst die bisherige Approximation der Pareto-Menge durch kleinere Boxen unterteilt. Innerhalb jeder Box werden dann stichprobenhaft Testpunkte gewählt. Anhand von Dominanztests aller Testpunkte untereinander wird anschließend entschieden, welche Boxen mit hoher Wahrscheinlichkeit keine Teile der Pareto-Menge enthalten. Diese Boxen werden verworfen. Durch rekursive Anwendung des Algorithmus wird die Approximation der Pareto-Menge schrittweise verfeinert. Bei ungünstiger Wahl der Testpunkte kann es allerdings vorkommen, dass auch Boxen entfernt werden, die Teile der Pareto-Menge enthalten. Die Approximation der Pareto-Menge ist also möglicherweise unvollständig. Daher wird zusätzlich der GAIO-Fortsetzungsalgorithmus eingesetzt. Dabei macht man sich die Tatsache zunutze, dass unter gewissen Voraussetzungen weitere Pareto-Punkte in der Nähe bereits bekannter Pareto-Punkte erwartet werden können. Eine Boxsammlung, die durch den Stichprobenalgorithmus bestimmt wurde, kann als Ausgangspunkt dienen. In Boxen, die in der Umgebung außerhalb dieser Boxsammlung liegen, wird wiederum durch Testpunkte überprüft, ob Teile der Pareto-Menge auch hier liegen können. Auf diese Weise kann die Approximation der Pareto-Menge schrittweise erweitert bzw. fehlende Boxen ergänzt werden.

SPEA

Evolutionäre Algorithmen sind Optimierungsverfahren nach dem Vorbild der biologischen Evolution. Bei geeigneter Formulierung lassen sie sich zur Lösung von Mehrzieloptimierungsproblemen verwenden, etwa in Form des SPEA. Auch hier erfolgt eine schrittweise Approximation der Pareto-Menge. Kandidaten zur Lösung des Optimierungsproblems werden als *Individuen* bezeichnet und in *Populationen* zusammengefasst. In jedem Durchlauf wird aus der bestehenden Population eine nachfolgende Population generiert, wobei eine bessere Approximation der Lösungsmenge angestrebt wird. Dabei wird das evolutionäre Prinzip verfolgt, gute Eigenschaften zu vererben und schlechte Eigenschaften verfallen zu lassen. Dazu

wird zunächst durch eine *Fitnessbewertung* jedem Individuum ein Fitnesswert zugeordnet, der die Eigenschaften dieses Individuums anhand der Pareto-Dominanz beurteilt. Anschließend erfolgt die *Selektion* von Individuen, die für eine Vererbung für die nachfolgende Population in Frage kommen. Die Vererbung selbst erfolgt durch *Rekombination* in den Teilschritten *Crossover* und *Mutation*. Dabei wird die Position von zwei zufällig gewählten Eltern-Individuen im Suchraum durch eine geeignete Operation „gemischt“ und mit einer gewissen Wahrscheinlichkeit zusätzlich geringfügig modifiziert. Auf diese Weise werden zwei neue Kinder erzeugt, die wiederum mit einer gewissen Wahrscheinlichkeit ihre Eltern ersetzen.

6.1.4 Voruntersuchungen

Für die Optimierung der Standardzellen müssen gewisse Rahmenbedingungen festgelegt werden, die den Betriebszustand der Schaltungen bestimmen und den verfügbaren Suchraum eingrenzen. Da die Betrachtung der schlechtesten Betriebsbedingungen im Subschwelligbereich unverhältnismäßig pessimistische Schaltungseigenschaften liefert, sollen typische Betriebsbedingungen, d.h. nominelle Prozessparameter und eine Betriebstemperatur von 25 °C bei der Optimierung angenommen werden. Die Versorgungsspannung soll etwa dort liegen, wo die geringste Energieaufnahme später implementierter Digitalschaltungen erwartet wird. Um diesen Spannungswert abzuschätzen, werden die Messergebnisse des ALU-Prototypen aus Abschnitt 5.5.3 erneut betrachtet. Der minimale Energiebedarf tritt dort bei einer Versorgungsspannung von 210 mV auf, die Schwellspannung der verwendeten 90-nm-Technologie beträgt etwa 300 mV. Damit liegt das Energieoptimum um etwa ein Drittel unterhalb der Schwellspannung. Überträgt man dieses Verhältnis auf die 65-nm-Technologie, so ergibt sich aus der Schwellspannung von ca. 450 mV eine Abschätzung der Lage des Energieoptimums von 300 mV, so dass diese Spannung als Entwurfspunkt der Standardzellen-Bibliothek festgelegt wurde. Die Schwellspannung der 65-nm-Technologie ist höher als die Schwellspannung der 90-nm-Technologie, da es sich bei ersterer um eine Low-Power-Technologie, bei zweiterer um eine General-Purpose-Technologie handelt.

Um einen ersten Eindruck davon zu bekommen, welchen Einfluss die Gegebenheiten des Subschwelligbereichs gegenüber einer nominellen Versorgungsspannung auf optimale Gatterauslegungen haben, wurde ein erster Optimierungslauf des Inverters jeweils für 1,2 V und 300 mV durchgeführt. Beispielhaft wurden als Zielfunktionen t_{pd} , E_{gate} sowie $-NM$ und als freie Parameter die Kanalweiten beider Transistoren

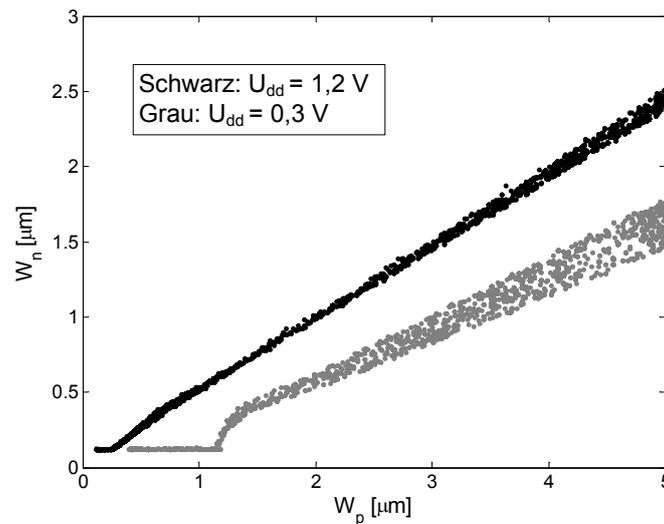


Abbildung 6.5: Approximation der Pareto-Menge des CMOS-Inverters bei 1,2 V und 300 mV [12, 178]

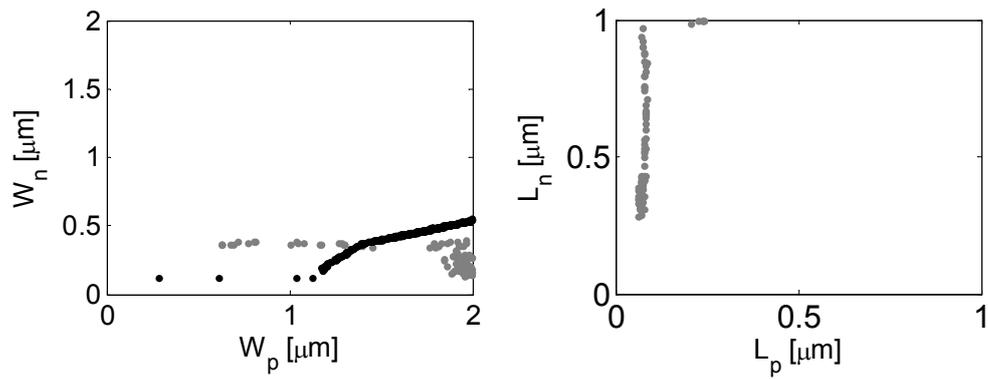
W_n , W_p bei minimalen Kanallängen von 60 nm festgelegt. In Abbildung 6.5 ist das Optimierungsergebnis als Approximation der Pareto-Mengen dargestellt. Es zeigt, dass die optimalen Kompromisse für 300 mV eine andere Dimensionierung aufweisen als die optimalen Kompromisse für nominelle Versorgungsspannung. Da die beiden Pareto-Mengen disjunkt sind, werden optimale Dimensionierungen für nominelle Versorgungsspannung im Subschwelligbetrieb dominiert, d.h. derart ausgelegte Gatter sind in Bezug auf die obigen Zielfunktionen im Subschwelligbetrieb nicht ressourceneffizient. Diese Erkenntnis unterstreicht, dass die Verwendung konventioneller Standardzellen-Bibliotheken zur Implementierung von Subschwelligschaltungen suboptimal ist, und dass stattdessen ein spezieller Gatterentwurf erforderlich ist.

Um den Optimierungsaufwand zu reduzieren, ist es sinnvoll, möglichst wenige freie Entwurfsparameter und Zielfunktionen bei der Optimierung zu berücksichtigen, ohne dabei jedoch wesentliche Wechselwirkungen zwischen den Größen zu vernachlässigen. Wie zuvor erläutert, sind die möglichen freien Parameter durch die Weiten und Längen aller Transistoren gegeben. Es stellt sich zunächst die Frage, inwieweit die Kanallängen als freie Parameter tatsächlich zu einer Verbesserung der Ressourcenmaße beitragen können. Dazu sind in Abbildung 6.6 die Ergebnisse zweier Optimierungsläufe des CMOS-Inverters bei 300 mV Versorgungsspannung dargestellt. Als Zielfunktionen wurden hier P_{stat} und t_{pd} ausgewählt. Teil 6.6a der Abbildung zeigt den Suchraum als Projektion nach Weiten und Längen getrennt, während in 6.6b der Bildraum abgebildet ist. Approximationen von Pareto-Menge

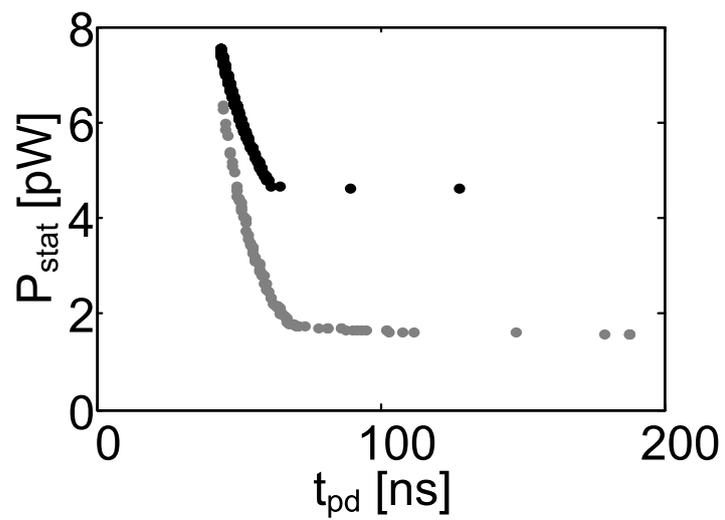
und Pareto-Front sind als schwarze Punkte für eine zweidimensionale Optimierung mit den Weiten W_n und W_p des NMOS- bzw. PMOS-Transistors als freie Parameter bei minimalen Kanallängen gegeben. Die grauen Punkte stellen hingegen die Ergebnisse einer vierdimensionalen Optimierung dar, wobei die Längen L_n und L_p von NMOS- bzw. PMOS-Transistor variabel sind. Im Bildraum ist zu sehen, dass durch die variablen Kanallängen zwar keine absolute Verbesserung der Verzögerungszeit möglich ist, die statische Verlustleistung jedoch bei gleicher Verzögerungszeit bis zu 60 % reduziert werden kann. Bei Betrachtung des vierdimensionalen Suchraums fällt auf, dass offenbar der optimale Wert von L_p der minimalen Kanallänge von 60 nm entspricht, während der Bereich optimaler Werte von L_n oberhalb von etwa 250 nm liegt. Optimale Weiten des NMOS-Transistors finden sich bei etwa 350 nm und darunter. Diese Gatterdimensionierungen haben jedoch einige entscheidende Nachteile: Zunächst sind die relativ großen Längen des NMOS-Transistors bei gleichzeitig geringen Weiten problematisch in Bezug auf ein flächeneffizientes Layout im Format einer Standardzelle. Weiterhin ist eine minimale PMOS-Kanallänge mit relativ großen Parameterstreuungen verbunden, so dass hier zumindest eine Anhebung der Untergrenze von L_p während der Optimierung erforderlich wäre. Schließlich ergeben sich durch die Ergebnisse der vierdimensionalen Suche stark asymmetrische Push-Pull-Verhältnisse (Verhältnis der Treiberstärke von Pull-Up- und Pull-Down-Zweig) beim Betrieb mit Versorgungsspannungen oberhalb der Transistor-Schwelspannung, was zu unverhältnismäßig hohen Verzögerungszeiten und Verlustleistungswerten in diesem Betriebsbereich führt. Derartige Implementierungen wären daher ausschließlich für den Subschwellsbetrieb und weniger für variable Versorgungsspannungen geeignet.

Aus diesen Gründen ist ein anderes Vorgehen zur Dimensionierung der Gatter erforderlich. Um übermäßig unsymmetrische Push-Pull-Verhältnisse zu vermeiden und ein homogeneres Layout zu ermöglichen, wird nun eine unabhängige Wahl von NMOS- und PMOS-Kanallänge unterbunden, indem eine gemeinsame Kanallänge L als freier Parameter eingeführt wird. Es ergibt sich auf diese Weise ein dreidimensionaler Suchraum. Ein Optimierungslauf in dieser Konstellation ergibt, dass trotz der nun erforderlichen Kompromissfindung zwischen NMOS- und PMOS-Länge der optimale Wert der gemeinsamen Länge beim Minimalwert von 60 nm liegt. Da dies, wie bereits erwähnt, aufgrund relativ großer Parameterstreuungen ungünstig im Subschwellsbetrieb ist, wird ein Aufschlag von 50 % als Kompromiss getroffen, d.h. die Kanallänge aller Transistoren wird auf 90 nm festgelegt.

Nach Festlegung der Kanallänge verbleiben noch die Kanalweiten als freie Parameter, wobei bei Gattern mit mehreren Eingängen prinzipiell unterschiedliche Weiten für



(a) Suchraum



(b) Bildraum

Abbildung 6.6: Pareto-Menge des CMOS-Inverters bei 2-dimensionalem (schwarz) und 4-dimensionalem (grau) Suchraum [12, 178]

individuelle Transistoren zugelassen werden können. Insbesondere bei Gattern mit mehr als zwei Eingängen steigt der Optimierungsaufwand dabei beträchtlich an, es lassen sich jedoch nur relativ geringe Verbesserungen gegenüber einer Gleichbehandlung aller NMOS- und aller PMOS-Transistoren erzielen. Zudem ist die Layout-Erstellung von Gattern mit unterschiedlichen Weiten der einzelnen Transistoren aufwändiger und der Flächenbedarf derartiger Implementierungen aufgrund spezieller Entwurfsregeln größer, als bei homogener Auslegung der Transistoren. Daher wird festgelegt, dass alle NMOS-Transistoren einer Gatterstufe die Weite W_n und alle PMOS-Transistoren die Weite W_p besitzen sollen. Es verbleibt auf diese Weise ein zweidimensionaler Suchraum für alle einstufigen Gattertypen, wodurch auch übermäßig lange Laufzeiten der Optimierungswerkzeuge vermieden werden.

Bezüglich der Zielfunktionen wird zunächst die Fläche, wie oben begründet, als Optimierungskriterium verworfen, da einleuchtend ist, dass größere Transistoren mit einem größeren Flächenbedarf einhergehen. Ferner wird die Entscheidung getroffen, bei Schaltzeiten und Störabständen nicht zwischen steigenden und fallenden Flanken bzw. High- und Low-Pegel zu unterscheiden, sondern stattdessen die jeweils ungünstigeren Werte t_{rf} , t_{pd} bzw. NM zu verwenden. Durch Betrachten der jeweils ungünstigeren Werte wird zugleich auch ein symmetrisches Schaltverhalten der gefundenen Optima sichergestellt. Voruntersuchungen ergaben außerdem, dass es nicht notwendig ist, mehrere Ressourcenmaße aus einer Kategorie (also beispielsweise t_{rf} und t_{pd} oder P_{stat} und E_{gate}) simultan bei der Optimierung zu berücksichtigen, da hier die Zielfunktionen kaum in Konkurrenz zueinander stehen und daher praktisch kein Bedarf zur Kompromissfindung besteht. Da jedoch die Ressourcenkategorien untereinander sehr wohl in Konkurrenz zueinander stehen, sollte aus jeder Kategorie eines der Bewertungsmaße zur Optimierung eingesetzt werden. Anhand dieser Überlegungen wurden die Zielfunktionen t_{pd} , E_{gate} und $-NM$ zur Dimensionierung der Gatter ausgewählt.

Der Suchraum der Optimierung muss eingegrenzt werden, da die Transistorweiten, die den Koordinaten im Suchraum entsprechen, einerseits die Minimalmaße der Technologie einhalten müssen und andererseits nicht zu groß werden sollen, damit sich Layouts im Standardzellenformat erstellen lassen. Das Minimalmaß der Transistorweiten beträgt 120 nm. Es wurden großzügige Obergrenzen von $1\ \mu\text{m}$ für die NMOS-Weite und $2,5\ \mu\text{m}$ für die PMOS-Weite als Grenzen des Suchraums festgelegt.

Nach Anwenden eines der im vorigen Abschnitt beschriebenen Algorithmen zur Lösung des MOP für ein Logikgatter steht eine Approximation der Pareto-Menge

zur Verfügung. Für eine konkrete Gatterrealisierung müssen aus dieser Menge ein oder mehrere geeignete Punkte als Dimensionierungsvorgabe ausgewählt werden. Dies kann anhand zusätzlicher Kriterien wie etwa der Fläche, wie zuvor beschrieben, geschehen. Eine andere Möglichkeit besteht darin, eine Priorisierung der bereits bei der Mehrzieloptimierung betrachteten Zielfunktionen vorzunehmen. Dieser Ansatz wurde bei der Entwicklung der hier vorgestellten Standardzellen-Bibliothek verfolgt. Dazu wurden aus der (approximierten) Pareto-Menge jedes Gattertyps zunächst die Punkte ausgewählt, die innerhalb der besten 5 % bezüglich des Störabstands NM liegen. Auf diese Weise wird der Robustheit die größte Priorität bei der Dimensionierung eingeräumt. Da Schaltungen mit möglichst niedrigem Energiebedarf implementiert werden sollen, werden aus der verbleibenden Menge wiederum die besten 5 % bezüglich der Energie E_{gate} ausgewählt, die damit den zweithöchsten Rang erhält. Die auf diese Weise selektierten Punkte befinden sich bereits in einem eng begrenzten Bereich in Such- und Bildraum. Es werden daher Weiten W_n und W_p für die Gatterdimensionierung ausgewählt, die in der Nähe des Schwerpunkts dieser Menge im Suchraum liegen und dem in der verwendeten Technologie zulässigen 5-nm-Raster genügen.

6.2 Begrenzung der Transistor-Stacks

Die Auswahl des Zellsatzes im Subschwelligbereich hat Auswirkungen auf die Robustheit der gesamten Schaltung, da bestimmte Zellentypen kritischer in Bezug auf Prozessparameter-Streuungen sind als andere [26]. Eine notwendige Bedingung für die korrekte Funktion einer Schaltung ist das Erreichen ausgeprägter Logikpegel aller Gatter, da ansonsten Fehlinterpretationen der zugrunde liegenden Boole'schen Funktionen auftreten können. Der Spannungswert des Low-Pegels eines Gatters ergibt sich durch den Spannungsteiler aus leitendem Pull-Down-NMOS-Zweig und sperrendem Pull-Up-PMOS-Zweig. Analoges gilt für den Spannungswert des High-Pegels. Je größer der Leitwert des leitenden Transistorpfades und je kleiner der Leitwert eines sperrenden Transistorpfades ist, desto näher liegt der Spannungspegel des Gatters am Idealwert von 0 bzw. V_{DD} . Diese Eigenschaft lässt sich durch das I_{on}/I_{off} -Verhältnis beschreiben, wie bereits in Abschnitt 3.2 erläutert wurde. Betrachtet man nun beispielsweise ein NAND4-Gatter, so befinden sich im Pull-Down-Pfad vier NMOS-Transistoren in Reihenschaltung, während der Pull-Up-Pfad aus vier parallel geschalteten PMOS-Transistoren besteht. Bei einem Lowpegel am Gatter-Ausgang steht also einem (in erster Näherung) auf $1/4$ reduzierten On-Strom ein

vierfacher Off-Strom gegenüber. Insgesamt verschlechtert sich also bei diesem Gatter das Verhältnis $I_{\text{on}}/I_{\text{off}}$ um den Faktor 16, was sich entsprechend ungünstig auf die im Subschwellschwellbetrieb erreichbaren Logikpegel auswirkt. Das obige Beispiel lässt sich zu der Aussage verallgemeinern, dass ein höherer Grad an Reihen- und Parallelschaltungen in einem Gatter eine Verschlechterung der Logikpegel zur Folge hat. Reihenschaltungen im Pull-Down-Zweig bewirken dabei eine Anhebung des Low-Spannungspegels, während Reihenschaltungen im Pull-Up-Zweig eine Senkung des High-Spannungspegels bewirken.

Darüber hinaus können die On- und Off-Ströme auch durch Prozessparameter beeinflusst sein. Der kritischste Fall tritt dann ein, wenn schwache bzw. langsame leitende Transistoren und starke bzw. schnelle sperrende Transistoren in einem Gatter zusammen treffen, also bei langsamem NMOS und schnellem PMOS bei einem NAND-Gatter und Low-Pegel am Ausgang sowie bei schnellem NMOS und langsamem PMOS bei einem NOR-Gatter und High-Pegel am Ausgang. Es stellt sich die Frage, welcher Grad an Reihen- und Parallelschaltung in den Gattern zugelassen werden kann, ohne dass sich in diesen kritischen Situationen nicht tolerierbar degenerierte Logikpegel ergeben. Zur Klärung dieser Frage sind in Abbildung 6.7 die auf die Versorgungsspannung normierten Spannungspegel von NAND- und NOR-Gattern mit 2, 3 und 4 Eingängen in den genannten kritischen Fällen angegeben. Die Gatter wurden für symmetrisches Schaltverhalten bei nominellen Prozessparametern ausgelegt. Da eine hohe Zuverlässigkeit der hier entwickelten Standardzellen-Bibliothek mit möglichst niedriger unterer Betriebsspannungsgrenze erreicht werden soll, wird die strenge Forderung gestellt, dass bei einer Spannung von 200 mV ein Low-Pegel von maximal 5 % und ein High-Pegel von mindestens 95 % erreicht werden sollen. Diese Forderung wird nur durch die Gattertypen NAND2 und NOR2 erfüllt. Als Verallgemeinerung wurde daher die Entscheidung getroffen, die Auswahl des Zellsatzes auf Gatter zu beschränken, die sich durch maximal zwei parallel oder in Reihe geschaltete Transistorpfade jeweils im Pull-Up- und Pull-Down-Zweig realisieren lassen.

6.3 Kombinatorische Gatter

Eine vielfältige Auswahl an Logikfunktionen für die zu implementierenden Standardzellen ist trotz der Beschränkung auf Gatter mit maximalem Grad zwei bezüglich Reihen- und Parallelschaltungen von Transistoren möglich. Für jede Logikfunk-

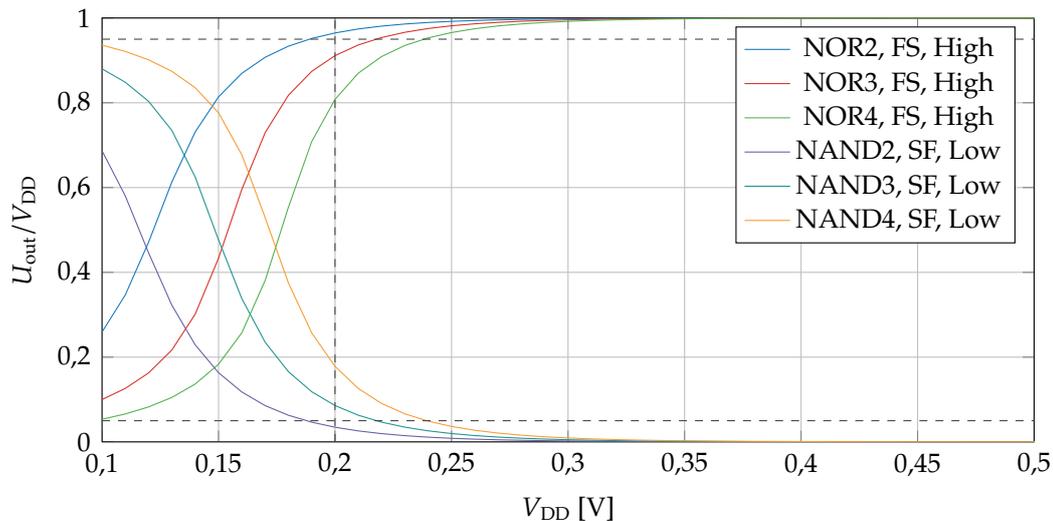


Abbildung 6.7: Relative Spannungspegel von NAND- und NOR-Gattern mit 2, 3 und 4 Eingängen

tion wurden Gatter mit unterschiedlichen Treiberstärken implementiert, um die Optimierungsmöglichkeiten der Synthese-Werkzeuge zu erweitern. Höhere Treiberstärken wurden dabei durch Parallelschalten entsprechend vieler Gatter der kleinsten Treiberstärke $X1$ realisiert. Die Dimensionierung der $X1$ -Gatter erfolgt durch das oben beschriebene Mehrzieloptimierungsverfahren. In Tabelle 6.1 sind alle gewählten Logikfunktionen sowie die jeweils realisierten Treiberstärken aufgelistet. Die Transistor-Verschaltungen der Gatter ergeben sich für die Zellen A0I22, INV, NAND2, NOR2 und OAI22 auf natürliche Weise durch die Bildungsregeln für CMOS-Gatter. Puffer (BUF) werden durch einen zweistufigen Aufbau realisiert, wobei jede Stufe aus ggf. parallel geschalteten Invertern besteht. Für das Minority3-Gatter (MIN3) sowie den (invertierenden) 2-zu-1-Multiplexer (MUXI2) werden die in Abbildung 6.8 dargestellten Schaltungsvarianten gewählt, die symmetrische NMOS- und PMOS-Teile besitzen. Diese Eigenschaft ist vorteilhaft für die Layout-Erstellung und wirkt sich darüber hinaus positiv auf die Gattereigenschaften aus. Im Falle des MIN3-Gatters wird diese Implementierung auch als *Mirrored Gate* bezeichnet. Für das MUXI2-Gatter ist zusätzlich ein Inverter für das Auswahlsignal S erforderlich, der als Inverter einfacher Treiberstärke dimensioniert ist.

Die approximierten Pareto-Mengen und -Fronten der Mehrzieloptimierung sind in Abbildung 6.9 dargestellt. Die hier gezeigten Ergebnisse wurden mit Hilfe des SPEA bestimmt. Man erkennt, dass die Pareto-Mengen der Gatter zum Teil sehr unterschiedlichen Charakter aufweisen. Insbesondere beim NAND2-Gatter fällt auf,

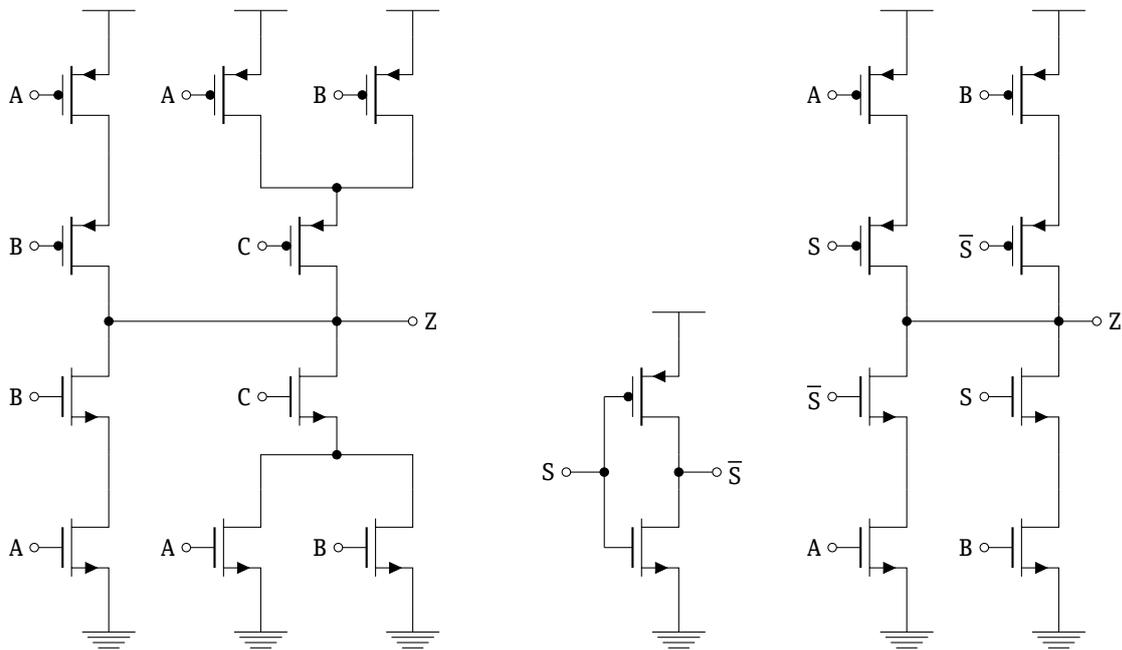


Abbildung 6.8: Schaltpläne der Gattertypen MIN3 (links) und MUXI2 (rechts)

dass es offenbar keine ressourceneffizienten Gatter-Implementierungen im Bereich $300 \text{ nm} \lesssim W_n \lesssim 600 \text{ nm}$ gibt. Dieses Phänomen lässt sich durch den RNCE erklären: Durch die Abnahme der Schwellspannung bei geringen Transistorweiten entsteht im Subschwelbereich ein Minimum des Drain-Stroms bei einer gewissen Kanalweite. Ein gegebener Drain-Strom (und damit beispielsweise auch eine gegebene Verzögerungszeit t_{pd}) lässt sich also sowohl mit einer größeren als auch mit einer kleineren Kanalweite erzielen. Da die größere Kanalweite jedoch mit einer höheren Energie E_{gate} verbunden ist, ist eine solche Dimensionierung nicht ressourceneffizient und wird daher durch das Mehrzieloptimierungsverfahren verworfen. In der hier verwendeten Technologie wirkt sich dieser Effekt lediglich für den NMOS-Transistor aus, so dass eine solche Lücke in der Pareto-Menge nur in Richtung von W_n auftritt. Außerdem hängt der genaue Einfluss des RNCE auch von der konkreten Schaltungskonstellation ab, wie der Vergleich der Pareto-Mengen in Abbildung 6.9 zeigt.

Die Schritte des Auswahlverfahrens für konkrete Gatterdimensionierungen sind durch unterschiedliche Grauabstufungen in Abbildung 6.9 gekennzeichnet. Der hellgrau hervorgehobene Teil der Mengen liegt jeweils außerhalb der besten 5% bezüglich des Störabstands NM , vom verbleibenden Rest liegt der dunkelgraue Teil außerhalb der besten 5% bezüglich der Energie E_{gate} . Schwarz markiert ist somit das nach beiden Kriterien gefilterte Optimierungsergebnis. Die konkreten

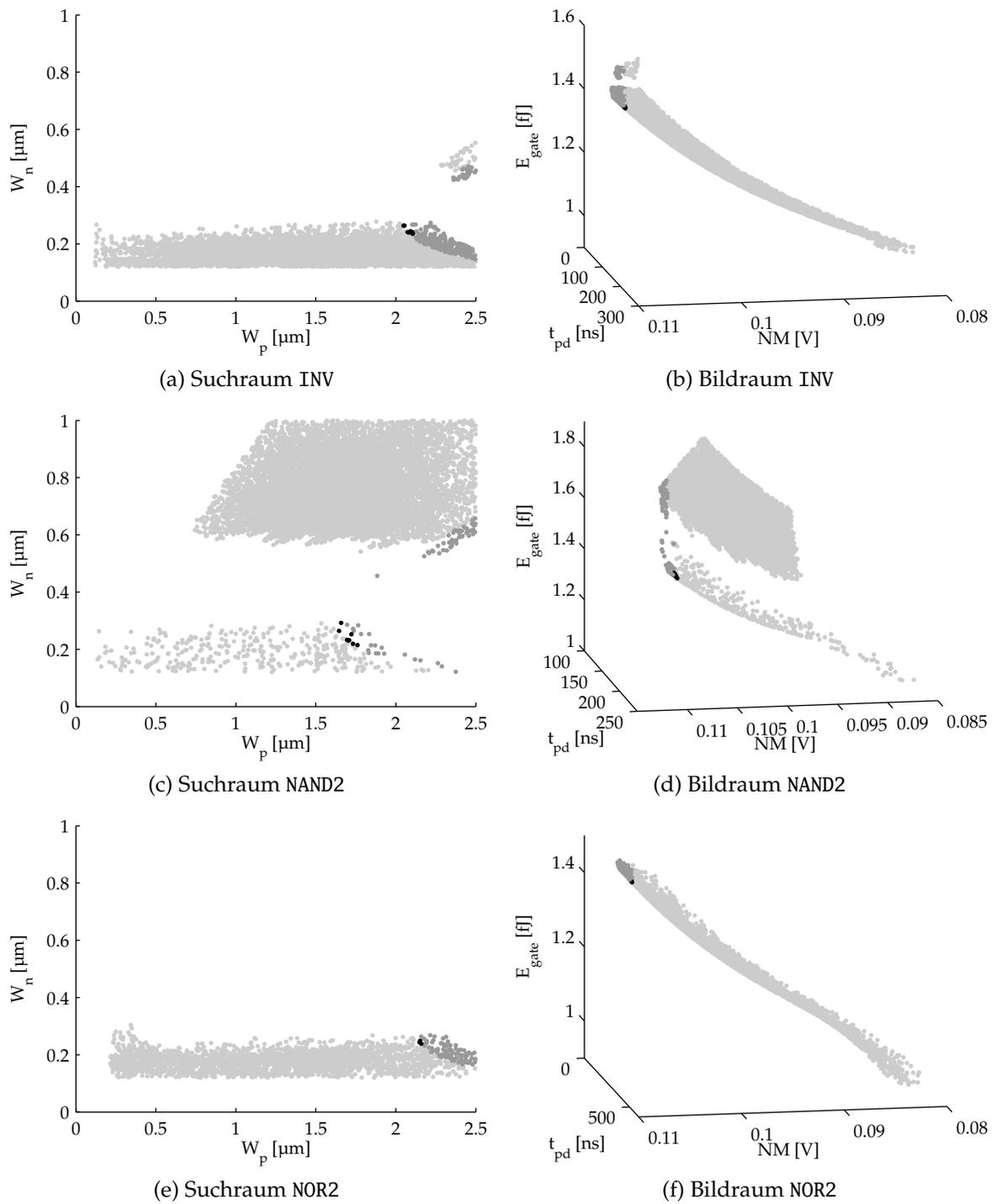
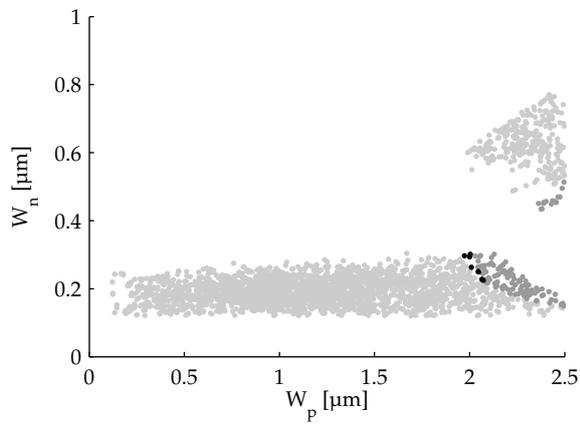
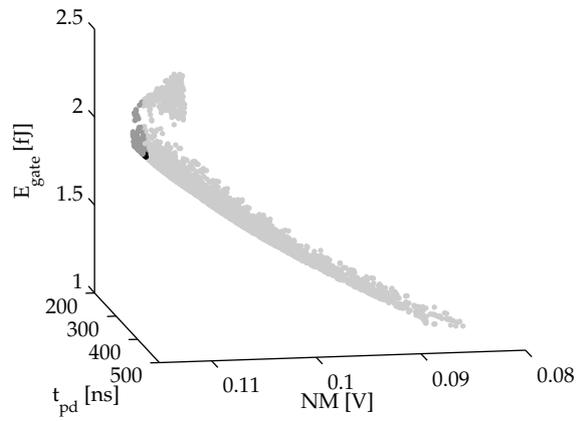


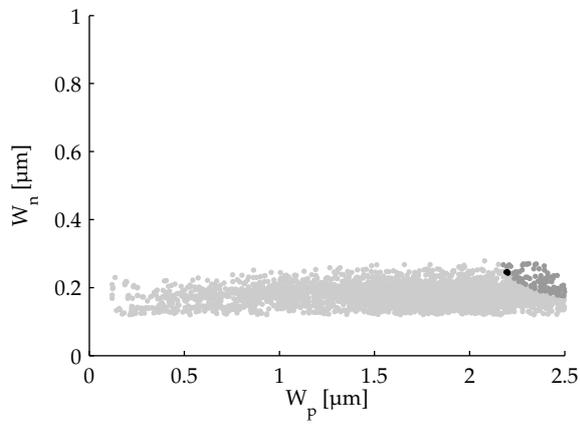
Abbildung 6.9: Pareto-Mengen und Pareto-Fronten der kombinatorischen Gatter



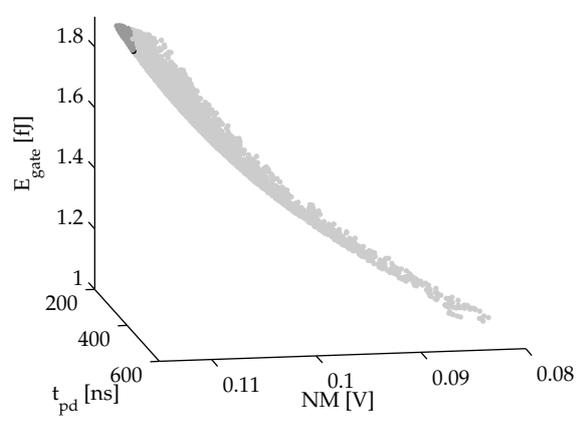
(g) Suchraum A0I22



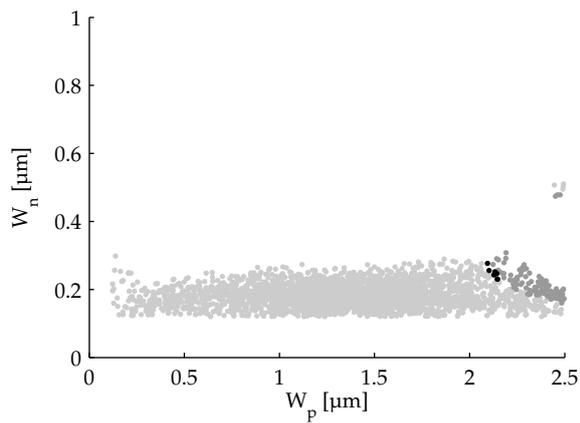
(h) Bildraum A0I22



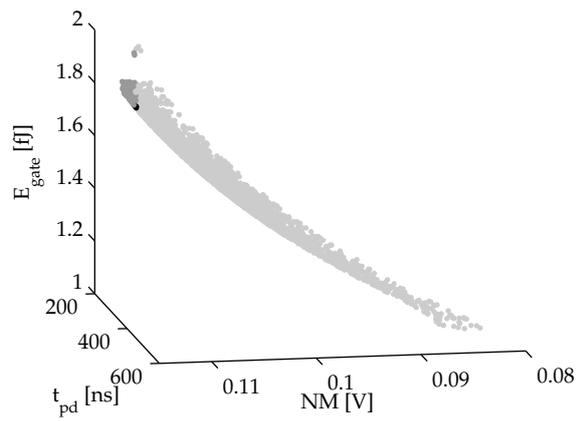
(i) Suchraum OAI22



(j) Bildraum OAI22



(k) Suchraum MUXI2 (Multiplexer-Stufe)



(l) Bildraum MUXI2 (Multiplexer-Stufe)

Abbildung 6.9: Pareto-Mengen und Pareto-Fronten der kombinatorischen Gatter (Fortsetzung)

Gatter	Boole'sche Funktion	Treiberstärken	W_n [μm]	W_p [μm]
AOI22	$(A \wedge B) \vee (C \wedge D)$	X1, X2	0,250	2,050
BUF	A	X1, X2, X3, X4, X6, X8	0,265	2,050
INV	\overline{A}	X1, X2, X3, X4, X6, X8	0,265	2,050
MIN3	$(A \wedge B) \vee (B \wedge C) \vee (C \wedge A)$	X1, X2	0,250	2,050
MUXI2	$(A \wedge \overline{S}) \vee (B \wedge S)$	X1, X2	0,250	2,150
NAND2	$\overline{A \wedge B}$	X1, X2, X4	0,230	1,710
NOR2	$\overline{A \vee B}$	X1, X2, X4	0,240	2,150
OAI22	$\overline{(A \vee B) \wedge (C \vee D)}$	X1, X2	0,245	2,200

Tabelle 6.1: Dimensionierung der kombinatorischen Gatter

Gatterdimensionierungen sind aus diesem eng eingegrenzten Bereich ausgewählt. Da die Lösungsmengen der ähnlich aufgebauten Gattertypen AOI22, OAI22 und MUXI2 sehr ähnlich sind, wurde auf eine separate Optimierung des ebenfalls ähnlich konstruierten MIN3-Gatters verzichtet und stattdessen die Dimensionierung des AOI22-Gatters übernommen. Die Ergebnisse der Gatterdimensionierung sind in Tabelle 6.1 angegeben.

Um den Einfluss von Prozessschwankungen auf die Eigenschaften der Gatter näher zu untersuchen, wurden *Post-Layout*¹-Monte-Carlo-Simulationen aller Gattertypen bei 300 mV vorgenommen. Abbildung 6.10 zeigt beispielhaft die Simulationsergebnisse des AOI22-Gatters der Treiberstärke X1. Für alle übrigen Gattertypen ergeben sich im Detail unterschiedliche Werte, jedoch prinzipiell ähnliche Zusammenhänge. Dargestellt sind Histogramme der auch bei der Optimierung verwendeten Ressourcenmaße t_{pd} , E_{gate} und NM sowie zusätzlich P_{stat} für 10 000 Simulationsläufe. Außerdem sind für 100 Simulationsläufe die sich ergebenden Spannungsübertragungskennlinien überlagert dargestellt. Relevant für eine korrekte Funktion einer aus den Standardzellen aufgebauten Digitalschaltung sind insbesondere die Streuungseigenschaften von t_{pd} und NM , da diese unmittelbar für die in Abschnitt 3.2 erläuterten primären Ausfallmechanismen, Hold-Zeit-Verletzungen und unzureichende Störabstände bzw. Logikpegel, verantwortlich sind. Die Variationen von E_{gate} und P_{stat} sind für eine korrekte Schaltungsfunktion von untergeordneter Bedeutung. Es sei an dieser Stelle lediglich erwähnt, dass der sehr geringe Variationskoeffizient σ/μ der Energie E_{gate} von 0,006 durch den überwiegenden Anteil E_{sw} der konstanten Lastkapazität von 10 fF für diese Simulation begründet ist. Der Mittelwert des Störabstands beträgt etwa 89,3 mV entsprechend ca. 30 % der Versorgungsspannung mit

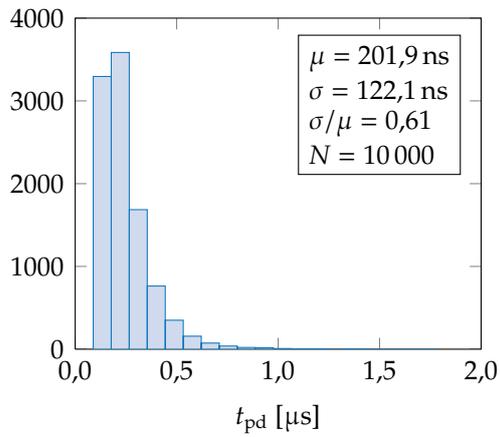
¹ Berücksichtigung aus dem Layout extrahierter parasitärer Kapazitäten

einem Variationskoeffizienten von 0,12. Diese Werte sind für sichere Ausprägungen der Logikpegel ausreichend. Der Variationskoeffizient der Verzögerungszeit von 0,61 erscheint auf den ersten Blick relativ groß. Zu beachten ist hierbei jedoch, dass dieser Wert globale und lokale Prozessschwankungen simultan abbildet. Innerhalb einer integrierten Schaltung entfällt der durch globale Prozessschwankungen verursachte Anteil, so dass sich eine wesentlich geringere Variation der zeitlichen Gattereigenschaften einstellt. Zudem nimmt die Variation der Verzögerung eines vollständigen Logikpfades mit der Anzahl der beteiligten Gatter ab, da sich Abweichungen der einzelnen Gatterverzögerungen teilweise aufheben. Zusätzlich können Variationen des zeitlichen Verhaltens einer Schaltung während der Optimierung durch Synthese- und Place-and-Route-Werkzeuge durch bestimmte Einstellungen berücksichtigt werden. Dies wird in Abschnitt 7.4 behandelt. Aus diesen Gründen sind auch die simulierten Streuungen der Verzögerungszeiten der Einzelgatter akzeptabel.

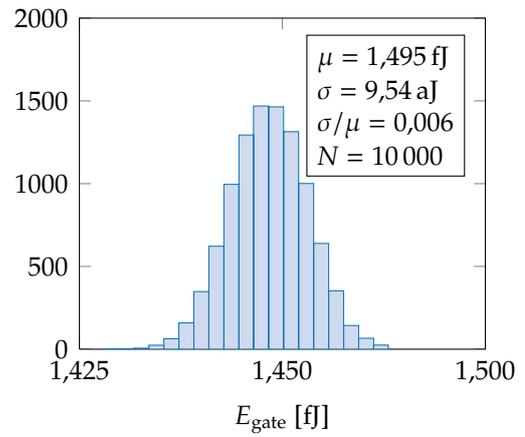
6.4 Flipflop und Latch

Es existiert eine Vielzahl verschiedener Implementierungsmöglichkeiten für Taktflanken-getriggerte Speicherelemente bzw. Flipflops. Viele Implementierungen wurden für spezielle Anwendungsfälle entworfen und bieten entsprechend individuelle Vor- und Nachteile. Da dabei jedoch oftmals Schaltungstechniken zum Einsatz kommen, die nicht als klassische, statische CMOS-Technik aufgefasst werden können, muss eine sorgfältige Untersuchung im Hinblick auf die Eignung für den Subschwellsbetrieb erfolgen.

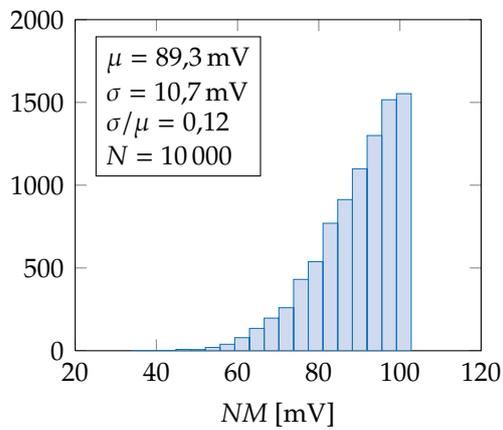
Generell lassen sich Flipflop-Schaltungen in statische und dynamische Typen unterteilen, wobei die Informationserhaltung bei statischen Typen durch Rückkopplungen und bei dynamischen Typen durch Ladungsspeicherung auf (parasitären) Kapazitäten erfolgt. Dynamische Flipflops besitzen im Allgemeinen den Vorteil einer gegenüber statischen Flipflops geringeren Verlustleistung aufgrund der fehlenden Rückkopplungs-Elemente bei gleichzeitig kleineren Verzögerungszeiten (*Clock to Output Delay* t_{CQ}). Im Subschwellsbereich ist allerdings die dynamisch gespeicherte Ladung aufgrund der niedrigen Versorgungsspannung sehr gering und kann daher leicht durch Rauscheffekte, Kopplung mit anderen Netzen, Leckströme oder radioaktive Strahlung verfälscht werden bzw. verloren gehen. Diese Anfälligkeit wird darüber hinaus durch Prozessparameter-Streuung weiter verstärkt [161]. Da die Robustheit ein wichtiges Entwurfskriterium der hier erstellten Subschwells-



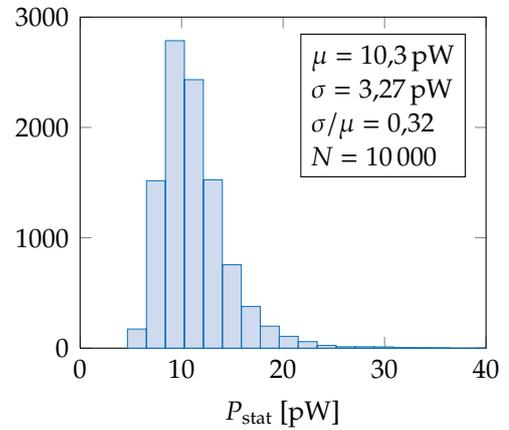
(a) Verzögerungszeit t_{pd}



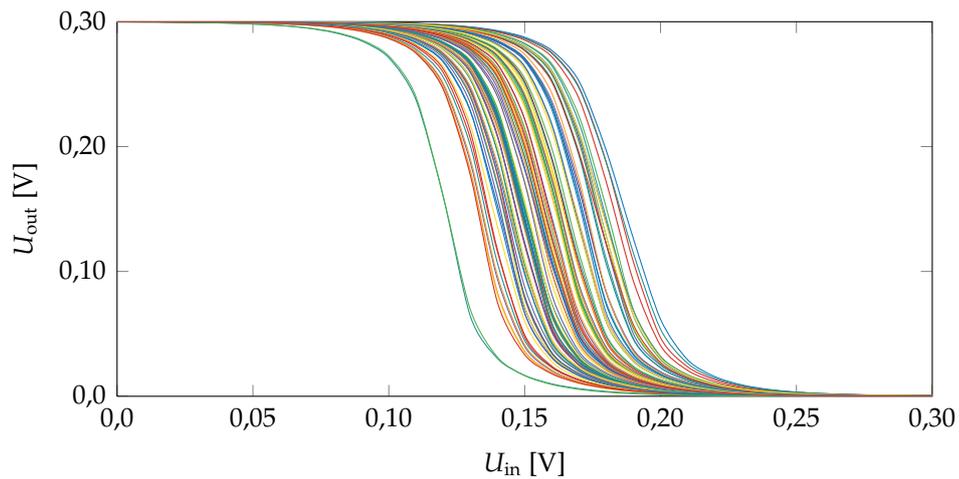
(b) Energie pro Umschaltvorgang E_{gate}



(c) Störabstand NM



(d) Statische Verlustleistung P_{stat}



(e) Spannungsübertragungskennlinie

Abbildung 6.10: Monte-Carlo-Simulationsergebnisse des A0I22-Gatters

Standardzellen-Bibliothek darstellt, wurden dynamische Flipflop-Schaltungen für die folgenden Betrachtungen von vornherein ausgeschlossen.

Eine weitere zum Teil in Flipflops zum Einsatz kommende Technik sind sogenannte *Ratioed Circuits*. Dabei wird die der Informationsspeicherung dienende Rückkopplung beim Schreiben eines neuen Zustands bzw. Datums nicht aufgetrennt. Stattdessen wird durch eine geeignete Transistordimensionierung sicher gestellt, dass das neue Datum durch einen starken Treiber den durch eine schwache Rückkopplung erhaltenen vorherigen Zustand überschreiben kann. In [161] wird jedoch gezeigt, dass Ratioed Circuits generell aufgrund des hohen Einflusses von Prozessparameter-Streuungen ungeeignet für den Betrieb im Subschwelligbereich sind und dass dies auch und insbesondere für Flipflops gilt. Eine analoge Problemstellung wurde auch in Abschnitt 5.3.1 bereits in Bezug auf Aufwärts-Pegelwandler betrachtet.

Untersuchungen zur Eignung weit verbreiteter Flipflop-Implementierungen für den Subschwelligbetrieb wurden in [3, 5, 36, 48] veröffentlicht. Die Veröffentlichungen umfassen u. a. folgende Flipflop-Typen: NAND2-basiert [147], *Clocked CMOS* (C2MOS) [143], *Transmission Gate Master Slave* (TGMS) [40], PowerPC 603 [51], *Hybrid Latch* (HL) [115], *Conditional Precharge Hybrid Latch* (CPHL) [111] und *Sense Amplifier* (SA) [104]. Auf weitere in den Veröffentlichungen behandelte Flipflops, die aufgrund der obigen Überlegungen auszuschließen sind, wird im Folgenden nicht eingegangen.

Die in [3] untersuchten Flipflops werden anhand von Simulationen in einer 90-nm-Technologie bezüglich der Eigenschaften im Subschwelligbetrieb bewertet. Dabei schneidet das PowerPC-Flipflop hinsichtlich Verzögerungszeiten und Leistungsaufnahme am besten ab. Mit 8 % höheren Verzögerungszeiten und etwa 20 % höherer Verlustleistung folgen das C2MOS- und das TGMS-Flipflop, mit leichten Vorteilen für das TGMS-Flipflop. Das NAND2-basierte Flipflop weist um ein Vielfaches schlechtere Werte auf. In [5] veröffentlichen die gleichen Autoren weitere Untersuchungen. Dabei erzielt in der gleichen 90-nm-Technologie ein SA-Flipflop nochmals etwa 15 % bessere Verzögerungszeiten als das PowerPC-Flipflop bei gleichzeitig um die Hälfte reduzierter Leistungsaufnahme. In einer dort ebenfalls betrachteten 65-nm-Technologie schneidet hinsichtlich der Verzögerungszeiten hingegen das PowerPC-Flipflop um den Faktor 2,8 besser als das SA-Flipflop ab. Die Verlustleistung des PowerPC-Flipflops ist dabei um knapp 70 % höher als die Verlustleistung des SA-Flipflops. Ein weiterer in [5] betrachteter Aspekt ist die niedrigste Versorgungsspannung, bei der die untersuchten Flipflops in allen relevanten Prozessparameter-Konstellationen noch korrekt arbeiten. Dabei liegen in der 90-nm-Technologie PowerPC- und SA-Flipflop

mit 175 mV gleichauf, während in der 65-nm-Technologie das PowerPC-Flipflop mit 125 mV einen Vorteil gegenüber dem SA-Flipflop mit 150 mV aufweist.

In [48] werden die Flipflop-Typen PowerPC, C2MOS, HL und SA durch Simulationen mit dem 90 nm *Predictive Technology Model* (PTM) [29] im Subschwelligbereich untersucht. Dabei erreicht das HL-Flipflop das beste zeitliche Verhalten, fällt allerdings bereits bei einer Versorgungsspannung von über 400 mV aus, während die übrigen Implementierungen bis unter 200 mV funktionsfähig sind. Dies ist dadurch zu erklären, dass der im HL-Flipflop erforderliche Impulsgenerator bei niedrigen Versorgungsspannungen nicht mehr einwandfrei funktioniert. Durch geeignete Modifikationen kann die Pulsdauer verlängert werden, wodurch das HL-Flipflop ebenfalls Versorgungsspannungen unter 200 mV erreicht. Das HL-Flipflop ist bei 200 mV etwa 36 % schneller als das PowerPC-Flipflop, gefolgt von C2MOS- und SA-Flipflop, die mit geringem Abstand etwa gleichauf liegen. Hinsichtlich des Energieverbrauchs schneiden das PowerPC-Flipflop und das C2MOS-Flipflop am besten ab.

In [36] werden bestehende Flipflop-Typen in einer 250-nm-SOI-Technologie simuliert und insbesondere im Hinblick auf ihre Empfindlichkeit gegenüber radioaktiver Strahlung untersucht. Dazu wird das Auftreffen eines hochenergetischen, elektrisch geladenen Teilchens simuliert, indem eine bestimmte Ladungsmenge in ein als sensibel identifiziertes Netz des jeweils untersuchten Flipflops injiziert wird. Es wird dann untersucht, ob das Flipflop den korrekten Zustand aufrecht erhalten kann. Die maximal injizierbare Ladung ohne Informationsverlust wird als kritische Ladung Q_{crit} bezeichnet. In Bezug auf die Verzögerungseigenschaften schneidet das CPHL-Flipflop am besten ab und besitzt in dieser Hinsicht im Subschwelligbereich einen Vorsprung gegenüber dem C2MOS-Flipflop von 25 %, gegenüber dem TGMS-Flipflop von 40 % und gegenüber dem SA-Flipflop von 50 %. Dieser Vorteil wird allerdings durch einen um 15 % bis 35 % erhöhten Energieverbrauch erkauft. Bezüglich der kritischen Ladung ist das C2MOS-Flipflop die beste Implementierung, gefolgt vom CPHL-Flipflop, dessen kritische Ladung bei nur etwa 55 % des Bestwertes liegt. TGMS-Flipflop und SA-Flipflop erreichen nur etwa 25 % der kritischen Ladung des C2MOS-Flipflops. In [36] werden darüber hinaus auch eine neue sowie zwei bestehende durch Redundanz strahlungsresistente Flipflop-Implementierungen untersucht. Diese weisen jedoch einen deutlich erhöhten Flächen- und Energiebedarf bei vergrößerten Verzögerungszeiten auf und sind daher eher für Anwendungen mit höchsten Anforderungen an Zuverlässigkeit oder Sicherheit sowie für Weltraum-Elektronik ausgelegt.

Die obigen Ergebnisse zeigen, dass je nach Technologie, Schaltungsdimensionierung und Bewertungskriterium unterschiedliche Flipflop-Varianten im Vorteil sind. Tendenziell bieten die miteinander verwandten Flipflops HL und CPHL das beste zeitliche Verhalten auf Kosten eines erhöhten Energieverbrauchs. Die korrekte Funktion dieser Flipflop-Typen hängt jedoch wesentlich von der exakten Impulserzeugung ab, was auch durch die gemäß [48] erforderlichen Modifikationen bestätigt wird. Da das zeitliche Schaltungsverhalten im Subschwelligebiet generell auch stark durch Prozessparameter-Streuung beeinflusst wird, ist für diese Flipflops von einer erhöhten Empfindlichkeit gegenüber derartigen Störeinflüssen auszugehen. In [48] und [36] wird der Einfluss von Prozessparameter-Streuung auf die korrekte Funktion des Impulsenergiegenerators nicht betrachtet. Darüber hinaus steht der erhöhte Energieverbrauch dem Einsatz in Anwendungen mit eingeschränktem Energie-Budget entgegen. Von den *Master-Slave*-Flipflops¹ C2MOS und TGMS weist tendenziell das C2MOS-Flipflop die besseren Werte in Bezug auf Verzögerungszeiten und Energieaufnahme auf. Gleichzeitig ist es um ein Vielfaches unempfindlicher gegenüber Ladungsinjektion. Aufgrund des ansonsten praktisch identischen Aufbaus kann die Anfälligkeit gegenüber injizierter Ladung den Transmission Gates des TGMS-Flipflop angelastet werden. Es ist daher für das PowerPC-Flipflop, das ebenfalls Transmission Gates enthält, von einer ähnlichen Empfindlichkeit auszugehen.

Für die im Rahmen dieser Arbeit entwickelte Subschwelligebiet-Standardzellen-Bibliothek wurde das C2MOS-Flipflop ausgewählt, da es einen sehr guten Kompromiss bezüglich Verzögerungszeiten, Energieaufnahme, unterer Betriebsspannungsgrenze und kritischer Ladung bietet. Darüber hinaus kommt in dieser Implementierung ausschließlich statische CMOS-Schaltungstechnik zum Einsatz, die sich im Subschwelligebiet bereits vielfach bewährt hat. Der Schaltplan des C2MOS-Flipflops ist in Abbildung 6.11 dargestellt. Je zwei C2MOS-Teilgatter bzw. Tristate-Inverter lassen sich auch als (invertierender) 2-zu-1-Multiplexer auffassen, der ebenfalls in der Subschwelligebietbibliothek implementiert wurde. Aus dem eingehenden Taktsignal CLK werden durch zwei Inverter die lokal benötigten invertierten und nicht-invertierten Kopien erzeugt, um die Belastung des Taktbaumes gering zu halten und die Taktverzögerung innerhalb des Flipflops besser optimieren zu können. Der Ausgang Q wird durch einen separaten Inverter gepuffert. Dadurch wird erreicht, dass Setup- und Hold-Zeiten des Flipflops unabhängig von der angeschlossenen Ausgangslast sind. Außerdem lassen sich höhere Treiberstärken auf diese Weise sehr einfach realisieren, da nur dieser ausgangsseitige Inverter zu vergrößern ist. Insgesamt benötigt das C2MOS-Flipflop 26 Transistoren. Indem die beiden Inverter in Master-

¹ Flipflop-Typen, die aus zwei hintereinander geschalteten Latches bestehen, dem *Master-Latch* und dem *Slave-Latch*

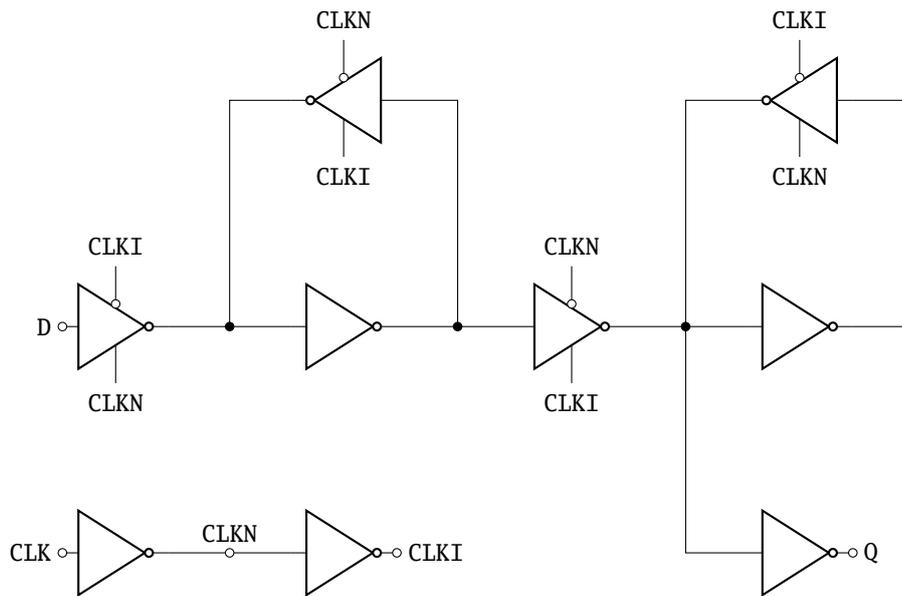


Abbildung 6.11: Schaltplan des verwendeten C2MOS-Flipflops

und Slave-Latch jeweils durch NOR2-Gatter ersetzt werden, lässt sich ein Flipflop mit asynchronem, high-aktivem *Reset*-Eingang (zur Initialisierung mit dem Logikpegel 0) realisieren. Setzt man stattdessen NAND2-Gatter an diesen Stellen ein, so erhält man ein Flipflop mit asynchronem, low-aktivem *Set*-Eingang (zur Initialisierung mit dem Logikpegel 1). Entfernt man den Master-Latch des Flipflops, so erhält man einen D-Latch, wobei auch dieser durch asynchrone Steuereingänge ergänzt werden kann. Diese Zellen wurden ebenfalls in die Subschwelligbibliothek aufgenommen.

Da Flipflop und Latch aus bereits vorhandenen kombinatorischen Gattern der Subschwelligbibliothek bestehen, nämlich den Gattertypen INV, MUX12, NAND2 und NOR2, wurden die entsprechenden Dimensionierungen der jeweils kleinsten Treiberstärke übernommen. Es wurden allerdings die PMOS-Weiten der Inverter zur Erzeugung der internen Taktsignale CLKN und CLKI durch zusätzliche Optimierungen modifiziert, um in diesen Elementen gleich große Verzögerungen für steigende und fallende Taktflanken sicherzustellen und somit die Symmetrie des Taktsignals nicht zu beeinträchtigen. Unterschiedliche Treiberstärken der Flipflops und Latches – X1, X2 und X4 – wurden durch Verwendung der entsprechenden ausgangsseitigen Puffer realisiert. Die Funktionalität und das Zeitverhalten der auf diese Weise dimensionierten sequenziellen Gatter wurde durch Monte-Carlo-Simulationen überprüft.

6.5 Taktbaum-Elemente

In synchronen Digitalschaltungen muss ein globales Taktsignal an alle vorhandenen Flipflops verteilt werden, um einheitliche Zeitintervalle für die Verarbeitung von Informationen zu gewährleisten. Da in der Regel eine große Anzahl an Flipflops zu versorgen ist, die Treiberfähigkeit eines einzelnen Takteingangs jedoch begrenzt ist, ist der Aufbau eines mehrstufigen Taktbaums aus Treiberbausteinen erforderlich. Die verwendeten Treiber müssen dabei so ausgelegt sein, dass Anstiegs- und Abfallzeiten bzw. die entsprechenden Verzögerungszeiten gleich groß sind, um auch über mehrere Stufen hinweg ein konstantes Tastverhältnis des Taktsignals von 50 % zu erhalten. Andernfalls können ggf. Fehlfunktionen in den angeschlossenen Flipflops auftreten. Des Weiteren soll die zur Synchronisierung typischerweise verwendete steigende Taktflanke an allen Flipflops in jedem Taktintervall möglichst gleichzeitig eintreffen, d.h. der Zeitversatz des Taktsignals (engl. *Clock Skew*) ist zu minimieren. Dieses Optimierungsproblem wird durch Softwarewerkzeuge zur Taktbaum-Synthese im Rahmen der Platzierung und Verdrahtung gelöst. Voraussetzung ist dafür eine Bibliothek von Takttreibern, die zum einen die obigen Anforderungen an Anstiegs- und Abfallzeiten erfüllen und zum anderen eine Reihe unterschiedlicher Treiberstärken aufweisen, um den Zeitversatz minimieren zu können.

Für die Subschwelligbibliothek wurden Takttreiber mit Treiberstärken von X1 bis X12 in zweistufigem Aufbau implementiert. Wie auch bei den Treibern der Logikbibliothek besteht die ausgangsseitige Stufe eines Takttreibers aus einer Anzahl entsprechend der Treiberstärke parallel geschalteter Inverter. Die eingangsseitige Stufe besteht ebenfalls aus parallel geschalteten Invertern, wobei die Anzahl so gewählt ist, dass das Verhältnis der Gesamt-Transistorweiten von Ausgangs- zu Eingangsstufe möglichst zwischen 2 und 3 und damit möglichst nah am theoretisch optimalen Wert von $e = 2,71828 \dots$ [92] liegt. Gegenüber den Treibern der Logikbibliothek wurden jedoch nachträglich die Weiten der PMOS-Transistoren so optimiert, dass die Verzögerungszeiten für steigende und fallende Flanken bei einer Versorgungsspannung von 300 mV übereinstimmen.

Eine häufig in integrierten Schaltungen angewandte Technik zur Reduktion der Leistungsaufnahme ist die Taktabschaltung (engl. *Clock Gating*), also das Ausblenden des Taktsignals in einzelnen Zweigen des Taktbaums. Dies kann entweder manuell oder automatisiert durch Synthese-Werkzeuge, grobgranular auf Blockebene durch Abschalten nicht benötigter Module oder feingranular auf Ebene weniger zusammengehöriger Flipflops erfolgen. Für diese Technik sind jedoch spezielle

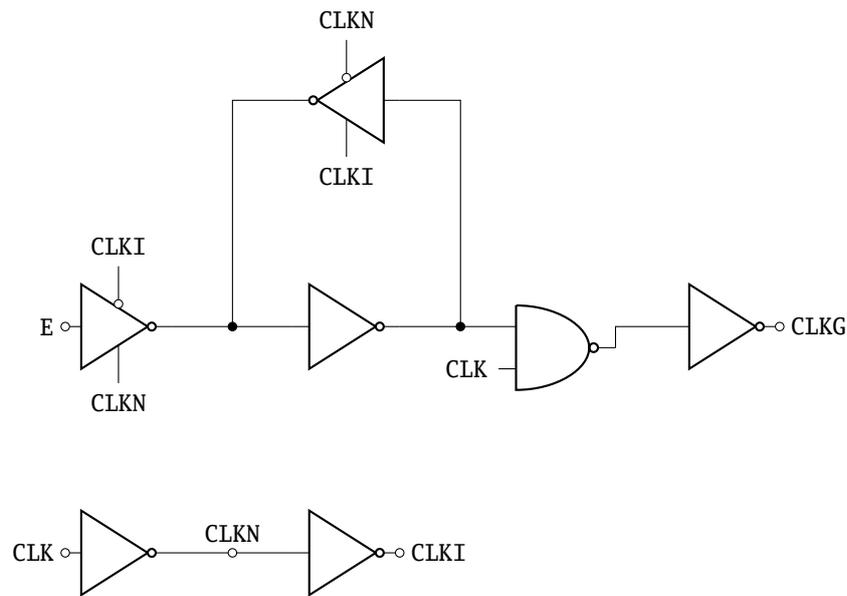


Abbildung 6.12: Schaltplan der Clock-Gating-Zelle

Clock-Gating-Zellen erforderlich, die in der Lage sind, das Taktsignal anhand einer gegebenen Bedingung störungsfrei zu aktivieren und zu deaktivieren. Dabei darf es nicht zu unerwünschten Taktflanken am Ausgang des Clock Gates kommen, da ansonsten Daten von Flipflops fälschlicherweise übernommen werden könnten. Eine verbreitete und sichere Methode ist die Verwendung eines Latch-basierten Clock Gates. Dies stellt sicher, dass das Ausgangssignal nur während der Low-Phase des Taktes aktiviert werden kann, so dass keine störenden Taktflanken entstehen können. Der Schaltplan des Clock Gates ist in Abbildung 6.12 dargestellt. Die Dimensionierung des Latches ist identisch zum Latch der Logikbibliothek. Die Weiten der PMOS-Transistoren von ausgangsseitigem NAND-Gatter und Inverter sind hingegen so optimiert, dass wie auch bei den Takttreibern ein symmetrisches Anstiegs- und Abfallverhalten erreicht wird. Der Ausgang besitzt durch die Parallelschaltung zweier Inverter die Treiberstärke X2.

Die Dimensionierungen der Takttreiber und des Clock Gates sind in Tabelle 6.2 gegeben.

Treiberstärke	n -mal	W_n [μm]	W_p [μm]	n -mal	W_n [μm]	W_p [μm]
Takttreiber	Stufe 1			Stufe 2		
X1	1	0,265	1,990	1	0,265	1,820
X2	1	0,265	1,965	2	0,265	1,835
X3	1	0,265	1,940	3	0,265	1,855
X4	2	0,265	1,965	4	0,265	1,855
X5	2	0,265	1,955	5	0,265	1,870
X6	2	0,265	1,940	6	0,265	1,890
X7	3	0,265	1,960	7	0,265	1,890
X8	3	0,265	1,945	8	0,265	1,905
X9	3	0,265	1,930	9	0,265	1,920
X10	4	0,265	1,945	10	0,265	1,915
X11	4	0,265	1,940	11	0,265	1,930
X12	4	0,265	1,935	12	0,265	1,945
Clock Gate	Ausgangsstufe NAND2			Ausgangsstufe INV		
X2	1	0,265	0,705	2	0,265	1,885

Tabelle 6.2: Dimensionierung der Taktbaum-Elemente

6.6 Pegelwandler

6.6.1 Aufwärts-Pegelwandler

In Abschnitt 5.3.1 wurden verschiedene Implementierungen für Aufwärts-Pegelwandler vorgestellt und verglichen. Alle Schaltungsvarianten wiesen dabei Nachteile in Bezug auf statische Querströme, Fläche, Robustheit oder schlechtes Skalierungsverhalten in Zusammenhang mit DVFS auf. Für den Aufbau der 65-nm-Subschwelligbibliothek wurde daher eine neue Pegelwandler-Schaltung entwickelt. Dabei wurde insbesondere ein Entwurf ohne statische Strompfade sowie ein gutes Skalierungsverhalten der Schaltzeiten bei einer Erhöhung der Versorgungsspannung gefordert. Da hierdurch eine niedrige Verlustleistung sowie ein geringer Einfluss auf die maximale Taktfrequenz beim Betrieb mit Versorgungsspannungen oberhalb der Schwellspannung erreicht wird, eignet sich dieser Pegelwandler besonders für DVFS-Systeme. Die Inhalte dieses Abschnitts wurden in [183] veröffentlicht.

Der Schaltplan des entwickelten Aufwärts-Pegelwandlers ist in Abbildung 6.13 dargestellt. Die einem Wilson-Stromspiegel entsprechende Transistoranordnung M4, M5, M7 stellt sicher, dass kein statischer Stromfluss durch M3 oder M6 stattfinden

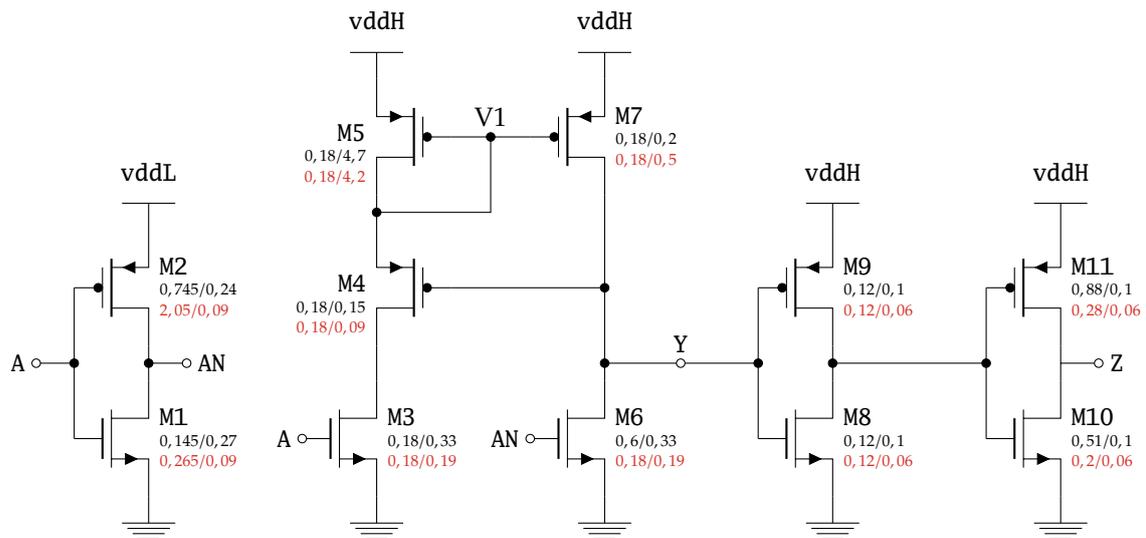


Abbildung 6.13: Pegelwandler mit Wilson-Stromspiegel mit optimierter Dimensionierung für 90 nm (schwarz) und 65 nm (rot)

kann, wenn einer dieser beiden Transistoren gesperrt ist. Wenn an A ein niedriger Logikpegel und an AN ein hoher Logikpegel anliegt, leitet M6 und zieht Y nach Masse. Da M3 sperrt, wird V1 durch M5 aufgeladen bis M5 und M7 ebenfalls sperren. Es kann dann kein statischer Strom durch einen der beiden Pfade auftreten. Liegt dagegen an A ein hoher und an AN ein niedriger Logikpegel, leitet M3, wodurch ein Stromfluss durch M3, M4 und M5 stattfindet. Da M5, M7 ein Stromspiegel ist, fließt auch in M7 ein Strom, der den Knoten Y auflädt. Mit steigendem Spannungspegel an Y sperrt M4 zunehmend, so dass wiederum kein statischer Stromfluss durch M3, M4, M5 auftreten kann.

Zu Vergleichszwecken wurde der Pegelwandler zunächst in der zuvor betrachteten 90 nm Technologie implementiert, simuliert und die Ergebnisse denen des nach [39] in Abschnitt 5.3.1.3 implementierten Pegelwandlers gegenüber gestellt. Die Dimensionierung des Pegelwandlers mit Wilson-Stromspiegel wurde ebenfalls mit dem Ziel einer Minimierung des EDP unter Verwendung der Software Cadence Virtuoso Analog Circuit Optimization durchgeführt. Es wurde wiederum eine Spannung von 200 mV auf der Eingangsseite sowie 1 V auf der Ausgangsseite des Pegelwandlers angelegt.

Zur Implementierung eines Pegelwandlers für die 65-nm-Subschwelligbibliothek wurde außerdem eine Optimierung für diese Technologie mit gleicher Zielsetzung

durchgeführt. Die niedrige Versorgungsspannung betrug dabei 300 mV und die hohe Versorgungsspannung 1,2 V.

Die gewonnenen Transistormaße des Pegelwandlers für die 90-nm- und 65-nm-Technologien sind in Abbildung 6.13 in der Form W/L in der Einheit μm angegeben.

6.6.1.1 Simulationsergebnisse

Die zunächst in diesem Abschnitt behandelten Ergebnisse beziehen sich auf die 90-nm-Implementierung des Pegelwandlers mit Wilson-Stromspiegel, um Vergleiche zu der bestehenden Implementierung nach Abschnitt 5.3.1.3 zu ziehen. Die wesentlichen Eigenschaften des Pegelwandlers sind prinzipiell auf die 65-nm-Technologie übertragbar, entsprechende Simulationsergebnisse werden am Ende dieses Abschnitts vorgestellt.

Beste und schlechteste PVT-Bedingungen für Subschwell-Pegelwandler entsprechen nicht den für allgemeine CMOS-Schaltungen üblichen Konstellationen aus Abschnitt 3.3. Die nominellen Bedingungen lauten: typischer NMOS, typischer PMOS, 200 mV niedrige Versorgungsspannung, 1 V hohe Versorgungsspannung sowie eine Temperatur von 25 °C. Wie in Abschnitt 5.3.1 erläutert, muss jeweils ein NMOS-Transistor die Treiberstärke des zugehörigen PMOS-Transistors überwinden, um den Zustand des Pegelwandlers zu kippen. Dies trifft sowohl auf den Pegelwandler mit Wilson-Stromspiegel als auch auf den Pegelwandler aus [39] zu. Daher ist die Konstellation langsamer NMOS, schneller PMOS am kritischsten für das Schaltverhalten des Pegelwandlers. Eine verringerte Versorgungsspannung auf der Eingangsseite und eine erhöhte Versorgungsspannung auf der Ausgangsseite führen zu einer weiteren Verschlechterung der Situation. Schließlich führt im Subschwellbereich eine niedrige Temperatur zu verringerten Transistorströmen, was das Schaltverhalten weiter verlangsamt. Daher wurde die schlechteste PVT-Bedingung als langsamer NMOS, schneller PMOS, -10% niedrige Versorgungsspannung, $+10\%$ hohe Versorgungsspannung bei einer Temperatur von -40 °C festgelegt. Durch Simulationen wurde festgestellt, dass die minimalen Verzögerungszeiten nicht bei entgegengesetzten Prozess-Bedingungen, also schneller NMOS, langsamer PMOS auftreten. Stattdessen wurde die Konstellation schneller NMOS, schneller PMOS, $+10\%$ niedrige Versorgungsspannung, -10% hohe Versorgungsspannung und eine Temperatur von 120 °C als beste PVT-Bedingung festgelegt.

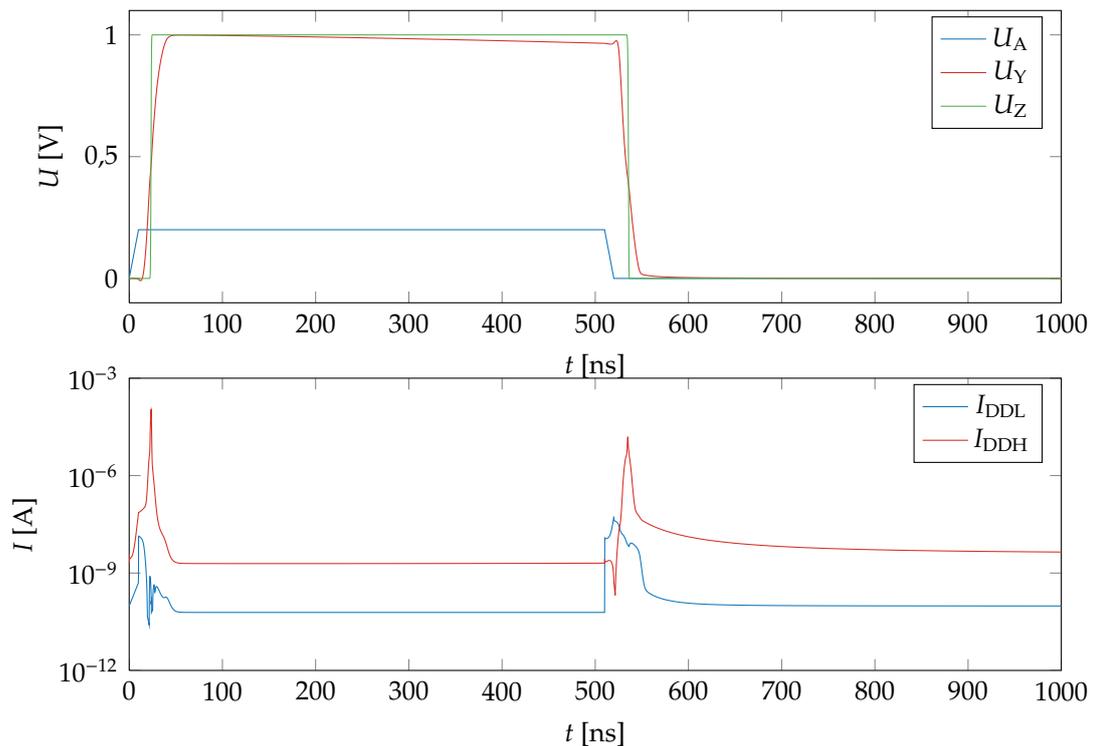


Abbildung 6.14: Transientenanalyse des Pegelwandlers mit Wilson-Stromspiegel

Das Ergebnis einer Transientenanalyse des Pegelwandlers ist in Abbildung 6.14 dargestellt. Das obere Diagramm zeigt die Spannungen am Eingangsknoten A, am Ausgangsknoten Y sowie am gepufferten Ausgang Z. Diese Simulation wurde unter nominellen PVT-Bedingungen durchgeführt. Am Eingang wurden Anstiegs- und Abfallzeiten von 10 ns sowie am gepufferten Ausgang eine Lastkapazität von 100 fF angenommen. Wie dem Diagramm zu entnehmen ist, erreicht der Pegelwandler am Ausgang einen maximalen Spannungshub (*Rail-to-Rail*-Charakteristik). Die Lastkapazität hat aufgrund des gepufferten Ausgangs nur einen geringen Einfluss auf die Verzögerungszeiten. Würde dagegen Knoten Y direkt mit einer Lastkapazität von 100 fF belastet, wären aufgrund der geringen Subschwellig-Treiberfähigkeit der zentralen Pegelwandler-Stufe sehr große Verzögerungszeiten zu beobachten.

Die Verzögerungszeiten der beiden untersuchten Implementierungen sind in Abhängigkeit von der niedrigen Versorgungsspannung in Abbildung 6.15a dargestellt. Danach arbeiten beide Pegelwandler bei Spannungen zwischen 100 mV und 1 V unter nominellen PVT-Bedingungen fehlerfrei. Damit eignen sich die Schaltungen in Bezug auf die Funktionalität auch für DVFS-Systeme. Für alle Versorgungsspannungen unterhalb von 1 V ist die Verzögerungszeit des Pegelwandlers mit

Wilson-Stromspiegel geringer als die der Schaltung nach [39]. Bei 200 mV beträgt die Verzögerungszeit des hier vorgestellten Pegelwandlers 18,4 ns und damit etwa 2,3-mal weniger als die Verzögerungszeit von 43,2 ns des Pegelwandlers aus [39].

Abbildung 6.15b zeigt die Verzögerungszeit des vorgeschlagenen Pegelwandlers für unterschiedliche PVT-Bedingungen. Bei 200 mV ist die schlechteste Verzögerungszeit etwa um den Faktor 140 größer als die beste Verzögerungszeit. Bei 100 mV erhöht sich dieser Faktor auf 1200. Dennoch arbeitet der Pegelwandler korrekt unter den betrachteten PVT-Konstellationen.

Um einen Eindruck für die Empfindlichkeit der Pegelwandler gegenüber Prozessschwankungen zu bekommen, wurde bei einer Versorgungsspannung von 200 mV eine Monte-Carlo-Simulation mit einem Stichprobenumfang von 10 000 Punkten unter Berücksichtigung von globalen und lokalen Streuungen durchgeführt. Die Ergebnisse sind in Form von Histogrammen in Abbildungen 6.15c und 6.15d dargestellt. Wie in [161] ausgeführt, folgt die Verzögerungszeit von CMOS-Gattern im Subschwellebetrieb einer logarithmischen Normalverteilung. Die in den Abbildungen 6.15c und 6.15d überlagerten angepassten logarithmischen Normalverteilungen deuten darauf hin, dass diese Approximation auch für die hier betrachteten Pegelwandler angenommen werden kann. Die Verzögerungszeit des Pegelwandlers mit Wilson-Stromspiegel weist eine geringere Empfindlichkeit gegenüber Prozessparameter-Variationen auf als der Pegelwandler nach [39], wie anhand der normalisierten Standardabweichungen σ/μ von 0,28 bzw. 0,46 zu erkennen ist. Falls erforderlich kann die Robustheit durch Vergrößern der Transistormaße weiter verbessert werden.

In Abbildung 6.15e werden die betrachteten Pegelwandler in Bezug auf die pro Schaltvorgang umgesetzte Energie bei einer Signalfrequenz von 1 MHz verglichen. Aufgrund des statischen Referenz-Strompfades des Pegelwandlers aus [39] ist dessen Energieumsatz vor allem für hohe Versorgungsspannungen größer als der Energieumsatz des Pegelwandlers mit Wilson-Stromspiegel. Wie zu erwarten, kann die Verlustenergie beider Pegelwandler durch Absenken der Versorgungsspannung im Bereich starker Inversion sowie zu Beginn des Subschwellebereichs reduziert werden. Im tiefen Subschwellebetrieb tritt jedoch eine Erhöhung der Energie pro Schaltvorgang auf, da in diesem Bereich die statische Verlustleistung dominiert. Das Diagramm zeigt für den hier vorgestellten Pegelwandler ein relativ flaches, breites Energieminimum zwischen 300 mV und 600 mV. Für eine Versorgungsspannung von 200 mV wurde ein Energieumsatz von 93,9 fJ ermittelt. Bei dieser Betriebsspannung benötigt der Pegelwandler nach [39] eine Energie von 182,7 fJ.

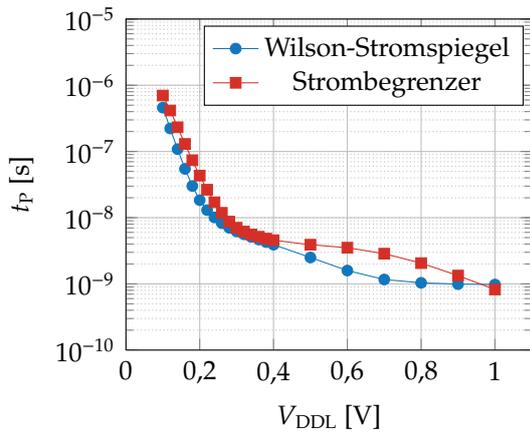
Implementierung	Strombegrenzer	Wilson-Stromspiegel		
		90 nm	65 nm	
Versorgungsspannung	200	200	300	mV
Gesamt-Kanalfläche	3,56	1,38	1,18	μm^2
Verzögerungszeit	43,2	18,4	28,2	ns
Gesamtenergie	182,7	93,9	162,7	fJ
statische Verlustleistung	8,6	6,6	0,26	nW

Tabelle 6.3: Zusammenfassung der Simulationsergebnisse

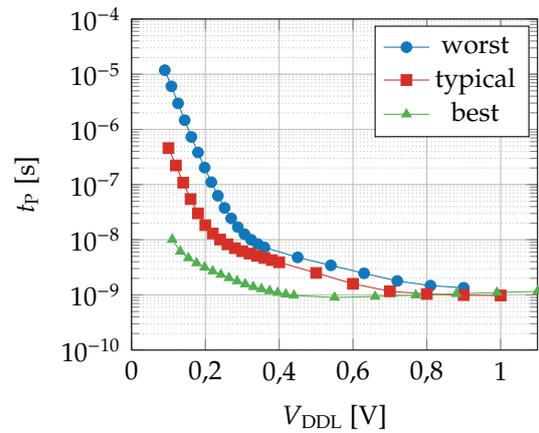
Die statische Verlustleistung der beiden Pegelwandler ist in Abbildung 6.15f dargestellt. Der vorgestellte Pegelwandler benötigt eine statische Verlustleistung von 6,6 nW bei 200 mV und 9,3 nW bei 1 V. Im Gegensatz dazu liegt die statische Verlustleistung des Pegelwandlers nach [39] zwischen 3,4 nW und 1,8 μW und ist dabei oberhalb einer Versorgungsspannung von 200 mV stets größer als die statische Verlustleistung des Pegelwandlers mit Wilson-Stromspiegel. Wie bereits erläutert, ist dies auf den statischen Referenzstrom-Pfad dieser Implementierung zurück zu führen. Bei 1 V Versorgungsspannung verbraucht der Pegelwandler aus [39] 190-mal mehr statische Verlustleistung als der hier vorgestellte Pegelwandler.

Das Verhalten des Pegelwandlers mit Wilson-Stromspiegel in der 65-nm-Implementierung kann den Diagrammen in Abbildung 6.16 entnommen werden. Qualitativ zeigen sich nur geringfügige Unterschiede gegenüber den vorangegangenen Simulationen. Hervorzuheben ist, dass durch die auf niedrige Verlustleistung optimierte 65-nm-Technologie trotz der von 200 mV auf 300 mV vergrößerten Versorgungsspannung leicht vergrößerte Verzögerungszeiten zu beobachten sind, während gleichzeitig die statische Verlustleistung um eine Größenordnung kleiner ist.

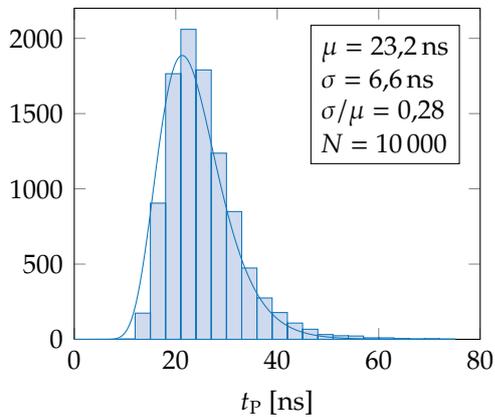
Die Eigenschaften der untersuchten Pegelwandler sind in Tabelle 6.3 zusammengefasst. Die hier angegebenen Werte gelten für nominelle PVT-Bedingungen. Für einen prinzipiellen Vergleich der Layoutfläche ist die Gesamtfläche aller Transistorkanäle angegeben. Aufgrund der großen Transistoren M3 und M4 im Referenzpfad des Pegelwandlers aus [39] ist dessen Gesamt-Kanalfläche mehr als zweimal so groß wie die Gesamt-Kanalfläche des hier vorgestellten Pegelwandlers. Dabei ist allerdings, wie schon oben erwähnt, anzumerken, dass mehrere Pegelwandler einen gemeinsamen Referenz-Pfad verwenden könnten, wodurch sich der Flächenbedarf reduzieren lässt.



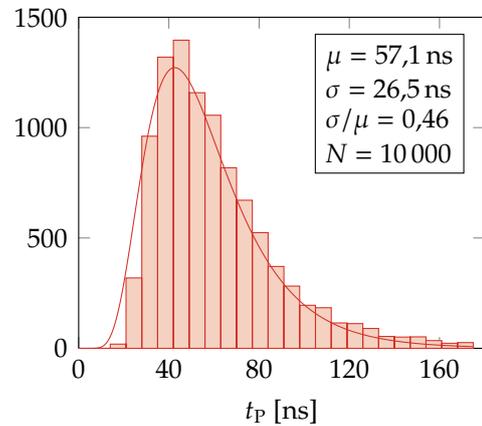
(a) Vergleich der Verzögerungszeiten unter nominellen PVT-Bedingungen



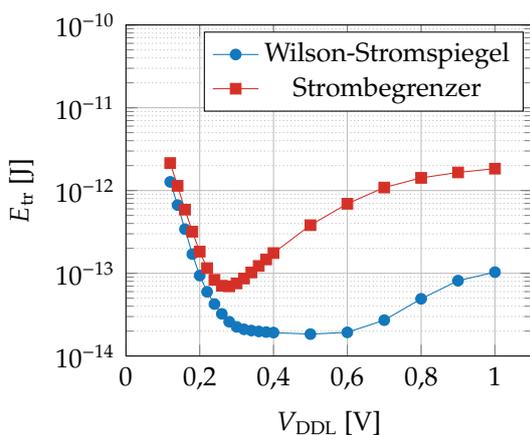
(b) Verzögerungszeit bei verschiedenen PVT-Bedingungen



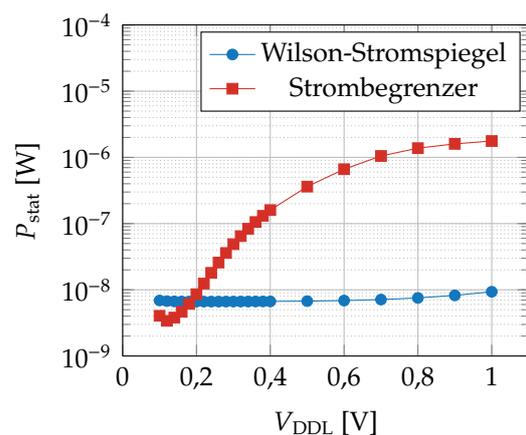
(c) Verteilung der Verzögerungszeit des vorgestellten Pegelwandlers



(d) Verteilung der Verzögerungszeit des Pegelwandlers nach [39]

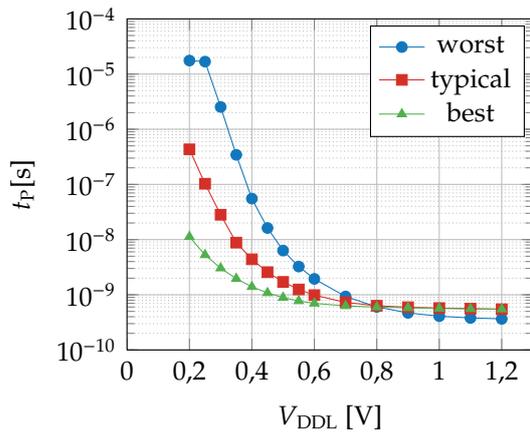


(e) Energie pro Umschaltvorgang bei 1 MHz

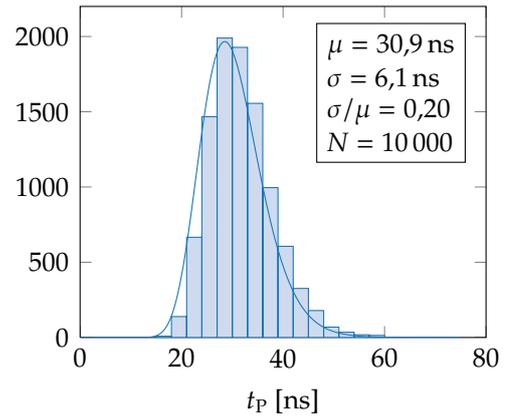


(f) Statische Verlustleistung

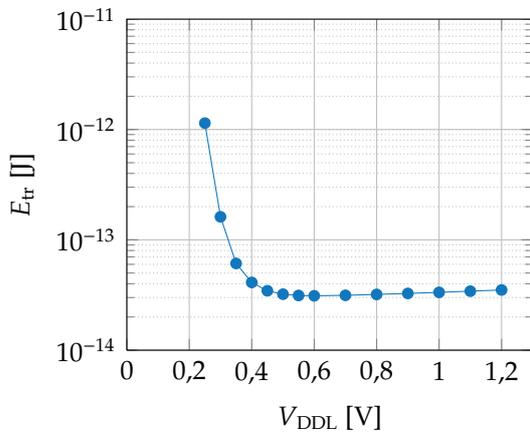
Abbildung 6.15: Simulationsergebnisse in der 90-nm-Technologie



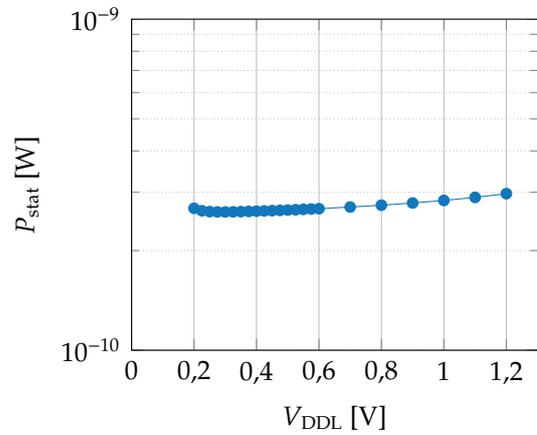
(a) Verzögerungszeit bei verschiedenen PVT-Bedingungen



(b) Verteilung der Verzögerungszeit



(c) Gesamtenergie pro Umschaltvorgang bei 1 MHz



(d) Statische Verlustleistung

Abbildung 6.16: Simulationsergebnisse in der 65-nm-Technologie

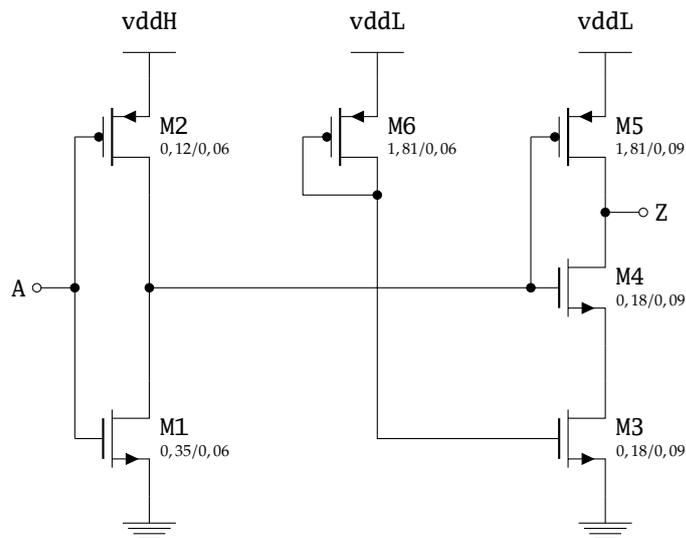


Abbildung 6.17: Schaltplan des Abwärts-Pegelwandlers

6.6.2 Abwärts-Pegelwandler

Für den Abwärts-Pegelwandler kommt die bereits in Abschnitt 5.3.2 vorgestellte Implementierung in leicht modifizierter Form zum Einsatz. Der Schaltplan mit für die 65-nm-Technologie optimierter Dimensionierung ist in Abbildung 6.17 angegeben. Die direkte Verbindung des Gates von M3 mit der Versorgungsspannung wird durch den zusätzlichen Transistor M6 vermieden. Die Motivation und Funktionsweise dieser Schaltungsänderung wird im nachfolgenden Abschnitt in Bezug auf TieHi- und TieLo-Zellen erläutert.

6.7 Place-and-Route-Zellen

Neben funktionalen Elementen enthält die Subschwelligbibliothek weitere Zellen, die vom Place-and-Route-Werkzeug benötigt werden. Hierbei handelt es sich um sogenannte *Filler-Zellen*, *Decap-Zellen* sowie *TieHi-* und *TieLo-Zellen*.

Filler-Zellen dienen zum Auffüllen nicht durch funktionale Zellen belegter Stellen im Layout. Filler-Zellen enthalten keine Transistoren, sondern sorgen für eine lückenlose Struktur der n-Wannen, Dotierungsgebiete, Substrat- und Wannenkontakte sowie der Versorgungsleitungen (*Power Rails*).

Bei größeren Leerstellen im Layout bietet sich die Möglichkeit, Entkopplungskapazitäten zur Stützung der Versorgungsspannung innerhalb des Schaltkreises zu integrieren. Diese Aufgabe haben Decapzellen. Als Kapazitäten kommen dabei die Gate-Kapazitäten großflächiger Transistoren zum Einsatz.

In den meisten digitalen Schaltungen existieren Signale, die einen konstanten Logikpegel besitzen. Eine Möglichkeit, Signale auf einen bestimmten Logikpegel zu legen, besteht in einer direkten Verbindung mit der positiven oder negativen Versorgungsspannung. Dieses Vorgehen ist zwar sehr einfach und erfordert keinerlei zusätzlichen Aufwand, hat allerdings einen Nachteil hinsichtlich der Robustheit der Schaltungen. Spannungsspitzen auf den Versorgungsnetzen wirken bei dieser Methode direkt auf die angeschlossenen Transistor-Gates der nachfolgenden Gattereingänge. Durch elektrostatische Entladungen (engl. *Electrostatic Discharge*, ESD) außerhalb einer integrierten Schaltung, beispielsweise durch Berühren leitfähiger Komponenten eines Schaltungsaufbaus, können Spannungsspitzen von mehreren 10 kV hervorgerufen werden. CMOS-Schaltungen weisen generell eine hohe Empfindlichkeit gegenüber ESD auf, da das nur wenige Nanometer dicke Gateoxid durch diese hohen Spannungen auch bei kurzzeitiger Einwirkung leicht zerstört werden kann. Die Durchbruchspannung des Gateoxids liegt in der hier verwendeten Technologie beispielsweise bei weniger als 2 V. Aus diesem Grund enthalten die Padzellen integrierter CMOS-Bausteine typischerweise umfangreiche Schutzmaßnahmen, die ESD-Ereignisse bis zu einem gewissen Grad tolerieren und ableiten können, so dass die eigentlichen Nutzschaltungen nicht beschädigt werden. Beim Betrieb mit Spannungen im Subschwelligbereich sind die ESD-Schutzfunktionen der entsprechenden Versorgungspads jedoch äußerst eingeschränkt oder unwirksam. Auf der anderen Seite reagieren die Subschwelligschaltungen aber ausgesprochen empfindlich gegenüber jeder Abweichung von der normalen Gateoxid-Charakteristik. Zum Schutz der Transistor-Gates sollten diese daher nicht direkt mit Versorgungsnetzen verbunden werden. Diese Maßnahme wird im Übrigen generell vom Hersteller der hier eingesetzten CMOS-Technologie empfohlen.

Eine Möglichkeit, Signale ohne direkte Verbindung zu Versorgungsnetzen auf konstante Logikpegel zu legen, bieten TieHi- und TieLo-Zellen. Die Schaltpläne dieser Schaltungen sind in Abbildung 6.18 dargestellt. Die Funktionsweise der TieHi-Zelle ergibt sich wie folgt: Ist die Spannung an den Transistor-Gates größer als die Schwellspannung von M1, so leitet M1 und die Spannung sinkt. Auch unterhalb der Schwellspannung sinkt die Spannung durch Subschwelligströme weiter und erreicht nach einiger Zeit einen Wert nahe 0 V. Dadurch leitet M2 und verbindet Ausgang Z mit der positiven Versorgungsspannung. Dabei besteht keine direkte

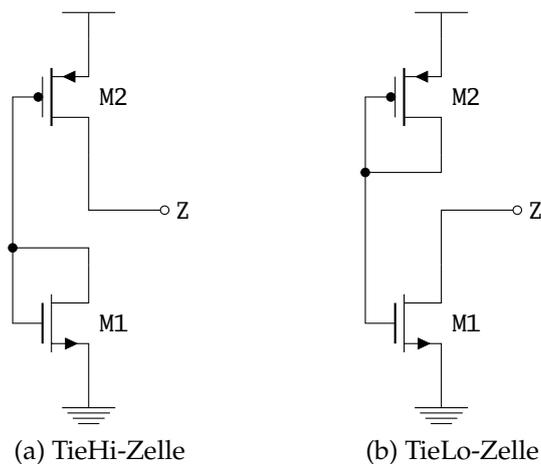


Abbildung 6.18: Schaltpläne der TieHi- und TieLo-Zellen

Verbindung zwischen einem Transistor-Gate und der positiven oder negativen Versorgungsspannung. Durch den endlichen Leitwert des Transistorkanals von M2 können kurzzeitige Spannungsspitzen sich kaum auf den Ausgang Z auswirken. Die Funktionsweise der TieLo-Zelle lässt sich analog erklären.

Die Dimensionierung der Tiezellen ergibt sich aus zwei Anforderungen: Erstens soll das Ausgangssignal nach Inbetriebnahme der Schaltung möglichst schnell den stabilen Endwert erreichen und zweitens soll die Ausgangsimpedanz auch nach Erreichen des Endwertes möglichst gering sein, um die Spannung auch bei Störeinflüssen wie etwa kapazitiven Einkopplungen anderer Netze stabil gehalten werden. Durch Verwenden großer Transistorweiten für M1 und M2 lassen sich diese Anforderungen erfüllen. Daher wurden die durch den Standardzellenrahmen maximal möglichen Weiten von $0,65\ \mu\text{m}$ für M1 und $2,3\ \mu\text{m}$ für M2 verwendet.

6.8 Layout

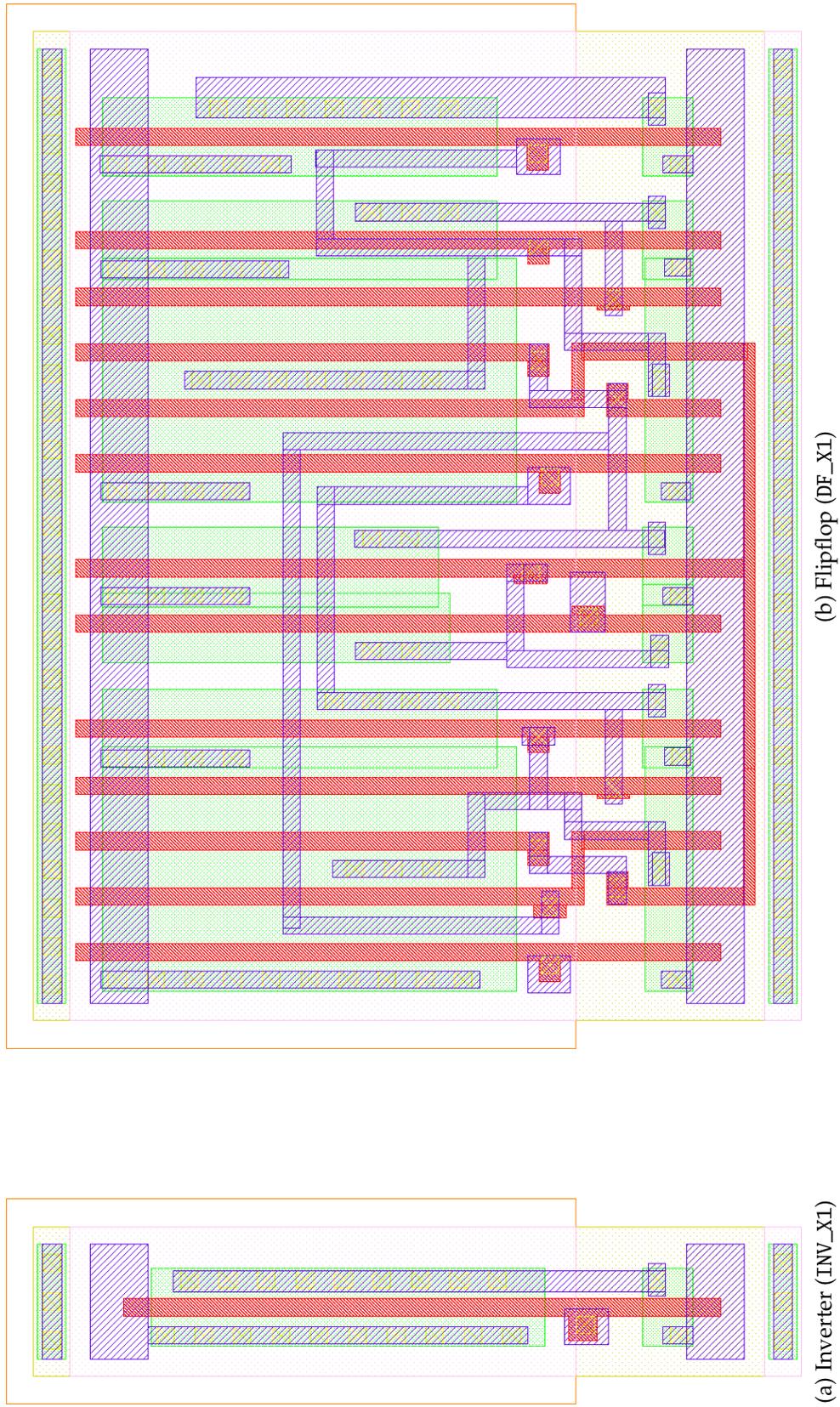
Da die Transistormaße der Subschwelligatter stark von denen konventioneller Standardzellen abweichen, wurde für die Layouts der Gatter ein angepasster Standardzellenrahmen entwickelt, um flächeneffiziente Implementierungen zu ermöglichen. Dieses Layout-Grundgerüst berücksichtigt insbesondere den Bedarf einer großen PMOS-Weite ohne Teilung der Transistoren. Diese kann bis zu $2,3\ \mu\text{m}$ betragen, die maximal mögliche NMOS-Weite beträgt $0,65\ \mu\text{m}$. Außerdem sind die Substrat- und Wannenschlüsse separat von den eigentlichen Versorgungsleitungen kontaktierbar,

so dass die Zellen unter Verwendung von Substrat- und Wannenvorspannungen betrieben werden können. Im Rahmen dieser Arbeit wird diese Möglichkeit allerdings vorerst nicht genutzt. Abbildung 6.19 zeigt beispielhaft die Layouts des Inverters und des Flipflops der jeweils kleinsten Treiberstärke X1.

6.9 Zusammenfassung

In diesem Kapitel wurde der Entwurf einer 62 Gatter umfassenden, mehrzieloptimierten Subschwellsbibliothek in einer 65-nm-Technologie vorgestellt. Die Auswahl des Zellsatzes erfolgte anhand grundsätzlicher Überlegungen zur Robustheit, wobei Simulationen besonders ungünstiger Prozess- und Gatterkonstellationen durchgeführt sowie Erkenntnisse existierender Publikationen herangezogen wurden. Die Standardzellen wurden simultan bezüglich der Störabstände, des Energiebedarfs und der Verzögerungszeiten optimiert. Nach einer numerischen Bestimmung der Pareto-Menge jedes Zellentyps wurde eine konkrete Gatterdimensionierung durch ein Auswahlverfahren ermittelt, wobei eine Gewichtung der drei Zielfunktionen mit absteigender Priorität nach vorstehender Reihenfolge zum Tragen kam. Für die Signalübertragung zu Schaltungsteilen, die bei maximaler Versorgungsspannung arbeiten, wurde eine neuartige Pegelwandlerzelle entwickelt, die besonders für Anwendungen mit einer dynamischen Spannungsskalierung geeignet ist und in diesem Bereich gegenüber bisherigen Implementierungen eine deutlich geringere statische Verlustleistung aufweist.

Die hier entwickelte Standardzellen-Bibliothek dient im folgenden Teil dieser Arbeit als Grundlage für die Implementierung eines adaptiven Subschwellsprozessors. Der Prozessorkern basiert auf einer leistungsfähigen 32-bit-Architektur und stellt damit einen der komplexesten Schaltungsentwürfen für den Subschwellsbetrieb dar. Eine weitere Besonderheit ist die Anpassungsfähigkeit des Prozessors an Schwankungen der Prozess- und Betriebsbedingungen, wodurch dynamisch zur Laufzeit eine Verbesserung der Robustheit erzielt wird.



(a) Inverter (INV_X1)

(b) Flipflop (DF_X1)

Abbildung 6.19: Layouts des Inverters und des Flipflops der 65-nm-Subschwelligbibliothek

Teil III

Adaptiver Subschwelligprozessor

7 Prototypische ASIC-Realisierung

Einige Inhalte dieses Teils der Arbeit wurden in [180, 181] veröffentlicht.

Die Kernkomponente eingebetteter Systeme ist zumeist ein Prozessor, der in Verbindung mit geeignetem Programmcode die Steuerungs- und Berechnungsaufgaben einer gegebenen Anwendung erfüllt. Für Anwendungen mit strikt begrenztem Energiebudget bietet sich eine Implementierung unter Verwendung von Subschwellechnik und einer Prozessorarchitektur mit hoher Ressourceneffizienz an. Eine Abschätzung der charakteristischen Eigenschaften einer solchen integrierten Schaltung wie maximale Taktfrequenz oder Verlustleistung bzw. Energieaufnahme ist durch Software-Werkzeuge mit hoher Genauigkeit möglich. Je nach Schaltungsart kann dies durch analoge Schaltungssimulationen auf Transistorebene oder durch Logiksimulationen auf Gatterebene mit Annotation des Zeitverhaltens und Aufnahme von Schaltaktivitäten erfolgen. Dennoch können Abweichungen der simulativ vorhergesagten Eigenschaften gegenüber der Realität auftreten. Dies gilt insbesondere für den Subschwelleitbetrieb integrierter Schaltungen, da nicht bekannt ist, mit welcher Genauigkeit die Simulationsmodelle des Technologieanbieters das Transistorverhalten und die Streuung desselben im Subschwelleitbereich nachbilden. Deshalb ist es gerade in diesem Fall sinnvoll, einen ASIC-Prototypen eines Schaltungsentwurfs zu fertigen und zu charakterisieren.

In diesem Kapitel wird die prototypische ASIC-Realisierung eines Prozessors in 65 nm Subschwelleittechnik als *System on a Chip* (SoC) unter dem Namen *CoreVA Ultra-Low Power* (CoreVA ULP) beschrieben. Dazu wird zunächst in Abschnitt 7.1 die dazu eingesetzte Prozessorarchitektur *CoreVA* vorgestellt. Abschnitt 7.2 erläutert die zur Inbetriebnahme des ASICs eingesetzte Rapid-Prototyping-Umgebung *RAPTOR* in Verbindung mit dem dafür neu entwickelten Erweiterungsmodul *DB-CoreVA*. Anschließend folgt in Abschnitt 7.3 eine Beschreibung der für den Subschwelleitprozessor angepassten Architekturkonfiguration sowie weiterer Komponenten des ASICs. Abschnitt 7.4 geht auf die Besonderheiten des zur ASIC-Implementierung verwendeten Entwurfsablaufs ein.

7.1 CoreVA-Architektur

Der im Rahmen dieser Arbeit realisierte Subschwellprozessor basiert auf der CoreVA¹-Prozessorarchitektur [66]. Das gesamte CoreVA-System liegt als RTL-Beschreibung in VHDL vor und kann somit für Implementierungen in anwendungsspezifischen integrierten Schaltungen (ASICs) und Feld-programmierbaren Gatter-Anordnungen (FPGAs) eingesetzt werden.

Ein Blockschaltbild des Prozessors ist in Abbildung 7.1 dargestellt. Es handelt es sich hierbei um eine modulare 32-bit-Architektur mit sechsstufiger Fließbandverarbeitung (engl. *Pipeline*) nach dem VLIW²-Konzept und getrenntem Daten- und Instruktionsspeicher (Harvard-Architektur). Der Prozessorkern enthält eine Registerbank (engl. *Register File*) mit 31 frei verfügbaren 32-bit-Registern sowie zwei 8-bit-Kontrollregister zur bedingten Ausführung von Instruktionen (*Condition-Register*). Aufgrund des VLIW-Konzeptes sind mehrere parallele Verarbeitungseinheiten (*Slots*) verfügbar, wobei deren Anzahl zur Entwurfszeit konfigurierbar ist. Ebenso können Multiplikations-Akkumulier-Einheiten, Divisionsschritt-Einheiten und Speicherzugriffs-Einheiten bezüglich ihrer Anzahl sowie ihrer Zuordnung zu den vorhandenen Slots konfiguriert werden. Der CoreVA-Prozessor verwendet eine Instruktionsspeicherung mittels *Stop-Bit*, so dass Leerinstruktionen in den hinteren Slots einer Instruktionsgruppe im Programmcode entfallen können. Dies ermöglicht eine Reduktion von Codegröße und Verlustleistung.

In der *Instruction-Fetch-Pipeline*-Stufe (FE) wird eine Instruktionsgruppe aus dem Instruktionsspeicher gelesen und unter Berücksichtigung der Kompression in einem *Alignment-Register* zwischengespeichert. Die *Instruction-Decode-Pipeline*-Stufe (DC) dekomprimiert die Instruktionsgruppe und dekodiert anschließend die einzelnen Teilinstruktionen. Außerdem enthält die DC-Stufe eine statische Sprungvorhersage, wobei spekulativ davon ausgegangen wird, dass Rücksprünge ausgeführt (*predict taken*) und Vorwärtssprünge nicht ausgeführt (*predict not taken*) werden. Ein Rückabwicklungsmechanismus sorgt für eine Korrektur des Pipeline-Zustands bei einem fehlerhaft vorhergesagten Sprung. Zur effizienten Pipeline-Nutzung bei Sprüngen wird zudem ein *Delay Slot* verwendet, d.h. eine auf einen Sprung folgende Instruktionsgruppe wird in jedem Fall ausgeführt. Die Bereitstellung von Registeroperanden erfolgt in der *Register-Read-Pipeline*-Stufe (RD), wobei alle Verarbeitungseinheiten Zugriff auf beliebige Registerinhalte haben. Die *Execute-*

¹ *Configurable Resource Efficient VLIW Architecture*

² *Very Long Instruction Word*

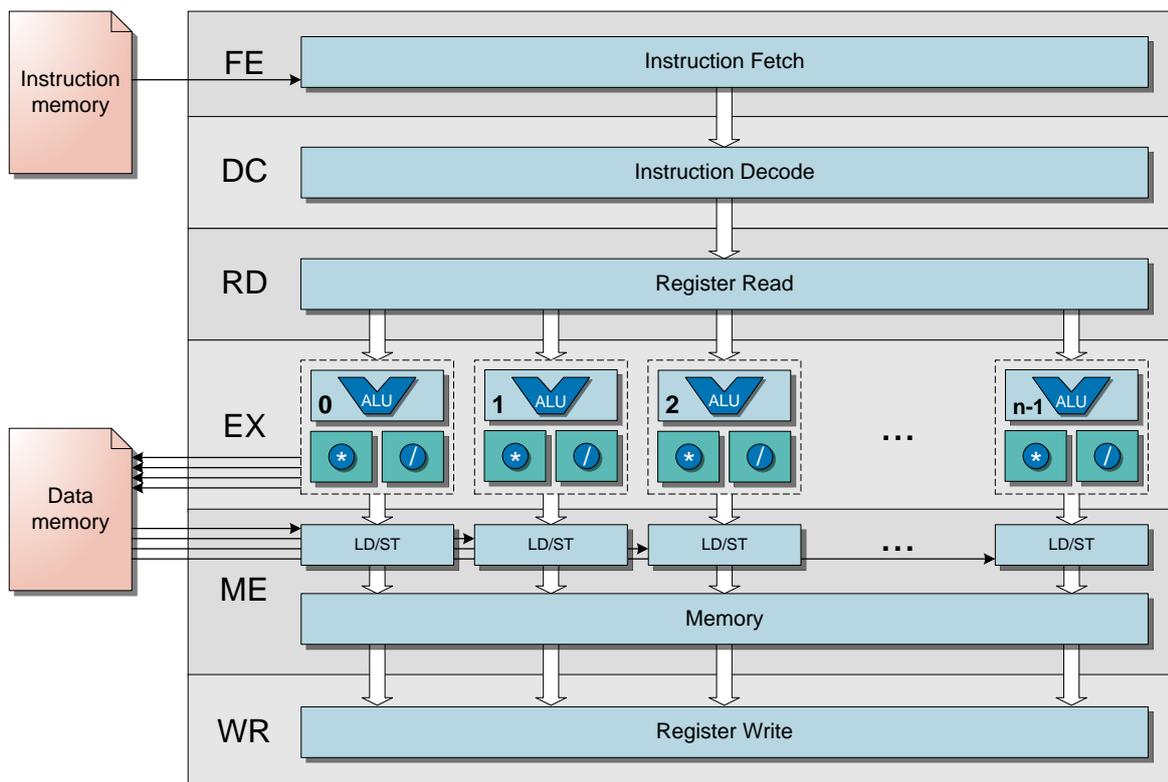


Abbildung 7.1: Generische Pipeline-Struktur des CoreVA-Prozessors [66]

Pipelinstufe (EX) enthält je Slot eine ALU, zusätzlich kann der Prozessor ein oder mehrere Multiplikations-Akkumulier- und Divisionsschritt-Einheiten enthalten. Für ALU- und Multiplizier-Instruktionen steht ein SIMD¹-Modus zur Verfügung, mit dem jede Verarbeitungseinheit zwei gleichartige 16-bit-Operationen gleichzeitig ausführen kann. Die EX-Stufe kann außerdem ein oder mehrere LD/ST²-Einheiten für die Verarbeitung von Datenspeicherzugriffen enthalten. Diese Einheiten reichen die Speicheradresse sowie bei Schreibzugriffen das zu schreibende Datum an den Datenspeicher weiter. Lesezugriffe auf den Datenspeicher werden in der *Memory-Access-Pipeline*-Stufe (ME) verarbeitet. Die Multiplikations-Akkumulier-Einheit erstreckt sich über EX- und ME-Stufe. Schließlich werden die Ergebnisse der ausgeführten Instruktionen in der *Register-Write*-Stufe (WR) in das Register-File zurück geschrieben. Aus der Pipeline-Struktur ergibt sich für die meisten Instruktionen eine Latenz von einem Takt, lediglich Multiplikations- und Speicherlese-Instruktionen weisen eine Latenz von zwei Takten auf. Instruktionssatzerweiterungen (engl. *In-*

¹ *Single Instruction, Multiple Data*

² *Load/Store*

struction Set Extension, ISE) bieten die Möglichkeit zur anwendungsspezifischen Optimierung des CoreVA-Prozessors.

Zur Vermeidung von Leerinstruktionen bzw. einem Anhalten der Prozessorpipeline (*Interlocking*) bei Datenabhängigkeiten zwischen aufeinander folgenden Instruktionen enthält der Prozessor einen vollständigen Pipeline-Bypass, d.h. die Register-Read-Pipeline-Stufe kann auf die Rechenergebnisse aller Verarbeitungseinheiten aus späteren Pipeline-Stufen vorgreifen. Jeder einzelne Bypass-Pfad kann sowohl statisch zur Entwurfszeit als auch dynamisch zur Laufzeit aktiviert bzw. deaktiviert werden. Durch das Deaktivieren eines Bypass-Pfades lässt sich der kritische Pfad des Prozessors verkürzen und somit die Taktfrequenz steigern. Erfolgt das Deaktivieren statisch, so können außerdem die zu den Logikpfaden gehörigen Gatter und Verdrahtungsleitungen entfallen, wodurch sich die Verlustleistung und die Fläche des Prozessors reduzieren lässt. Allerdings erhöht sich die Ausführungslatenz durch Strafzyklen, wenn ein Datenkonflikt aufgrund eines deaktivierten Bypass-Pfades auftritt.

Die Instruktions- und Datenspeicher können wahlweise als gewöhnliche Speicher (*On-Chip-Speichermodus*) oder als *Caches* (*Cachemodus*) konfiguriert werden [68], wobei die Wahl des Speichermodus vor der Initialisierung des Prozessors erfolgt. Es werden Datentransfers mit Breiten von 8, 16 und 32 bit unterstützt, wobei eine *Little-Endian-Byte-Reihenfolge* verwendet wird.

Abbildung 7.2 zeigt ein Blockschaltbild der CoreVA-Architektur auf Systemebene. Im Hinblick auf prototypische ASIC-Realisierungen des CoreVA-Prozessors ist das System in einen ASIC-Teil und einen FPGA-Teil unterteilt. Der Prozessorkern sowie Instruktions- und Datencache bilden die Hauptbestandteile des ASIC-Teils. Zusätzlich können Hardware-Erweiterungen mittels *Memory Mapped Input/Output* (MMIO) über die Datenspeicher-Schnittstelle des Prozessors angesprochen werden, hierfür stehen Peripherie-Module wie z.B. ein UART (*Universal Asynchronous Receiver Transmitter*) oder eine Ethernet-Schnittstelle sowie Hardwarebeschleuniger etwa für CRC (*Cyclic Redundancy Check*), AES (*Advanced Encryption Standard*) oder ECC (*Elliptic Curve Cryptography*) zur Verfügung. ASIC-Teil und FPGA-Teil des CoreVA-Systems sind durch einen Systembus miteinander verbunden, wobei ein Arbitrer den Buszugriff den verschiedenen Teilnehmern zuteilt. Der FPGA-Teil ist vergleichbar mit dem Chipsatz eines PC-Systems. Über einen Speichercontroller lässt sich SDRAM (*Synchronous Dynamic Random Access Memory*) oder DDR2-SDRAM (*Double Data Rate SDRAM*) als Hauptspeicher des Systems nutzen. Ferner können externe Hardware-Erweiterungen oder -Beschleuniger integriert werden. Das

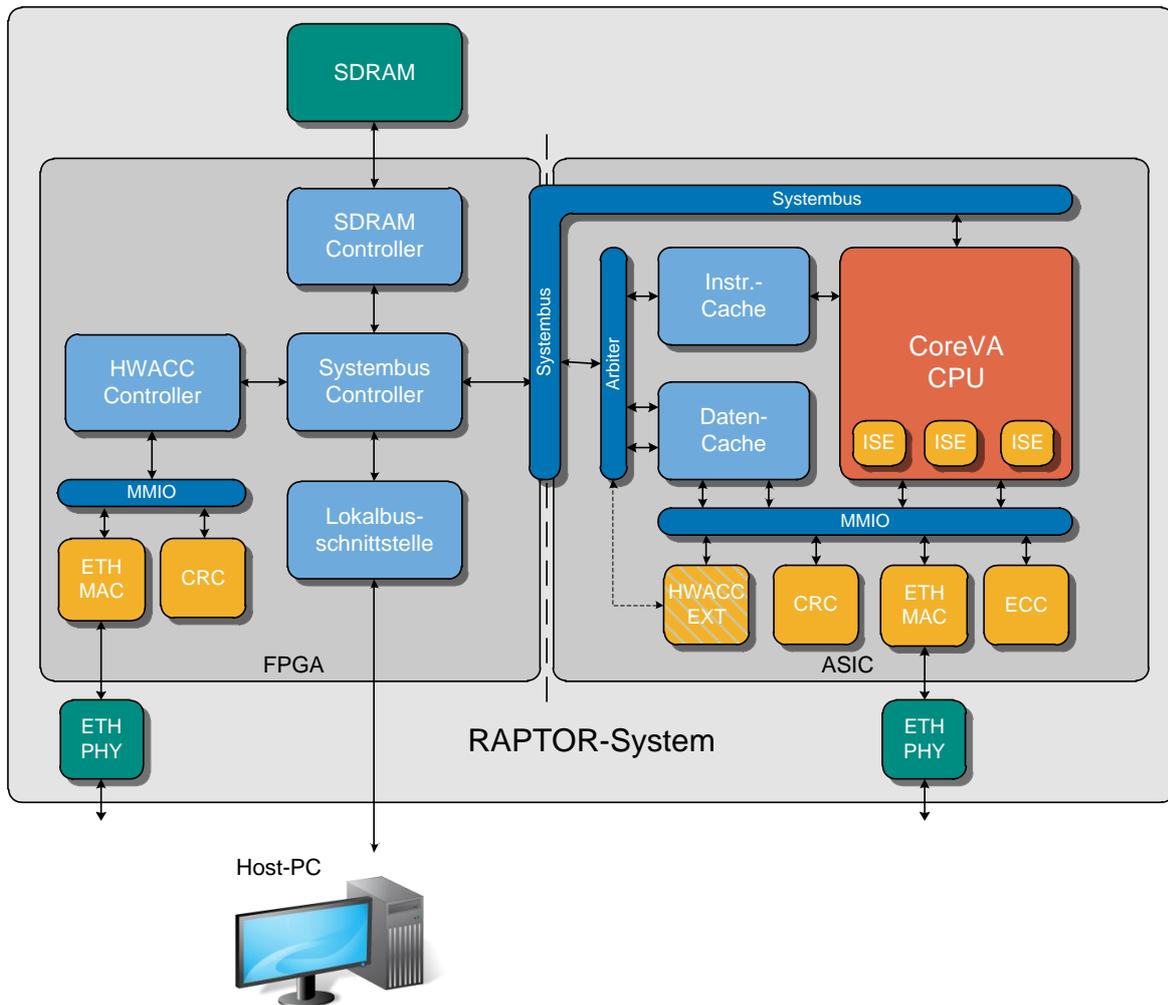


Abbildung 7.2: Architektur des CoreVA-Systems [66]

CoreVA-System kann über eine Lokalbus-Schnittstelle an übergeordnete Systemumgebungen angebunden werden, insbesondere an das im folgenden Abschnitt beschriebene RAPTOR-System.

7.2 Rapid-Prototyping-System RAPTOR

Das in der Fachgruppe Schaltungstechnik der Universität Paderborn entwickelte, modulare, FPGA-basierte *Rapid-Prototyping*-System RAPTOR [70, 118, 119] ermöglicht durch eine Vielzahl verfügbarer Erweiterungsmodule einen schnellen, flexiblen

Aufbau komplexer Systementwürfe, etwa zur funktionalen Prüfung, als prototypische Testumgebung oder zu Demonstratorzwecken. Mit Hilfe der FPGA-Module lassen sich beispielsweise Systeme mit bis zu 200 Millionen Transistoren bereits in einem frühen Entwurfsstadium emulieren. Gegenüber einer RTL-Simulation liegt die Emulationsgeschwindigkeit dabei um Größenordnungen höher. Schnittstellen-Module erweitern das System um Kommunikationsfähigkeiten, wodurch die Einbettung in reale Betriebsszenarien möglich ist. Durch das modulare Konzept können FPGA-Entwürfe auf einfache Weise durch prototypische ASIC-Realisierungen ausgetauscht werden. Die Steuerung und Analyse der Prototypsysteme erfolgt durch die Anbindung an einen *Host*-Rechner. Je nach verwendetem Basissystem ist die Host-Anbindung durch PCI¹, PCI-X oder PCI Express realisiert. Zur Kommunikation der Erweiterungsmodule untereinander sowie mit dem RAPTOR-Basissystem stehen ein Lokalbus, ein *Broadcast*-Bus, Direktverbindungen zwischen benachbarten Modulen sowie je nach verwendeten FPGA-Typen serielle *High-Speed*-Verbindungen zur Verfügung. In Abbildung 7.3 ist beispielhaft das PCI-X-basierte Basissystem RAPTOR-X64 dargestellt.

7.2.1 ASIC-Evaluationsmodul DB-CoreVA

Als Evaluationsplattform für CoreVA-ASICs wurde das in Abbildung 7.4 dargestellte Erweiterungsmodul DB-CoreVA für das RAPTOR-System entwickelt. Kernkomponente des Moduls ist ein FPGA des Typs Xilinx Spartan-6 LX XC6SLX100 oder XC6SLX150 [165]. Dieser ist einerseits über den Lokalbus mit dem RAPTOR-Basissystem und andererseits über den CoreVA-Systembus mit dem ASIC-Prototypen verbunden. Außerdem greift das FPGA auf einen DDR2-SDRAM-Baustein mit einer Kapazität von bis zu 512 MB bei einem Bustakt von maximal 400 MHz und einer Datenbusbreite von 16 bit zu. Dieser steht als Hauptspeicher für das CoreVA-System zur Verfügung. Zur Fehlersuche und als einfache Ein- und Ausgabemöglichkeiten besitzt das DB-CoreVA sechs frei verwendbare GPIO²s sowie sieben LED³s. Weiterhin sind 76 Signalleitungen über RAPTOR-konforme *Debug*-Steckverbinder verfügbar. Außerdem sind jeweils im ASIC und im FPGA implementierte UARTs über einen UART-USB⁴-Adapter verfügbar, um Textein- und -ausgaben in den Anwendungsszenarien realisieren zu können. Als Basistaktquellen für das FPGA sind

¹ *Peripheral Component Interconnect*

² *General Purpose I/O*

³ *Light-Emitting Diode*

⁴ *Universal Serial Bus*

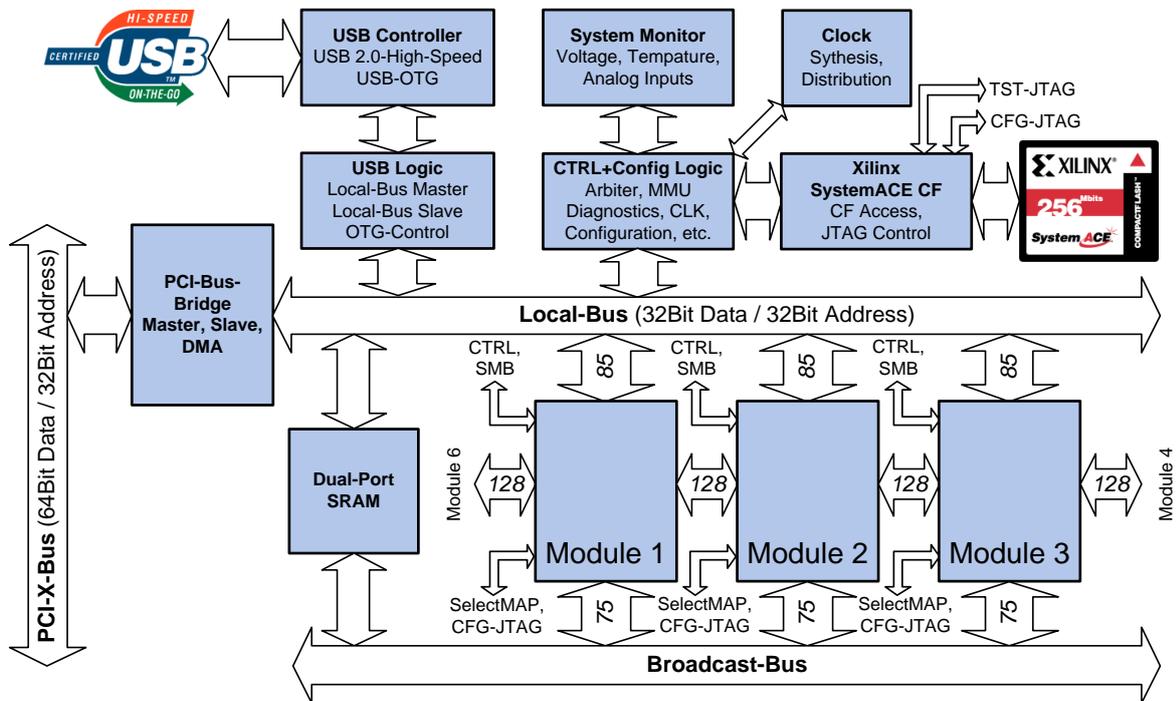


Abbildung 7.3: Das RAPTOR-X64 Basissystem

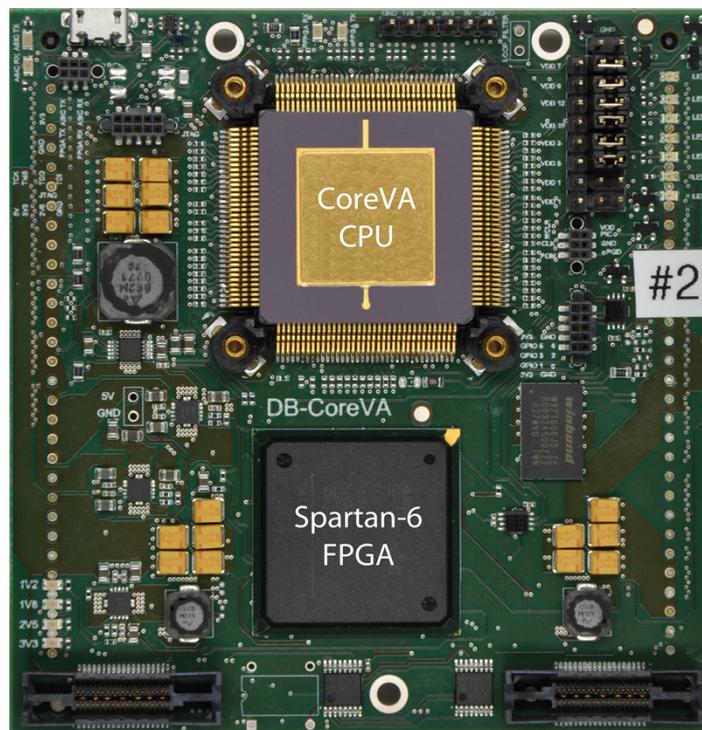
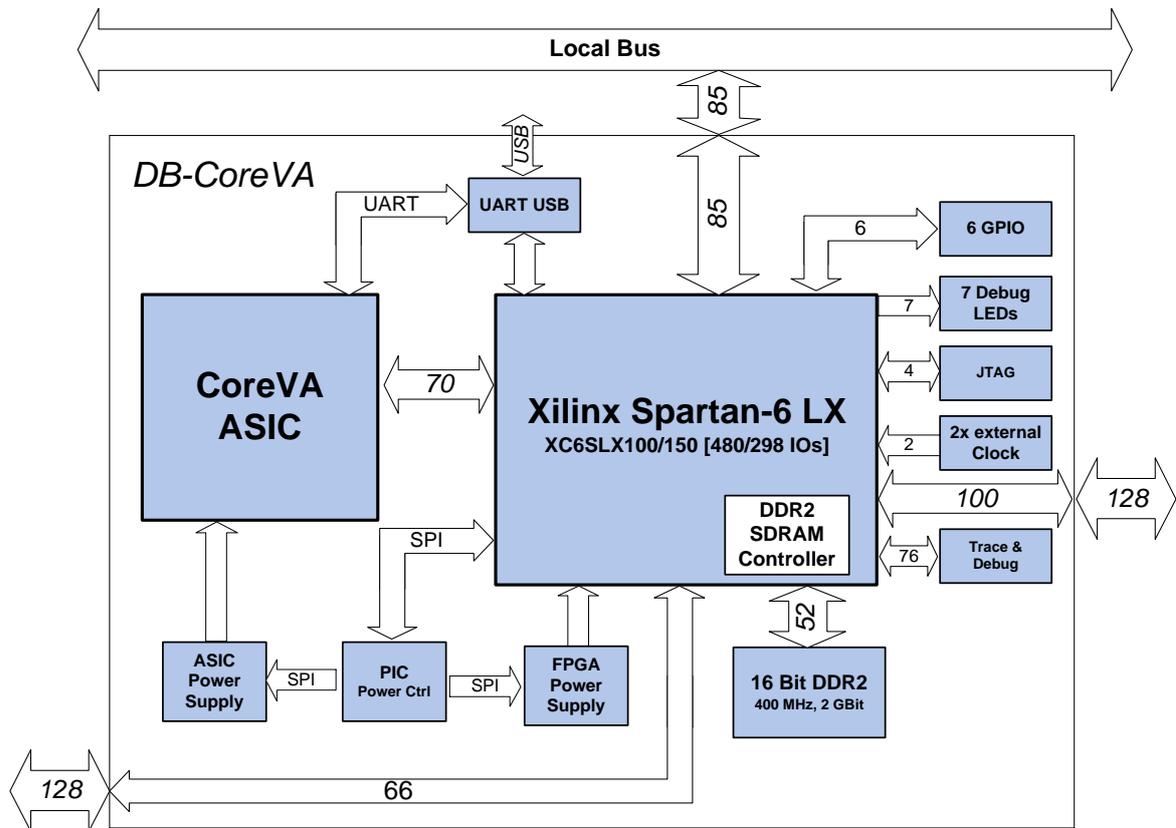


Abbildung 7.4: Das Erweiterungsmodul DB-CoreVA

ein Quarzoszillator mit einer Frequenz von 100 MHz sowie ein Sockel zur Aufnahme eines weiteren Quarzoszillators vorhanden. Auch der Lokalbus-Takt kann als Taktquelle genutzt werden. Mit Hilfe der Taktressourcen des FPGAs können beliebige weitere Taktsignale für das CoreVA-System und insbesondere den CoreVA-ASIC generiert werden. Die Konfiguration des FPGAs kann durch das RAPTOR-System oder über einen JTAG¹-Steckverbinder erfolgen.

Die Spannungsversorgung der verschiedenen Komponenten erfolgt durch mehrere Spannungsregler, die von einem Mikrocontroller konfiguriert und überwacht werden. Es werden feste Spannungen von 1,2 V, 1,8 V, 2,5 V und 3,3 V erzeugt. Außerdem werden zwei variable I/O²-Spannungen für die Direktverbindungen zu den benachbarten RAPTOR-Modulen generiert. Diese sind über den Mikrocontroller durch das RAPTOR-System konfigurierbar, so dass unterschiedliche Übertragungsverfahren für die Direktverbindungen genutzt werden können und eine Kompatibilität mit den meisten verfügbaren RAPTOR-Erweiterungsmodulen besteht. Ein weiterer einstellbarer Spannungsregler erzeugt eine variable Versorgungsspannung für den CoreVA-ASIC zur Realisierung einer DVFS-Unterstützung. In Verbindung mit dem in dieser Arbeit entwickelten Subschwelligprozessor ermöglicht dieser Spannungsregler zudem eine adaptive Spannungserzeugung zur Kompensation von PVT-Schwankungen. Dies wird detailliert in Abschnitt 10.2 behandelt.

Das DB-CoreVA kann CoreVA-ASICs im CQFP³-144-Gehäuse aufnehmen. Diese können wahlweise direkt auf die Platine aufgelötet oder in einen entsprechenden Sockel bestückt werden. Die Anschlussbelegung wurde so ausgelegt, dass eine gewisse Flexibilität bezüglich des Funktionsumfangs verschiedener, auch zukünftiger ASIC-Varianten besteht. Von den 144 Anschlüssen ist auf jeder ASIC-Seite je einer nicht verbunden, um den Padring um 50 µm in horizontaler und vertikaler Richtung verkleinern und damit Fläche und Fertigungskosten einsparen zu können. Zwölf mal je zwei benachbarte Pins, also insgesamt 24 Pins, sind fest für die Versorgung der Kernlogik vorgesehen. Davon sind sechs Pinpaare fest mit der 1,2-V-Schiene verbunden. Die übrigen sechs Pinpaare können über Kurzschlussbrücken (engl. *Jumper*) wahlweise ebenfalls mit der 1,2-V-Schiene, mit der variablen Versorgungsspannung des DB-CoreVA oder mit einer externen Spannungsquelle verbunden werden. Dieses Versorgungskonzept ermöglicht die unabhängige Versorgung oder messtechnische Charakterisierung verschiedener Module des ASICs. Die I/O-Versorgungsspannung von 1,8 V wird über 14 Pinpaare, d.h. 28 Pins, zugeführt. Von den übrigen Anschlüssen

¹ *Joint Test Action Group*

² *Input/Output*

³ *Ceramic Quad Flatpack*

sen besitzen 80 eine vorgegebene Funktion. Hierzu zählen der CoreVA-Systembus, das Takt- und das Rücksetzsignal sowie das Sende- und das Empfangssignal des UARTs. Drei weitere Anschlüsse sind aus technologischen Gründen belegt. Die verbleibenden fünf Signale sind zu verschiedenen Zwecken nutzbar, etwa als zusätzliche Signale zwischen ASIC und FPGA oder im Rahmen der bereits erwähnten adaptiven Spannungserzeugung.

Die DB-CoreVA-Platine besitzt einen 14-lagigen Aufbau mit acht Signallagen, zwei geteilten Versorgungslagen und vier Masselagen. Es werden ausschließlich durchgehende Vias (Durchkontaktierungen) verwendet. Die minimale Strukturgröße beträgt 100 μm . Die Leitungsführung der Kommunikationsbusse mit Impedanzkontrolle, Längenausgleich und teilweise differenziellen Signalpaaren ermöglicht die Nutzung von Kommunikationsstandards mit hohen Datenübertragungsraten.

7.3 Konfiguration und Komponenten des ASICs

Das hohe Maß an Modularität und Konfigurierbarkeit des CoreVA-Systems ermöglicht die optimale Anpassung an ein gegebenes Anwendungsszenario. Hauptsächlich Einsatzzweck von CoreVA ULP sind Anwendungen, bei denen nur eine begrenzte, geringe Menge elektrischer Energie zur Verfügung steht, während die Performanzanforderungen zweitrangig und zumindest zeitweise moderat sind. Da die Leistungsaufnahme des CoreVA-Prozessors linear mit der Anzahl an Verarbeitungseinheiten steigt [66], erscheint es sinnvoll eine Prozessorvariante mit einer Verarbeitungseinheit für diesen Einsatzzweck zu wählen. Auf diese Weise reduziert sich die VLIW-Architektur zu einer gewöhnlichen RISC-Architektur mit Multiplikations-Akkumulier- und Divisionsschritt-Einheit.

Wie in Abschnitt 7.1 beschrieben, kann die Taktfrequenz des CoreVA-Prozessors durch Deaktivieren bestimmter Bypass-Pfade gesteigert werden, wobei jedoch zusätzliche Strafzyklen auftreten können, wenn Datenkonflikte aufgrund eines deaktivierten Pfades bestehen. In [66, 67] wurde gezeigt, dass das Deaktivieren bestimmter Bypass-Pfade je nach Anwendung dennoch zu einer Verringerung der Laufzeit und/oder des für die Ausführung benötigten Energiebedarfs führt. Es wurde festgestellt, dass für fast alle Anwendungen das Deaktivieren der Kontroll-, MCR- und Condition-Bypässe aus der EX-Stufe sinnvoll ist. Aus diesem Grund bietet der Subschwellschwellprozessor des CoreVA-ULP-ASICs die Möglichkeit, diese Bypass-Pfade dynamisch zu deaktivieren.

Eine Folge der Entwurfsentscheidung, eine RISC-Prozessorvariante für CoreVA ULP zu verwenden, ist ein relativ geringer Flächenbedarf des Prozessorkerns. Im Rahmen der prototypischen ASIC-Realisierung wird jedoch die minimale Chipfläche durch das festgelegte Anschlussschema bzw. durch die Anzahl der Pad-Zellen bestimmt, so dass die verfügbare Fläche für einen einzelnen Prozessorkern deutlich überdimensioniert wäre. Da gerade für den Subschwellschwellprozessor eine Untersuchung von Exemplarstreuungen interessant ist, wurde das CoreVA-System um die Möglichkeit, mehrere Prozessorkerne auf einem ASIC unterbringen zu können, erweitert. Somit stehen bei gleicher Anzahl gefertigter ASICs mehr testbare Prozessorexemplare zur Verfügung und die Chipfläche kann besser ausgenutzt werden. Da alle übrigen Systemkomponenten jedoch weiterhin nur einfach vorhanden sind, kann jeweils nur ein Prozessorkern gleichzeitig aktiv sein. Über ein extern zu konfigurierendes Steuerregister wird der jeweils aktive Prozessorkern durch Multiplexer ausgewählt und alle übrigen Prozessorkerne zurückgesetzt und angehalten.

Abbildung 7.5 zeigt ein Blockschaltbild des CoreVA-ULP-ASIC-Prototypen einschließlich dessen Einbettung in die Systemumgebung des DB-CoreVA. Hauptbestandteil sind zwei Prozessorkerne auf Basis der in Kapitel 6 vorgestellten Subschwellschwellbibliothek sowie zu Vergleichszwecken eine zusätzliche, auf konventionellen Standardzellen basierende Instanz. Im Interesse einer besseren Vergleichbarkeit wurden für diese Instanz analog zur Subschwellschwellbibliothek ebenfalls nur Standardzellen mit LPSVT-Transistoren verwendet. Da der primäre Zweck dieses ASIC-Prototypen der Funktionsnachweis und die Charakterisierung des CoreVA-Prozessorkerns auf Basis der Subschwellschwellbibliothek ist, sind die Schaltungsteile außerhalb der Prozessorkerne ebenfalls mit konventionellen Standardzellen aufgebaut und werden mit einer Spannung von 1,2 V betrieben. Hierdurch wird ein maximaler Grad an Testbarkeit garantiert, selbst wenn einige Subschwellschaltungsteile fehlerhaft arbeiten sollten. Alle drei Prozessorinstanzen verfügen über individuelle Versorgungszuleitungen, um getrennte Messungen der Stromaufnahme bzw. des Energiebedarfs zu ermöglichen.

Das Speichersubsystem besteht aus Daten- und Instruktioncaches von jeweils 16 kB. Für die Caches kommen aus den geschilderten Gründen konventionelle Speicherblöcke des Technologieanbieters zum Einsatz. Zur Untersuchung von im Subschwellschwellbereich betriebenen Speicher enthält CoreVA ULP unabhängig von den Caches mehrere Subschwellschwell-SRAM-Blöcke mit einer Gesamtkapazität von ca. 0,6 kB, die als Hardware-Erweiterung an das CoreVA-System angebunden sind. Außer zu Testzwecken können diese Speicherblöcke auch als *Scratchpad*-Speicher genutzt werden. Der Subschwellschwell-SRAM wird in Kapitel 9 thematisiert.

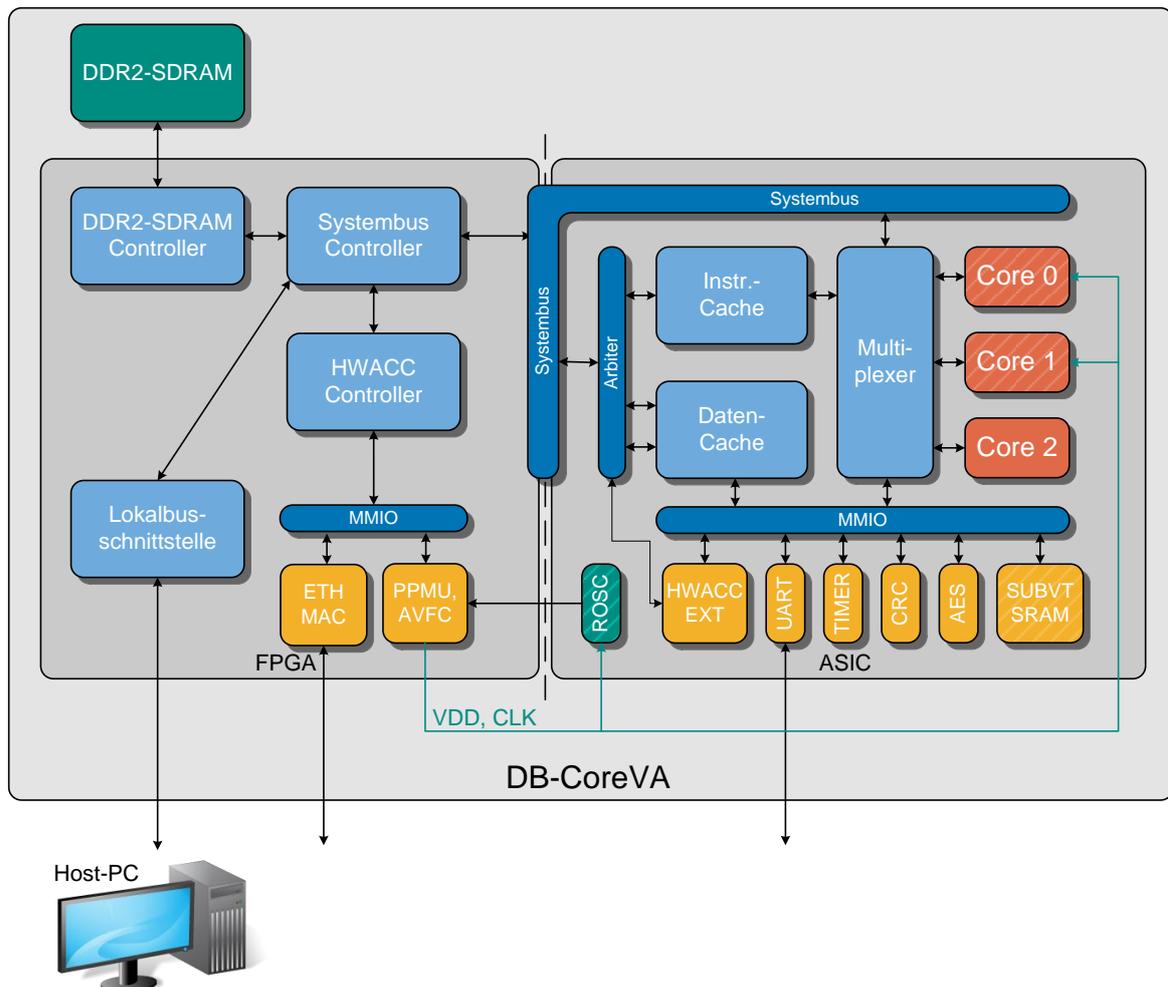


Abbildung 7.5: Blockschaltbild des Gesamtsystems von CoreVA ULP (Subschwellemodule schraffiert)

Des Weiteren enthält der ASIC-Prototyp einen UART, einen *Timer* sowie CRC- und AES-Hardwarebeschleuniger. Zusätzliche Hardwareerweiterungen wie etwa eine Ethernet-Schnittstelle können extern im FPGA-Teil des Gesamtsystems implementiert werden. Insbesondere ist auf diese Weise ein Takt- und Energieverwaltungsmodul (*Performance and Power Management Unit*, PPMU) angebunden. Es ist eng mit einem Modul zur adaptiven Spannungs- und Frequenzsteuerung (*Adaptive Voltage and Frequency Control*, AVFC) verknüpft, zu dem auch ein auf dem ASIC befindlicher Ringoszillator gehört. Diese Einheiten werden in Kapitel 10 behandelt.

7.4 Entwurfsablauf

Die Synthese sowie das Platzieren und Verdrahten (engl. *Place and Route*, PaR) des CoreVA-ULP-ASICs erfordert, ähnlich wie in Abschnitt 5.5.2 für den ALU-Testbaustein beschrieben, die Beachtung einiger Besonderheiten. Zunächst wurde das für den ALU-Prototypen entwickelte, für Subschwellschaltungen geeignete Implementierungskonzept in den in der Fachgruppe Schaltungstechnik der Universität Paderborn vorhandenen automatisierten Hardware-Entwurfsablauf integriert und weiterentwickelt [179]. Die resultierende Werkzeugkette basiert auf den Softwarewerkzeugen Cadence RTL Compiler und Cadence SoC Encounter. Sie bietet Unterstützung für MSV, DVFS und mehrere simultan analysierte Betriebsmodi und Betriebsbedingungen (engl. *Multi-Mode, Multi-Corner, MMMC*) als Grundlage für die Implementierung von Schaltungen mit variabler Versorgungsspannung vom tiefen Subschwellsbereich bis zur nominellen Versorgungsspannung der Technologie. Dabei werden auch hier Pegelwandler durch das Synthese-Werkzeug automatisch an Signalübergängen zwischen Spannungsdomänen in die Netzliste eingefügt.

Bei der Implementierung von CoreVA ULP kam ein hierarchischer Entwurfsablauf basierend auf der automatisierten Werkzeugkette zum Einsatz. Zunächst wurden vollständige Hard-Macros der Prozessorkerne auf Basis der Subschwellsbibliothek und der konventionellen Standardzellen-Bibliothek implementiert. Anschließend wurden diese Hard-Macros auf Chip-Ebene in die CoreVA-Systemumgebung eingebettet. Dieser hierarchische Ansatz bietet gegenüber einem einzigen Entwurfsablauf zur Implementierung des Gesamtsystems einen wesentlichen Geschwindigkeitsvorteil. Zudem kann das Hard-Macro des Subschwellsprozessors unmittelbar für neue ASIC-Generationen in der gleichen Technologie wiederverwendet werden. Eine Herausforderung des hierarchischen Entwurfs ist jedoch, dass zu Beginn bei der Implementierung des Prozessor-Hard-Macros die zeitlichen Anforderungen an die Signale der Schnittstelle zum Gesamtsystem noch nicht bekannt sind. In Abbildung 7.6 ist beispielhaft ein Logikpfad dargestellt, der von einem Flipflop innerhalb des Hard-Macros zu einem Flipflop außerhalb des Hard-Macros führt. Der Pfad beginnt in einem Pipeline-Register zwischen RD- und EX-Stufe und verläuft innerhalb des Hard-Macros über die ALU, die Auswertung der Sprungbedingungen, den Bypass, die Sprungvorhersage sowie einen Aufwärtspegelwandler und verlässt den Prozessorkern an der Hierarchiegrenze über das Signal der Instruktionsadresse `imem_a`. Außerhalb des Hard-Macros setzt sich der Pfad innerhalb des Instruktionscaches fort und endet dort wiederum in einem Register. In der Abbildung sind die zeitlichen Anteile des Logikpfades bei einer Versorgungsspannung von 1,2 V und

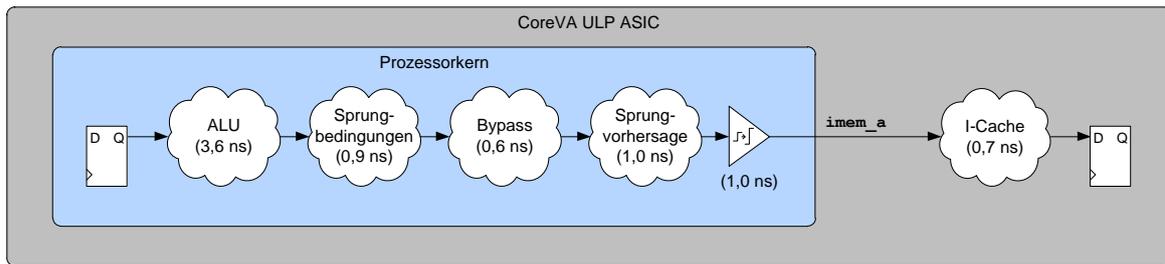


Abbildung 7.6: Verzögerungszeiten eines Logikpfades über Hierarchiegrenzen hinweg (für 1,2 V)

aktiviertem Bypass angegeben. Bei der Implementierung des Hard-Macros ist die Verzögerungszeit des Pfades außerhalb des Hard-Macros, d.h. durch den Instruktionscache, noch unbekannt. Ohne Angabe dieser externen Verzögerungszeit würde jedoch das Synthese-Werkzeug davon ausgehen, dass die gesamte Taktperiode für die Signallaufzeit innerhalb des Hard-Macros ausgenutzt werden kann. Als Folge wäre es bei der anschließenden Implementierung auf Chip-Ebene nicht möglich, die Zeitvorgabe für diesen Logikpfad zu erfüllen. Eine möglichst genaue Abschätzung des Zeitverhaltens des Gesamtsystems ist daher bereits bei der Implementierung der Hard-Macros erforderlich. Ausgangspunkt hierfür waren beim Entwurf von CoreVA ULP Analysen eines probeweise durchgeführten nicht-hierarchischen Entwurfs. Die so gewonnene Abschätzung des Zeitverhaltens wurde anschließend durch mehrere Iterationen des hierarchischen Entwurfsablaufs jeweils mit Anpassung der externen Zeitvorgaben für die Hard-Macro-Implementierungen verfeinert.

Eine weitere Besonderheit bezüglich der zeitlichen Schaltungsanalyse und -optimierung (engl. *Timing*-Analyse bzw. -Optimierung) ergibt sich aus der starken Streuung der Verzögerungszeiten der Standardzellen durch PVT-Schwankungen im Subschwelligbereich, insbesondere durch Prozessschwankungen innerhalb einer integrierten Schaltung (engl. *On-Chip Variation*, OCV). OCV führt also dazu, dass manche Gatterinstanzen eines bestimmten Typs schneller als andere Gatterinstanzen des gleichen Typs sind. Überträgt man dies auf die zur Timing-Analyse herangezogenen Logikpfade, so gibt es also durch OCV frühe/schnelle Pfad-Anteile und späte/langsame Pfad-Anteile (siehe Abbildung 7.7). Für die Analyse der Setup-Bedingungen tritt der ungünstigste Fall dann ein, wenn der eigentliche Signalpfad und der zu dessen Anfangspunkt führende Taktpfad als spät und der aufnehmende Taktpfad als früh angenommen werden. In diesem Fall wirkt OCV als Taktfrequenz-reduzierender Effekt. Die ungünstigste Situation für die Hold-Analyse tritt genau im umgekehrten Fall für frühe und späte Teilpfade ein. Dieser Fall ist besonders problematisch, da Hold-Verletzungen in jedem Fall unabhängig von der angelegten

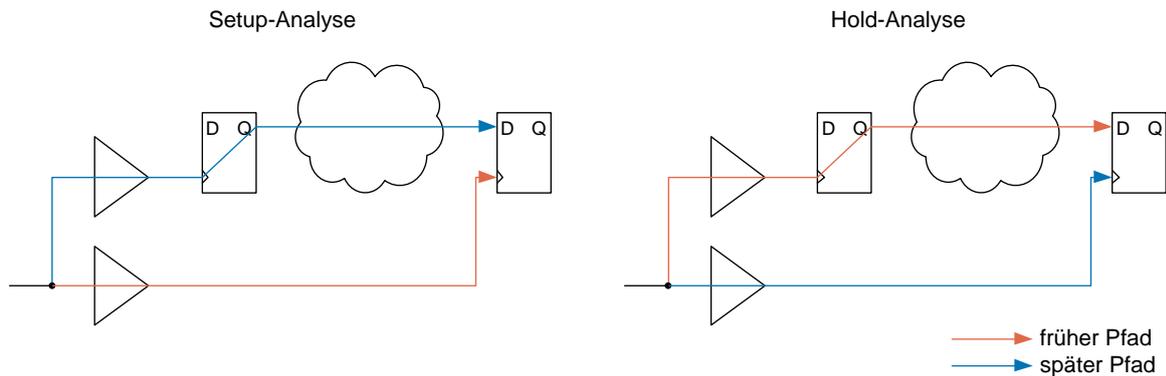


Abbildung 7.7: Frühe und späte Timing-Pfade zur Berücksichtigung von OCV

Taktfrequenz zu einer fehlerhaften Schaltungsfunktion führen, wie bereits in Abschnitt 3.2 erwähnt. Daher ist OCV für Subschwellschaltungen besonders kritisch. In den Softwarewerkzeugen für Synthese und PaR wird OCV durch gewisse Skalierungsfaktoren berücksichtigt¹. Dabei wird der ungünstigste Fall angenommen, dass alle Zellen des frühen Pfades um einen bestimmten Faktor schneller, und dass alle Zellen des späten Pfades um einen bestimmten Faktor langsamer sind, als in der Timing-Bibliothek angegeben. Diese Skalierungsfaktoren müssen also für Subschwellschaltungen größer als üblich gewählt werden, um den stärkeren OCV-Einfluss zu berücksichtigen.

Durch erste Probesynthesen des Subschwellsprozessors wurde die Lage des minimalen Energiebedarfs bei einer Versorgungsspannung von ca. 400 mV abgeschätzt. Um hier eine bestmögliche Schaltungsoptimierung etwa bezüglich des Energiebedarfs zu erreichen, wurde dieser Arbeitspunkt im Rahmen des MMMC-Entwurfs als primär betrachteter Betriebsmodus gewählt. Es wurde eine Zieltaktfrequenz von 1 MHz ohne Berücksichtigung von OCV vorgegeben. Bezüglich des Setup-Zeitverhaltens wird OCV durch die in Abschnitt 10.2 behandelte adaptive Spannungs- und Frequenzsteuerung berücksichtigt. Der gewählte Frequenzwert liegt etwa 10 % unter der maximal erreichbaren Taktrate, wodurch wie bereits in Abschnitt 5.5.2 erläutert ein niedrigerer Flächen- und Energiebedarf erzielt wird. Als zusätzliche Stromspartechniken kommen Clock Gating und *Operand Isolation*² zum Einsatz. Für alle anderen Spannungsschritte erfolgte keine zusätzliche Frequenzvorgabe, sondern lediglich eine Analyse der dort möglichen Taktfrequenz und Energieaufnahme. Potentielle Hold-Verletzungen wurden bei den Versorgungsspannungen 400 mV

¹ Alternativ könnten auch separate Timing-Bibliotheken für frühe und späte OCV-Bedingungen eingebunden werden. Diese sind jedoch für die eingesetzte Technologie nicht verfügbar.

² Methode, um das Propagieren von Signalfanken in inaktive Schaltungsteile zu verhindern

und 1,2 V analysiert und korrigiert. Dabei wurden OCV-Skalierungsfaktoren von $\pm 20\%$ bei 400 mV und $\pm 3\%$ bei 1,2 V gewählt. Mit dieser Konfiguration traten bei keinem analysierten Spannungsschritt Hold-Verletzungen auf. Die minimale Hold-Reserve beträgt bei 400 mV unter Berücksichtigung der OCV-Skalierungsfaktoren gemäß statischer Timing-Analyse +0,148 ns. Um zusätzliche Sicherheit in Bezug auf das Hold-Zeitverhalten zu gewinnen, wurden außerdem analoge Monte-Carlo-Schaltungssimulationen des durch die Timing-Analyse ermittelten kritischen Hold-Pfades unter dem Einfluss lokaler Prozessschwankungen durchgeführt. Die für 10 000 Simulationsläufe ermittelte Fehlerhäufigkeit des kritischen Hold-Pfades von 0,11 % erscheint hinreichend gering. Alle weiteren Hold-Pfade besitzen eine größere Hold-Reserve und damit eine nochmals geringere Fehlerwahrscheinlichkeit.

Ein Chipfoto des ASIC-Prototypen mit Kennzeichnung einzelner Teilbereiche ist in Abbildung 7.8 dargestellt. Der Baustein besitzt eine Fläche von $2,64\text{ mm}^2$, wobei jeder Subschwellsprozessorkern $0,36\text{ mm}^2$ und der konventionelle Prozessorkern $0,18\text{ mm}^2$ beiträgt. Neben der vorgegebenen Anschlussbelegung war ein wesentlicher Einflussfaktor beim *Floorplan*¹-Entwurf die räumliche Trennung der Subschwellsmodule (rot) von anderen Schaltungsteilen (blau), um eine Beeinträchtigung der Integrität schwacher Subschwellsignale durch Einkopplung steiflankiger Störungen zu vermeiden. Durch entsprechende Regeln wurde das Place-and-Route-Werkzeug angewiesen, keine Signale der 1,2-V-Domäne durch oder über die Subschwellschaltungsbereiche zu verlegen. Umgekehrt sind Subschwellsignale ausschließlich auf den Bereich innerhalb der jeweiligen Subschwellsmodule beschränkt. Zudem sind alle Pegelwandler jeweils an den physikalischen Grenzen zwischen den unterschiedlichen Spannungsdomänen angeordnet. Die Separierung der Signale wurde mittels *Crosstalk*-Analyse² und -Optimierung durch das Place-and-Route-Werkzeug sichergestellt.

7.5 Zusammenfassung

Dieses Kapitel dokumentierte die prototypische ASIC-Realisierung des in dieser Arbeit entwickelten adaptiven Subschwellsprozessors CoreVA ULP. Dazu wurde zunächst die zum Einsatz kommende VLIW-Architektur CoreVA beschrieben. Dabei handelt es sich um eine leistungsfähige 32-bit-Architektur mit sechs Pipelinestufen sowie hardwareseitiger Unterstützung für Berechnungen von Multiplikationen

¹ Grundlegende Anordnung der wesentlichen Elemente einer integrierten Schaltung

² Analyse des Übersprechens

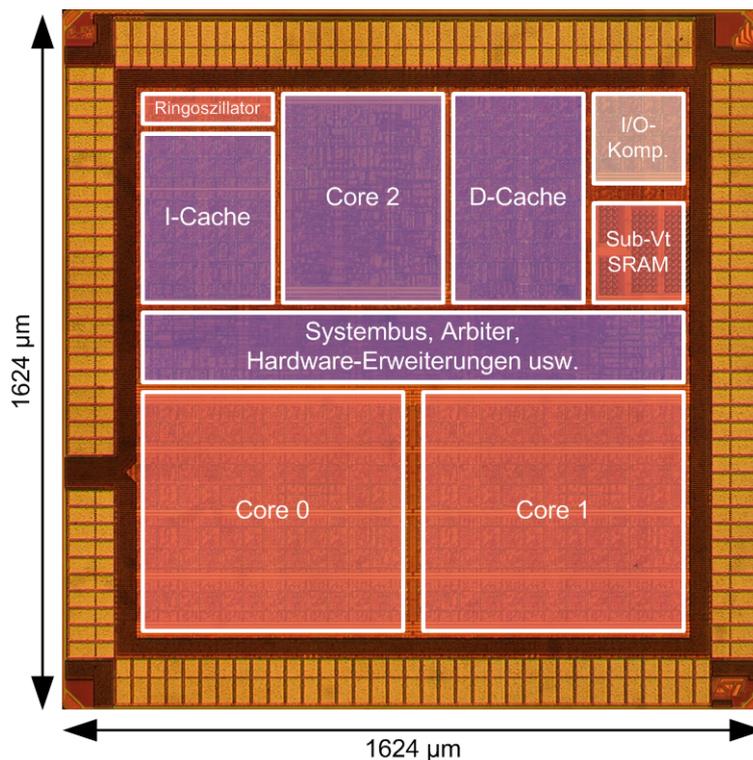


Abbildung 7.8: Chipfoto des CoreVA-ULP-ASICs

und Divisionsschritten. Zusätzlich bietet der Prozessor einen SIMD-Modus zur Verdopplung des Durchsatzes bei 16-bit-Berechnungen. Anschließend wurde das Rapid-Prototyping-System RAPTOR vorgestellt und die Entwicklung des RAPTOR-Erweiterungsmoduls DB-CoreVA beschrieben, das zur prototypischen Inbetriebnahme von CoreVA-ASICs dient. Das DB-CoreVA ist mit einem FPGA des Typs Xilinx Spartan-6 LX XC6SLX100 oder XC6SLX150 und maximal 512 MB DDR2-SDRAM ausgestattet. Anschließend wurde die Konfiguration der CoreVA-Architektur als RISC-Prozessor für die Implementierung von CoreVA ULP sowie weitere Bestandteile des ASICs erläutert. Schließlich folgte eine Beschreibung des Entwurfsablaufs, wobei die sich aus dem Subschwellebetrieb der Schaltung ergebenden Besonderheiten der Timing-Optimierung im Fokus standen.

Im nachfolgenden Kapitel wird die Charakterisierung der gefertigten ASIC-Exemplare geschildert und es werden die so gewonnenen Messergebnisse interpretiert. Außerdem erfolgt der Vergleich mit ähnlichen Implementierungen und mit einer nicht für den Subschwellebetrieb optimierten Prozessorvariante.

8 Eigenschaften des Subschwellsprozessors

Um die Ressourceneffizienz des Subschwellsprozessors zu analysieren, wurden einerseits entsprechende Simulationen und Abschätzungen während der Implementierungsphase des ASICs auf Basis der Informationen aus der charakterisierten Standardzellen-Bibliothek sowie andererseits Messungen anhand aller gefertigter Schaltungsexemplare vorgenommen. In diesem Kapitel werden die eingesetzten Simulations- und Messverfahren erläutert, die damit ermittelten Eigenschaften des Prozessors interpretiert und mit ähnlichen Implementierungen verglichen.

8.1 Simulations- und Messverfahren

Die Angabe der Eigenschaften des Subschwellsprozessors erfordert zunächst eine Diskussion der Verfahren, durch die diese Eigenschaften bestimmt wurden. Da der Schwerpunkt der Evaluierung auf dem Prozessorkern selbst liegt, wurde bei den Simulationen und Messungen der On-Chip-Speichermodus zur Programmausführung gewählt, d.h. es treten keine Strafzyklen durch Fehlzugriffe des Caches auf. Dies ist insbesondere für die energetische Analyse relevant. Dazu darf allerdings der Speicherbedarf der Testprogramme die auf dem ASIC vorhandene Speicherkapazität nicht übersteigen. Dennoch ergibt sich eine umfangreiche Sammlung geeigneter Testprogramme, die eine Vielfalt unterschiedlicher Anwendungsgebiete abdecken. Die eingesetzten Testprogramme sind in Tabelle 8.1 zusammengefasst und sind in [66] näher erläutert. Es wurde ferner untersucht, welchen Einfluss das dynamische Deaktivieren von Bypass-Pfaden auf die Eigenschaften des Prozessors hat. Dabei wurden zwei Bypass-Konfigurationen für jedes Testprogramm analysiert. In der ersten Variante sind alle Bypass-Pfade aktiv, während in der zweiten Variante die Kontroll-, MCR- und Condition-Bypässe aus der EX-Stufe deaktiviert sind. Der

Name	Kategorie	Kurzbeschreibung, Referenz
Coremark	Benchmark	Synthetischer Benchmark [49]
Dhrystone	Benchmark	Synthetischer Benchmark [163]
Division	Mikrobenchmark	Ganzzahlige Berechnungen von Division und Rest
Fibonacci	Mikrobenchmark	Berechnung der Fibonacci-Folge
CRC	Codierung	Zyklische Redundanzprüfung [117]
Viterbi	Codierung	Dekodierung von Faltungscodes [156]
AES	Kryptographie	Symmetrisches Verschlüsselungsverfahren [41]
ECC	Kryptographie	Asymmetrisches Verschlüsselungsverfahrens mit elliptischen Kurven [100]
SNOW	Kryptographie	Stromverschlüsselung [46]
Faltung	Signalverarbeitung	Bitweise Faltungsoperation
IEEE 802.11b	Signalverarbeitung	Teilimplementierung eines WLAN-Senders [44]
SATD	Bildverarbeitung	Teilalgorithmus der Videokompression des H.264-Standards [128]

Tabelle 8.1: Testprogramme zur Charakterisierung des Subschwellprozessors

Einfachheit halber werden diese Varianten im Folgenden als *Bypass aktiviert* bzw. *Bypass deaktiviert* bezeichnet.

Die Leistungsaufnahme bzw. der Energiebedarf ist generell abhängig von den Schalthäufigkeiten der Signale innerhalb einer integrierten Schaltung. Dies ist sowohl bei der Abschätzung durch Synthese- und Place-and-Route-Werkzeuge als auch bei einer messtechnischen Bestimmung zu berücksichtigen. Bei der Abschätzung während der Implementierungsphase sind unterschiedliche Vorgehensweisen möglich.

Die erste Möglichkeit besteht darin, die durchschnittlichen Schaltwahrscheinlichkeiten der primären Schaltungseingänge anzugeben. Dies können Standard-Werte der Implementierungswerkzeuge (typischerweise 10 % oder 20 %) oder Abschätzungen auf Basis bekannter Signalfunktionen sein. Durch statistische Methoden können diese Wahrscheinlichkeitswerte durch die einzelnen Logikgatter propagiert werden, um so eine Abschätzung der Schalthäufigkeiten aller übrigen Signale im Inneren der Schaltung zu erhalten. Alternativ können Schaltwahrscheinlichkeiten auch global für alle Signale vorgegeben werden. Aufbauend auf diesen Werten kann dann die Leistungsaufnahme bzw. der Energiebedarf der Schaltung ermittelt werden. Ein Nachteil dieses Vorgehens ist die damit verbundene relativ große Ungenauigkeit, so dass es oftmals nur als erste grobe Abschätzung in einem frühen Entwurfsstadium eingesetzt wird.

Eine höhere Genauigkeit lässt sich durch digitale Schaltungssimulationen erreichen, wobei die sich ergebenden realen Schalthäufigkeiten aller Signale aufgezeichnet und anschließend für die Bestimmung von Verlustleistung bzw. Energie herangezogen werden. Hierbei ist es auch möglich, verschiedene Anwendungsszenarien zu analysieren. Bei der Simulation ist zwischen unterschiedlichen Detaillierungsgraden zu unterscheiden. Eine RTL-Simulation kann nur die Schalthäufigkeiten der in der HDL-Beschreibung vorhandenen Signale ermitteln. Bei der Verlustleistungsabschätzung auf Basis solcher RTL-Schalthäufigkeiten können zudem meist nur die Eingänge von Submodulen annotiert werden, da durch die Optimierungsschritte von Synthese und Place-and-Route keine Zuordnung zwischen den übrigen Signalen auf RTL-Ebene und Gatter-Ebene möglich ist. Die fehlenden Schalthäufigkeiten werden wiederum durch Propagieren der bekannten Schalthäufigkeiten bestimmt. Das genaue Zeitverhalten der Gatter wird durch die RTL-Simulation nicht berücksichtigt. Damit weist auch dieses Verfahren einige Quellen für Ungenauigkeiten auf.

Die höchste Genauigkeit der Schalthäufigkeiten und damit der Abschätzungen von Verlustleistung und Energie kann durch Simulationen auf Gatter-Ebene mit Annotation des durch Synthese oder Place-and-Route ermittelten Zeitverhaltens der Gatter erzielt werden. Hierbei können die tatsächlichen Aktivitätswerte aller Signale aufgezeichnet werden, ein Propagieren von Schalthäufigkeiten entfällt. Auch kurzzeitige Signaländerungen innerhalb einer Taktperiode (*Glitches*) werden dabei berücksichtigt. Dieses Verfahren wurde bei der simulativen Bestimmung des Energiebedarfs eingesetzt, wobei die Programmsammlung aus Tabelle 8.1 genutzt wurde.

Die messtechnische Charakterisierung der Prozessorexemplare besteht aus einem an die binäre Suche angelehnten Bisektionsverfahren zur Bestimmung der maximalen Taktfrequenz für einen Spannungsschritt und der anschließenden Ermittlung des Energiebedarfs bei dieser Frequenz. Dabei wird zunächst der abzudeckende Frequenzbereich von 10 kHz bis 280 MHz¹ durch ein geeignetes Raster diskretisiert, indem jede Frequenzdekade in 32 näherungsweise logarithmische Frequenzschritte unterteilt wird. Zwischen 100 MHz und 200 MHz beträgt die Schrittweite 5 MHz und zwischen 200 MHz und 280 MHz beträgt die Schrittweite 10 MHz. Es ergeben sich also 157 mögliche Frequenzschritte. Die maximale Taktfrequenz bei einer gegebenen Spannung wird nun durch Intervallschachtelung in maximal $\lfloor \log_2 157 \rfloor + 1 = 8$ Schritten ermittelt. Das Testverfahren beginnt mit der maximalen Spannung von 1,2 V und dem Testintervall von 10 kHz bis 280 MHz. In jedem Schritt wird nun als Testfrequenz das mittlere Element zwischen Ober- und Untergrenze des Testinter-

¹ Die maximale testbare Taktfrequenz ist durch den FPGA-Entwurf begrenzt.

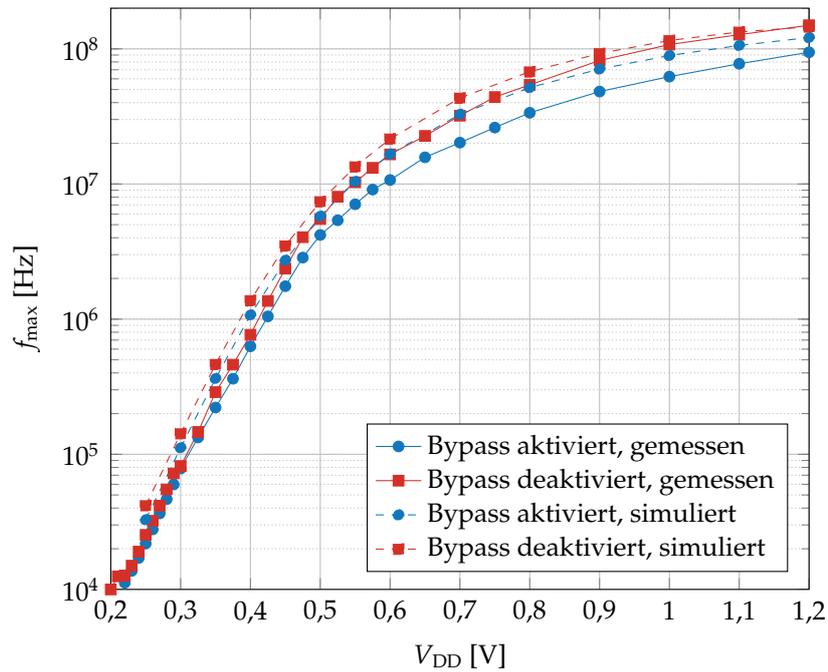
valls ermittelt und das Testprogramm ECC mit dieser Kombination aus Frequenz und Spannung gestartet. Dieses Testprogramm beansprucht alle relevanten, potentiell kritischen Pfade des Prozessors und ist daher sehr gut zur Bestimmung der maximalen Taktfrequenz geeignet¹. Mit Hilfe des CRC-Hardwarebeschleunigers wird während der Ausführung eine zyklische Redundanzprüfung über Zwischenergebnisse des Algorithmus gebildet und durch die externe Testumgebung geprüft. Bei korrekter Prüfsumme wird das Testintervall auf die obere Hälfte eingeschränkt und die Testfrequenz als neue maximale Frequenz gespeichert. Andernfalls wird das Testintervall auf die untere Hälfte eingeschränkt. Das Verfahren endet, wenn das Testintervall leer ist. Mit der gefundenen maximalen Taktfrequenz werden dann Energiemessungen für alle Testprogramme aus Tabelle 8.1 durchgeführt. Verlieh der Test für keine Frequenz erfolgreich, so wird das Exemplar für diese Spannung als nicht mehr funktionsfähig markiert. Andernfalls beginnt das Bisektionsverfahren für den nächst niedrigeren Spannungsschritt mit dem Testintervall von 10 kHz bis zur maximalen Taktfrequenz des vorherigen (höheren) Spannungsschrittes.

In den nachfolgenden Abschnitten werden die Analyse-Ergebnisse bzgl. maximaler Taktfrequenz, Energiebedarf und minimaler Versorgungsspannung diskutiert.

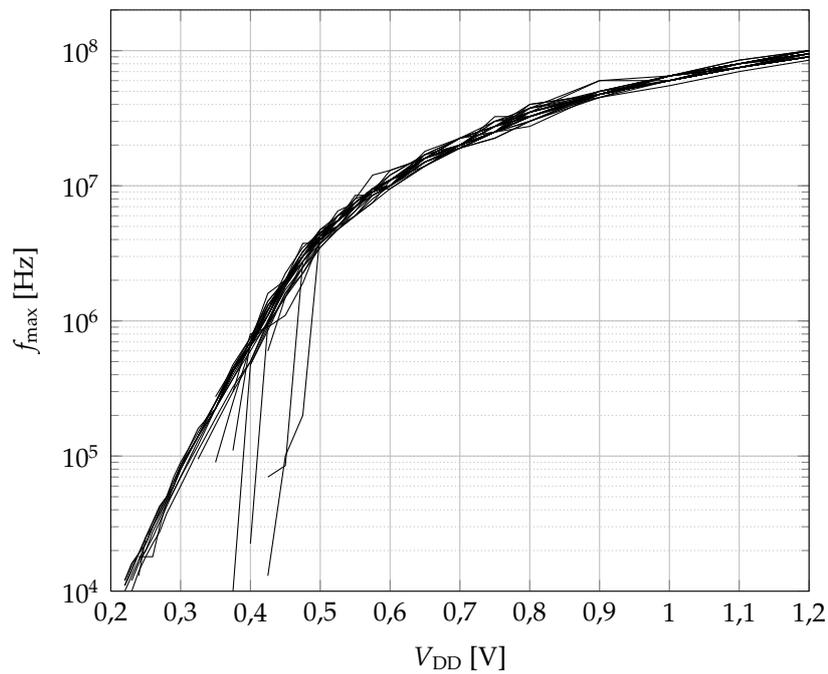
8.2 Maximale Taktfrequenz

Die Abhängigkeit der maximalen Taktfrequenz von der Versorgungsspannung ist in Abbildung 8.1 dargestellt. Während Abbildung 8.1a die Frequenzwerte zwischen aktiviertem und deaktiviertem Bypass sowie zwischen Simulation und dem Stichprobenmittelwert der Messung vergleicht, zeigt Abbildung 8.1b die Messkurven der einzelnen Prozessorexemplare bei aktiviertem Bypass. Bei 400 mV liegt die simulierte Taktfrequenz mit aktiviertem Bypass bei 1,07 MHz, d.h. geringfügig oberhalb der Synthesevorgabe von 1 MHz. Bei deaktiviertem Bypass lässt sich die Taktfrequenz auf 1,37 MHz steigern. Die entsprechenden Mittelwerte der Messungen liegen bei 629 kHz bzw. 766 kHz. Bei der maximalen Versorgungsspannung von 1,2 V beträgt die simulierte Taktfrequenz 121,2 MHz bei aktiviertem Bypass und 145,5 MHz bei deaktiviertem Bypass. Die Messwerte liegen hier bei 94,3 MHz bzw. 149,5 MHz. Die

¹ Es wurde für einige Schaltungsexemplare untersucht, ob neben ECC weitere Testprogramme berücksichtigt werden müssen, um ggf. zusätzliche kritische Pfade abzudecken. Es ergaben sich jedoch stets für alle Testprogramme die gleichen Maximalwerte der Frequenz, so dass für alle weiteren Exemplare die Analyse auf ECC beschränkt wurde.



(a) Vergleich zwischen aktiviertem und deaktiviertem Bypass sowie zwischen Messung (Stichprobenmittelwert) und Simulation



(b) Messergebnisse der einzelnen Exemplare bei aktiviertem Bypass

Abbildung 8.1: Maximale Taktfrequenz des Subschwellsprozessors

Auswertung der Exemplarstreuungen der Messwerte ergibt bei 1,2 V und aktiviertem Bypass einen Variationskoeffizienten σ/μ von 4,4 %. Bei 325 mV, der Spannung niedrigster Energieaufnahme (siehe Abschnitt 8.3), beträgt der Variationskoeffizient 13,9%. Hieran erkennt man die Zunahme der Parameterstreuungen mit sinkender Versorgungsspannung. Es wurde festgestellt, dass es dabei statistisch unerheblich ist, ob Prozessorexemplare innerhalb des gleichen Bausteins oder auf unterschiedlichen Bausteinen verglichen werden. Dies lässt darauf schließen, dass sich wahrscheinlich alle Exemplare bei der prototypischen Fertigung in unmittelbarer Nähe auf dem produzierten *Wafer* befunden haben.

Beim genaueren Vergleich der Kurvenverläufe fällt zunächst auf, dass die Messwerte im Mittel um ca. 28 % unter den jeweiligen Simulationsergebnissen für typische PVT-Bedingungen liegen. Da die konkreten Prozessbedingungen für den Produktionswafer nicht bekannt sind, kann dies jedoch als gute Übereinstimmung angesehen werden. Ferner ist zu bemerken, dass der Geschwindigkeitszuwachs durch das Deaktivieren des Bypasses abhängig von der Versorgungsspannung variiert. Eine Erklärung hierfür liefert der Verlauf des kritischen Pfades. Bei aktiviertem Bypass verläuft der kritische Pfad von den Operanden-Registern zwischen RD- und EX-Stufe durch den Addierer der ALU, die Auswertungseinheit der Sprungbedingungen, den Kontroll-Bypass und die Sprungvorhersage über den Adresseingang zum Instruktionscache (vgl. Abbildung 7.6 auf Seite 126). Da der Instruktionscache außerhalb der variablen Spannungsdomäne des Prozessorkerns liegt, ist dessen Anteil an der Pfadverzögerung konstant, d.h. es ändert sich nur die Verzögerungszeit des verbleibenden Pfadanteils bei einer Skalierung der Versorgungsspannung. Bei deaktiviertem Bypass hängt der Verlauf des kritischen Pfades von der Versorgungsspannung ab. Während oberhalb von 600 mV ebenfalls kritische Pfade zum Tragen kommen, die außerhalb des Prozessorkerns beginnen oder enden, liegt der kritische Pfad bei geringeren Spannungen vollständig innerhalb der variablen Spannungsdomäne. Bei 400 mV beispielsweise verläuft der kritische Pfad vom Instruktionsregister zwischen RD- und EX-Stufe durch den Multiplizierer zum Zwischenergebnis-Register der Multiplikation zwischen EX- und ME-Stufe. Daher skaliert das Zeitverhalten des Prozessors bei aktiviertem oder deaktiviertem Bypass unterschiedlich mit der Versorgungsspannung.

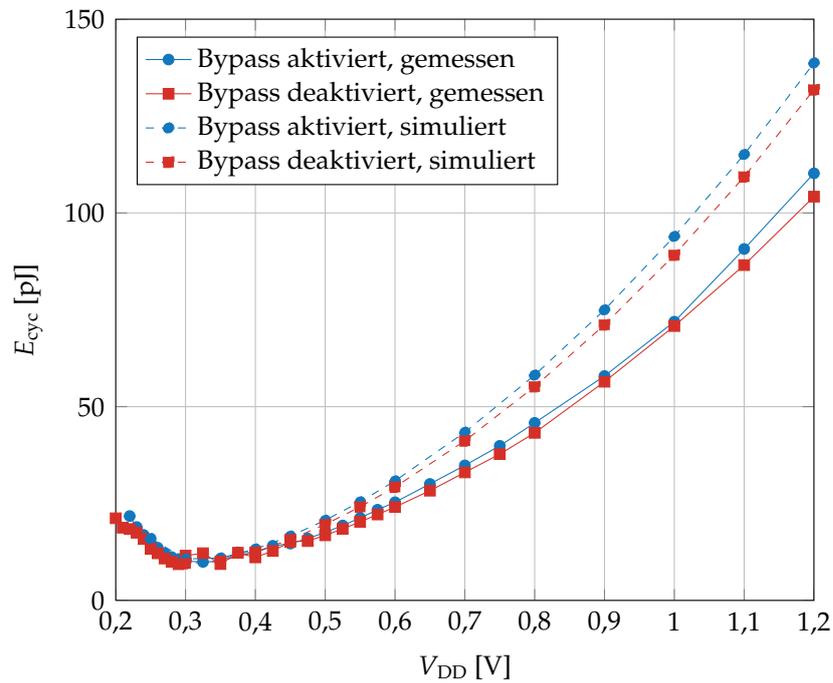
Anhand von Abbildung 8.1b erkennt man, dass bei einigen Exemplaren unterhalb von etwa 500 mV die Taktfrequenz stärker abfällt, als bei den meisten übrigen Exemplaren. Hierbei handelt es sich um die Exemplare, die bereits bei einer relativ hohen Versorgungsspannung ein Fehlverhalten zeigen. Eine Diskussion der Ursachen folgt in Abschnitt 8.4.

8.3 Energiebedarf

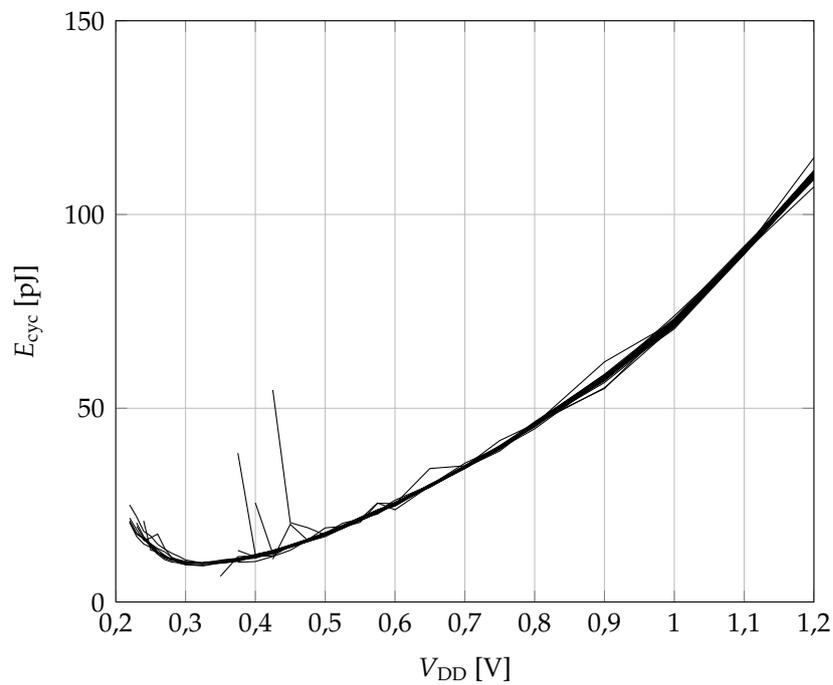
Abbildung 8.2 zeigt die Energieaufnahme des Prozessors pro Takt als Durchschnitt über alle Testprogramme. Wie zuvor gibt Abbildung 8.1a einen Vergleich zwischen den beiden Bypass-Konfigurationen sowie zwischen Messung und Simulation. Die Messwerte der einzelnen Prozessorexemplare bei aktiviertem Bypass sind in Abbildung 8.1b dargestellt. Der simulierte minimale Energiebedarf tritt bei 300 mV mit 10,69 pJ für aktivierten Bypass und 9,55 pJ für deaktivierten Bypass auf. Bei den Messungen ergibt sich für aktivierten Bypass ein Energieminimum von 9,94 pJ bei 325 mV und 133,18 kHz bzw. für deaktivierten Bypass ein Energieminimum von 9,35 pJ bei 290 mV und 72,22 kHz. Es ist anzumerken, dass für die Messungen ein feineres Spannungsraster als bei den Simulationen verwendet wird, was eine genauere Lokalisierung des Energieminimums ermöglicht. Bei 1,2 V beträgt der simulierte Energiebedarf 138,70 pJ bzw. 131,79 pJ bei aktiviertem bzw. deaktiviertem Bypass. Die gemessenen Werte betragen 110,22 pJ und 104,20 pJ. Dies entspricht einem Einsparungspotential für den Subschwellebetrieb im Energieminimum um den Faktor 11 gegenüber einem Betrieb bei maximaler Versorgungsspannung. Die Verlustleistung reduziert sich dabei um fast vier Größenordnungen von 10,4 mW auf 1,32 μ W bei aktiviertem Bypass.

Vergleicht man die simulativ und messtechnisch ermittelten Werte des Energiebedarfs, so ergeben sich insbesondere bei niedrigen Versorgungsspannungen sehr geringe Abweichungen von wenigen Prozent. Bei höheren Versorgungsspannungen liegt der simulierte Energiebedarf um etwa 25 % über den Messwerten. Auch dies ist jedoch angesichts der unbekanntenen Prozessbedingungen als gute Übereinstimmung zu werten.

Die in Abbildung 8.2b vom durchschnittlichen Kurvenverlauf abweichenden Graphen stammen von den Exemplaren mit den größten Werten der minimalen Versorgungsspannung. Da diese Exemplare, wie in Abschnitt 8.2 beschrieben, unmittelbar oberhalb ihrer minimalen Versorgungsspannung eine deutliche Abnahme der maximalen Taktfrequenz aufweisen, kommt es zu einer verstärkten Akkumulation statischer Verlustleistungsanteile während einer Taktperiode und damit zu einer Zunahme des Energiebedarfs.



(a) Vergleich zwischen aktiviertem und deaktiviertem Bypass sowie zwischen Messung (Stichprobenmittelwert) und Simulation



(b) Messergebnisse der einzelnen Exemplare bei aktiviertem Bypass

Abbildung 8.2: Energiebedarf des Subschwellsprozessors

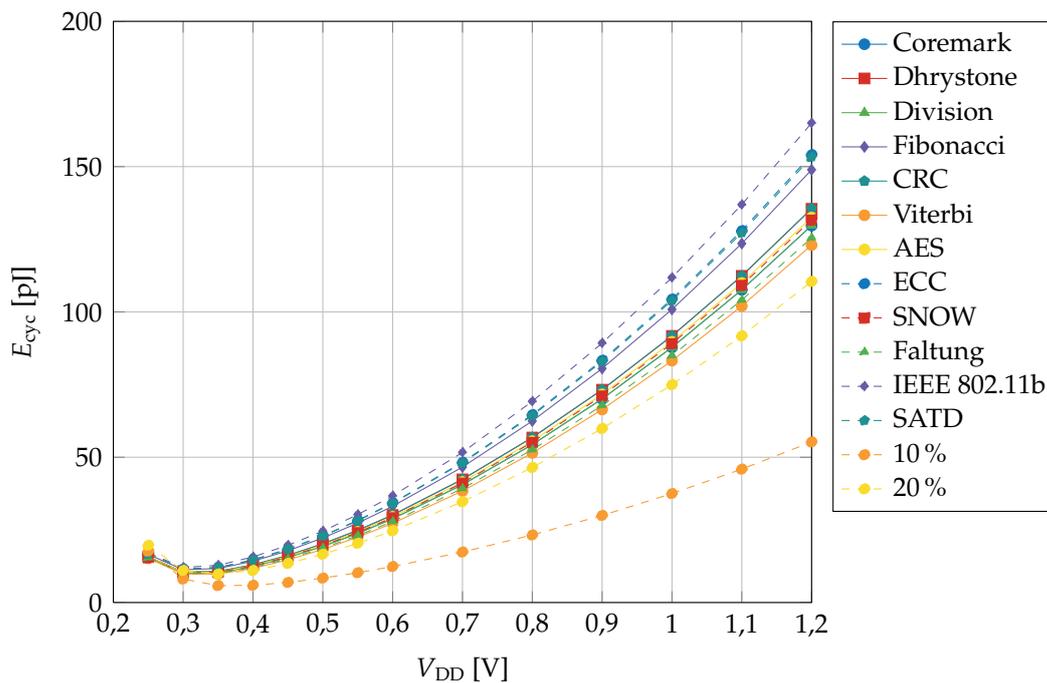


Abbildung 8.3: Energiebedarf pro Takt bei aktiviertem Bypass für unterschiedliche Testprogramme (simuliert)

8.3.1 Vergleich unterschiedlicher Testprogramme

Abbildung 8.3 zeigt die simulierte Energieaufnahme des Prozessors pro Takt bei aktiviertem Bypass für die einzelnen Testprogramme sowie für die Standardwerte 10 % und 20 % globaler Schaltwahrscheinlichkeiten aller Signale des Prozessorkerns. Zunächst ist festzustellen, dass die Kurven für die beiden zuletzt genannten Fälle unterhalb aller realen Testprogramme liegt, d.h. die mittlere Schaltwahrscheinlichkeit innerhalb des Prozessorkerns beträgt mehr als 20 %. Für alle betrachteten Testprogramme tritt der minimale Energiebedarf bei der bereits zuvor genannten Spannung von 300 mV auf. Das Testprogramm IEEE 802.11b weist den größten Wert von 12,1 pJ auf, während der kleinste Wert von 9,8 pJ mit dem Testprogramm Viterbi auftritt. Bei 1,2 V liegt der Energiebedarf zwischen 123,0 pJ und 165,0 pJ. Mit den Messdaten ergibt sich ein sehr ähnliches Bild bezüglich der unterschiedlichen Testprogramme, so dass auf eine separate Darstellung und Angabe entsprechender Werte verzichtet werden soll.

8.3.2 Vergleich der Bypass-Konfigurationen

Für einen fairen Vergleich der Bypass-Konfigurationen bezüglich der Ausführungszeit und des Energiebedarfs ist eine Berücksichtigung der durch Datenkonflikte bei deaktiviertem Bypass verursachten zusätzlichen Strafzyklen erforderlich, deren Anzahl von dem jeweiligen Testprogramm abhängt. Sei $f_{\max,1}$ die (maximale) Taktfrequenz, $E_{\text{cyc},1}$ der Energiebedarf pro Takt und N_1 die Anzahl an Taktzyklen für den Durchlauf eines Testprogramms bei aktiviertem Bypass. $f_{\max,2}$, $E_{\text{cyc},2}$ und N_2 seien die entsprechenden Werte bei deaktiviertem Bypass. Die Ausführungszeit und der Gesamtenergiebedarf eines Testprogramms sind durch

$$T_{\text{gesamt},i} = \frac{N_i}{f_{\max,i}} \quad (8.1)$$

$$E_{\text{gesamt},i} = N_i E_{\text{cyc},i} \quad (8.2)$$

mit $i \in \{1,2\}$ gegeben. Es werden die Verhältnisse

$$\rho_{\text{CLK}} = \frac{1/f_{\max,2}}{1/f_{\max,1}} = \frac{f_{\max,1}}{f_{\max,2}} \quad (8.3)$$

$$\rho_{\text{N}} = \frac{N_2}{N_1} \quad (8.4)$$

$$\rho_{\text{T}} = \frac{T_{\text{gesamt},2}}{T_{\text{gesamt},1}} = \rho_{\text{N}} \rho_{\text{CLK}} \quad (8.5)$$

$$\rho_{\text{E}} = \frac{E_{\text{gesamt},2}}{E_{\text{gesamt},1}} \quad (8.6)$$

definiert.

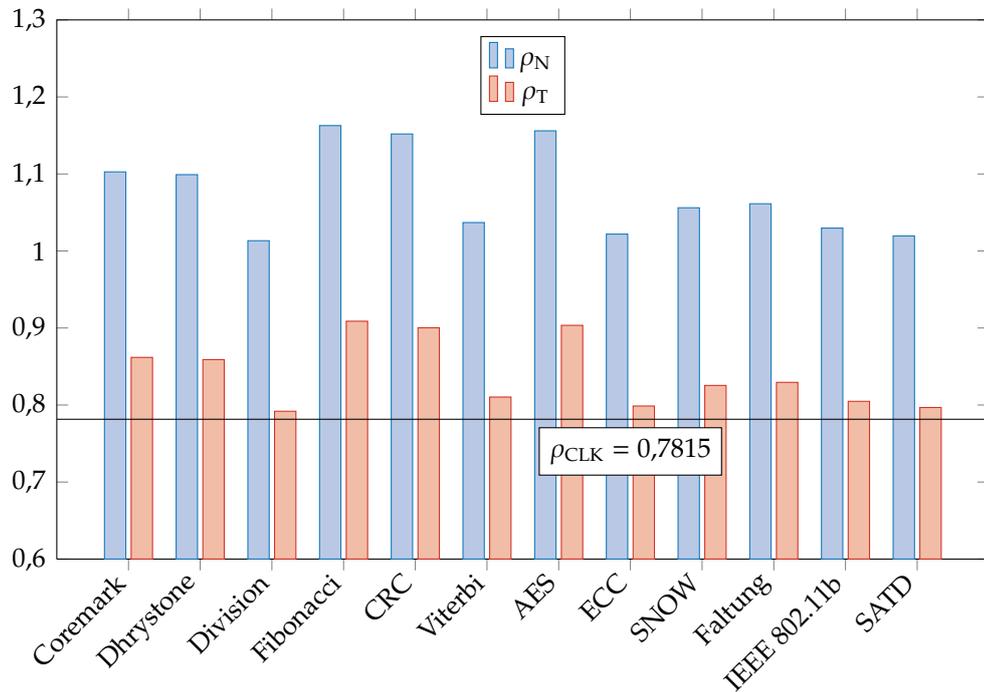
Abbildung 8.4a zeigt die relative Änderung der Anzahl an Taktzyklen ρ_{N} und der Ausführungszeit ρ_{T} der Testprogramme sowie der Periodendauer ρ_{CLK} bei deaktiviertem Bypass. Der für ρ_{CLK} angegebene Wert von ca. 78 % ist hier als Durchschnitt über alle simulierten Spannungsschritte zu verstehen. Je nach Testprogramm steigt die Anzahl an Taktzyklen durch zusätzliche Strafzyklen bei deaktiviertem Bypass um 1,3 % bis 16,3 %. Dennoch erkennt man, dass für alle Testprogramme aufgrund der höheren Taktfrequenz eine Verringerung der Ausführungszeit zwischen 10 % und 20 % erzielt werden kann. Dies ist konsistent zu den Analyse-Ergebnissen in [66, 67].

Die Änderung des Gesamtenergiebedarfs bei deaktiviertem Bypass ρ_E ist in Abbildung 8.4b dargestellt. Die größten Werte für ρ_E sind für das Testprogramm Fibonacci zu beobachten, während die geringsten Werte auf das Testprogramm SATD entfallen. Oberhalb von etwa 450 mV ist ρ_E nahezu konstant und liegt je nach Testprogramm zwischen 1,001 und 1,050. Das bedeutet, dass in diesem Spannungsbereich ein dynamisches Deaktivieren des Bypasses aus energetischer Sicht nicht sinnvoll ist, da der Gesamtenergiebedarf zur Erfüllung einer gegebenen Aufgabe steigt. Darunter ist eine Abnahme von ρ_E zu beobachten, so dass sich bei einer Spannung von 300 mV für alle Testprogramme eine Energieersparnis durch das Deaktivieren des Bypasses ergibt. Das dynamische Deaktivieren von Bypass-Pfaden ist also insbesondere im Subschwellbetrieb des Prozessors vorteilhaft. Das Verhalten von ρ_E lässt sich folgendermaßen erklären: Da das Abschalten der Bypass-Pfade dynamisch erfolgt, sind keine Energieeinsparungen durch wegfallende Logikgatter und Verdrahtung möglich, wie dies bei einer statischen Änderung der Bypass-Konfiguration der Fall wäre [66, 67]. Durch die veränderte Taktfrequenz und die zusätzlichen Strafzyklen kommt es jedoch zu einer Verschiebung zwischen statischen und dynamischen Verlusten. Während die insgesamt verkürzte Ausführungszeit der Testprogramme bei deaktiviertem Bypass zu einer Abnahme des statischen Energiebedarfs führt, bewirken die zusätzlichen Strafzyklen eine Erhöhung des dynamischen Energiebedarfs. Da bei hohen Versorgungsspannungen die dynamischen Anteile überwiegen, kommt es hier insgesamt zu einer Zunahme des Energiebedarfs. Im tiefen Subschwellbereich hingegen dominieren die statischen Anteile, so dass sich hier durch die verkürzte Ausführungszeit insgesamt eine Energieersparnis ergibt.

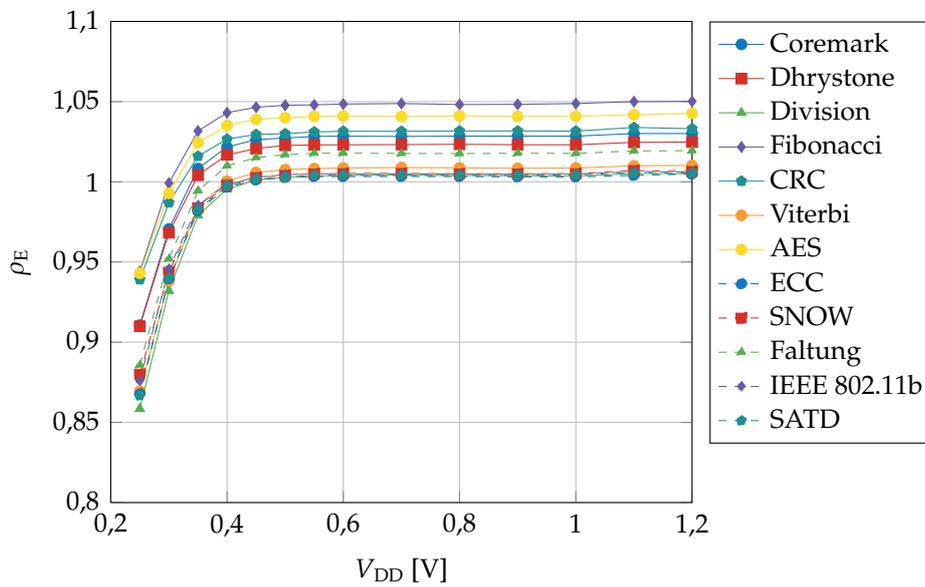
8.4 Minimale Versorgungsspannung

Ein Histogramm der minimalen Versorgungsspannung für einen fehlerfreien Betrieb der Prozessorexemplare ist in Abbildung 8.5 dargestellt. Die besten Exemplare sind noch mit einer Spannung von 200 mV bei einer Taktfrequenz von 10 kHz funktionsfähig, bei den schlechtesten Exemplaren treten Fehler bereits bei 450 mV auf. Die Stichprobe besitzt einen Mittelwert von 346,9 mV und eine Standardabweichung von 87,1 mV. Dies entspricht einem Variationskoeffizienten $\sigma/\mu \approx 0,2512$.

Wie bereits in Abschnitt 3.2 erläutert, wird die minimale Versorgungsspannung einer Schaltung im Subschwellbetrieb in erster Linie durch unzureichende Störabstände und Hold-Zeit-Verletzungen bestimmt. Da die im Rahmen dieser Arbeit



(a) Relative Änderung der Periodendauer sowie der Anzahl an Taktzyklen und der Ausführungszeit



(b) Relative Änderung des Gesamtenergiebedarfs

Abbildung 8.4: Einfluss des dynamisch deaktivierten Bypass für unterschiedliche Testprogramme (simuliert)

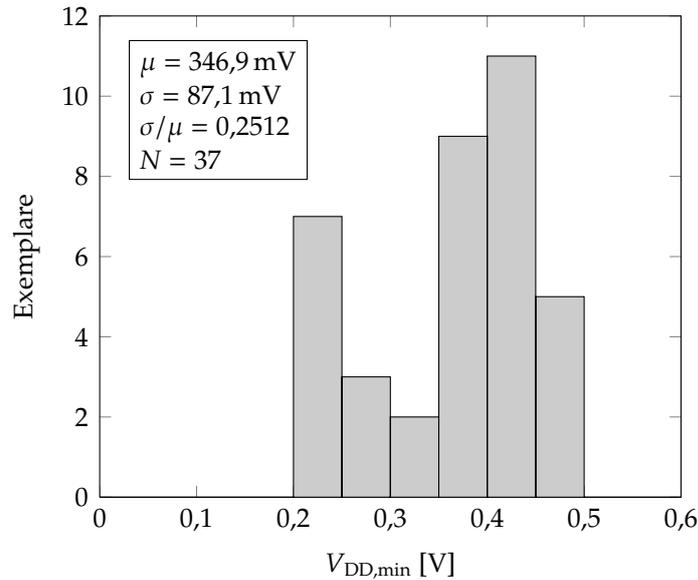


Abbildung 8.5: Minimale Versorgungsspannung der Prozessorexemplare

entwickelten Standardzellen explizit auf maximale Störabstände im Subschwelligbereich optimiert wurden, ist diese Fehlerursache zumindest bei den vorstehenden Spannungswerten sehr unwahrscheinlich. Es ist daher anzunehmen, dass die minimale Versorgungsspannung des Prozessors primär durch Hold-Zeit-Verletzungen begründet ist. Eine Verbesserung ist daher eventuell durch eine Modifikation der OCV-Skalierungsfaktoren bei der Implementierung oder durch alternative Takt-Konzepte mit verringerter Streuung der Verzögerungen innerhalb des Taktbaums möglich. In [15, 135] wird beispielsweise für Subschwellschaltungen vorgeschlagen, die Taktverteilung auf einen einzigen, leistungsstarken Takttreiber zu reduzieren und das Taktsignal lediglich durch die Verdrahtung mit allen Flipflops zu verbinden. Variationen der Takttreiber-Verzögerung wirken daher gleichermaßen auf alle Flipflops. Zudem ist der Einfluss der Prozessschwankungen auf die Leitungsverzögerungen insbesondere bei geringen Taktraten im Subschwelligbereich zu vernachlässigen, so dass sich ein besseres Hold-Zeit-Verhalten ergibt. Ein Nachteil dieses Ansatzes ist, dass Clock-Gating nicht oder nur eingeschränkt einsetzbar ist und in jedem Fall zu einer Reduktion der Hold-Zeit-Reserven führt. Ein weiterer möglicher Ansatz ist die Verwendung von Takttreibern mit programmierbarer Verzögerung, um deren variierende Charakteristika nach der Fertigung zu kompensieren [63].

8.5 Vergleich mit ähnlichen Implementierungen

Um die vorstehend diskutierten Eigenschaften des Subschwellprozessors im Vergleich zu ähnlichen Implementierungen bewerten zu können, gibt Tabelle 8.2 eine entsprechende Übersicht. Da in den genannten Veröffentlichungen für einen sinnvollen Vergleich teilweise wichtige Daten nicht als Zahlenwerte angegeben sind, wurde versucht, solche Daten aus Diagrammen oder durch Berechnungen zu gewinnen. Solche Daten sind mit einer gewissen Unsicherheit behaftet und daher in der Tabelle entsprechend markiert. Die angegebene Speicherkapazität bezieht sich bei Harvard-Architekturen jeweils auf den insgesamt vorhandenen Instruktion- und Datenspeicher. Die in der Tabelle genannte Fläche umfasst, sofern den Veröffentlichungen entsprechende Angaben entnommen werden konnten, auf die tatsächlich von Prozessorkern und ggf. Speicher belegte Fläche. Die Untergrenze des Betriebsspannungsbereichs ist der jeweils niedrigste beobachtete Wert.

In der oberen Hälfte der Tabelle sind Daten für Gesamtsysteme bestehend aus Prozessorkern und Speicher angegeben, während die untere Tabellenhälfte Daten für separate Prozessorkerne enthält. Für den im Rahmen dieser Arbeit realisierten Subschwellprozessor sind in der Tabelle, neben den rein auf den Prozessorkern bezogenen Daten, zusätzlich rechnerisch ermittelte Werte für ein Gesamtsystem in Verbindung mit 8 kB Datenspeicher und 8 kB Instruktionsspeicher auf Basis des im nachfolgenden Kapitel 9 vorgestellten Subschwell-SRAM angegeben. Diese Werte basieren auf der Annahme, dass diese Speicherkapazität durch 32 Blöcke zu je 64×32 bit realisiert wird, wobei in jedem Takt jeweils ein Block aktiv ist und alle übrigen Blöcke inaktiv sind. Dies lässt sich durch eine entsprechende Speicherschnittstelle sicherstellen.

Anhand der in der Tabelle gegebenen Vergleichsdaten fällt zunächst auf, dass mit Ausnahme von [63, 65] alle übrigen Implementierungen offenbar nicht für den Betrieb bei hohen Spannungen, d.h. bis zur nominellen Versorgungsspannung der jeweiligen Technologie, geeignet oder ausgelegt sind. Dementsprechend erreichen diese Implementierungen im Vergleich zu dem hier entwickelten Subschwellprozessor keine vergleichbare maximale Taktfrequenz und sind daher nicht für Anwendungen geeignet, die zeitweise hohe Performanzanforderungen aufweisen. Bezogen auf die Taktfrequenz im Energieminimum sind die aufgeführten Implementierungen einschließlich CoreVA ULP in etwa vergleichbar, mit Ausnahme der Implementierungen [15, 63] mit Taktfrequenzen von 25 MHz bzw. 60 MHz, die durch die Verwendung von Transistoren mit geringer Schwellspannung bei moderater Inversion (engl.

near-threshold) erreicht werden. Diese Betriebsart ist mit deutlich erhöhter statischer Verlustleistung verbunden [15, 63].

Bezüglich des minimalen Energiebedarfs sowie der Fläche weist der Prozessorkern von CoreVA ULP im Vergleich auf den ersten Blick ungewöhnlich hohe Werte auf. Der Energiebedarf liegt um einen Faktor 4 bis 19 über den übrigen Implementierungen. Auch hier gilt es jedoch die Architektur-bezogenen Unterschiede zu berücksichtigen. Während CoreVA ULP innerhalb eines Taktes eine 32-bit-Operation oder im SIMD-Modus zwei 16-bit-Operationen durchführen kann, sind die übrigen Implementierungen je nach Architektur lediglich in der Lage, eine 8-bit- oder 16-bit-Operation auszuführen. Zur Berechnung einer 32-bit-Operation benötigen die Vergleichs-Implementierungen daher mehrere Takte und entsprechend mehr Energie. Besonders deutlich wird dieser Unterschied etwa bei der Berechnung einer 32-bit-Multiplikation. Dazu müssen in einer 8-bit-Architektur ohne Hardware-Multiplizierer 256 Additionen, 128 Schiebeoperationen sowie einige Vergleichsoperationen und Sprünge ausgeführt werden. Daher benötigt eine solche Architektur für diese Berechnung mehr als das 300-fache des Energie- und Zeitbedarfs einer einzelnen Taktperiode. Für derartige Berechnungen ist daher die CoreVA-Architektur mit Hardware-Multiplizierer effektiv wesentlich energieeffizienter und leistungsfähiger. Die Multiplizier-Akkumulier-Einheit und die Divisionsschritt-Einheit von CoreVA ULP sind für ca. 15 % des Energiebedarfs des Prozessorkerns verantwortlich. Dank der Konfigurierbarkeit der Architektur kann dieser Anteil durch Entfernen der Module zur Entwurfszeit im Wesentlichen eliminiert werden, wenn sie für eine bestimmte Anwendung nicht notwendig sind.

Beim Vergleich von Fläche und Energiebedarf der Implementierungen einschließlich Speicher liegt CoreVA ULP im Mittelfeld. Der Energiebedarf von [63] liegt aufgrund der sehr komplexen superskalaren Architektur und dem Betrieb bei moderater Inversion um mehr als das Zehnfache über dem Energiebedarf von CoreVA ULP. Diese Realisierung zielt jedoch eher auf Anwendungen mit höheren Performanzanforderungen ab.

8.6 Vergleich mit konventionellen Standardzellen

Wie in Abschnitt 7.3 erwähnt, enthält die prototypische ASIC-Realisierung von CoreVA ULP neben zwei Instanzen des Subschwellsprozessors eine zusätzliche Instanz, die auf der konventionellen Standardzellen-Bibliothek von STMicroelectronics

Referenzen	Architektur	Tech. [nm]	Speicher [kB]	Typ	Fläche [mm ²]	Spannungsbereich [V]	Energieminimum [V]	[MHz]	[pJ]
[57, 58, 109]	8 bit RISC, 3-stufig	130	0,315	Latch-Mux	0,18	0,15 – 0,5	0,35	0,354	3,5
[175, 176]	8 bit CISC, 2-stufig, <i>Subliminal</i>	130	0,25	Latch-Mux	0,085	0,2 – 0,625	0,36	0,833	3,0
[55, 136]	8 bit, 2-stufig, <i>Phoenix</i>	180	0,41	14T	0,84 ³	0,45 ¹ – 0,9 ¹	0,5 ²	0,106	2,8
[89, 90]	16 bit RISC, MSP430-basiert	65	16	8T	1,5	0,3 – 0,6	0,5	0,434	27,2
[14]	16 bit RISC, MSP430-basiert	65	18	konventionell	0,42 ³	0,4	0,4	25,0	7,0
[63]	32 bit superskalar, IA-32	32	16	10T	2,0	0,28 – 1,2	0,45	60,0 ¹	170,0
Diese Arbeit	32 bit RISC, <i>CoreVA LLP</i>	65	16	9T	0,96 ¹	0,27 ¹ – 1,2	0,325 ¹	0,133 ¹	16,02 ¹
[57, 58, 109]	8 bit RISC, 3-stufig	130			0,025	0,15 – 0,5	0,29	0,077	0,515
[175, 176]	8 bit CISC, 2-stufig, <i>Subliminal</i>	130		ohne	0,03	0,2 – 0,625	0,28	0,07 ¹	0,85
[55, 136]	8 bit, 2-stufig, <i>Phoenix</i>	180				0,45 ¹ – 0,9 ¹	0,5 ²	0,106	2,46 ¹
[65]	8 bit, 2-stufig, PIC16C5X-basiert	130				0,24 – 1,2	0,28	0,475	1,51
Diese Arbeit	32 bit RISC, <i>CoreVA LLP</i>	65		ohne	0,36	0,2 – 1,2	0,325	0,133	9,94

¹ Aus Diagrammen und/oder Berechnungen abgeleitet, ungefährender Wert

² Nennbetriebspunkt statt Energieminimum

³ Einschließlich Peripherie

Tabelle 8.2: Vergleich des Subschwellprozessors mit ähnlichen Implementierungen

basiert. In Abbildung 8.6 werden die Messergebnisse dieses Prozessorkerns mit den entsprechenden Eigenschaften des Subschwellsprozessorkerns verglichen.

Der konventionelle Prozessor erreicht bei 1,2 V und aktiviertem Bypass eine Taktfrequenz von 260 MHz bei einem Energiebedarf von 66,13 pJ pro Takt. Dies entspricht Verbesserungen von 160 % bezüglich der Geschwindigkeit und 40 % bezüglich der Energie gegenüber dem Subschwellsprozessor bei gleicher Spannung. Bei diesem Vergleich ist allerdings zu beachten, dass sich die jeweils verwendeten Standardzellen-Bibliotheken stark in ihrer Größe unterscheiden. So nutzt die konventionelle Implementierung 720 unterschiedliche Zelltypen, während die Subschwellsimplementierung lediglich 43 Gattertypen nutzt (jeweils ohne Place-and-Route-Zellen). Um zu untersuchen, welchen Einfluss diese Tatsache auf die Eigenschaften des konventionellen Prozessors hat, wurden zusätzliche Synthesen durchgeführt, bei denen die konventionelle Bibliothek auf einen der Subschwellsbibliothek äquivalenten Zellensatz reduziert wurde. Gegenüber der tatsächlich für die ASIC-Realisierung verwendeten Synthese ergab sich dabei eine um 48 % geringere Taktfrequenz und ein um 9 % höherer Energiebedarf. Geht man davon aus, dass sich diese relativen Änderungen näherungsweise auf die Messergebnisse eines entsprechenden gefertigten Prozessors übertragen lassen, so ergäbe sich eine Taktfrequenz von 135 MHz und ein Energiebedarf von 72,08 pJ. Dies entspricht im Vergleich mit dem Subschwellsprozessor nur noch Vorteilen von 43 % bezüglich der Taktfrequenz bzw. 35 % bezüglich der Energieaufnahme. Umgekehrt sind bei einer Erweiterung der Subschwellsbibliothek Verbesserungen der Eigenschaften von CoreVA ULP zu erwarten.

Die minimale Versorgungsspannung des konventionellen Prozessors beträgt 700 mV und liegt damit deutlich oberhalb des Subschwellsbereichs und deutlich höher als die minimale Versorgungsspannung des Subschwellsprozessors. Dies unterstreicht die Notwendigkeit einer speziell für den Subschwellsbetrieb optimierten Standardzellen-Bibliothek. Die Taktfrequenz des konventionellen Prozessors beträgt in diesem Betriebspunkt 42,5 MHz und der Energiebedarf 21,63 pJ. Damit erreicht der konventionelle Prozessor nicht die gleiche Energieeffizienz wie der Subschwellsprozessor in seinem optimalen Betriebspunkt.

8.7 Zusammenfassung

In diesem Kapitel wurden die Messverfahren und Messergebnisse des Subschwellsprozessors CoreVA ULP vorgestellt und diskutiert. Im Stichprobenmittel und bei

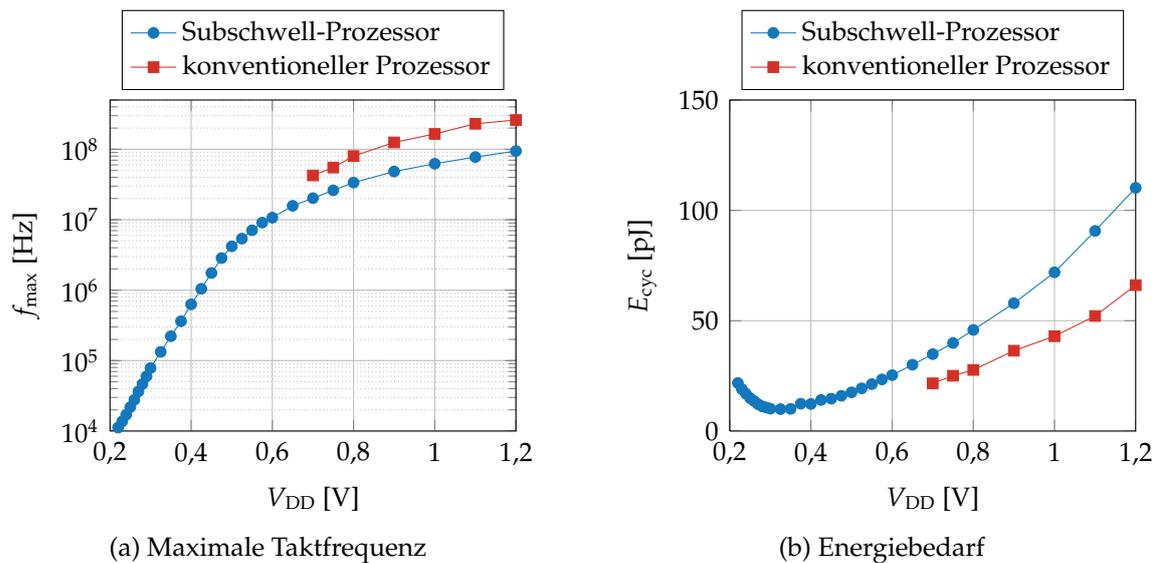


Abbildung 8.6: Vergleich der gemessenen Eigenschaften des Subschwellsprozessors und des konventionellen Prozessors

aktiviertem Bypass erreicht der Prozessorkern einen minimalen Energiebedarf von 9,94 pJ bei 325 mV und 133,18 kHz. Bei der maximalen Versorgungsspannung von 1,2 V erhöhen sich Energiebedarf und Taktfrequenz auf 110,22 pJ bzw. 94,3 MHz. Der Subschwellsbetrieb gestattet also Energieeinsparungen um etwa den Faktor 11. Unter Berücksichtigung der unbekanntenen Prozessbedingungen ergibt sich eine gute Übereinstimmung zwischen den gemessenen und simulierten Schaltungseigenschaften. Die minimale Versorgungsspannung des adaptiven Prozessors liegt im Mittel bei 346,9 mV. Sieben Exemplare bzw. 18,9 % arbeiten noch bei 200 mV korrekt. Durch den Einsatz alternativer Taktkonzepte ist an dieser Stelle weiteres Potential durch eine Verbesserung des Hold-Zeitverhaltens vorhanden. Im Vergleich mit ähnlichen Implementierungen sticht der Subschwellsprozessor durch seine architekturbedingt hohe Leistungsfähigkeit hervor. Bei multiplikationslastigen Anwendungen besitzt CoreVA ULP deutliche Vorteile bezüglich Laufzeit und Energiebedarf. Weiterhin ist CoreVA ULP eine von wenigen Realisierungen, die auch für einen Betrieb bei der nominellen Versorgungsspannung der jeweiligen Technologie ausgelegt sind und damit kurzzeitig hohen Performanzanforderungen genügen können. Eine konventionelle Prozessorvariante besitzt bei 1,2 V eine Taktfrequenz von 260 MHz und einen Energiebedarf von 66,13 pJ pro Takt. Der Energiebedarf ist damit um etwa 40 % geringer als der des Subschwellsprozessors bei maximaler Spannung. Allerdings erreicht der konventionelle Prozessor lediglich eine minimale Versorgungsspannung von 700 mV mit einem Energiebedarf von 21,63 pJ. Damit kommt dieser Prozessor dennoch nicht an die Energieeffizienz des Subschwellsprozessors heran.

Neben einem Prozessorkern ist üblicherweise ein Speicher mit wahlfreiem Zugriff bzw. RAM als Komponente eines eingebetteten Systems erforderlich. Der Speicher dient als Ablageort für den Programmcode und die zu verarbeitende Daten. Im Rahmen der prototypischen ASIC-Realisierung wurden die eigentlichen Programm- und Datenspeicher zunächst als konventionelle Speicher realisiert, da der Entwurf von Speicher für den Subschwellsbetrieb mit besonderen Herausforderungen verbunden ist. Der Entwurf und die Charakterisierung eines solchen Subschwellspeichers, der als separate Komponente ebenfalls Teil des ASICs ist, wird im nachfolgenden Kapitel behandelt.

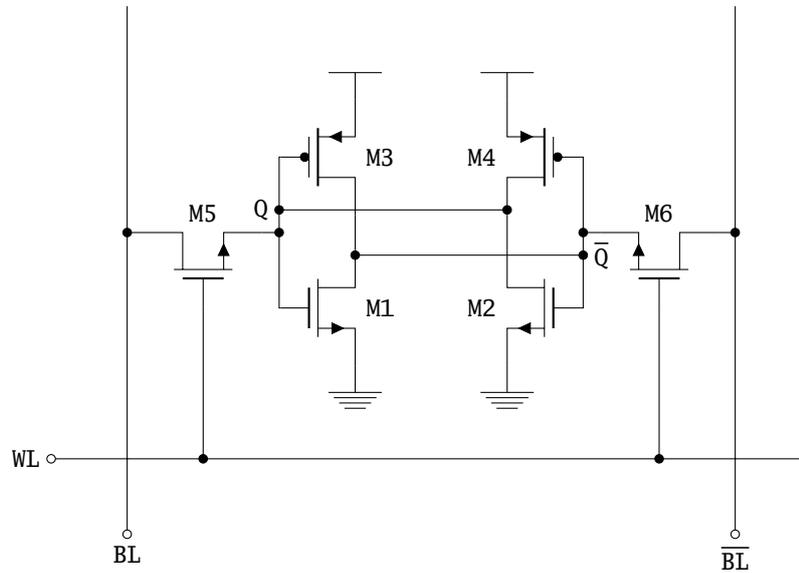
9 Subschwellig-SRAM

Der zu dem in Kapitel 7 vorgestellten ASIC gehörende Subschwellig-SRAM entstand im Rahmen einer Kooperation mit der Fachgruppe Nanoelektronik der Universität Oslo, die für den Entwurf der zugehörigen Schaltkreise auf Schaltplan- und Layout-Ebene verantwortlich war. Die Dimensionierung der Speicherzelle durch ein Mehrzieloptimierungsverfahren wird in der Dissertation [12] sowie in [8] behandelt. Schwerpunkte der Tätigkeiten im Zusammenhang mit dieser Arbeit waren die Entwicklung einer Schnittstelle zur Integration des SRAMs in das CoreVA-System sowie die messtechnische Charakterisierung der gefertigten Speicherexemplare.

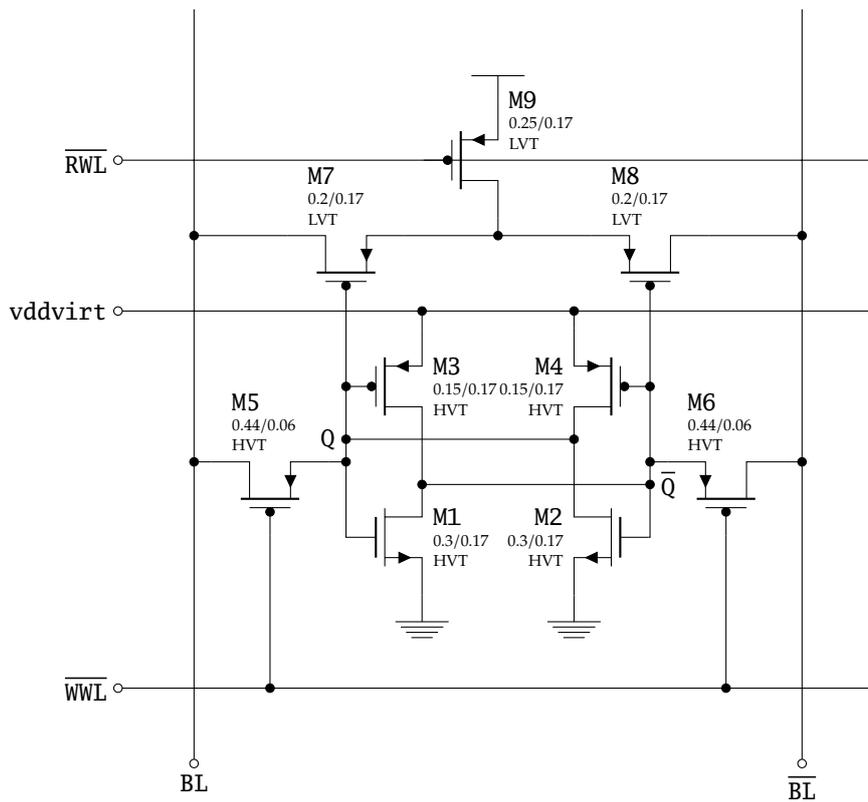
Der in diesem Kapitel vorgestellte SRAM ist für den Betrieb mit niedrigen Versorgungsspannungen im Subschwelligbereich ausgelegt. In diesem Betriebsbereich sind gewöhnliche SRAM-Implementierungen auf Basis konventioneller 6T-Speicherzellen, wie in Abbildung 9.1a dargestellt, nicht zuverlässig funktionsfähig, da widersprechende Anforderungen an die statischen Störabstände für Lesen und Schreiben SNM_{read} , SNM_{write} [133] auftreten [18, 22, 34]. Während für ein zerstörungsfreies Auslesen einer Zelle eine Begrenzung des maximalen Auslesestroms notwendig ist, muss für das Beschreiben ein hinreichend großer Umschaltstrom gewährleistet sein. Dies führt zu einem Konflikt bezüglich der Dimensionierung der Passtransistoren M5, M6: Für ein zerstörungsfreies Auslesen müssten diese mit möglichst kleinem Verhältnis W/L ausgelegt werden. Demgegenüber erfordert eine zuverlässige Beschreibbarkeit einen großen Wert für W/L . Bei nomineller Versorgungsspannung lässt sich hier ein Kompromiss finden, der beiden Anforderungen genügt, da das innere Latch-Element der 6T-Zelle M1 – M4 eine hohe Treiberstärke besitzt, so dass die Gefahr des zerstörenden Auslesens relativ gering ist. Mit sinkender Versorgungsspannung nimmt diese Treiberstärke jedoch ab, so dass die Dimensionierung der Passtransistoren immer kritischer wird. Im Subschwelligbereich ist selbst bei optimaler Dimensionierung ein zerstörungsfreies Auslesen nicht mehr sichergestellt. Eine robuste Funktion der SRAM-Zellen wird durch Prozess- und Temperaturschwankungen zusätzlich erschwert [19, 37, 38]. Die Grenze der Versorgungsspannung, unterhalb der ein 6T-basierter SRAM nicht mehr praktikabel ist, liegt je nach Technologie etwa im

Bereich 0,5 V bis 0,7 V [18, 162]. Bei kleineren Spannungen sind daher Speicherzellen mit zusätzlichen Transistoren erforderlich, um die Probleme bezüglich Stabilität und Variabilität zu lösen. In der Regel werden dabei Ansätze gewählt, die durch zusätzliche Lesebuffer externe Verbindungen mit dem inneren Latch-Element während des Auslesens vermeiden, so dass Auslesen und Beschreiben für Betrieb und Zelldimensionierung entkoppelt sind, z.B. [18, 34, 79].

Für das hier behandelte Subschwelle-SoC wurde ein SRAM auf Basis einer 9T-Speicherzelle realisiert. Der Schaltplan dieser Speicherzelle ist in Abbildung 9.1b einschließlich Dimensionierungsangaben und Transistortypen dargestellt. Die Topologie der Zelle ist an [94] angelehnt, es wurden jedoch Modifikationen zur Verbesserung der Eigenschaften bei niedrigen Versorgungsspannungen vorgenommen. Die Transistoren M1 – M6 entsprechen der konventionellen 6T-Speicherzelle, während mit M7 – M9 ein zusätzlicher Puffer für differenzielles Auslesen aufgebaut ist. Die Passtransistoren M5, M6 werden nur für den Schreibvorgang genutzt. Es existieren daher separate Wortleitungen für Schreibzugriffe (\overline{WWL}) und Lesezugriffe (\overline{RWL}). Die Dimensionierung der Transistorgröße erfolgte durch ein Mehrzieloptimierungsverfahren für eine Versorgungsspannung von 300 mV [8, 12]. Es kommen Transistortypen verschiedener Schwellenspannungen zum Einsatz, um weiteres Optimierungspotential auszuschöpfen. Das innere Latch-Element M1 – M4 ist aus HVT-Transistoren mit einer Gatelänge von 170 nm aufgebaut, um die hier besonders wichtige statische Verlustleistung zu reduzieren. Die vergrößerte Gatelänge wirkt sich zudem positiv auf das I_{on}/I_{off} -Verhältnis aus [1] und die Streuung der Transistorparameter nimmt durch die vergrößerte Gatefläche ab (vgl. Gleichung 2.21). Zur Optimierung der Beschreibbarkeit ist das innere Latch-Element mit einer virtuellen Spannungsversorgung $V_{DD, virt}$ verbunden, die während des Schreibvorgangs deaktiviert wird, so dass sich praktisch $SNM_{write} = -V_{DD}$ ergibt. Gemäß Monte-Carlo-Simulationen weisen in der verwendeten 65-nm-Technologie die PMOS-Transistoren bei gleichen Abmessungen eine deutlich geringere Streuung des Drain-Stromes als die NMOS-Transistoren auf. Im Fall der Passtransistoren M5, M6 ergibt sich bei einer Versorgungsspannung von 300 mV bei vergleichbarem Drain-Strom eine Reduktion des Variationskoeffizienten σ/μ von 42 % durch die Verwendung eines PMOS-Transistors anstelle eines NMOS-Transistors. Im Interesse geringer Leckströme sind diese Transistoren als PMOS-HVT-Typ ausgeführt. Die Transistoren M7 bis M9 stellen ein differenzielles Lesesignal auf den Bitleitungen für ein zerstörungsfreies Auslesen zur Verfügung. Diese Transistoren sind zur Beschleunigung des Lesevorgangs vom PMOS-LVT-Typ. Gegenüber NMOS-LVT-Transistoren ergibt sich eine Verringerung des Variationskoeffizienten um 22 % sowie eine absolute Abnahme des Drain-Stroms von 78 %. Dieser Strom reicht jedoch aus, um eine der Schreibgeschwindigkeit



(a) konventionelle 6T-Zelle



(b) 9T-Zelle

Abbildung 9.1: Schaltpläne der SRAM-Speicherzellen

ähnliche Lesegeschwindigkeit zu realisieren, so dass der Vorteil des verringerten Variationskoeffizienten überwiegt. Auf die statische Verlustleistung hat die Wahl des LVT-Transistortyps an dieser Stelle nur einen untergeordneten Einfluss, da die Leckströme über die Bitleitungen durch periphere Pull-Down-Transistoren vom SVT-Typ im Leseverstärker begrenzt werden.

Das Subschwelle-SoC enthält jeweils zwei 9T-SRAM-Blöcke der Größen 16×32 und 64×32 bit, so dass jedes Exemplar des Bausteins 5120 9T-Speicherzellen enthält. Die relativ geringe Speichergröße ist unter anderem durch die hierfür zur Verfügung stehende Chipfläche begründet. Zudem ergeben sich beim Auslesen um so ausgeprägtere Spannungspegel auf den Bitleitungen, je weniger Speicherzellen an diesen angeschlossen sind. Daher ist bei einer geringeren Speichergröße mit einer höheren Robustheit zu rechnen. Die umgesetzten Konzepte lassen sich jedoch ungeachtet dessen auch auf größere Speicherblöcke anwenden. Die Speicherzelle besitzt eine Nettofläche von $2,86 \mu\text{m}^2$. Einschließlich peripherer Schaltkreise wie Adressdekoder sowie Schreib- und Leseverstärker ergibt sich eine effektive Fläche von $4,61 \mu\text{m}^2$ pro Bit. Diese Werte sind vergleichbar mit vielen aktuellen Subschwelle-SRAM-Entwürfen, z.B. [99, 150]. Der gegenüber konventionellem 6T-SRAM deutlich höhere Flächenbedarf ist zur Unterbringung der zusätzlichen Transistoren und größeren Gateflächen jedoch unvermeidbar, um die Herausforderungen bezüglich Stabilität und Variabilität zu lösen.

Bei der Anbindung der 9T-SRAM-Blöcke an das Subschwelle-SoC sind zwei wesentliche Anforderungen zu erfüllen. Einerseits soll eine detaillierte Charakterisierung der einzelnen Komponenten des Speichers bezüglich zeitlichem und energetischem Verhalten möglich sein, andererseits der Speicher jedoch auch durch die Subschwelleprozessoren beispielsweise als Scratchpad-Speicher genutzt werden können. Diese Anforderungen werden dadurch erfüllt, dass der Speicher über Pegelwandler sowie Daten- und Kontrollregister als Hardwareerweiterung an das CoreVA-System des SoC angebunden ist. Die Ablaufsteuerung der Lese- und Schreibzugriffe erfolgt dabei programmorientiert durch den jeweils aktiven Prozessorkern des SoC. Die Steuersequenz eines Lesevorgangs besteht aus folgenden Schritten:

1. Setzen des Bits zum Vorladen der Bitleitungen im Kontrollregister und Schreiben der Speicheradresse in das Adressregister
2. Löschen des Bits zum Vorladen der Bitleitungen und Setzen des Bits zum Aktivieren der ausgewählten Lese-Wortleitung im Kontrollregister
3. Setzen des Bits zum Aktivieren des Leseverstärkers im Kontrollregister

-
4. Auslesen des Datenregisters
 5. Löschen der Bits zum Aktivieren der Lese-Wortleitung und des Leseverstärkers im Kontrollregister

Ein Schreibvorgang wird durch folgende Steuersequenz ausgeführt:

1. Schreiben der Speicheradresse in das Adressregister und der zu schreibenden Daten in das Datenregister
2. Setzen des Bits zum Aktivieren der ausgewählten Schreib-Wortleitung im Kontrollregister
3. Löschen des Bits zum Aktivieren der ausgewählten Schreib-Wortleitung im Kontrollregister

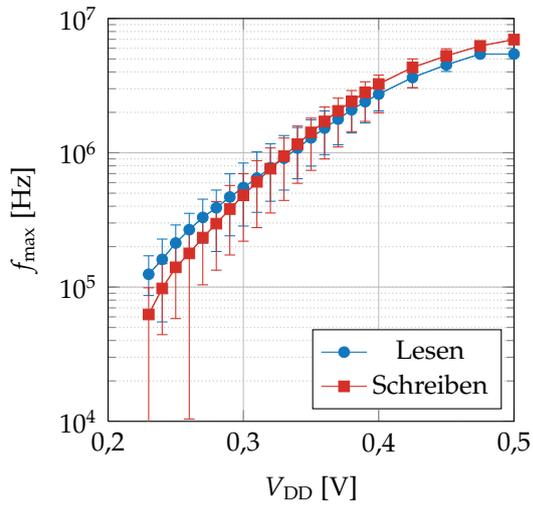
Das zeitliche Verhalten des Subschwellspeichers wird durch das Einfügen von Warteschleifen zwischen den einzelnen Teilschritten berücksichtigt. Zur Ermittlung der minimal benötigten Verzögerungen wird ein Lese- oder Schreibzugriff probeweise ausgeführt und das Resultat anschließend auf Korrektheit überprüft. Bei einem Lesetest wird dazu zunächst ein Datum mit als sicher anzusehenden Verzögerungen in den Speicher geschrieben und dann ein Lesezugriff mit der zu testenden Verzögerungskonfiguration ausgeführt. Bei einem Schreibtest wird das Datum mit der zu testenden Verzögerungskonfiguration in den Speicher geschrieben und dann durch einen Lesezugriff mit sicheren Verzögerungszeiten wieder ausgelesen. Es wird jeweils verglichen, ob geschriebenes und gelesenes Datum übereinstimmen, wobei nacheinander verschiedene Bitmuster getestet werden. Die Bestimmung der minimalen Verzögerungszeiten erfolgt ähnlich der Charakterisierung der Prozessorkerne nach Abschnitt 8.1 durch ein Bisektionsverfahren. Die Auflösung der zeitlichen Charakterisierung ist durch die Taktfrequenz des Prozessors, auf dem das Testverfahren ausgeführt wird, begrenzt, so dass sich nur etwa unterhalb von 500 mV genaue Angaben ermitteln lassen. Da dieser Spannungsbereich jedoch von primärem Interesse für den Betrieb des Subschwellspeichers ist, ist diese Einschränkung akzeptabel. Der Speicher ist mit den schnellstmöglichen Schreib- und Lesesequenzen jedoch bis hinauf zur nominellen Versorgungsspannung von 1,2 V korrekt funktionsfähig. Die Bestimmung des Energiebedarfs erfolgt bei jeweils maximaler Lese- bzw. Schreibgeschwindigkeit und abwechselnd invertiertem Bitmuster zur Erzeugung maximaler Schaltaktivität. Zur Ermittlung der Häufigkeit von Bitfehlern wird bei ausreichend hoher (sicherer) Spannung ein Datum in den Speicher geschrieben und anschließend die Spannung für 10 Sekunden auf einen niedrigeren Wert abgesenkt. Anschließend

wird die Spannung wieder angehoben, das Datum ausgelesen und auf Korrektheit überprüft.

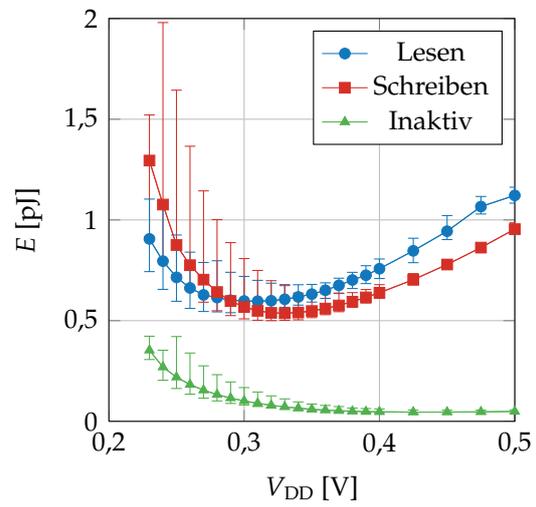
Die nachfolgend behandelten Messergebnisse beziehen sich der Einfachheit halber, mit Ausnahme der Bitfehler-Betrachtung, auf den 64×32 -Speicherblock. Der 16×32 -Block zeigt prinzipiell vergleichbare Eigenschaften, wobei erwartungsgemäß eine etwas höhere Geschwindigkeit und Energieaufnahme zu beobachten ist. Die maximale Frequenz für Schreib- und Lesevorgänge ist in Abbildung 9.2a und der zugehörige Energiebedarf pro 32-bit-Operation in Abbildung 9.2b dargestellt. Der minimale Energiebedarf von 529 fJ bis 620 fJ, gemittelt für Schreib- und Lesezugriffe, tritt je nach Exemplar im Bereich 300 mV bis 350 mV mit einer Häufung bei 320 mV auf, und damit nahe des Energieminimums des Subschwellprozessors (vgl. Kapitel 8.3). Bei 320 mV beträgt die maximale Frequenz 330 kHz bis 1016 kHz und die aktive Leistungsaufnahme 246 nW bis 605 nW. Im inaktiven Zustand beträgt die durchschnittliche statische Verlustleistung pro Bit 17,8 pW bei 300 mV und 25,4 pW bei 400 mV.

Abbildung 9.2c zeigt die relative Häufigkeit von Bitfehlern bezogen auf die Gesamtheit der Bits aller getesteten Exemplare im inaktiven Zustand bei abgesenkter Versorgungsspannung. Der erste Bitfehler tritt bei einer Spannung von 273 mV auf, während die nächsten Bitfehler nach einer Lücke bei 242 mV auftreten. Bei 171 mV weisen 1 % der Speicherzellen Bitfehler auf. Geht man von der oben genannten minimalen Versorgungsspannung von 500 mV bis 700 mV für eine 6T-Speicherzelle aus, so bietet die 9T-Speicherzelle eine Verbesserung der minimalen Versorgungsspannung von 180 mV bis 380 mV gegenüber einer 6T-Speicherzelle, wenn man davon ausgeht, dass im Energieminimum bei 320 mV auch bei einem größeren Speicherblock keine Bitfehler auftreten. Eine Verbesserung der minimalen Versorgungsspannung kann durch den Einsatz fehlerkorrigierender Codes zur Rekonstruktion verfälschter Bits erzielt werden.

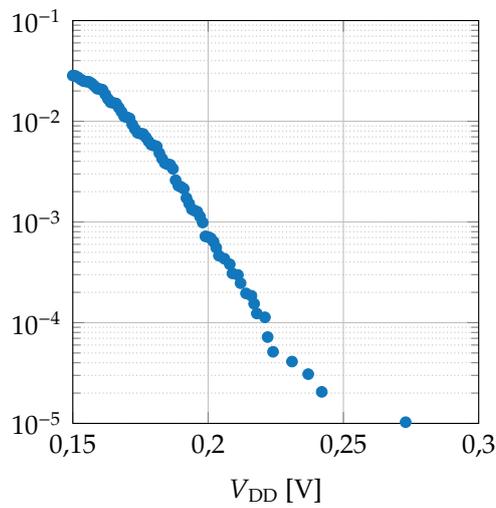
Es existiert bereits eine größere Zahl an veröffentlichten Subschwell-SRAM-Implementierungen, in denen unterschiedliche Ansätze zum Lösen der oben beschriebenen Herausforderungen eingesetzt werden. Tabelle 9.1 stellt die Eigenschaften des hier präsentierten Subschwell-SRAMs sechs ähnlichen, repräsentativen Implementierungen gegenüber. Weitere Realisierungen finden sich etwa in [85, 86, 105, 138, 140, 151–153, 155]. Es fällt zunächst auf, dass einige Implementierungen aus der Tabelle eine relativ hohe Taktfrequenz auf Kosten höherer Verlustleistung und Energie erreichen. Das kann z.B. auf unterschiedliche eingesetzte Transistortypen (LP/GP oder unterschiedliche Schwellspannungen) zurückzuführen sein. Eine



(a) Maximale Frequenz für Lese- und Schreibzugriffe



(b) Energieaufnahme pro Takt bei jeweils maximaler Frequenz



(c) Relative Häufigkeit von Bithaltefehlern

Abbildung 9.2: Gemessene Eigenschaften des Subschwells-SRAMs

Sonderposition nimmt [173] ein, der einzige aufgeführte Speicher auf Basis einer (modifizierten) 6T-Bitzelle. Zwar erreicht diese Implementierung interessanterweise den geringsten Wert bezüglich der minimalen Versorgungsspannung, allerdings wird dies durch die größte effektive Zellfläche, die höchste Energie pro Zugriff, die zweithöchste statische Verlustleistung und die geringste Taktfrequenz im Energieminimum erkauft. Zudem setzt die Implementierung eine Regelung der Substratspannung zur Angleichung von NMOS- und PMOS-Leitfähigkeit ein. Dies unterstützt die obige Aussage, dass eine 6T-Bitzelle im Subschwellebereich nicht praktikabel ist, also bezüglich der meisten Eigenschaften Nachteile gegenüber anderen Implementierungen aufweist.

Der hier präsentierte Subschwelle-SRAM besitzt im Vergleich die zweitkleinste reale und effektive Zellfläche. Dabei ist zu berücksichtigen, dass die Implementierung eine relativ kleine, nämlich die zweitkleinste Blockgröße und dadurch einen relativ großen Flächenbedarf durch periphere Schaltkreise (Leseverstärker, Adressdekoder) aufweist. Die effektive Zellfläche ließe sich daher durch eine größere Blockkapazität reduzieren. Bezüglich des minimalen Energiebedarfs pro Bit und Zugriff weist der Subschwelle-SRAM den zweitbesten Wert aus dem Vergleichsfeld auf, wobei allerdings zu bedenken ist, dass in drei Veröffentlichungen anstelle von Daten über den minimalen Energiebedarf lediglich Werte bei einer bestimmten Versorgungsspannung angegeben sind. Die Versorgungsspannung, bei der dieser minimale Energiebedarf erzielt wird, ist geringer als bei den übrigen Implementierungen. Dennoch ist die Taktfrequenz des Subschwelle-SRAMs bei dieser Spannung höher als bei der Hälfte der Referenzen. Die statische Verlustleistung bei 300 mV bewegt sich etwa im Mittelfeld der SRAM-Realisierungen, ebenso wie die minimale Versorgungsspannung.

9.1 Zusammenfassung

In diesem Kapitel wurde ein in Kooperation mit der Fachgruppe Nanoelektronik der Universität Oslo entwickelter Subschwelle-SRAM betrachtet, der auf einer 9T-Zelle basiert. Der reale Flächenbedarf dieser SRAM-Zelle beträgt $2,83 \mu\text{m}^2$ und ist im Vergleich mit ähnlichen Implementierungen gering. Für einen SRAM-Block der Größe 64×32 ergibt sich eine minimale Energieaufnahme von 17,8 fJ pro Bit und Zugriff bei 320 mV und 761 kHz. In einem Vergleich mit sechs weiteren Implementierungen ist dies der zweitniedrigste Energiebedarf. Der 9T-SRAM besitzt eine minimale

Referenzen	Diese Arbeit	[150]	[139, 140]	[18, 21]	[81, 82]	[79, 80]	[171, 173]
Technologie	65 nm	65 nm	65 nm	65 nm	130 nm	130 nm	130 nm
Zelltyp	9T	9T	8T	10T	8T	10T	6T
Blockgröße	64 × 32	64 × 72	64 × 128	256 × 128	512 × 64	1024 × 256	16 × 16
Zellfläche, real [μm^2]	2,83	1,77			6,36	7,50	4,79
Zellfläche, effektiv [μm^2]	4,61	3,04			9,34		13,96
Spannung für min. Energie [mV]	320	500	400	400 ²	400 ²	400 ²	340
Frequenz [kHz]	761	5000 ¹	500 ¹	475	6700	1900	400 ¹
Energie [f/bit/Zugriff]	17,8	62,5	<100	14,1 ¹	25,2	32,9 ¹	138
statische Verlustleistung bei 300 mV [pW/bit]	17,8	40,7 ¹	12,2 ¹	11,8 ¹	64,1 ¹	7,3 ¹	54,9
Minimale Versorgungsspannung [mV]	273	275	250	380	230	200	193

¹ Aus Diagrammen und/oder Berechnungen abgeleiteter, ungefährender Wert

² In der jeweiligen Veröffentlichung gewählter Betriebspunkt statt Energieminimum

Tabelle 9.1: Vergleich des Subschwellig-SRAMs mit ähnlichen Implementierungen

Versorgungsspannung von 273 mV und eine statische Verlustleistung von 17,8 pW pro Bit bei 300 mV. Diese Werte sind vergleichbar mit ähnlichen Implementierungen. Eine weitere Verbesserung der Energie- und Verlustleistungswerte lässt sich durch eine Erhöhung der Blockgröße erzielen, da sich hierdurch der Anteil der peripheren Schaltkreise (Adressdekoder, Leseverstärker usw.) reduzieren lässt.

Das nachfolgende Kapitel befasst sich mit einer anderen Teilkomponente der prototypischen ASIC-Realisierung, der Takt- und Energieverwaltung des Subschwellprozessors. Außerdem wird eine adaptive Spannungs- und Frequenzsteuerung vorgestellt, die die dynamische Kompensation von Prozess- und Betriebsbedingungen erlaubt.

10 Takt- und Energieverwaltung

Zur Minimierung des Energiebedarfs einer integrierten Digitalschaltung ist es vorteilhaft, die Versorgungsspannung und die Taktfrequenz zur Laufzeit durch DVFS so niedrig wie möglich zu wählen. Der Einsatz von DVFS stellt höhere Anforderungen an Hilfsschaltungen als bei konventionellem Schaltungsbetrieb mit konstanter Frequenz und Spannung. So müssen Spannungsregler und Frequenzgenerator in der Lage sein, ihre Ausgangssignale dynamisch den Anforderungen anzupassen, wobei eine Schaltungsfehlfunktion durch ein ungültiges Taktsignal, durch eine unzulässige Versorgungsspannung oder durch eine unzulässige Kombination aus Taktfrequenz und Versorgungsspannung während des gesamten Anpassungsvorgangs verhindert werden muss. Eine weitere Reduktion des Energiebedarfs ist durch Bereitschaftsmodi möglich, indem während Phasen ohne Rechenbedarf Schaltungsteile durch Abschalten des Taktsignals (*Clock Gating*) oder der Spannungsversorgung (*Power Gating*) deaktiviert werden. Hierbei muss sicher gestellt sein, dass sich die Schaltung nach dem Verlassen eines Bereitschaftsmodus in einem gültigen Zustand befindet. In Abschnitt 10.1 wird ein Modul vorgestellt, das das CoreVA-Prozessorsystem um solche Möglichkeiten zur dynamischen Energieeinsparung bzw. Performanzanpassung erweitert.

Eine besondere Herausforderung bei der Steuerung von Versorgungsspannung und Taktfrequenz ergibt sich beim Subschwellebetrieb einer Schaltung. Durch die hohe Empfindlichkeit der Schaltungseigenschaften gegenüber Prozess- und Betriebsparameterschwankungen in diesem Bereich kommt es zu einer starken Variation der maximalen Taktfrequenz bei einer gegebenen Spannung. Die für DVFS verwendeten Spannungs-Frequenz-Kombinationen müssen also in diesem Fall stets die ungünstigste Parameterkonstellation abdecken d.h. große Sicherheitsreserven einkalkulieren. Dies wirkt sich jedoch nachteilig auf die Energieeffizienz der Schaltung aus. Aus diesem Grund wurde eine adaptive Spannungs- und Frequenzsteuerung entwickelt, die eine Kompensation solcher Parameterschwankungen zur Laufzeit ermöglicht. Dieses System wird in Abschnitt 10.2 erläutert.

10.1 Hardware-Erweiterung PPMU

Zur Realisierung einer DVFS-Unterstützung für das CoreVA-System wurde ein Takt- und Energieverwaltungsmodul (*Performance and Power Management Unit*, PPMU) entwickelt [185]. Die PPMU kann wahlweise als interne oder externe Hardware-Erweiterung an das CoreVA-System angebunden werden und ist nicht an die CoreVA-ULP-Systemumgebung oder das DB-CoreVA-Modul gebunden. Daher ist der Kontrollpfad plattformunabhängig, so dass eine einfache Portierung auf verschiedene Zielplattformen möglich ist. Im Rahmen dieser Arbeit wurde eine Unterstützung für die FPGA-Familien Xilinx Spartan-6 [165] und Virtex-II [170] implementiert.

Abbildung 10.1 zeigt vereinfacht den schematischen Aufbau der PPMU. Die Erzeugung von Taktsignalen erfolgt plattformspezifisch durch entsprechende dedizierte FPGA-Ressourcen. In der Virtex-II-Implementierung stehen hierfür zwei statisch konfigurierte digitale Taktmanager (engl. *Digital Clock Manager*, DCM) [170] zur Verfügung. Für die Spartan-6-Implementierung kommen wahlweise ein oder zwei DCMs oder eine Phasenregelschleife (engl. *Phase-Locked Loop*, PLL) [166] zum Einsatz. Für FPGAs mit einem größeren Umfang nutzbarer Taktressourcen wie etwa Xilinx Virtex-5 [169] ist auch die Verwendung mehrerer PLLs und/oder DCMs denkbar. Die Parameter der Frequenzsynthese – Multiplikator und Divisor – lassen sich zur Laufzeit modifizieren, was im Falle der DCMs durch eine SPI¹-Schnittstelle [166], im Falle der PLL durch ein spezielles Rekonfigurations-Protokoll [87] erfolgt. Die Nutzung der PLL bietet gegenüber den DCMs den Vorteil eines größeren möglichen Frequenzbereichs, geringerer *Jitter*-Werte² sowie der simultanen Erzeugung mehrerer Takte mit feinschrittigen Phasenverschiebungen [165–167]. Die letztgenannte Möglichkeit wird im Falle einer ASIC-Realisierung zur Kompensation von Signallaufzeiten durch Pad-Zellen und Leiterbahnen auf dem externen Systembus des CoreVA-Systems genutzt. Nach dem Rekonfigurieren der DCMs bzw. PLLs ist ein erneutes Einrasten derer internen Regelschleifen erforderlich, währenddessen jeweils kein stabiles Taktsignal verfügbar ist und das CoreVA-System daher angehalten werden muss. Der simultane Einsatz mehrerer DCMs oder PLLs gestattet dagegen das schnelle Umschalten zwischen zwei vorkonfigurierten Frequenzen innerhalb einer Taktperiode durch einen Taktmultiplexer, wodurch derartige Verzögerungen vermieden werden. Eine Rekonfiguration der jeweils nicht durch den Taktmultiplexer

¹ *Serial Peripheral Interface*

² Zufällige sowie deterministische Abweichungen zwischen idealer und realer Periodendauer und/oder Phasenlage

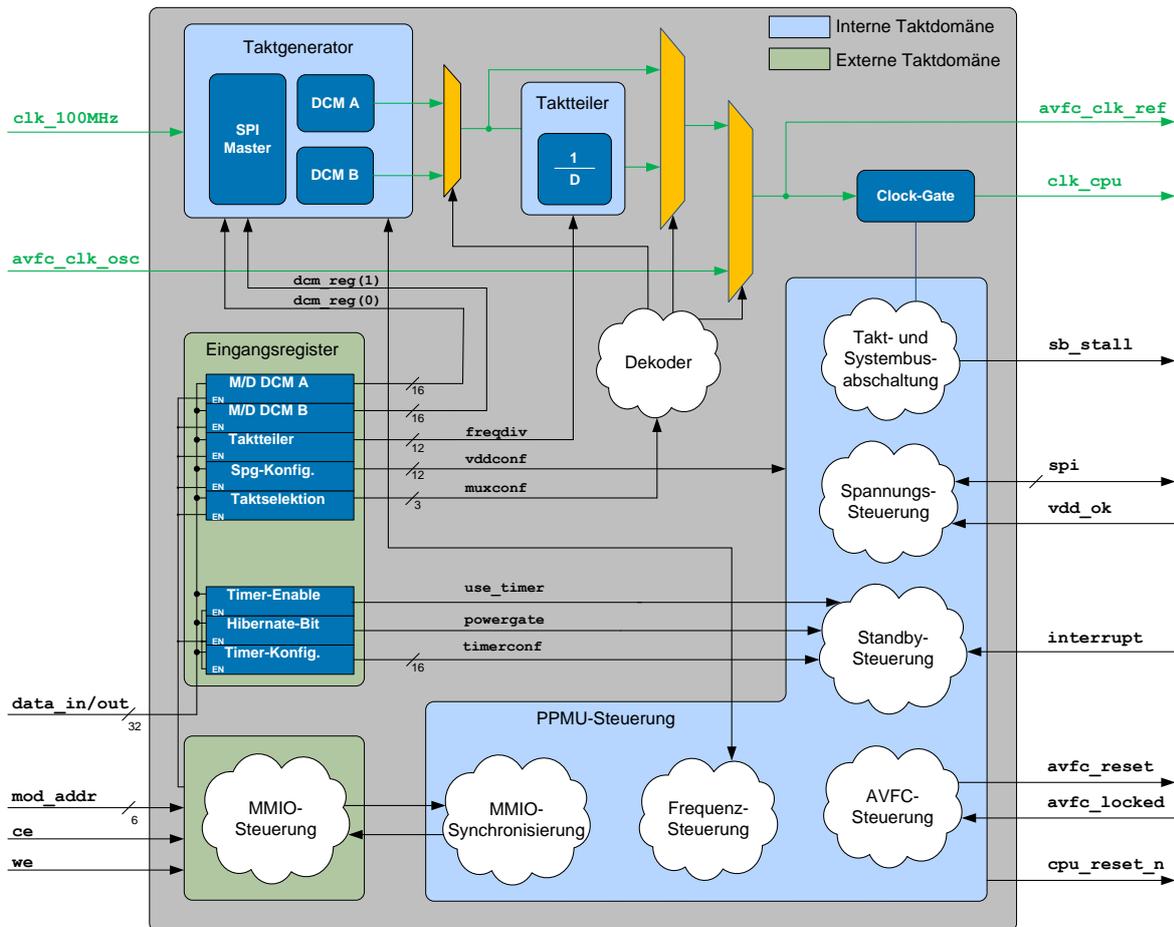


Abbildung 10.1: Blockschaltbild der PPMU

selektierten DCM oder PLL erfordert kein Anhalten des CoreVA-Systems, wodurch derartige Rekonfigurationen im Hintergrund durchgeführt werden können. Das Eingangstaktsignal der DCMs oder PLLs kann von einer beliebigen geeigneten Taktquelle stammen. Im Falle des DB-CoreVA kommt hierfür ein dedizierter 100-MHz-Quarzoszillator zum Einsatz.

Da die DCMs und PLLs der verwendeten FPGA-Familien bezüglich der erzeugten Frequenzen Minimalvorgaben von einigen MHz aufweisen, kann dem vorstehend beschriebenen Basistaktgenerator durch einen weiteren Taktmultiplexer ein Taktteiler nachgeschaltet werden, um niedrigere Taktfrequenzen für den Subschwellprozessor generieren zu können. Die zählerbasierte Implementierung des Taktteilers ist in Abbildung 10.2 dargestellt. Der Zähler wird mit jeder steigenden Flanke des Eingangstaktsignals inkrementiert und mit dem eingestellten Schwellwert dv verglichen. Bei Erreichen des Schwellwertes wird im nächsten Takt der Zähler zurück gesetzt

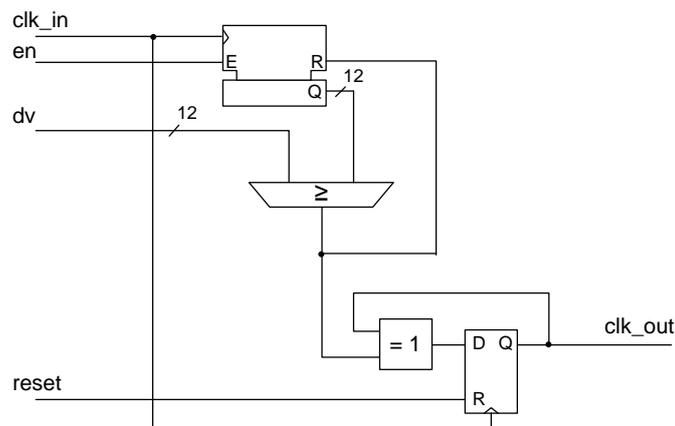


Abbildung 10.2: Aufbau des zählerbasierten Taktteilers

und der Zustand des Ausgangstaktes invertiert. Das Teilungsverhältnis beträgt damit $1 : 2(dv + 1)$. Da das Ausgangstaktsignal direkt durch ein Flipflop generiert wird, können keine Störimpulse auftreten. Bei der Verwendung des Taktteilers ist allerdings aufgrund der Implementierung unter Nutzung der normalen FPGA-Logikressourcen mit verschlechterten Jitter-Eigenschaften zu rechnen [166]. Dies ist angesichts der in diesem Fall niedrigen erzeugten Taktfrequenzen jedoch unproblematisch. Über einen weiteren Taktmultiplexer besteht zusätzlich die Möglichkeit ein externes Taktsignal einzuspeisen. Dies kann etwa das Ringoszillator-Signal der in Abschnitt 10.2 vorgestellten adaptiven Spannungs- und Frequenzsteuerung sein.

Ein Clock-Gate sorgt dafür, dass während der Rekonfiguration das Ausgangstaktsignal deaktiviert ist, da es wie oben erwähnt andernfalls durch einen instabilen Takt zu einem Fehlverhalten der angeschlossenen Schaltung kommen kann. Die Kontrolllogik der PPMU sorgt währenddessen auch dafür, dass sich der CoreVA-Systembus sowie die daran angeschlossenen Einheiten in einem konsistenten Wartezustand befinden.

Neben den Möglichkeiten zur dynamischen Takterzeugung bietet die PPMU auch Mechanismen zur dynamischen Spannungsskalierung. Dazu ist eine SPI-Schnittstelle vorhanden, über die ein geeigneter externer Spannungsregler konfiguriert werden kann. Im Falle der DB-CoreVA-Realisierung erfolgt die Einstellung des variablen Spannungsreglers durch ein digital konfigurierbares Potentiometer. Dieses Konzept ist sowohl auf Linearregler als auch auf Schaltregler anwendbar. Während einer aktiven Spannungsänderung wird das Ausgangstaktsignal deaktiviert, um Schaltungsfehlfunktionen in Folge des Reglerverhaltens zu vermeiden. Ein Statussignal

vom Spannungsregler teilt der PPMU mit, ob eine stabile Versorgungsspannung verfügbar ist.

Zur Energieeinsparung während Phasen ohne Rechenbedarf stellt die PPMU Bereitschaftsmodi durch Clock-Gating und Power-Gating zur Verfügung. Im ersten Bereitschaftsmodus (*Sleep-Modus*) wird das Ausgangstaktsignal durch das vorhandene Clock-Gate deaktiviert, um alle dynamischen Verluste der angeschlossenen Schaltung zu eliminieren. Da zum Reaktivieren der Schaltung lediglich das Taktsignal wieder freigegeben werden muss, ist ein schnelles Aufwachen aus dem Sleep-Modus möglich, der Prozessor setzt den Betrieb dann unmittelbar an derjenigen Programmstelle fort, an der er bei Eintritt in den Sleep-Modus zuvor angehalten wurde. Im zweiten Bereitschaftsmodus (*Hibernate-Modus*) wird neben dem Taktsignal zusätzlich die Versorgungsspannung deaktiviert, so dass auch statische Verluste praktisch nicht mehr vorhanden sind. In diesem Fall treten längeren Aufwachzeiten auf, da zunächst die Spannungsversorgung stabil wiederhergestellt sein muss. Da CoreVA ULP keine integrierte Power-Gating-Funktionalität besitzt, wird das Abschalten in diesem Fall durch eine entsprechende Konfiguration des Spannungsreglers für eine Versorgungsspannung von 0 V erreicht. Daher wird die Aufwachzeit in diesem Fall maßgeblich durch das Einschalt- bzw. Regelverhalten des Spannungsreglers bestimmt. Der Prozessor muss nach dem Verlassen des Hibernate-Modus die Programmausführung ab Programmbeginn starten. Dazu wird durch die PPMU kurzzeitig das Rücksetzsignal des Prozessors aktiviert. Durch eine geeignete Startsequenz kann ggf. der vorherige Prozessorzustand programm basiert rekonstruiert werden. Alternativ ist eine Absenkung der Versorgungsspannung auf einen geeigneten Wert größer 0 V möglich, so dass der Prozessorzustand erhalten bleibt (*Retention-Modus*). Das Aufwachen kann innerhalb der PPMU durch zwei Ereignisse ausgelöst werden. Einerseits ist ein Zeitgeber vorhanden, durch den ein Bereitschaftsmodus automatisch nach Ablauf einer bestimmten Zeitspanne beendet wird. Alternativ ist die Reaktivierung auch durch eine Unterbrechungsanforderung (*Interrupt*) möglich, um auf beliebige Ereignisse reagieren zu können.

Die Konfiguration aller PPMU-Einstellungen erfolgt durch Kontroll- und Statusregister anwendungsgesteuert über die interne oder externe Hardwareerweiterungsschnittstelle des CoreVA-Systems. Für diese Schnittstelle ist die Verwendung verschiedener Taktquellen möglich, bei Verwendung der PPMU als interne Hardwareerweiterung handelt es hierbei im Normalfall etwa um das von der PPMU selbst erzeugte CPU-Taktsignal. Da die Steuerlogik der PPMU jedoch mit einem kontinuierlichen, definierten Takt betrieben werden muss, ist eine Synchronisation zwischen der MMIO-Schnittstelle und den PPMU-internen Komponenten

erforderlich. Dies wird durch entsprechende Quittierungssignale zwischen dem zentralen PPMU-Zustandsautomaten und einem eingangsseitigen Synchronisations-Zustandsautomaten realisiert.

Die PPMU benötigt nur eine relativ geringe Menge an FPGA-Ressourcen. In der Spartan-6-Implementierung belegt die Kontrolllogik der PPMU 230 *Slices*, dies entspricht etwa 1,5 % der verfügbaren *Slices* des verwendeten FPGA-Typs. Zusätzlich werden vier von 16 Taktpuffern sowie eine von sechs vorhandenen PLLs zur Takterzeugung benötigt.

Während die PPMU die grundlegenden Mechanismen für DVFS und Bereitschaftsmodi auf Hardwareebene bereit stellt, entfällt die übergeordnete Koordination der Performanz- und Energieverwaltung auf die Anwendungssoftware. Dies kann etwa auf Betriebssystem-Ebene erfolgen. Hier kann auch entschieden werden, ob durch die dynamische Bypass-Abschaltung für die jeweils aktive Anwendung eine Verbesserung der Ressourceneffizienz erzielt werden kann. Bei der Modifikation von PPMU- und/oder Bypass-Konfiguration ist die Einhaltung einer festgelegten Reihenfolge bezüglich der Änderung der einzelnen Parameter erforderlich, um eine korrekte Schaltungsfunktion zu garantieren. Insbesondere muss sichergestellt sein, dass die eingestellte Taktperiode zu jedem Zeitpunkt größer als die Verzögerungszeit des durch die Bypass-Konfiguration bestimmten kritischen Pfades bei der jeweiligen Versorgungsspannung ist. Daher muss bei einer Erhöhung der Taktfrequenz zuvor die Versorgungsspannung erhöht und/oder der Bypass deaktiviert werden. Bei einer Verringerung der Taktfrequenz darf die Reduktion der Versorgungsspannung und/oder die Aktivierung des Bypass erst nach der Frequenzänderung erfolgen.

10.2 Adaptive Spannungs- und Frequenzsteuerung

Die hohe Empfindlichkeit von Subschwellschaltungen bezüglich Fertigungs- und Betriebsparametern stellt besondere Herausforderungen nicht nur an den Entwurf sondern auch an den Betrieb dieser Schaltungen. Der klassische Ansatz zur Handhabung von Parameterschwankungen bei integrierten Digitalschaltungen ist die Annahme einer hinreichend geringen Taktfrequenz, mit der das System auch unter ungünstigsten Bedingungen noch fehlerfrei betrieben werden kann. Wenngleich dieser Ansatz prinzipiell auch auf Subschwellschaltungen anwendbar ist, so ist er in diesem Fall doch übermäßig pessimistisch. Gerade wegen der großen Streuung der Schaltungseigenschaften im Subschwellbetrieb sind die meisten gefertigten

Exemplare in der Lage, mit wesentlich größeren Taktfrequenzen zu arbeiten als unter der Annahme der schlechtesten Betriebsbedingungen. Zur Einschätzung dieses Effektes sollen erneut die Messergebnisse des Subschwellsprozessors betrachtet werden: Bei 1,2 V betragen die minimal und maximal gemessenen Frequenzen 85 MHz und 100 MHz. Dies entspricht Abweichungen von $-9,9\%$ und $+6,0\%$ vom Stichprobenmittel. Bei 325 mV dagegen liegen minimale und maximale Frequenzen bei 95 kHz und 160 kHz, entsprechend Abweichungen von $-28,7\%$ und $+20,1\%$ vom Mittelwert. Würden alle Exemplare bei dieser Spannung mit 95 kHz betrieben, so wäre bei den besten Exemplaren die Performanz um den Faktor 1,7 niedriger als notwendig. Neben der Verringerung der Performanz wäre ein weiterer Nachteil eine unnötige Erhöhung der pro Takt benötigten Energie, da mehr statische Verluste während der verlängerten Taktperiode akkumuliert würden.

Eine Möglichkeit zur Kompensation von Parameterschwankungen ist die Verwendung einer variablen Substratvorspannung (auch *Back-Gate*-Steuerung oder *Body Biasing* genannt). Aufgrund des Substrateffekts hängt die Schwellspannung von der Source-Bulk-Spannung U_{SB} eines Transistors ab. Da nach Gleichung 2.3 die Schwellspannung exponentiellen Einfluss auf den Drain-Strom im Subschwellsbereich hat, genügt hierdurch bereits eine relativ geringe Substratvorspannung, um eine signifikante Veränderung der Leitfähigkeit zu erzielen und somit Prozess- oder Betriebsparameterabweichungen zu kompensieren. Geht man davon aus, dass diese Effekte gleichermaßen auf alle Transistoren einer integrierten Schaltung wirken, so genügt eine einheitliche Substratvorspannung für alle NMOS- und/oder PMOS-Transistoren. Eine Vorwärtsvorspannung (d.h. $U_{SB} > 0\text{ V}$ für einen NMOS-Transistor) führt dabei zu einer Reduktion der Schwellspannung und damit zu einer Leitfähigkeitszunahme. Analog ergibt eine Rückwärtsvorspannung eine Abnahme der Leitfähigkeit. Zu beachten ist, dass eine Substratvorspannung zu einem Stromfluss durch die Source-Bulk-Dioden aller Transistoren führt und damit zur statischen Verlustleistung der Schaltung beiträgt. Insbesondere der Einsatz einer Vorwärtsvorspannung ist daher nur eingeschränkt sinnvoll, da in diesem Fall die Source-Bulk-Dioden in Flussrichtung betrieben werden, wodurch zusätzlich die Gefahr eines *Latch-Ups*¹ besteht.

Eine variable Substratvorspannung kann prinzipiell auf unterschiedliche Ziele ausgerichtet sein. Eine Möglichkeit besteht darin, durch Prozessschwankungen verursachte Unterschiede zwischen NMOS- und PMOS-Transistoren bezüglich der Leitfähigkeit zu korrigieren (schneller NMOS, langsamer PMOS, oder umgekehrt)

¹ Ausbildung eines niederohmigen Strompfades zwischen Versorgungsspannung und Masse, wodurch das Halbleiterbauelement ohne geeignete Schutzmaßnahmen thermisch zerstört wird.

[54, 120, 121]. Hierdurch lässt sich ein symmetrisches Schaltverhalten für steigende und fallende Signalflanken sowie maximale Störabstände auch unter ungünstigen Prozessbedingungen erreichen und somit die Robustheit optimieren.

Ein anderer Ansatz zielt auf die Minimierung der statischen Verlustleistung insbesondere im Bereitschaftsmodus ab [69, 76, 110]. Während eine größere Rückwärtsvorspannung die Leitfähigkeit der Transistoren herabsetzt und dadurch den durch Subschwelligströme verursachten Anteil der statischen Verlustleistung reduziert, kommt es durch Band-zu-Band-Tunneln zu einem zunehmenden Stromfluss durch die in Sperrrichtung betriebenen Source-Bulk-Dioden. Es existiert daher eine optimale Rückwärtsvorspannung, bei der die gesamte statische Verlustleistung minimal ist. Dieser Betriebspunkt kann durch geeignete Mess- und Regelschaltungen erreicht werden.

Die Verwendung einer adaptiven Substratvorspannung kann auch der Korrektur der dynamischen Eigenschaften einer Schaltung, insbesondere der maximalen Taktfrequenz unter dem Einfluss von Parameterschwankungen dienen [71, 72, 103, 148, 149]. Die nachfolgende Diskussion bezieht sich auf diese Art der Kompensation von Parameterschwankungen. Dabei kommen PLL- oder DLL¹-Regelstrukturen zum Einsatz, die eine optimale Substratvorspannung zum Erreichen einer vorgegebenen Taktfrequenz generieren. In bisherigen Ansätzen kommen dazu Kopien des kritischen Pfades (Replikatschaltung) zum Einsatz, um dessen zeitliches Verhalten unter den aktuellen Prozess- und Betriebsbedingungen zu ermitteln. Die Regelschleife passt nun die Substratvorspannung so an, dass es zu einer Übereinstimmung zwischen der Verzögerungszeit der Replikatschaltung und der vorgegebenen Taktperiode kommt. Ein Nachteil dieser Herangehensweise ist die Erfordernis eines zusätzlichen Spannungsreglers zur Bereitstellung der Substratvorspannung, wodurch der Schaltungsaufwand und damit die Kosten steigen. Zudem muss der Regler der Substratvorspannung eine hohe Effizienz aufweisen, um den Leistungs- oder Energiebedarf nicht insgesamt zu verschlechtern. Ein weiteres Problem betrifft die Replikatschaltung: Aufgrund von OCV kann es zu einer Abweichung des zeitlichen Verhaltens gegenüber dem echten kritischen Pfad kommen. Dies muss durch entsprechende Sicherheitsreserven berücksichtigt werden, um ein korrektes Regelverhalten zu gewährleisten, etwa indem die Verzögerungszeit der Replikatschaltung durch Hinzufügen von Puffern bewusst verlängert wird. Dies wiederum hat die bereits eingangs erläuterten Nachteile.

¹ *Delay-Locked Loop*

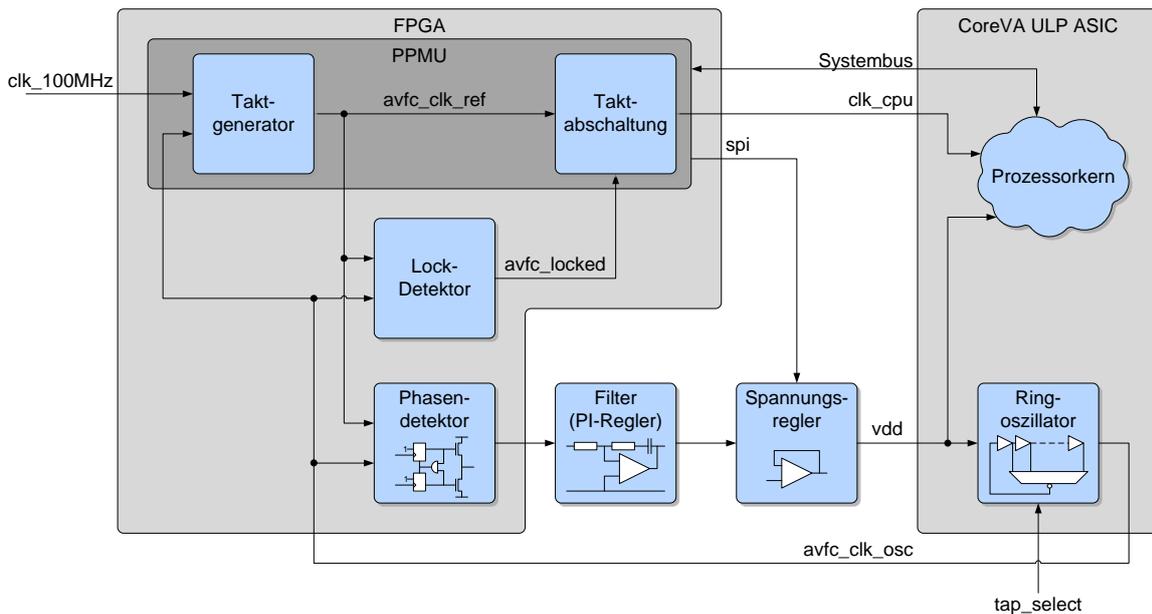


Abbildung 10.3: Blockschaltbild der adaptiven Spannungs- und Frequenzsteuerung

Ein adaptiver Umgang mit Parameterschwankungen ist jedoch auch ohne Substratvorspannung möglich, indem direkt Einfluss auf die Versorgungsspannung genommen wird. Da ein Spannungsregler für die Versorgungsspannung ohnehin vorhanden ist, ist der zusätzliche Aufwand für einen solchen Ansatz gering und beschränkt sich im Wesentlichen auf eine geeignete Hilfsschaltung für den Spannungsregler. Ein solche Lösung wurde für CoreVA ULP durch das AVFC-System realisiert. Es ist in der Lage, Parameterschwankungen durch eine geeignete Anpassung der Spannung oder der Frequenz dynamisch zu kompensieren. In [126, 127] wird ein ähnlicher Ansatz vorgestellt, der allerdings auf den ständigen Betrieb der Schaltung im Energieminimum durch eine entsprechende Spannungsnachführung ausgerichtet ist. Das hier vorgestellte AVFC-System erlaubt dagegen eine flexible Wahl von Spannung oder Taktfrequenz bei adaptiver Anpassung der jeweils freien Größe. In Abbildung 10.3 ist ein Blockschaltbild des Systems sowie die Interaktion mit der PPMU dargestellt. Es besteht aus einem Phasen/Frequenz-Detektor (engl. *Phase/Frequency Detector*, PFD), einem *Lock*-Detektor, einem Schleifenfilter, einem Spannungsregler und einem konfigurierbaren Ringoszillator, der die Replikatschaltung bisheriger Lösungen ersetzt. Die Komponenten bilden zusammen eine PLL-ähnliche Regelschleife.

Der gemeinsam mit dem CoreVA-Prozessor auf dem CoreVA-ULP-ASIC befindliche Ringoszillator dient als Referenz für das zeitliche Verhalten des kritischen

Pfades unter Berücksichtigung von Prozess- und Betriebsbedingungen. Bezüglich der Versorgungsspannung lässt sich der Ringoszillator als (nichtlinearer) VCO (*Voltage Controlled Oscillator*) auffassen. Durch eine digitale Schnittstelle kann die Periodendauer des Oszillators an die Verzögerungszeit des kritischen Pfades angepasst werden, wodurch für jedes Exemplar individuell eine OCV-Kompensation vorgenommen werden kann. Der kritische Pfad hängt zusätzlich von der Bypasskonfiguration des Prozessors ab, so dass für jede Bypasskonfiguration eine andere Ringoszillator-Einstellung ermittelt werden muss. Da lokale Prozessparameterabweichungen statisch sind, muss dies jedoch nur einmalig für jedes Schaltkreis-Exemplar durchgeführt werden, indem zunächst die maximale Taktfrequenz des Prozessors in Abhängigkeit von der Versorgungsspannung bestimmt wird und anschließend eine Ringoszillator-Einstellung gesucht wird, bei der die Oszillationsfrequenz für alle Spannungswerte geringer als die jeweilige maximale Taktfrequenz ist. Effekte wie eine unterschiedliche Temperaturabhängigkeit von Oszillator und Prozessor können durch eine geeignete Sicherheitsreserve berücksichtigt werden. Eine Reserve von 10% der Periodendauer über den gesamten Spannungsbereich stellte sich im praktischen Betrieb des AVFC-Systems als ausreichend heraus. Alterungseffekte können durch eine nachträgliche Modifikation der Ringoszillator-Einstellung kompensiert werden.

Das AVFC-System unterstützt drei verschiedene Betriebsmodi. Im klassischen DVFS-Betrieb ist die AVFC-Funktion deaktiviert und die PPMU ist allein für die Einstellung von Taktfrequenz und Spannung zuständig. In diesem Modus kann das System nicht dynamisch auf sich ändernde Betriebsparameter, d.h. Spannungs- oder Temperaturschwankungen, reagieren.

Im Betriebsmodus mit adaptiver Taktfrequenz wird die Versorgungsspannung weiterhin durch die PPMU vorgegeben, während die Taktfrequenz durch den Ringoszillator erzeugt wird. Wegen der obigen Argumentation ist der Oszillatortakt unmittelbar als Taktquelle für den Prozessor geeignet. Treten nun Abweichungen der Spannung oder der Temperatur von den Nominalwerten auf, so sind hiervon Ringoszillator und Prozessor gleichermaßen betroffen und es ergibt sich automatisch eine entsprechende Anpassung der Taktfrequenz. Die Anwendung muss daher mit variabler Taktfrequenz bzw. Performanz arbeiten können.

Für die Einhaltung fester Performanzvorgaben bietet sich schließlich der Betriebsmodus mit adaptiver Versorgungsspannung an. Dabei wird das Taktsignal anwendungsgesteuert durch die PPMU bereitgestellt, während die Versorgungsspannung adaptiv durch die AVFC-Regelschleife erzeugt wird. Die genaue Funktionsweise ergibt sich

wie folgt: Im eingeschwungenen Zustand stimmen Phasenlage und Frequenz von Referenz- und Oszillatortakt überein. Der Ausgang des PFD ist damit im zeitlichen Mittel inaktiv und die Ausgangsspannung des Schleifenfilters ist konstant. Damit ist auch die Versorgungsspannung für Prozessor und Ringoszillator konstant. Da der Ringoszillator bei korrekter Konfiguration das zeitliche Verhalten des kritischen Pfades widerspiegelt, ist der Wert der Versorgungsspannung optimal bezüglich der gegebenen Referenzfrequenz und der aktuellen PVT-Bedingungen. Tritt nun eine Temperaturänderung oder eine Störung der Versorgungsspannung ein, so ändert sich dadurch das zeitliche Verhalten des Prozessors und des Ringoszillators und damit die Oszillationsfrequenz. Je nach Vorzeichen der Frequenzänderung ergeben sich damit am PFD-Ausgang überwiegend positive oder negative Impulse, die nach Filterung durch das Schleifenfilter am Spannungsreglerausgang zu einer Erhöhung oder Verringerung der Versorgungsspannung und damit auch der Oszillatorfrequenz führen, bis es wieder zu einer Übereinstimmung zwischen Referenz- und Oszillatortakt bezüglich Frequenz und Phasenlage kommt. Die optimale Versorgungsspannung ist damit wieder hergestellt.

Im Kontext der auf dem DB-CoreVA basierenden prototypischen Testumgebung ist das AVFC-System wie folgt implementiert. Der konfigurierbare Ringoszillator ist Teil des CoreVA-ULP-ASICs, um dort als Referenz für das zeitliche Verhalten des Subschwellsprozessors zu dienen. PFD und Lock-Detektor befinden sich gemeinsam mit der PPMU als synthetisierte Digitalschaltungen auf dem Spartan-6-FPGA des DB-CoreVA. Die FPGA-Implementierung wurde gewählt, um ggf. verschiedene Detektionsverfahren im realen Betrieb prototypisch evaluieren zu können. Die analogen Komponenten der Regelschleife – Schleifenfilter und Spannungsregler – befinden sich als Operationsverstärker-Schaltungen auf der DB-CoreVA-Platine.

Abbildung 10.4 zeigt eine Schaltplandarstellung des Phasendetektors und des Lock-Detektors. Als Phasendetektor kommt aufgrund des großen abzudeckenden Frequenzbereichs lediglich ein PFD in Betracht, da durch dessen unendlichen Ziehbereich sichergestellt ist, dass die Regelschleife unter allen Bedingungen auf die korrekte Frequenz einrastet. Zudem ergibt sich gegenüber anderen Phasendetektortypen eine wesentlich kürzere Ziehzeit. PFDs können wahlweise mit Tristate- oder Ladungspumpen-Ausgang realisiert werden. Aufgrund der FPGA-Implementierung bietet sich die erste Variante an, da ein Tristate-Ausgang durch einen *Input/Output Block* (IOB) des FPGAs realisiert werden kann [168]. Ein Ladungspumpen-Ausgang müsste dagegen durch externe Zusatzschaltungen umgesetzt werden. Der Tristate-Ausgang wird durch die Signale *up* und *down* angesteuert. Sind beide Signale inaktiv, so ist der Ausgang hochohmig. Ein Nachteil der Tristate-Variante ist, dass sich bei

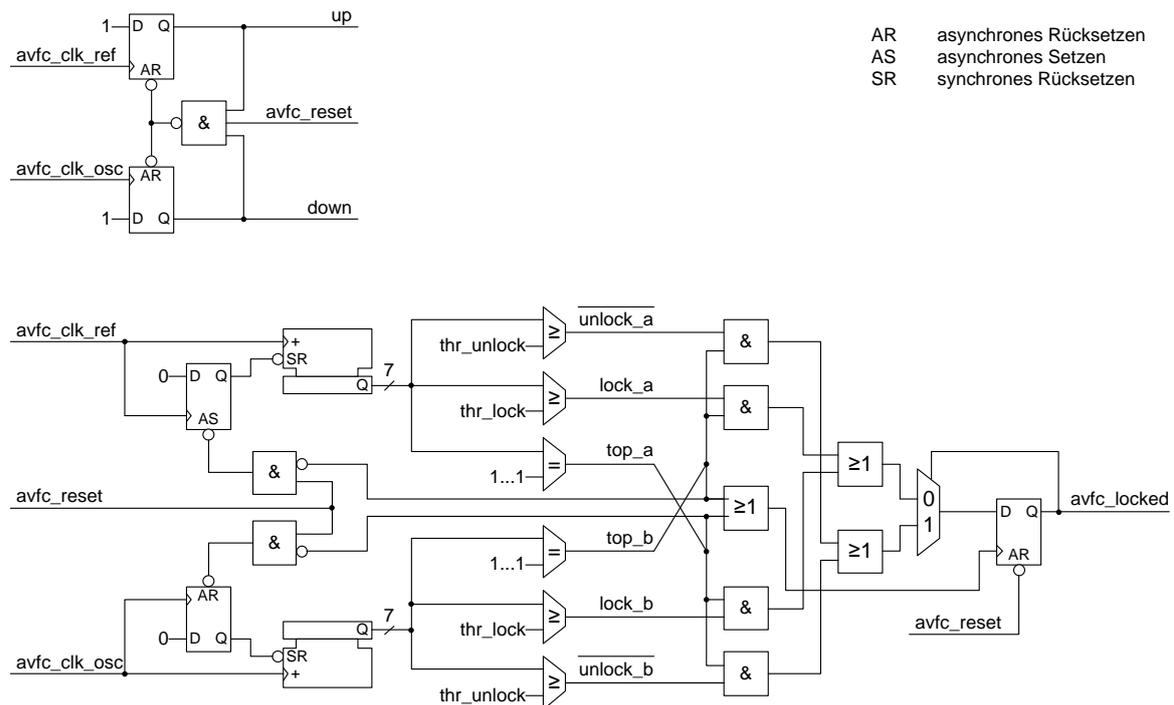


Abbildung 10.4: Schaltplan des Phasendetektors sowie des Lock-Detektors

Verwendung eines passiven Schleifenfilters ein unsymmetrischer Phasendetektorgewinn für Aufwärts- und Abwärts-Pulse ergibt, wenn die Gleichspannung über dem Filter von der Hälfte der Versorgungsspannung des PFD-Ausgangs abweicht. Da die AVFC-Regelschleife diesbezüglich einen weiten Betriebsbereich abdecken muss, kommt ein aktives PI-Filter zum Einsatz, das diesen Nachteil nicht aufweist. Der als nicht-invertierender Verstärker realisierte Spannungsregler liefert eine niedrige Ausgangsimpedanz und sorgt gleichzeitig für eine Anpassung der Ausgangsspannung des Schleifenfilters an den zulässigen Versorgungsspannungsbereich des CoreVA-ULP-ASICs. Der Ringoszillator enthält eine aus Invertern bestehende Verzögerungskette mit 128 Signalabgriffen sowie einen Multiplexerbaum, über den der Rückkopplungspfad durch das digitale Konfigurationswort festgelegt werden kann. Über einen zusätzlichen Steuereingang kann der Oszillator angehalten und in einen definierten Anfangszustand versetzt werden. Durch die generische Auslegung und den weiten Konfigurationsbereich kann der Ringoszillator unverändert in nachfolgenden Schaltungsentwürfen der gleichen Technologie wiederverwendet werden.

Für den Lock-Detektor wird ein rein digitales, zählerbasiertes Verfahren eingesetzt, das den Einrastzustand der Regelschleife anhand der Frequenzdifferenz zwischen

Referenz- und Oszillatortakt erkennt. Ein eventueller Phasenunterschied ist für die Regelung der Versorgungsspannung nicht relevant und wird daher durch den Lock-Detektor nicht berücksichtigt. Die Erkennung des Einrastzustands erfolgt folgendermaßen. Für die beiden zu vergleichenden Taktsignale `avfc_clk_ref` und `avfc_clk_osc` ist jeweils ein Zähler mit wählbarer Breite vorhanden. In der hier eingesetzten Realisierung besitzen die Zähler eine Breite von 7 bit, d.h. die Einrasterkennung erfolgt über ein Intervall von 128 Taktperioden. Erreicht einer der beiden Zähler den Maximalwert, hier 127, so wird geprüft, ob der jeweils andere Zähler einen zuvor definierten Schwellwert überschritten hat, d.h. das entsprechende Taktsignal eine bestimmte Mindestanzahl an Taktperioden im aktuellen Intervall aufweist. Ist dies der Fall, so befindet sich die Regelschleife im eingerasteten Zustand, andernfalls im nicht-eingerasteten Zustand. Der jeweils zum Vergleich herangezogene Schwellwert hängt vom Einrastzustand der Regelschleife im vorangegangenen Prüfintervall ab. Für einen Wechsel vom nicht-eingerasteten in den eingerasteten Zustand muss der Schwellwert `thr_lock` überschritten werden. Dieser Schwellwert ist so gewählt, dass dieser Wechsel erst nach dem vollständigen Abklingen des Einschwingvorgangs der Regelschleife erfolgt. Sobald sich die Regelschleife im eingerasteten Zustand befindet, wird der (toleranter) Schwellwert `thr_unlock` verwendet, so dass ein Wechsel in den nicht-eingerasteten Zustand nur bei größeren Frequenzabweichungen, die die oben genannte zeitliche Reserve überschreiten, stattfindet. Dazu darf also `thr_unlock` um nicht mehr als 10 % unter dem Maximalwert von 127 liegen. Der jeweils nicht überlaufende Zähler wird am Ende eines Prüfintervalls zurück gesetzt. Ein Flipflop dient dabei der Synchronisation des Deaktivierens des Rücksetzsignals in die jeweilige Taktdomäne, um metastabile Zustände beim erneuten Anlaufen des Zählers zu verhindern.

Phasen- und Lock-Detektor können durch das Signal `avfc_reset` zurück gesetzt werden. Ist das Signal durch einen Low-Pegel aktiviert, so befindet sich das AVFC-System in einem Haltezustand. Der Phasendetektor erzeugt in dieser Situation keine Aufwärts- oder Abwärtsimpulse, so dass die erzeugte Versorgungsspannung unter Vernachlässigung eventueller Leckströmen im Schleifenfilter konstant bleibt. Der Lock-Detektor zeigt einen nicht-eingerasteten Zustand an. Der Haltezustand wird insbesondere dann benötigt, wenn im Betriebsmodus mit adaptiver Versorgungsspannung eine Rekonfiguration des Taktgenerators erfolgt. In diesem Fall kann der AVFC-Referenztakt ggf. instabil sein, was direkte Auswirkungen auf die Versorgungsspannung hätte. Blicke etwa der Referenztakt für längere Zeit aus, so würden durch den Phasendetektor kontinuierlich Abwärtsimpulse generiert, wodurch die Versorgungsspannung bis auf kritisch geringe Werte absinken könnte, so dass Speicherelemente unter Umständen ihren Zustand verlieren könnten.

Um dies zu vermeiden, wird das AVFC-System während der Rekonfiguration des Taktgenerators in den Haltezustand versetzt.

10.2.1 Dynamische Eigenschaften der Regelschleife

Ein wesentlicher Unterschied der AVFC-Regelschleife gegenüber einer gewöhnlichen PLL ist das nichtlineare Verhalten des VCO, d.h. des Ringoszillators. Im Subschwelligebiet ergibt sich ein exponentieller Zusammenhang, darüber ein linearer bis quadratischer Zusammenhang zwischen Versorgungsspannung und Oszillationsfrequenz. Im Folgenden sollen die Auswirkungen dieser Besonderheit auf das dynamische Verhalten der Regelschleife untersucht werden. Die nachfolgend aufgeführten Betrachtungen und Formeln wurden vollständig [9] entnommen und an die Besonderheiten der AVFC-Regelschleife angepasst. Die Anpassungen betreffen die Hinzunahme des Verstärkungsfaktors des Spannungsreglers K_v sowie die Linearisierung des Ringoszillator-Verhaltens in zwei unterschiedlichen Betriebsbereichen durch $K_{0,L}$ und $K_{0,P}$.

Gegeben sei eine PLL mit einem PFD als Phasendetektor und einem aktiven PI-Schleifenfilter, wobei im Rückkopplungspfad kein Frequenzteiler vorhanden ist. Sei V_{PFD} die positive Versorgungsspannung des PFD-Ausgangs. Dann ist

$$K_d = \frac{V_{\text{PFD}}}{4\pi} \quad (10.1)$$

der Phasendetektorgewinn des PFD. Ferner sei $\omega_{\text{OSC}}(V_{\text{DD}})$ die Kreisfrequenz des Ringoszillator-Ausgangs in Abhängigkeit von der variablen Versorgungsspannung V_{DD} . Dann ist

$$K_0 = \frac{d\omega_{\text{OSC}}}{dV_{\text{DD}}} \quad (10.2)$$

der VCO-Gewinn, der jedoch im Falle der AVFC-Regelschleife aufgrund der Nichtlinearität vom Arbeitspunkt des Oszillators abhängt. Zusätzlich muss der Verstärkungsfaktor des Spannungsreglers K_v berücksichtigt werden. Für die DB-CoreVA-Realisierung gilt

$$K_v \approx 0.26 \quad \text{und} \quad V_{\text{PFD}} = 1,8 \text{ V} \quad (10.3)$$

Es werden nun zunächst die Spannungssignale des Referenztaktes U_{REF} und des Oszillatortaktes U_{OSC} unter der Annahme sinusförmiger Taktsignale¹ definiert:

$$U_{\text{REF}}(t) = \hat{U}_{\text{REF}} \sin(\omega_{\text{REF}}t + \theta_{\text{REF}}(t)) \quad (10.4)$$

$$U_{\text{OSC}}(t) = \hat{U}_{\text{OSC}} \sin(\omega_{\text{OSC}}t + \theta_{\text{OSC}}(t)) \quad (10.5)$$

Im eingerasteten Zustand ist $\omega_{\text{OSC}} \approx \omega_{\text{REF}}$ und es kann eine Linearisierung des Ringoszillatorverhaltens vorgenommen werden:

$$\omega_{\text{OSC,L}} = \omega_0 + K_{0,L}(V_{\text{DD}} - V_{\text{DD},0}) \quad (10.6)$$

$$\omega_0 = \omega_{\text{REF}} \quad (10.7)$$

$$V_{\text{DD},0} = V_{\text{DD}}|_{\omega_{\text{OSC}}=\omega_{\text{REF}}} \quad (10.8)$$

$$K_{0,L} = \left. \frac{d\omega_{\text{OSC}}}{dV_{\text{DD}}} \right|_{\omega_{\text{OSC}}=\omega_{\text{REF}}} \quad (10.9)$$

Das zeitliche Verhalten der eingerasteten Regelschleife wird nun durch die Phasensignale $\theta_{\text{REF}}(t)$, $\theta_{\text{OSC}}(t)$ bzw. deren Laplace-Transformierte $\Theta_{\text{REF}}(s)$, $\Theta_{\text{OSC}}(s)$ bestimmt. Es gilt

$$\Theta_{\text{OSC}}(s) = H(s)\Theta_{\text{REF}}(s) \quad (10.10)$$

mit der Phasenübertragungsfunktion der geschlossenen Regelschleife $H(s)$:

$$H(s) = \frac{K_{0,L}K_dK_vF(s)}{s + K_{0,L}K_dK_vF(s)} \quad (10.11)$$

Dabei ist $F(s)$ die Übertragungsfunktion des Schleifenfilters. Für ein aktives PI-Filter mit den Zeitkonstanten τ_1, τ_2 gilt

$$F(s) = \frac{1 + s\tau_2}{s\tau_1} \quad (10.12)$$

Die Phasenübertragungsfunktion $H(s)$ kann in normalisierter Form

$$H(s) = \frac{2s\zeta\omega_n + \omega_n^2}{s^2 + 2s\zeta\omega_n + \omega_n^2} \quad (10.13)$$

$$\omega_n = \sqrt{\frac{K_{0,L}K_dK_v}{\tau_1}} \quad (10.14)$$

$$\zeta = \frac{\omega_n}{2}\tau_2 \quad (10.15)$$

¹ Für die nachfolgenden Betrachtungen ist die konkrete Signalform der Taktsignale ohne Bedeutung. Es wäre ebenso die Annahme von Rechtecksignalen möglich.

angegeben werden. ω_n heißt *Eigenfrequenz* und ζ *Dämpfungsgrad* der Regelschleife.

Es sollen nun einige dynamische Eigenschaften der AVFC-Regelschleife untersucht werden. Dabei ist zu beachten, dass im nicht-eingerasteten Zustand $\omega_{OSC} \neq \omega_{REF}$ und die Linearisierung nach Gleichung 10.9 daher nicht gültig ist. Stattdessen muss K_0 ggf. als variabel angesehen werden, was damit ebenso für ω_n und ζ gilt. Der Betrieb einer PLL lässt sich in verschiedene Betriebsbereiche $\omega_0 \pm \Delta\omega$ unterteilen, die durch unterschiedliche Frequenzabstände $\Delta\omega$ von einem gegebenen Arbeitspunkt ω_0 bestimmt sind. Dabei wird angenommen, dass sich die Regelschleife bei ω_0 im eingerasteten Zustand befindet oder dort einrastet.

Haltebereich

Als *Haltebereich* (engl. *Hold Range*) $\Delta\omega_H$ wird der Frequenzbereich bezeichnet, in dem die PLL bei hinreichend langsamer Änderung der Referenzfrequenz im eingerasteten Zustand verbleibt. Bei Verwendung eines PFDs als Phasendetektor ergibt sich für den Haltebereich der Ausdruck [9]

$$\Delta\omega_H = K_0 K_d K_v F(0) 2\pi \quad (10.16)$$

Da für ein aktives PI-Schleifenfilter $F(0) = \infty$ gilt, folgt unabhängig von der Variabilität von K_0

$$\Delta\omega_H = \infty \quad (10.17)$$

Fangbereich und Fangzeit

Der *Fangbereich* (engl. *Lock Range*) $\Delta\omega_L$ ist dadurch gekennzeichnet, dass die PLL in diesem Bereich in der Lage ist, innerhalb einer Schwebungsperiode zwischen Referenztakt und Oszillatortakt aus dem nicht-eingerasteten Zustand in den eingerasteten Zustand zu wechseln. Die Zeitdauer dieses Vorgangs wird als *Fangzeit* (engl. *Lock Time*) t_L bezeichnet. Unter der Annahme, dass die Linearisierung nach Gleichung 10.9 im gesamten Fangbereich ihre Gültigkeit behält, ergibt sich [9]

$$\Delta\omega_L \approx 4\pi\zeta\omega_n \quad (10.18)$$

$$t_L \approx \frac{2\pi}{\omega_n} \quad (10.19)$$

Ziehbereich und Ziehzeit

Im *Ziehbereich* (engl. *Pull-in Range*) $\Delta\omega_P$ ist die PLL in der Lage, ggf. mit Überspringen von Takten einzurasten. Dabei wird ein Frequenzsprung $\omega_0 \rightarrow \omega_0 \pm \Delta\omega_0$ mit $\Delta\omega_0 > \Delta\omega_L$ am Eingang der PLL angenommen. Eine PLL mit PFD weist immer einen unendlichen Ziehbereich auf [9]:

$$\Delta\omega_P = \infty \quad (10.20)$$

Die Dauer des Ziehprozesses wird durch die *Ziehzeit* (engl. *Pull-in Time*) t_P ausgedrückt. Für ein aktives PI-Schleifenfilter lässt sich die Abschätzung [9]

$$t_P \approx \frac{4\tau_1\Delta\omega_0}{K_{0,P}K_V V_{PFD}} \quad (10.21)$$

ermitteln, wobei zunächst ein konstanter VCO-Gewinn $K_{0,P}$ angenommen werde. Tatsächlich ist bei der AVFC-Regelschleife im Ziehbereich der VCO-Gewinn jedoch variabel. Dennoch kann ein geeigneter konstanter Wert für $K_{0,P}$ angegeben werden. Der Teilausdruck

$$\frac{\Delta\omega_0}{K_{0,P}} =: \Delta V_{DD} \quad (10.22)$$

aus Gleichung 10.21 gibt die Spannungsdifferenz an, um die die Versorgungsspannung des Ringoszillators angehoben bzw. abgesenkt werden muss, um zu einem Einrasten der Regelschleife zu führen, d.h. eine Änderung der Oszillationsfrequenz um $\Delta\omega_0$ hervorzurufen. Für eine gegebene Frequenzsprungweite $\Delta\omega_0$ kann ΔV_{DD} direkt aus dem (bekannten, nichtlinearen) Verhalten des Ringoszillators $\omega_{OSC}(V_{DD})$ ermittelt werden. Es gilt dann

$$K_{0,P} = \frac{\Delta\omega_0}{\Delta V_{DD}} \quad (10.23)$$

Ausrastbereich

Der *Ausrastbereich* (engl. *Pull-out Range*) $\Delta\omega_{PO}$ gibt die maximale Weite eines Frequenzsprungs $\omega_0 \rightarrow \omega_0 \pm \Delta\omega_{PO}$ an, dem die eingerastete Regelschleife ohne Überspringen von Taktperioden folgen kann. Für eine PLL mit PFD und aktivem PI-Filter lässt sich der Ausrastbereich durch

$$\Delta\omega_{PO} \approx 11.55\omega_n(\zeta + 0.5) \quad (10.24)$$

approximieren [9]. Da im eingerasteten Zustand die Linearisierung nach Gleichung 10.9 gilt, ist diese Approximation auch für die AVFC-Regelschleife gültig.

Eigenfrequenz	ω_n	$\sqrt{\frac{K_{0,L}K_dK_v}{\tau_1}}$
Dämpfungsgrad	ζ	$\frac{\omega_n}{2}\tau_2$
Haltebereich	$\Delta\omega_H$	∞
Fangbereich	$\Delta\omega_L$	$4\pi\zeta\omega_n$
Fangzeit	t_L	$\frac{2\pi}{\omega_n}$
Ziehbereich	$\Delta\omega_P$	∞
Ziehzeit	t_P	$\frac{4\tau_1\Delta\omega_0}{K_{0,P}K_vV_{PFD}}$
Ausrastbereich	$\Delta\omega_{PO}$	$11.55\omega_n(\zeta + 0.5)$

Tabelle 10.1: Dynamische Eigenschaften der AVFC-Regelschleife (nach [9])

Rauschbandbreite

Das Verhalten der Regelschleife in Anwesenheit von Rauschen wird durch die (zweiseitige) Rauschbandbreite B_L beschrieben. Es gilt [9]

$$B_L = \frac{\omega_n}{2} \left(\zeta + \frac{1}{4\zeta} \right) \quad (10.25)$$

Es ist zu beachten, dass B_L die Einheit Hz besitzt, während ω_n , $\Delta\omega_H$, $\Delta\omega_L$, $\Delta\omega_P$ und $\Delta\omega_{PO}$ als Kreisfrequenz in rad s^{-1} angegeben sind.

Die betrachteten dynamischen Eigenschaften sind nochmals in Tabelle 10.1 zusammengefasst.

10.2.2 Einfluss der Oszillator-Nichtlinearität

Die Nichtlinearität des Ringoszillators hat signifikante Auswirkungen auf das dynamische Verhalten der AVFC-Regelung. Legt man die Regelschleife für einen bestimmten Arbeitspunkt aus, so kann daher das Regelverhalten in anderen Arbeitspunkten erheblich hiervon abweichen. Um diesen Einfluss genauer einschätzen zu können, ist zunächst eine entsprechende Analyse des Ringoszillatorverhaltens erforderlich.

Abbildung 10.5a zeigt den durch Messung und Mittelung anhand mehrerer Schaltungsexemplare ermittelten VCO-Gewinn K_0 in Abhängigkeit von der Versorgungsspannung bei Selektion des Rückkopplungspfades am 32. Abgriff der Verzögerungskette. Dieser Abgriff stellt bei diesen Exemplaren die an den kritischen Pfad angepasste Konfiguration (einschließlich der oben genannten Reserve) für die AVFC-Regelung dar. Man erkennt, dass oberhalb von etwa 0,7 V ein nahezu lineares Verhalten des Ringoszillators mit $K_0 \approx 8 \cdot 10^8 \text{ rad V}^{-1} \text{ s}^{-1}$ vorliegt. Zu kleineren Spannungen hin nimmt K_0 ab, wobei im Subschwelligbereich offenbar ein näherungsweise exponentieller Zusammenhang vorliegt. Betrachtet man nun erneut die dynamischen Eigenschaften nach Tabelle 10.1 und vernachlässigt den geringfügigen Abfall von K_0 knapp unterhalb der maximalen Versorgungsspannung, so lässt sich feststellen, dass bei einer Verringerung der Versorgungsspannung

- ω_n abnimmt, d.h. die Eigenschwingungen des Systems bei geringerer Spannung eine größere Periodendauer aufweisen,
- ζ sinkt, d.h. das System bei kleineren Spannungen zu Überschwingen und bei größeren Spannungen zu einem aperiodischen Verhalten neigt,
- $\Delta\omega_L$ und $\Delta\omega_{PO}$ abnehmen, d.h. die Stabilitätsbereiche absolut betrachtet kleiner werden.

Die Abnahme von ω_n , $\Delta\omega_L$ und $\Delta\omega_{PO}$ kann prinzipiell toleriert werden, da zugleich auch die Oszillationsfrequenz ω_{OSC} abnimmt. Kritisch ist jedoch die Änderung des Dämpfungsgrades ζ . Durch die Auslegung des Schleifenfilters muss sichergestellt sein, dass ζ in allen Arbeitspunkten in einem akzeptablen Bereich liegt. Dies gelingt durch Wahl einer mittleren Spannung $V_{DD} = 400 \text{ mV}$ und Vorgabe eines Dämpfungsgrades $\zeta = 0,7$ (Butterworth-Charakteristik) sowie einer Fangzeit von 100 Taktzyklen in diesem Arbeitspunkt, d.h. $t_L \approx 175 \mu\text{s}$. Löst man die im vorangegangenen Abschnitt angegebenen Gleichungen nach den Zeitkonstanten des PI-Filters, so erhält man $\tau_1 = 2,138 \text{ ms}$ und $\tau_2 = 0,039 \text{ ms}$. Es ergeben sich die in den Abbildungen 10.5b bis 10.5d dargestellten dynamischen Eigenschaften. Man erkennt, dass der Dämpfungsgrad bei einer Spannung von 325 mV, die dem Punkt minimaler Energieaufnahme des Prozessors entspricht, einen Wert von 0,33 aufweist und bei höheren Spannungen bis zu einem Maximalwert von 2,6 steigt. Der Spannungsbereich unterhalb von 325 mV sollte, zumindest bei Verwendung der adaptiven Spannungsregelung, aufgrund übermäßigen Überschwingens und damit einem zunehmend instabilen Regelverhalten vermieden werden. Im verbleibenden Spannungsintervall variiert der Fangbereich zwischen 0,6 % und 10 %, der

Ziehbereich zwischen 0,7 % und 24 % und die Rauschbandbreite zwischen 0,2 % und 8 % der Ringoszillatorfrequenz. Die Fangzeit sinkt bei einer Erhöhung der Versorgungsspannung von 368 μs auf 47 μs . Für Frequenzschritte von 0 Hz auf die jeweilige Frequenz bei einer bestimmten Spannung (z.B. bei der Inbetriebnahme des Systems) liegt die Ziehzeit zwischen 6 ms und 22 ms. Diese Werte sind akzeptabel für die adaptive Spannungsregelung.

Neben einer theoretischen Untersuchung der dynamischen Eigenschaften wurde das Regelverhalten auch mit Hilfe von Simulationen sowie durch Messungen untersucht. Durch die große Anzahl der zur Regelschaltung gehörenden Bauteile kommt es im Falle einer vollständig analogen Schaltungssimulation zu außerordentlich langen Laufzeiten von mehreren Tagen für einen einzelnen Simulationslauf. Zur Beschleunigung wurde daher eine VHDL-basierte *Mixed-Signal-Modellierung*¹ der Regelschleife erstellt. Für die digitalen Bestandteile – PFD und Lock-Detektor – kann dabei der für die FPGA-Synthese verwendete Quellcode direkt genutzt werden. Für den Ringoszillator wurde der Zusammenhang zwischen Versorgungsspannung und erzeugter Frequenz durch eine abschnittsweise Polynominterpolation dritten Grades mit 29 Stützstellen basierend auf gemittelten Messwerten approximiert. Der PI-Filter wird durch eine entsprechende zeitdiskrete Realisierung simuliert. Als Anregung für die Simulation dienen Frequenzschritte des Referenztaktes im Abstand von jeweils 25 ms. Die gleiche Anregung wird auch für eine messtechnische Analyse des Regelverhaltens genutzt. Das Regelverhalten ist in Abbildung 10.6 dargestellt. Es besteht eine gute Übereinstimmung zwischen den simulierten und den gemessenen Signalverläufen. Unterschiede der Spannungswerte jeweils bei eingerasteter Regelschleife sind darin begründet, dass der VCO-Gewinn des verwendeten Exemplars von dem in der Simulation verwendeten über alle Exemplare gemittelten VCO-Gewinn geringfügig abweicht. Man erkennt, dass die Regelung in der Größenordnung einiger Millisekunden benötigt, um nach einem Frequenzschritt die erforderliche Versorgungsspannung zu erreichen. Dies ist konsistent zur obigen theoretischen Voruntersuchung bezüglich der Ziehzeit. Der Einschwingvorgang bei 10 kHz und ca. 240 mV ist im rechten Teil der Abbildung vergrößert dargestellt. Es ist hier ein deutliches Überschwingen sichtbar, so dass, wie bereits zuvor erwähnt, dieser Betriebsbereich möglichst vermieden werden sollte, was jedoch auch aus energetischen Gründen aufgrund der Lage des Punktes minimaler Energieaufnahme des Subschwellsprozessors bei 325 mV geboten ist.

¹ gemischte analoge und digitale Modellierung

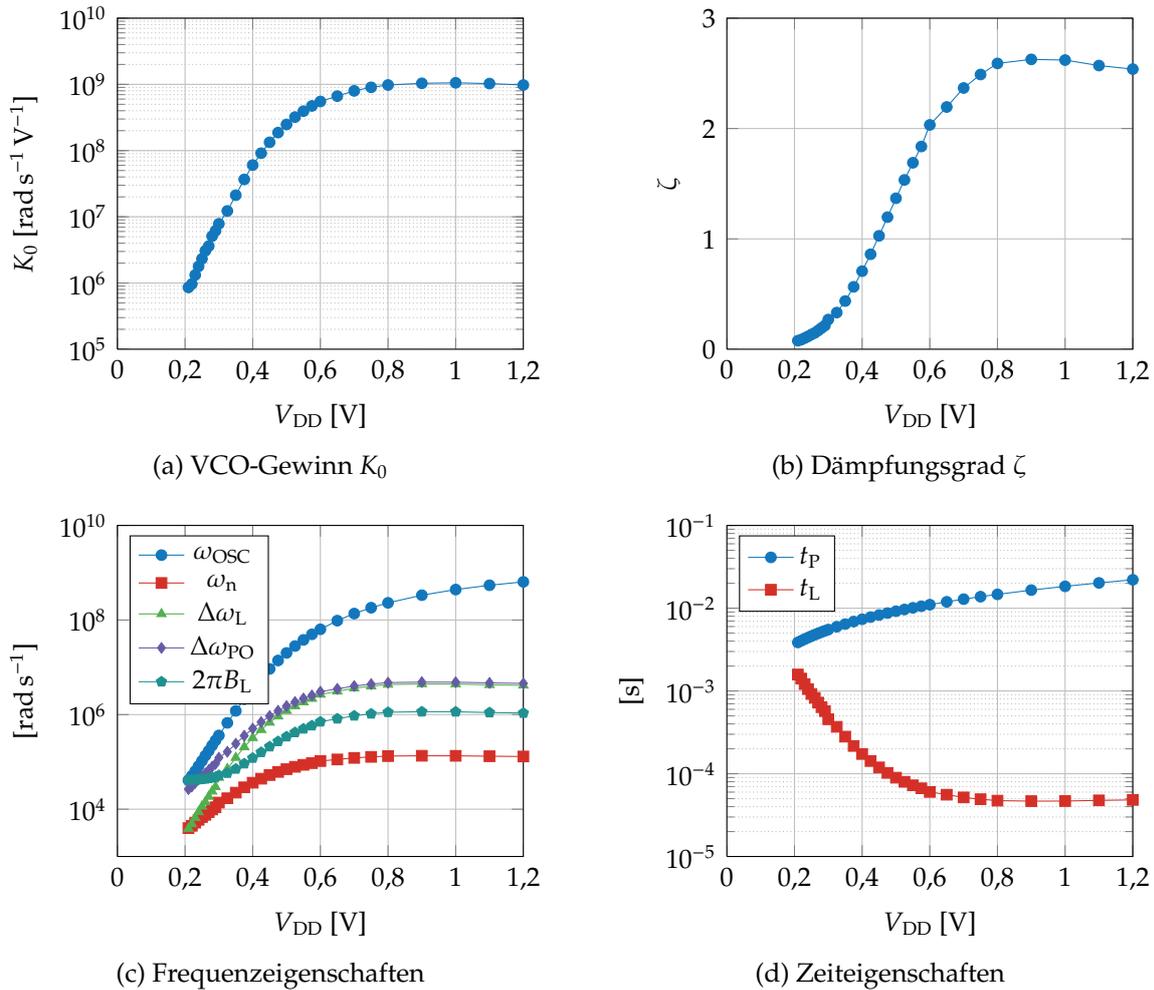


Abbildung 10.5: Dynamische Eigenschaften der AVFC-Regelschleife

10.3 Zusammenfassung

In diesem Kapitel wurden Methoden zur Minimierung des Energiebedarfs durch die Wahl eines optimalen Betriebspunktes betrachtet. Dazu wurde mit der PPMU zunächst ein Modul zur Takt- und EnergiEVERWALTUNG für das CoreVA-Prozessorsystem vorgestellt. Eine Anwendung hat damit die Möglichkeit, die Versorgungsspannung und die Taktfrequenz dynamisch zu variieren oder sogar zu deaktivieren und so den Performanzanforderungen oder dem Energiebudget anzupassen. Die PPMU nutzt in der konkreten Realisierung Taktverwaltungskomponenten von Xilinx-FPGAs, lässt sich allerdings ebenso auf andere Plattformen portieren. Außerdem wurde das AVFC-System zur adaptiven Spannungs- und Frequenzsteuerung vorgestellt, das

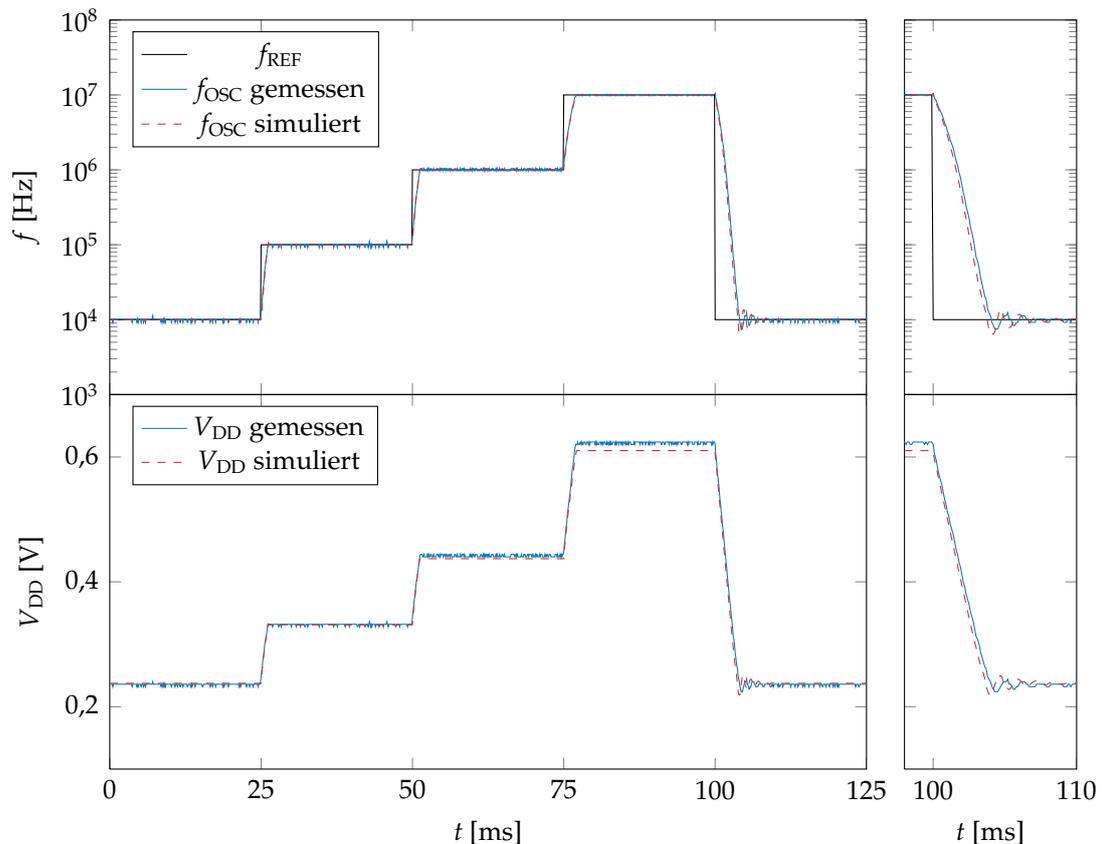


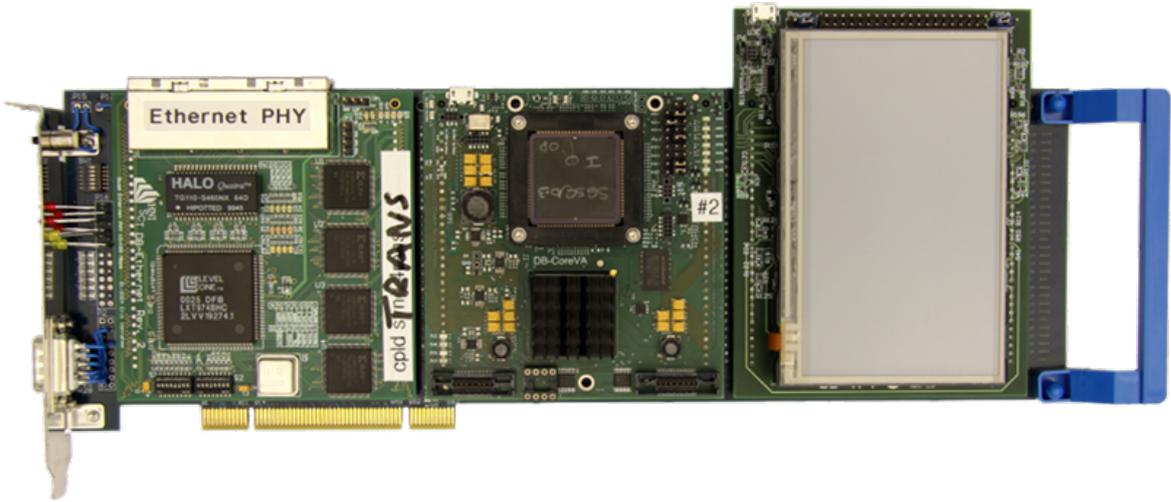
Abbildung 10.6: Zeitverhalten der AVFC-Regelschleife

eine dynamische Kompensation von Prozess- und Betriebsparameter-Schwankungen nach der Fertigung ermöglicht. Der Vorteil dieses Verfahrens besteht darin, dass insbesondere im Subschwelligbereich Sicherheitsreserven bezüglich der Taktfrequenz bei einer gegebenen Spannung reduziert werden können, wodurch sich die Energieeffizienz einer Schaltung weiter steigern lässt. Dabei dient ein auf dem Baustein befindlicher konfigurierbarer Ringoszillator als Sensor für Prozess- und Betriebsbedingungen. Eine PLL-ähnliche Regelung generiert eine diesen Einflüssen angepasste, optimale Versorgungsspannung oder Taktfrequenz. Die dynamischen Eigenschaften der AVFC-Regelung wurden theoretisch, simulativ und messtechnisch untersucht. Die adaptive Anpassung der Versorgungsspannung nach einem Frequenzschritt dauert je nach Schrittgröße in der Größenordnung einiger Millisekunden.

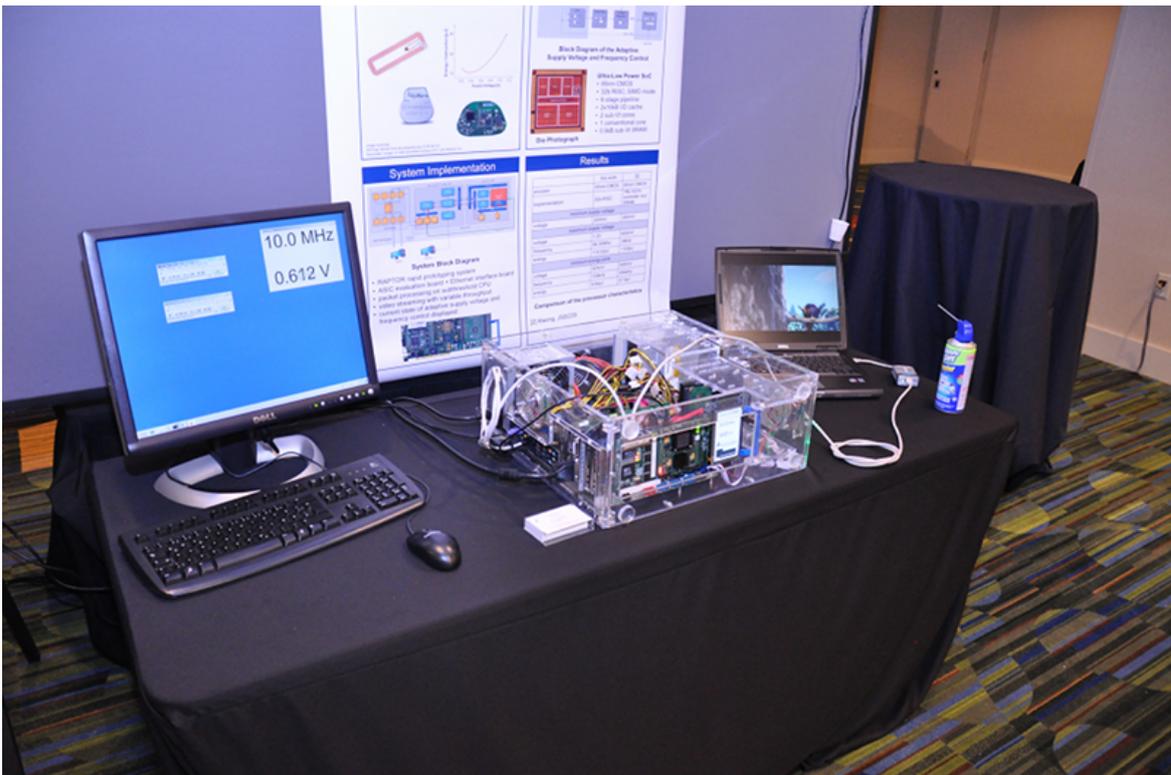
11 Demonstrator

Um die Funktionsweise des CoreVA-ULP-Prozessors in Verbindung mit der PPMU sowie der adaptiven Spannungs- und Frequenzsteuerung in einer realen Anwendung zu präsentieren, wurde ein Demonstrator auf Basis des RAPTOR-Systems realisiert. Neben dem DB-CoreVA kommen dabei die Erweiterungsmodule DB-Ethernet und DB-TFT zum Einsatz. Der Demonstrator wurde einem internationalen Fachpublikum mit großem Interesse auf der IEEE International Solid-State Circuits Conference (ISSCC) 2012 im Rahmen der Academic Demonstration Session als ausgewählte Veröffentlichung [181] vorgeführt. Abbildung 11.1 zeigt den Aufbau des vorgestellten Demonstrators.

In der Demo-Anwendung verarbeitet der Subschwellschwellprozessor den Paket-Datenstrom einer Video-Übertragung zwischen zwei PCs über ein Ethernet-Netzwerk. Während der Verarbeitung der Netzwerkpakete finden dabei unter anderem Berechnungen und Kontrollen von Prüfsummen auf unterschiedlichen Protokoll-Ebenen statt. Für die Video-Übertragung wird eine Codierung mit variabler Bitrate verwendet, so dass der Prozessor je nach Bildinhalten einer variierenden Auslastung ausgesetzt ist. Die Erkennung der aktuellen Auslastung wird durch den auf dem Prozessor ausgeführten Programmcode anhand der Puffer-Füllstände der Ethernet-Schnittstellen erkannt. Liegen die Füllstände oberhalb eines bestimmten Schwellwertes, so sendet der Prozessor eine Anforderung zur Erhöhung der Taktfrequenz an die PPMU. Bei leeren Paket-Puffern wird die Taktfrequenz wiederum reduziert. Auf diese Weise wählt der Prozessor selbständig und kontinuierlich eine optimale Taktrate in einem weiten Frequenzbereich zwischen 100 kHz und 80 MHz. Die Untergrenze dieses Bereichs entspricht etwa der Taktfrequenz im Energieminimum des Prozessors. Eine weitere Reduktion der Taktfrequenz ist daher aus energetischer Sicht nicht sinnvoll. Zur Erzeugung einer optimalen Versorgungsspannung in Abhängigkeit von der Taktfrequenz kommt die adaptive Spannungs- und Frequenzsteuerung im Betriebsmodus mit adaptiver Spannungsregelung zum Einsatz. Die aktuellen Betriebsparameter, Taktfrequenz und Versorgungsspannung, werden auf der Anzeige des DB-TFT dargestellt.



(a) RAPTOR-System mit DB-Ethernet, DB-CoreVA und DB-TFT



(b) Vorführung auf der ISSCC 2012

Abbildung 11.1: Aufbau des CoreVA-ULP-Demonstrators

Während der Demonstration kommen zwei Video-Datenströme mit verschiedenen durchschnittlichen Datenraten zum Einsatz. Da bei Szenewechseln oder anderen stark variierenden Bildinhalten durch die variable Kompression kurzzeitig größere Datenmengen übertragen werden, beziehen sich die nachfolgend angegebenen Frequenz- und Spannungswerte auf Videosequenzen mit annähernd konstanter Datenrate. Bei der Übertragung des mit 3500 kbit s^{-1} codierten Videos beträgt die Taktfrequenz des Prozessors 25 MHz bei einer Spannung von etwa 750 mV. Das zweite Video besitzt eine Datenrate von 1000 kbit s^{-1} und wird durch den Prozessor bei 8 MHz und 584 mV übertragen. Werden beide Videos gleichzeitig über dieselbe Verbindung übermittelt, so ergibt sich eine Taktfrequenz von 40 MHz und eine Spannung von 860 mV.

Die Funktion der AVFC-Regelung kann mit dem Demonstrator durch eine mittels Kälte-Spray verursachte Temperaturänderung des CoreVA-ULP-ASICs veranschaulicht werden. Zur Kompensation des Kälteeinflusses erhöht die Regelung im Subschwelligbereich die Versorgungsspannung, was durch die Anzeige auf dem DB-TFT sichtbar ist. Währenddessen setzt der Prozessor die Paketverarbeitung fehlerfrei fort. Am deutlichsten lässt sich dieser Effekt ohne laufende Datenübertragung beobachten, da hier Frequenz und Spannung am niedrigsten und die Schaltungseigenschaften damit am empfindlichsten gegenüber Temperaturänderungen sind. Der Prozessor arbeitet dann bei Raumtemperatur mit 100 kHz bei einer Spannung von ca. 328 mV. Beim Abkühlen des ASICs erhöht die AVFC-Regelschleife die Versorgungsspannung je nach erreichter Temperatur bis auf etwa 360 mV.

12 Zusammenfassung und Ausblick

Für eine wachsende Anzahl an Anwendungen integrierter Digitalschaltungen stellt der Energie- oder Leistungsbedarf die kritischste Ressource dar, während gleichzeitig die Performanzanforderungen mäßig sind. Beispiele solcher Anwendungen sind RFID-Transponder, drahtlose Sensornetzwerke oder biomedizinische Implantate. Die Subschwellechnik ist in der Lage, diese Anforderungen sehr gut zu erfüllen. Dabei stellt sich allerdings die Herausforderung, dass derartige Schaltungen eine erhöhte Sensibilität gegenüber verschiedenen Störeinflüssen sowohl der Fertigungs- als auch der Betriebsparameter aufweisen. Es sind daher spezielle Entwurfsverfahren notwendig, um eine ressourceneffiziente und robuste Implementierung zu erhalten.

Im Rahmen dieser Arbeit wurden für zwei Halbleitertechnologien mit Strukturgrößen von 90 nm und 65 nm für den Subschwellebetrieb ausgelegte Standardzellenbibliotheken entwickelt. Beide Bibliotheken gestatten eine dynamische Spannungskalierung während des Betriebs bis hinauf zur nominellen Versorgungsspannung der jeweiligen Technologie. Die 90-nm-Bibliothek enthält 15 Gatter, die für eine optimale Symmetrie des Spannungsübertragungsverhaltens bei einer Versorgungsspannung von 200 mV ausgelegt sind. Hierdurch weisen die Gatter eine hohe Toleranz gegenüber Störeinflüssen im Subschwellebereich auf. Bei der 65-nm-Bibliothek, deren Zellsatz insgesamt 62 Gatter umfasst, wurde ein Mehrzieloptimierungsverfahren beim Zellenentwurf eingesetzt, das die Berücksichtigung weiterer Zielfunktionen gestattet. Dadurch ist es möglich, neben der Robustheit simultan den Energiebedarf sowie die Verzögerungszeiten bereits auf Gatter-Ebene zu optimieren. Für Signalverbindungen zwischen Schaltungsteilen, die im Subschwellebereich arbeiten, und solchen, die auf konventionellem Spannungsniveau betrieben werden, wurde ein neuartiger Pegelwandler entwickelt, dessen zentrale Wandler-Stufe aus lediglich fünf Transistoren und damit nur einem Transistor mehr als der klassische DCVS-Pegelwandler besteht. Dieser Pegelwandler besitzt gegenüber alternativen Lösungen für den Subschwellebetrieb den Vorteil, dass keine statisch leitenden Strompfade existieren, so dass auch bei einer Erhöhung der Versorgungsspannung nur geringe Leckverluste anfallen. Bei maximaler Versorgungsspannung lässt sich dadurch die

Verlustleistung um etwa den Faktor 200 reduzieren. Gleichzeitig weist der Pegelwandler gegenüber einer mit gleichem Aufwand optimierten Vergleichsimplementierung eine mehr als doppelt so hohe Schaltgeschwindigkeit im Subschwelligbereich auf.

Basierend auf den Subschwellig-Standardzellen wurden zwei ASICs entwickelt, gefertigt und messtechnisch charakterisiert. Ein integrierter Schaltkreis mit vier 32-bit-ALUs in der 90-nm-Technologie erreicht im Stichprobenmittel einen minimalen Energiebedarf von 0,45 pJ pro Takt bei 210 mV und 3 MHz sowie eine minimale Versorgungsspannung von 143 mV. Fünf Exemplare sind noch bei einer Spannung von 115 mV voll funktionsfähig. Dieses, für eine Schaltung entsprechender Komplexität beachtliche Ergebnis wird ohne eine Regelung der Substratspannung zur Kompensation von Prozessschwankungen erreicht.

Die zweite prototypische ASIC-Realisierung enthält zwei Kopien eines vollständigen Subschwelligprozessors, der als Grundbaustein energieeffizienter eingebetteter Systeme etwa für die eingangs genannten Anwendungen dienen kann. Der Subschwelligprozessor basiert auf der CoreVA-Architektur, einer leistungsfähigen 32-bit-Architektur mit sechs Pipelinestufen sowie hardwareseitiger Unterstützung für Berechnungen von Multiplikationen und Divisionsschritten. Zusätzlich bietet der Prozessor einen SIMD-Modus zur Verdopplung des Durchsatzes bei 16-bit-Berechnungen. Der minimale Energiebedarf des Subschwelligprozessors von 9,94 pJ pro Takt tritt bei 325 mV und 133,18 kHz auf. Bei maximaler Versorgungsspannung beträgt der Energiebedarf 110,22 pJ und die Taktfrequenz 94,3 MHz. Im Subschwelligbetrieb kann der Energieumsatz also um den Faktor 11 reduziert werden. Die minimale Versorgungsspannung beträgt im Stichprobenmittel 346,9 mV, die besten Exemplare arbeiten noch bei 200 mV korrekt. Durch den Einsatz alternativer Taktkonzepte ließe sich die untere Versorgungsspannungsgrenze voraussichtlich weiter absenken. Ein identischer Prozessorkern auf Basis konventioneller Standardzellen auf demselben ASIC erreicht eine Taktfrequenz von 260 MHz bei 66,13 pJ pro Takt. Der Energiebedarf ist damit bei maximaler Spannung um etwa 40 % geringer als der des Subschwelligprozessors. Daher sollte dieser Betriebsbereich des Subschwelligprozessors möglichst nur eingesetzt werden, wenn dazu die Notwendigkeit, d.h. eine kurzfristige, entsprechend hohe Performanzanforderung besteht. Die minimale Versorgungsspannung des konventionellen Prozessors liegt mit 700 mV deutlich oberhalb des Subschwelligbereichs.

Neben den Prozessorkernen enthält der ASIC außerdem spezielle Subschwellig-SRAM-Blöcke auf Basis einer 9T-Zelle. Dieser Speicher wurde in Kooperation mit der Fachgruppe Nanoelektronik der Universität Oslo entwickelt. Die SRAM-Zelle

besitzt eine Fläche von $2,83 \mu\text{m}^2$ und liegt damit im unteren Bereich vergleichbarer Implementierungen. Der minimale Energiebedarf von $17,8 \text{ fJ}$ pro Bit und Zugriff tritt bei 320 mV und 761 kHz auf und stellt den zweitbesten Wert in einem Vergleich mit sechs weiteren Realisierungen dar. Die minimale Versorgungsspannung von 273 mV sowie die statische Verlustleistung von $17,8 \text{ pW}$ pro Bit bei 300 mV liegen im mittleren Bereich des Vergleichsfeldes.

Zur dynamischen Anpassung von Versorgungsspannung und Taktfrequenz wurde ein entsprechendes Modul, die PPMU, als Hardwareerweiterung für das Prozessorsystem entwickelt. Zusätzlich ist die PPMU mit dem AVFC-Modul zur adaptiven Spannungs- und Frequenzsteuerung gekoppelt, die eine dynamische Kompensation von Prozess- und Betriebsparameter-Schwankungen nach der Fertigung gestattet. Dabei kommt ein auf dem Chip befindlicher konfigurierbarer Ringoszillator als Sensor für derartige Störeinflüsse zum Einsatz. Eine den Störeinflüssen angepasste, optimale Versorgungsspannung oder Taktfrequenz wird dann durch eine PLL-ähnliche Regelstruktur bereitgestellt. Gegenüber einer konventionellen Phasenregelschleife weist die AVFC-Regelung Besonderheiten bezüglich des dynamischen Verhaltens auf, die theoretisch, simulativ und messtechnisch untersucht wurden. Dabei wurde eine gute Übereinstimmung zwischen den drei Untersuchungsebenen festgestellt. Die adaptive Anpassung der Versorgungsspannung nach einem Frequenzschritt dauert je nach Schrittgröße in der Größenordnung einiger Millisekunden.

Zur Einbettung des Subschwellsprozessors und kompatibler ASICs in das Rapid-Prototyping-System RAPTOR wurde das Erweiterungsmodul DB-CoreVA entwickelt. Es ist mit einem FPGA des Typs Xilinx Spartan-6 LX XC6SLX100 oder XC6SLX150 sowie bis zu 512 MB DDR2-SDRAM ausgestattet. Außerdem enthält es einige analoge Komponenten der AVFC-Regelung, insbesondere den Schleifenfilter sowie den einstellbaren Spannungsregler. Auf Basis dieser Plattform wurde zudem ein Demonstrator realisiert, der die Funktionsweise und die Vorteile des Subschwellsprozessors in Verbindung mit der adaptiven Spannungs- und Frequenzsteuerung in einer realen Anwendung veranschaulicht. Hierbei handelt es sich um die Übertragung eines Video-Datenstroms mit variabler Datenrate über ein Ethernet-Netzwerk. Der Subschwellsprozessor passt dabei selbstständig seine Taktfrequenz und Versorgungsspannung adaptiv an die Netzwerklast sowie an die Umgebungstemperatur als Störeinfluss an.

Aufbauend auf den Ergebnissen dieser Arbeit sind verschiedene weiterführende Untersuchungen denkbar. So wurde etwa gezeigt, dass der Entwurf einer Schaltung für den Subschwellsbetrieb gegenüber einem konventionellen Schaltungsentwurf

unter anderem zu einem gesteigerten Energiebedarf bei hoher Versorgungsspannung führt. Für eine konkrete Anwendung mit gegebener zeitlicher Verteilung des Performanzbedarfs lassen sich also Kosten und Nutzen des Subschwellschaltens etwa in energetischer Sicht bestimmen. Eine äquivalente Fragestellung lautet, wie häufig und wie lange eine Subschwellschaltung im hohen Spannungsbereich betrieben werden darf, um die dabei auftretende suboptimale Energieaufnahme während der übrigen Zeit im Subschwellschaltensbetrieb wieder aufholen zu können. Zwischen Schaltungsentwürfen für tatsächlichen Subschwellschaltensbetrieb auf der einen Seite und konventionellen Betrieb auf der anderen Seite sind außerdem Kompromisse möglich, d.h. die Optimierung von Schaltungen für den Bereich moderater Inversion (near-threshold). Zudem lassen sich Schaltungen aller Entwurfsschwerpunkte mit weiteren Verfahren zur Energieeinsparung, insbesondere Power Gating, also das Abschalten der Spannungsversorgung bei Nichtbenutzung, kombinieren. Für einen vollständigen Überblick über die Vor- und Nachteile solcher Verfahrenskombinationen sind umfassende Untersuchungen verschiedener konkreter Anwendungen erforderlich. Auch die Verwendung von Substratvorspannungen, wie sie bei der zweiten in dieser Arbeit entwickelten Subschwellschaltensbibliothek bereits möglich ist, zu verschiedenen Zwecken – Reduktion von Parameterstreuungen, Erhöhung der Performanz, Verringerung der Verlustleistung oder Absenken der minimalen Versorgungsspannung – ist zu evaluieren.

Andere weiterführende Arbeiten sind unter dem Gesichtspunkt der Leistungssteigerung von Subschwellschaltungen möglich. Neben einer Erhöhung von Versorgungsspannung und Taktfrequenz ist hierzu auch die Erhöhung der Parallelität geeignet. Anstatt ein einzelnes Schaltungsexemplar also bei höherer Spannung zu betreiben, könnten auch mehrere parallele Exemplare bei geringerer Spannung betrieben und somit ein energetischer Vorteil erzielt werden. Bezogen auf die hier eingesetzte CoreVA-Prozessorarchitektur besteht etwa die Möglichkeit der Verwendung mehrerer Ausführungseinheiten, also der Übergang zu einer tatsächlichen VLIW-Architektur. Eine andere Variante ist die Nutzung mehrerer paralleler Prozessorkerne in einem SoC (engl. *Multi-Processor System on a Chip*, MPSoC). Solche Ansätze erscheinen insbesondere für Anwendungen mit einem hohen Grad an Parallelisierbarkeit auf Instruktionsebene oder, im zweiten Fall, bei einer Vielzahl unabhängig voneinander gleichzeitig auszuführender Aufgaben sinnvoll. Auch hier ergeben sich interessante Fragestellungen. So ist es beispielsweise möglich, alle Prozessorkerne mit gemeinsamer oder individueller Taktfrequenz und Versorgungsspannung zu betreiben. Auch der Einsatz von Power Gating auf Ebene der Prozessorkerne ist denkbar. Eine Herausforderung ergibt sich möglicherweise auch aus der globalen Kommunikation in Verbindung mit den starken Schwankungen

des Zeitverhaltens von Subschwellschaltungen. Hier könnte sich ein global asynchrones, lokal synchrones Kommunikationsschema (engl. *Global Asynchronous, Local Synchronous*, GALS) als vorteilhaft erweisen. Auch der Energiebedarf der Kommunikationsstrukturen in einem solchen Subschwell-MPSoC stellt einen interessanten Untersuchungsaspekt dar.

Abkürzungsverzeichnis

AES	Advanced Encryption Standard
ALU	Arithmetic Logic Unit
ASIC	Application Specific Integrated Circuit
AVFC	Adaptive Voltage and Frequency Control
C2MOS	Clocked CMOS
CAD	Computer-Aided Design
CISC	Complex Instruction Set Computer
CMOS	Complementary Metal Oxide Semiconductor
CoreVA	Configurable Ressource Efficient VLIW Architec- ture
CoreVA ULP	CoreVA Ultra-Low Power
CPHL	Conditional Precharge Hybrid Latch
CQFP	Ceramic Quad Flatpack
CRC	Cyclic Redundancy Check
DC	Instruction Decode
DCM	Digital Clock Manager
DCVS	Differential Cascode Voltage Switch
DDR2-SDRAM	Double Data Rate SDRAM
DIBL	Drain Induce Barrier Lowering
DLL	Delay-Locked Loop
DVFS	Dynamic Voltage and Frequency Scaling
ECC	Elliptic Curve Cryptography
EDP	Energy Delay Product
EEG	Elektroenzephalogramm
EKG	Elektrokardiogramm
ESD	Electrostatic Discharge
EX	Execute

FE	Instruction Fetch
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Array
GAIO	Global Analysis of Invariant Objects
GALS	Global Asynchronous, Local Synchronous
GP	General Purpose
GPIO	General Purpose I/O
HL	Hybrid Latch
HVT	High Vt
I/O	Input/Output
IOB	Input/Output Block
IP	Intellectual Property
ISE	Instruction Set Extension
JPEG	Joint Photographic Experts Group
JTAG	Joint Test Action Group
LD/ST	Load/Store
LED	Light-Emitting Diode
LOCOS	Local Oxidation of Silicon
LP	Low Power
LVT	Low Vt
ME	Memory Access
MIS	Metal Insulator Semiconductor
MMIO	Memory Mapped Input/Output
MMMC	Multi-Mode, Multi-Corner
MOP	Mehrzieloptimierungsproblem
MOS	Metal Oxide Semiconductor
MPSoC	Multi-Processor System on a Chip
MSV	Multiple Supply Voltage
MTCMOS	Multithreshold CMOS
NCE	Narrow Channel Effect

OCV	On-Chip Variation
PaR	Place and Route
PCI	Peripheral Component Interconnect
PFD	Phase/Frequency Detector
PLL	Phase-Locked Loop
PPMU	Performance and Power Management Unit
PTM	Predictive Technology Model
PVT	Process, Voltage, Temperature
RAM	Random Access Memory
RD	Register Read
RFID	Radio-Frequency Identification
RISC	Reduced Instruction Set Computer
RNCE	Reverse Narrow Channel Effect
ROM	Read-Only Memory
RSCE	Reverse Short Channel Effect
RTL	Register Transfer Level
SA	Sense Amplifier
SCE	Short Channel Effect
SDRAM	Synchronous Dynamic Random Access Memory
SIMD	Single Instruction, Multiple Data
SoC	System on a Chip
SOI	Silicon on Insulator
SPEA	Strength Pareto Evolutionary Algorithm
SPI	Serial Peripheral Interface
SRAM	Static Random Access Memory
STI	Surface Trench Isolation
SVT	Standard V _t
TGMS	Transmission Gate Master Slave
UART	Universal Asynchronous Receiver Transmitter
USB	Universal Serial Bus
VCO	Voltage Controlled Oscillator
VHDL	Very High Speed Integrated Circuit Hardware Description Language

Abkürzungsverzeichnis

VLIW	Very Long Instruction Word
WR	Register Write

Referenzen

- [1] M. Alioto. „Understanding DC Behavior of Subthreshold CMOS Logic Through Closed-Form Analysis“. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 57.7 (Juli 2010), S. 1597–1607. ISSN: 1549-8328. DOI: 10.1109/TCSI.2009.2034233.
- [2] H. P. Alstad und S. Aunet. „Improving Circuit Security against Power Analysis Attacks with Subthreshold Operation“. In: *Proceedings of the IEEE Workshop Design and Diagnostics of Electronic Circuits and Systems*. 2008, S. 1–2. DOI: 10.1109/DDECS.2008.4538746.
- [3] H. P. Alstad und S. Aunet. „Seven subthreshold flip-flop cells“. In: *Proceedings of the NORCHIP Conference*. 2007, S. 1–4. DOI: 10.1109/NORCHP.2007.4481061.
- [4] H. P. Alstad und S. Aunet. „Subthreshold AES S-Box with Increased Power Analysis Resistance“. In: *Proceedings of the NORCHIP Conference*. 2008, S. 13–16. DOI: 10.1109/NORCHP.2008.4738273.
- [5] H. P. Alstad und S. Aunet. „Three Subthreshold Flip-Flop Cells Characterized in 90 nm and 65 nm CMOS Technology“. In: *Proceedings of the IEEE Workshop Design and Diagnostics of Electronic Circuits and Systems*. 2008, S. 1–4. DOI: 10.1109/DDECS.2008.4538745.
- [6] R. J. Baker. *CMOS: Circuit Design, Layout, and Simulation*. 3. Aufl. Wiley – IEEE Press, 2010. ISBN: 978-0470881323.
- [7] H. K. O. Berge und S. Aunet. „Benefits of decomposing wide CMOS transistors into minimum-size gates“. In: *Proceedings of the NORCHIP Conference*. 2009, S. 1–4. DOI: 10.1109/NORCHP.2009.5397795.
- [8] H. K. O. Berge, M. Blesken, S. Aunet und U. Rückert. „Design of 9T SRAM for dynamic voltage supplies by a multiobjective optimization approach“. In: *Proceedings of the IEEE International Electronics, Circuits, and Systems Conference*. 2010, S. 319–322. DOI: 10.1109/ICECS.2010.5724517.
- [9] R. E. Best. *Phase-Locked Loops: Design, Simulation, and Applications*. 6. Aufl. McGraw-Hill Professional, 2007. ISBN: 9780071499262.

- [10] K. Binder. *Monte Carlo methods in statistical physics*. Springer, 1979. ISBN: 3-540-09018-5.
- [11] D. Blaauw und B. Zhai. „Energy efficient design for subthreshold supply voltage operation“. In: *Proceedings of the IEEE International Symposium Circuits and Systems*. 2006. DOI: 10.1109/ISCAS.2006.1692514.
- [12] M. Blesken. *Ein Mehrzieloptimierungsansatz zur Dimensionierung ressourcen-effizienter integrierter Schaltungen*. Dissertation. HNI-Verlagsschriftenreihe, Paderborn. Heinz Nixdorf Institut, 2012. URL: <http://digital.ub.uni-paderborn.de/hs/content/titleinfo/349866>.
- [13] M. Blesken, U. Rückert, D. Steenzen, K. Witting und M. Dellnitz. „Multiobjective optimization for transistor sizing of CMOS logic standard cells using set-oriented numerical techniques“. In: *Proceedings of the NORCHIP Conference*. 2009, S. 1–4. DOI: 10.1109/NORCHIP.2009.5397800.
- [14] D. Bol, J. De Vos, C. Hocquet, F. Botman, F. Durvaux u. a. „A 25MHz 7uW/MHz ultra-low-voltage microcontroller SoC in 65nm LP/GP CMOS for low-carbon wireless sensor nodes“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 2012, S. 490–492. DOI: 10.1109/ISSCC.2012.6177104.
- [15] D. Bol. „Robust and Energy-Efficient Ultra-Low-Voltage Circuit Design under Timing Constraints in 65/45 nm CMOS“. In: *Journal of Low Power Electronics and Applications* 1.1 (2011), S. 1–19. ISSN: 2079-9268. DOI: 10.3390/jlpea1010001. URL: <http://www.mdpi.com/2079-9268/1/1/1>.
- [16] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. Ellis-Monaghan u. a. „Low-power CMOS at $V_{dd} = 4kT/q$ “. In: *Proceedings of the Device Research Conference*. Juni 2001, S. 22–23. DOI: 10.1109/DRC.2001.937856.
- [17] M. Bucher, C. Lallement, C. Enz, F. Théodoloz und F. Krummenacher. *The EPFL-EKV MOSFET Model Equations for Simulation*. Techn. Ber. Lausanne, Schweiz: Electronics Laboratories, Swiss Federal Institute of Technology (EPFL), Juli 1998. URL: http://ekv.epfl.ch/files/content/sites/ekv/files/pdf/ekv_v262.pdf.
- [18] B. H. Calhoun und A. P. Chandrakasan. „A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation“. In: *IEEE Journal of Solid-State Circuits* 42.3 (2007), S. 680–688. DOI: 10.1109/JSSC.2006.891726.
- [19] B. H. Calhoun und A. P. Chandrakasan. „Static noise margin variation for sub-threshold SRAM in 65-nm CMOS“. In: *IEEE Journal of Solid-State Circuits* 41.7 (2006), S. 1673–1679. DOI: 10.1109/JSSC.2006.873215.

-
- [20] B. H. Calhoun und A. P. Chandrakasan. „Ultra-dynamic Voltage scaling (UDVS) using sub-threshold operation and local Voltage dithering“. In: *IEEE Journal of Solid-State Circuits* 41.1 (2006), S. 238–245. doi: 10.1109/JSSC.2005.859886.
- [21] B. H. Calhoun und A. Chandrakasan. „A 256kb Sub-threshold SRAM in 65nm CMOS“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2006, S. 2592–2601. doi: 10.1109/ISSCC.2006.1696325.
- [22] B. H. Calhoun und A. Chandrakasan. „Analyzing static noise margin for sub-threshold SRAM in 65nm CMOS“. In: *Proceedings of the European Solid-State Circuits Conference*. 2005, S. 363–366. doi: 10.1109/ESSCIR.2005.1541635.
- [23] B. H. Calhoun und A. Chandrakasan. „Characterizing and modeling minimum energy operation for subthreshold circuits“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2004, S. 90–95. doi: 10.1109/LPE.2004.1349316.
- [24] B. H. Calhoun und A. Chandrakasan. „Ultra-dynamic voltage scaling using sub-threshold operation and local voltage dithering in 90nm CMOS“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2005, S. 300–599. doi: 10.1109/ISSCC.2005.1493988.
- [25] B. H. Calhoun, D. C. Daly, N. Verma, D. F. Finchelstein, D. D. Wentzloff u. a. „Design considerations for ultra-low energy wireless microsensor nodes“. In: *IEEE Transactions on Computers* 54.6 (2005), S. 727–740. doi: 10.1109/TC.2005.98.
- [26] B. H. Calhoun, A. Wang und A. Chandrakasan. „Device sizing for minimum energy operation in subthreshold circuits“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 2004, S. 95–98. doi: 10.1109/CICC.2004.1358745.
- [27] B. H. Calhoun, A. Wang und A. Chandrakasan. „Modeling and sizing for minimum energy operation in subthreshold circuits“. In: *IEEE Journal of Solid-State Circuits* 40.9 (2005), S. 1778–1786. doi: 10.1109/JSSC.2005.852162.
- [28] B. H. Calhoun, A. Wang, N. Verma und A. Chandrakasan. „Sub-Threshold Design: The Challenges of Minimizing Circuit Energy“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2006, S. 366–368. doi: 10.1109/LPE.2006.4271869.

- [29] Y. Cao, T. Sato, M. Orshansky, D. Sylvester und C. Hu. „New paradigm of predictive MOSFET and interconnect modeling for early circuit simulation“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 2000, S. 201–204. doi: 10.1109/CICC.2000.852648.
- [30] S. Chakraborty, A. Mallik, C. K. Sarkar und V. R. Rao. „Impact of Halo Doping on the Subthreshold Performance of Deep-Submicrometer CMOS Devices and Circuits for Ultralow Power Analog/Mixed-Signal Applications“. In: *IEEE Transactions on Electron Devices* 54.2 (2007), S. 241–248. doi: 10.1109/TED.2006.888630.
- [31] A. Chandrakasan, R. Amirtharajah, S. Cho, J. Goodman, G. Konduri u. a. „Design considerations for distributed microsensor systems“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 1999, S. 279–286. doi: 10.1109/CICC.1999.777291.
- [32] A. Chandrakasan, R. Min, M. Bhardwaj, S. Cho und A. Wang. „Power aware wireless microsensor systems“. In: *Proceedings of the European Solid-State Circuits Conference*. 2002, S. 47–54. doi: 10.1109/ESSCIR.2002.186709.
- [33] I. J. Chang, J.-J. Kim und K. Roy. „Robust Level Converter Design For Sub-Threshold Logic“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. New York, NY, USA: ACM, 2006, S. 14–19. ISBN: 1-59593-462-6. doi: <http://doi.acm.org/10.1145/1165573.1165579>.
- [34] M.-H. Chang, Y.-T. Chiu und W. Hwang. „Design and Iso-Area V_{min} Analysis of 9T Subthreshold SRAM With Bit-Interleaving Scheme in 65-nm CMOS“. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 59.7 (Juli 2012), S. 429–433. ISSN: 1549-7747. doi: 10.1109/TCSII.2012.2198984.
- [35] A. Chavan und E. MacDonald. „Ultra Low Voltage Level Shifters to Interface Sub and Super Threshold Reconfigurable Logic Cells“. In: *Proceedings of the IEEE Aerospace Conference*. März 2008, S. 1–6. doi: 10.1109/AERO.2008.4526473.
- [36] A. Chavan, G. Dukle, B. Gianiello und E. MacDonald. „Robust Ultra-Low Power Subthreshold Logic Flip-Flop Design for Reconfigurable Architectures“. In: *Proceedings of the IEEE International Conference Reconfigurable Computing and FPGAs*. 2006, S. 1–7. doi: 10.1109/RECONF.2006.307764.
- [37] G. K. Chen, D. Blaauw, T. Mudge, D. Sylvester und N. S. Kim. „Yield-driven near-threshold SRAM design“. In: *Proceedings of the IEEE/ACM International Conference Computer-Aided Design*. 2007, S. 660–666. doi: 10.1109/ICCAD.2007.4397341.

- [38] G. Chen, D. Sylvester, D. Blaauw und T. Mudge. „Yield-Driven Near-Threshold SRAM Design“. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 18.11 (2010), S. 1590–1598. DOI: 10.1109/TVLSI.2009.2025766.
- [39] T.-H. Chen, J. Chen und L. T. Clark. „Subthreshold to Above Threshold Level Shifter Design“. In: *Journal of Low Power Electronics* 2.2 (2006), S. 251–258. DOI: 10.1166/jolpe.2006.071.
- [40] K.-H. Cordes, A. Waag und N. Heuck. *Integrierte Schaltungen: Grundlagen – Prozesse – Design – Layout*. Pearson Studium, 2011. ISBN: 978-3-86894-011-4.
- [41] J. Daemen und V. Rijmen. *The Design of Rijndael: AES – The Advanced Encryption Standard*. Springer, 2002. ISBN: 3-540-42580-2.
- [42] J. De Boeck. „Game-changing opportunities for wireless personal healthcare and lifestyle“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 2011, S. 15–21. DOI: 10.1109/ISSCC.2011.5746204.
- [43] M. Dellnitz, O. Schütze und T. Hestermeyer. „Covering Pareto sets by multi-level subdivision techniques“. In: *Journal of optimization theory and applications* 124 (2005), S. 113–136. DOI: 10.1007/s10957-004-6468-7.
- [44] S. D. Deshpande. „Software Implementation of IEEE 802.11b Wireless LAN Standard“. In: *Proceedings of the SDR Technical Conference and Product Exposition*. 2004. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.114.9204>.
- [45] R. G. Dreslinski, B. Zhai, T. Mudge, D. Blaauw und D. Sylvester. „An Energy Efficient Parallel Architecture Using Near Threshold Operation“. In: *Proceedings of the International Conference Parallel Architecture and Compilation Techniques*. 2007, S. 175–188. DOI: 10.1109/PACT.2007.4336210.
- [46] P. Ekdahl und T. Johansson. „SNOW – a new stream cipher“. In: *Proceedings of the First Open Nessie Workshop, KU-Leuven*. 2000. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.6566>.
- [47] C. C. Enz, F. Krummenacher und E. A. Vittoz. „An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications“. In: *Analog Integrated Circuits Signal Processing* 8 (1 Juli 1995), S. 83–114. ISSN: 0925-1030. DOI: 10.1007/BF01239381. URL: <http://portal.acm.org/citation.cfm?id=209019.209030>.
- [48] B. Fu und P. Ampadu. „Comparative Analysis of Ultra-Low Voltage Flip-Flops for Energy Efficiency“. In: *Proceedings of the IEEE International Symposium Circuits and Systems*. 2007, S. 1173–1176. DOI: 10.1109/ISCAS.2007.378259.

- [49] S. Gal-On und M. Levy. *Exploring CoreMark™ – A Benchmark Maximizing Simplicity and Efficacy (Whitepaper)*. URL: <http://www.eembc.org/techlit/whitepaper.php#coremark>.
- [50] C. G. B. Garrett und W. H. Brattain. „Physical Theory of Semiconductor Surfaces“. In: *Physical Review* 99.2 (Juli 1955), S. 376–387. DOI: 10.1103/PhysRev.99.376.
- [51] G. Gerosa, S. Gary, C. Dietz, D. Pham, K. Hoover u. a. „A 2.2 W, 80 MHz superscalar RISC microprocessor“. In: *IEEE Journal of Solid-State Circuits* 29.12 (Dez. 1994), S. 1440–1454. ISSN: 0018-9200. DOI: 10.1109/4.340417.
- [52] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha u. a. „PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation“. In: *IEEE Transactions on Electron Devices* 53.9 (2006), S. 1979–1993. DOI: 10.1109/TED.2005.881006.
- [53] K. Granhaug und S. Aunet. „Six subthreshold full adder cells characterized in 90 nm CMOS technology“. In: *Proceedings of the IEEE Design and Diagnostics of Electronic Circuits and Systems*. 2006, S. 25–30. DOI: 10.1109/DDECS.2006.1649565.
- [54] K. Granhaug, S. Aunet und T. S. Lande. „Body-bias regulator for ultra low power multifunction CMOS gates“. In: *Proceedings of the IEEE International Symposium Circuits and Systems*. 2006. DOI: 10.1109/ISCAS.2006.1692820.
- [55] S. Hanson, M. Seok, Y.-S. Lin, Z. Y. Foo, D. Kim u. a. „A Low-Voltage Processor for Sensing Applications With Picowatt Standby Mode“. In: *IEEE Journal of Solid-State Circuits* 44.4 (2009), S. 1145–1155. DOI: 10.1109/JSSC.2009.2014205.
- [56] S. Hanson, B. Zhai, D. Blaauw, D. Sylvester, A. Bryant u. a. „Energy Optimality and Variability in Subthreshold Design“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2006, S. 363–365. DOI: 10.1109/LPE.2006.4271868.
- [57] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou u. a. „Exploring Variability and Performance in a Sub-200-mV Processor“. In: *IEEE Journal of Solid-State Circuits* 43.4 (2008), S. 881–891. DOI: 10.1109/JSSC.2008.917505.
- [58] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou u. a. „Performance and Variability Optimization Strategies in a Sub-200mV, 3.5pJ/inst, 11nW Subthreshold Processor“. In: *Proceedings of the IEEE Symposium VLSI Circuits*. 2007, S. 152–153. DOI: 10.1109/VLSIC.2007.4342694.

- [59] A. Hasanbegovic und S. Aunet. „Low-power subthreshold to above threshold level shifter in 90 nm process“. In: *Proceedings of the NORCHIP Conference*. 2009, S. 1–4. DOI: 10.1109/NORCHIP.2009.5397793.
- [60] U. Hilleringmann. *Silizium-Halbleitertechnologie*. 5. Auflage. ISBN 978-3-8351-0245-3. Vieweg+Teubner Verlag, Stuttgart, 2008.
- [61] K. Hoffmann. *Systemintegration – Vom Transistor zur großintegrierten Schaltung*. 2. korrigierte und erweiterte Auflage. Oldenbourg Wissenschaftsverlag GmbH, 2006. DOI: 10.1524/9783486593372. URL: <http://www.oldenbourg-link.com/isbn/9783486578942>.
- [62] M.-E. Hwang, A. Raychowdhury, K. Kim und K. Roy. „A 85mV 40nW Process-Tolerant Subthreshold 8x8 FIR Filter in 130nm Technology“. In: *Proceedings of the IEEE Symposium VLSI Circuits*. 2007, S. 154–155. DOI: 10.1109/VLSIC.2007.4342695.
- [63] S. Jain, S. Khare, S. Yada, V. Ambili, P. Salihundam u. a. „A 280mV-to-1.2V wide-operating-range IA-32 processor in 32nm CMOS“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 2012, S. 66–68. DOI: 10.1109/ISSCC.2012.6176932.
- [64] M. Jamal Deen, M. Kazemeini und S. Naseh. „Ultra-low power VCOs - performance characteristics and modeling (invited)“. In: *Proceedings of the IEEE International Caracas Conference on Devices, Circuits and Systems*. 2002, S. C033–1–C033–8. DOI: 10.1109/ICDCS.2002.1004021.
- [65] S. C. Jocke, J. F. Bolus, S. N. Wooters, T. N. Blalock und B. H. Calhoun. „A 2.6uW sub-threshold mixed-signal ECG SoC“. In: *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*. ISLPED '09. San Francisco, CA, USA: ACM, 2009, S. 117–118. ISBN: 978-1-60558-684-7. DOI: 10.1145/1594233.1594260. URL: <http://doi.acm.org/10.1145/1594233.1594260>.
- [66] T. Jungeblut. *Entwurfsraumexploration ressourceneffizienter VLIW-Prozessoren*. Dissertation. Universität Bielefeld, 2011.
- [67] T. Jungeblut, B. Hübener, M. Pormann und U. Rückert. „A Systematic Approach for Optimized Bypass Configurations for Application-specific Embedded Processors“. In: *ACM Transactions on Embedded Computing Systems* (2013). Zur Veröffentlichung angenommen.

- [68] T. Jungeblut, G. Sievers, M. Pormann und U. Rückert. „Design Space Exploration for Memory Subsystems of VLIW Architectures“. In: *Proceedings of the IEEE International Networking, Architecture and Storage Conference*. 2010, S. 377–385. DOI: 10.1109/NAS.2010.14.
- [69] M. Kakoei und L. Benini. „Fine-Grained Power and Body-Bias Control for Near-Threshold Deep Sub-Micron CMOS Circuits“. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 1.2 (Juni 2011), S. 131–140. ISSN: 2156-3357. DOI: 10.1109/JETCAS.2011.2159285.
- [70] H. Kalte, M. Pormann und U. Rückert. „A Prototyping Platform for Dynamically Reconfigurable System on Chip Designs“. In: *Proceedings of the IEEE Workshop Heterogeneous Reconfigurable Systems on Chip*. 2002.
- [71] K. Kang, S. P. Park, K. Kim und K. Roy. „On-Chip Variability Sensor Using Phase-Locked Loop for Detecting and Correcting Parametric Timing Failures“. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 18.2 (2010), S. 270–280. DOI: 10.1109/TVLSI.2008.2010399.
- [72] J. T. Kao, M. Miyazaki und A. R. Chandrakasan. „A 175-MV multiply-accumulate unit using an adaptive supply voltage and body bias architecture“. In: *IEEE Journal of Solid-State Circuits* 37.11 (2002), S. 1545–1554. DOI: 10.1109/JSSC.2002.803957.
- [73] T. Kaulmann. *Ressourceneffiziente Realisierung Pulsodierter Neuronaler Netze*. Dissertation. Bd. 266. HNI-Verlagsschriftenreihe, Paderborn. Heinz Nixdorf Institut, 2009.
- [74] R. Keyes und T. Watson. „On Power Dissipation in Semiconductor Computing Elements“. In: *Proceedings of the IRE*. 50 Bde. 1962, S. 2485.
- [75] R. Keyes. „Fundamental limits in digital information processing“. In: *Proceedings of the IEEE* 69.2 (Feb. 1981), S. 267–278. ISSN: 0018-9219. DOI: 10.1109/PROC.1981.11959.
- [76] K. K. Kim und Y.-B. Kim. „Optimal Body Biasing for Minimum Leakage Power in Standby Mode“. In: *Proceedings of the IEEE International Symposium Circuits and Systems*. 2007, S. 1161–1164. DOI: 10.1109/ISCAS.2007.378256.
- [77] T.-H. Kim, H. Eom, J. Keane und C. Kim. „Utilizing Reverse Short Channel Effect for Optimal Subthreshold Circuit Design“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2006, S. 127–130. DOI: 10.1109/LPE.2006.4271820.

-
- [78] T.-H. Kim, J. Keane, H. Eom und C. H. Kim. „Utilizing Reverse Short-Channel Effect for Optimal Subthreshold Circuit Design“. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 15.7 (2007), S. 821–829. doi: 10.1109/TVLSI.2007.899239.
- [79] T.-H. Kim, J. Liu, J. Keane und C. H. Kim. „A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing“. In: *IEEE Journal of Solid-State Circuits* 43.2 (2008), S. 518–529. doi: 10.1109/JSSC.2007.914328.
- [80] T.-H. Kim, J. Liu, J. Keane und C. H. Kim. „A High-Density Subthreshold SRAM with Data-Independent Bitline Leakage and Virtual Ground Replica Scheme“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2007, S. 330–606. doi: 10.1109/ISSCC.2007.373428.
- [81] T.-H. Kim, J. Liu und C. H. Kim. „A Voltage Scalable 0.26 V, 64 kb 8T SRAM With V_t Lowering Techniques and Deep Sleep Mode“. In: *IEEE Journal of Solid-State Circuits* 44.6 (2009), S. 1785–1795. doi: 10.1109/JSSC.2009.2020201.
- [82] T.-H. Kim, J. Liu und C. H. Kim. „A voltage scalable 0.26V, 64kb 8T SRAM with V_{min} lowering techniques and deep sleep mode“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 2008, S. 407–410. doi: 10.1109/CICC.2008.4672106.
- [83] T.-H. Kim, J. Liu und C. H. Kim. „An 8T Subthreshold SRAM Cell Utilizing Reverse Short Channel Effect for Write Margin and Read Performance Improvement“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 2007, S. 241–244. doi: 10.1109/CICC.2007.4405723.
- [84] P. M. Kogge und H. S. Stone. „A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations“. In: *IEEE Transactions on Computers* C-22.8 (Aug. 1973), S. 786–793. ISSN: 0018-9340. doi: 10.1109/TC.1973.5009159.
- [85] J. P. Kulkarni, K. Kim und K. Roy. „A 160 mV Robust Schmitt Trigger Based Subthreshold SRAM“. In: *IEEE Journal of Solid-State Circuits* 42.10 (2007), S. 2303–2313. doi: 10.1109/JSSC.2007.897148.
- [86] J. P. Kulkarni, K. Kim und K. Roy. „A 160 mV, fully differential, robust schmitt trigger based sub-threshold SRAM“. In: *Proceedings of the ACM/IEEE International Low Power Electronics and Design*. 2007, S. 171–176. doi: 10.1145/1283780.1283818.
- [87] K. Kurbjun und C. Ribbing. *Xilinx: PLL Dynamic Reconfiguration, XAPP879 (v1.1)*. Okt. 2011.

- [88] J. Kwong und A. P. Chandrakasan. „Variation-Driven Device Sizing for Minimum Energy Sub-threshold Circuits“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2006, S. 8–13. doi: 10.1109/LPE.2006.4271799.
- [89] J. Kwong, Y. K. Ramadass, N. Verma und A. P. Chandrakasan. „A 65 nm Sub-Vt Microcontroller With Integrated SRAM and Switched Capacitor DC-DC Converter“. In: *IEEE Journal of Solid-State Circuits* 44.1 (2009), S. 115–126. doi: 10.1109/JSSC.2008.2007160.
- [90] J. Kwong, Y. Ramadass, N. Verma, M. Koesler, K. Huber u. a. „A 65nm Sub-Vt Microcontroller with Integrated SRAM and Switched-Capacitor DC-DC Converter“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2008, S. 318–616. doi: 10.1109/ISSCC.2008.4523185.
- [91] F. Leuenberger und E. Vittoz. „Complementary-MOS low-power low-voltage integrated binary counter“. In: *Proceedings of the IEEE* 57.9 (Sep. 1969), S. 1528–1532. issn: 0018-9219. doi: 10.1109/PROC.1969.7331.
- [92] N. C. Li, G. L. Haviland und A. A. Tuszynski. „CMOS tapered buffer“. In: *IEEE Journal of Solid-State Circuits* 25.4 (1990), S. 1005–1008. doi: 10.1109/4.58293.
- [93] D. Liu und C. Svensson. „Trading speed for low power by choice of supply and threshold voltages“. In: *IEEE Journal of Solid-State Circuits* 28.1 (Jan. 1993), S. 10–17. issn: 0018-9200. doi: 10.1109/4.179198.
- [94] Z. Liu und V. Kursun. „Characterization of a Novel Nine-Transistor SRAM Cell“. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 16.4 (Apr. 2008), S. 488–492. issn: 1063-8210. doi: 10.1109/TVLSI.2007.915499.
- [95] N. Lotze, J. Goppert und Y. Manoli. „Timing modeling for digital sub-threshold circuits“. In: *Proceedings of the Design Automation & Test in Europe Conference & Exhibition*. 2010, S. 299–302.
- [96] N. Lotze und Y. Manoli. „A 62mV 0.13 um CMOS standard-cell-based design technique using schmitt-trigger logic“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 2011, S. 340–342. doi: 10.1109/ISSCC.2011.5746345.
- [97] N. Lotze, M. Ortmanns und Y. Manoli. „Variability of flip-flop timing at sub-threshold voltages“. In: *Proceedings of the ACM/IEEE International Low Power Electronics and Design*. 2008, S. 221–224. doi: 10.1145/1393921.1393979.
- [98] C.-Y. Lu und J. M. Sung. „Reverse short-channel effects on threshold voltage in submicrometer salicide devices“. In: *IEEE Electron Device Letters* 10.10 (1989), S. 446–448. doi: 10.1109/55.43095.

-
- [99] P. Meinerzhagen, S. Sherazi, A. Burg und J. Rodrigues. „Benchmarking of Standard-Cell Based Memories in the Sub- VT Domain in 65-nm CMOS Technology“. In: *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on* 1.2 (Juni 2011), S. 173–182. ISSN: 2156-3357. DOI: 10.1109/JETCAS.2011.2162159.
- [100] V. S. Miller. „Use of elliptic curves in cryptography“. In: *Lecture notes in computer sciences; 218 on Advances in cryptology – CRYPTO 85*. New York, NY, USA: Springer-Verlag New York, Inc., 1986, S. 417–426. ISBN: 0-387-16463-4. URL: <http://dl.acm.org/citation.cfm?id=18262.25413>.
- [101] R. Min, M. Bhardwaj, S.-H. Cho, E. Shih, A. Sinha u. a. „Low-power wireless sensor networks“. In: *Proceedings of the IEEE International VLSI Design Conference*. 2001, S. 205–210. DOI: 10.1109/ICVD.2001.902661.
- [102] M. Miura-Mattausch, M. Suetake, H. J. Mattausch, S. Kumashiro, N. Shigyo u. a. „Physical modeling of the reverse-short-channel effect for circuit simulation“. In: *IEEE Transactions on Electron Devices* 48.10 (2001), S. 2449–2452. DOI: 10.1109/16.954493.
- [103] M. Miyazaki, J. Kao und A. P. Chandrakasan. „A 175 mV multiply-accumulate unit using an adaptive supply voltage and body bias (ASB) architecture“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Bd. 1. 2002, S. 58–444. DOI: 10.1109/ISSCC.2002.992937.
- [104] J. Montanaro, R. Witek, K. Anne, A. Black, E. Cooper u. a. „A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor“. In: *IEEE Journal of Solid-State Circuits* 31.11 (Nov. 1996), S. 1703–1714. ISSN: 0018-9200. DOI: 10.1109/JSSC.1996.542315.
- [105] F. Moradi, D. T. Wisland, S. Aunet, H. Mahmoodi und T. V. Cao. „65NM sub-threshold 1T-SRAM for ultra low voltage applications“. In: *Proceedings of the IEEE International SoC Conference*. 2008, S. 113–118. DOI: 10.1109/SOCC.2008.4641491.
- [106] F. Moradi, D. T. Wisland, H. Mahmoodi, S. Aunet, T. V. Cao u. a. „Ultra low power full adder topologies“. In: *Proceedings of the IEEE International Symposium Circuits and Systems*. 2009, S. 3158–3161. DOI: 10.1109/ISCAS.2009.5118473.
- [107] F. Moradi, D. T. Wisland, H. Mahmoodi, A. Peiravi, S. Aunet u. a. „New subthreshold concepts in 65nm CMOS technology“. In: *Proceedings of the Quality of Electronic Design Conference*. 2009, S. 162–166. DOI: 10.1109/ISQED.2009.4810287.

- [108] T. H. Morshed, W. Yang, M. V. Dunga, X. Xi, J. He u. a. *BSIM4.6.4 MOSFET Model*. Techn. Ber. Department of Electrical Engineering und Computer Sciences, University of California, Berkeley, 2009.
- [109] L. Nazhandali, B. Zhai, A. Olson, A. Reeves, M. Minuth u. a. „Energy optimization of subthreshold-voltage sensor network processors“. In: *Proceedings of the International Symposium Computer Architecture*. 2005, S. 197–207. DOI: 10.1109/ISCA.2005.26.
- [110] C. Neau und K. Roy. „Optimal body bias selection for leakage improvement and process compensation over different technology generations“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2003, S. 116–121. DOI: 10.1109/LPE.2003.1231846.
- [111] N. Nedovic und V. Oklobdzija. „Hybrid latch flip-flop with improved power efficiency“. In: *Proceedings of the IEEE Symposium on Integrated Circuits and Systems Design*. 2000, S. 211–215. DOI: 10.1109/SBCCI.2000.876032.
- [112] W. Oberschelp und G. Vossen. *Rechneraufbau und Rechnerstrukturen*. Oldenbourg-Verlag, 2000.
- [113] S. Oesterle, P. Gerrish und P. Cong. „New interfaces to the body through implantable-system integration“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 2011, S. 9–14. DOI: 10.1109/ISSCC.2011.5746203.
- [114] A. Ono, R. Ueno und I. Sakai. „TED control technology for suppression of reverse narrow channel effect in 0.1 μm MOS devices“. In: *Proceedings of the International Electron Devices Meeting*. 1997, S. 227–230. DOI: 10.1109/IEDM.1997.650353.
- [115] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio u. a. „Flow-through latch and edge-triggered flip-flop hybrid elements“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Feb. 1996, S. 138–139. DOI: 10.1109/ISSCC.1996.488543.
- [116] M. Pelgrom, A. Duinmaijer und A. Welbers. „Matching properties of MOS transistors“. In: *IEEE Journal of Solid-State Circuits* 24.5 (Okt. 1989), S. 1433–1439. ISSN: 0018-9200. DOI: 10.1109/JSSC.1989.572629.
- [117] W. Peterson und D. Brown. „Cyclic Codes for Error Detection“. In: *Proceedings of the IRE* 49.1 (Jan. 1961), S. 228–235. ISSN: 0096-8390. DOI: 10.1109/JRPROC.1961.287814.

-
- [118] M. Porrman, J. Hagemeyer, C. Pohl, J. Romoth und M. Strugholtz. „RAPTOR — A Scalable Platform for Rapid Prototyping and FPGA-based Cluster Computing“. In: *Parallel Computing: From Multicores and GPU's to Petascale, Advances in Parallel Computing*. Bd. 19. IOS press, 2010, S. 592–599. ISBN: 978-1-60750-529-7.
- [119] M. Porrman, J. Hagemeyer, J. Romoth und M. Strugholtz. „Rapid Prototyping of Next-Generation Multiprocessor SoCs“. In: *Proceedings of the Semiconductor Conference Dresden*. Dresden, 2009.
- [120] Y. Pu, J. P. de Gyvez, H. Corporaal und Y. Ha. „An ultra-low-energy/frame multi-standard JPEG co-processor in 65nm CMOS with sub/near-threshold power supply“. In: *Proceedings of the IEEE International Solid-State Circuits Conference*. 2009, S. 146–147. DOI: 10.1109/ISSCC.2009.4977350.
- [121] Y. Pu, J. Pineda de Gyvez, H. Corporaal und Y. Ha. „An Ultra-Low-Energy Multi-Standard JPEG Co-Processor in 65 nm CMOS With Sub/Near Threshold Supply Voltage“. In: *IEEE Journal of Solid-State Circuits* 45.3 (2010), S. 668–680. DOI: 10.1109/JSSC.2009.2039684.
- [122] C. S. Rafferty, H.-H. Vuong, S. A. Eshraghi, M. D. Giles, M. R. Pinto u. a. „Explanation of reverse short channel effect by defect gradients“. In: *Proceedings of the International Electron Devices Meeting*. 1993, S. 311–314. DOI: 10.1109/IEDM.1993.347345.
- [123] Y. K. Ramadass und A. P. Chandrakasan. „A Battery-Less Thermoelectric Energy Harvesting Interface Circuit With 35 mV Startup Voltage“. In: *IEEE Journal of Solid-State Circuits* 46.1 (2011), S. 333–341. DOI: 10.1109/JSSC.2010.2074090.
- [124] Y. K. Ramadass und A. P. Chandrakasan. „A batteryless thermoelectric energy-harvesting interface circuit with 35mV startup voltage“. In: *Proceedings of the IEEE International Solid-State Circuits Conference*. 2010, S. 486–487. DOI: 10.1109/ISSCC.2010.5433835.
- [125] Y. K. Ramadass und A. P. Chandrakasan. „An Efficient Piezoelectric Energy Harvesting Interface Circuit Using a Bias-Flip Rectifier and Shared Inductor“. In: *IEEE Journal of Solid-State Circuits* 45.1 (2010), S. 189–204. DOI: 10.1109/JSSC.2009.2034442.
- [126] Y. K. Ramadass und A. P. Chandrakasan. „Minimum Energy Tracking Loop With Embedded DC–DC Converter Enabling Ultra-Low-Voltage Operation Down to 250 mV in 65 nm CMOS“. In: *IEEE Journal of Solid-State Circuits* 43.1 (2008), S. 256–265. DOI: 10.1109/JSSC.2007.914720.

- [127] Y. K. Ramadass und A. P. Chandrakasan. „Minimum Energy Tracking Loop with Embedded DC-DC Converter Delivering Voltages down to 250mV in 65nm CMOS“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2007, S. 64–587. doi: 10.1109/ISSCC.2007.373589.
- [128] L. Richardson. *The H.264 Advanced Video Compression Standard*. Wiley, 2010.
- [129] J. F. Ryan und B. H. Calhoun. „A sub-threshold FPGA with low-swing dual-VDD interconnect in 90nm CMOS“. In: *Proceedings of the IEEE Custom Integrated Circuits Conference*. 2010, S. 1–4. doi: 10.1109/CICC.2010.5617466.
- [130] O. Schütze. „A New Data Structure for the Nondominance Problem in Multi-Objective Optimization“. In: *Proceedings of EMO 2003*. 2003.
- [131] O. Schütze. „Set Oriented Methods for Global Optimization“. Diss. University of Paderborn, Germany, 2004. url: <http://digital.lib.uni-paderborn.de/ubpb/urn/urn:nbn:de:hbz:466-20040101418>.
- [132] O. Schütze, A. Dell’Aere und M. Dellnitz. „On Continuation Methods for the Numerical Treatment of Multi-Objective Optimization Problems“. In: *Practical Approaches to Multi-Objective Optimization*. Hrsg. von J. Branke, K. Deb, K. Miettinen und R. E. Steuer. Dagstuhl Seminar Proceedings 04461. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2005.
- [133] E. Seevinck, F. J. List und J. Lohstroh. „Static-noise margin analysis of MOS SRAM cells“. In: *IEEE Journal of Solid-State Circuits* 22.5 (1987), S. 748–754. doi: 10.1109/JSSC.1987.1052809.
- [134] M. Seok, G. Chen, S. Hanson, M. Wieckowski, D. Blaauw u. a. „CAS-FEST 2010: Mitigating Variability in Near-Threshold Computing“. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 1.1 (März 2011), S. 42–49.
- [135] M. Seok, D. Blaauw und D. Sylvester. „Robust Clock Network Design Methodology for Ultra-Low Voltage Operations“. In: *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on* 1.2 (Juni 2011), S. 120–130. issn: 2156-3357. doi: 10.1109/JETCAS.2011.2160753.
- [136] M. Seok, S. Hanson, Y.-S. Lin, Z. Foo, D. Kim u. a. „The Phoenix Processor: A 30pW platform for sensor applications“. In: *Proceedings of the IEEE Symposium VLSI Circuits*. 2008, S. 188–189. doi: 10.1109/VLSIC.2008.4586001.
- [137] H. M. Sheffer. „A set of five independent postulates for Boolean algebras, with application to logical constants“. In: *Transactions of the American Mathematical Society* 14 (1913), S. 481–488.

-
- [138] M. E. Sinangil, N. Verma und A. P. Chandrakasan. „A 45nm 0.5V 8T column-interleaved SRAM with on-chip reference selection loop for sense-amplifier“. In: *Proceedings of the IEEE Asian Solid-State Circuits Conference*. 2009, S. 225–228. doi: 10.1109/ASSCC.2009.5357219.
- [139] M. E. Sinangil, N. Verma und A. P. Chandrakasan. „A reconfigurable 65nm SRAM achieving voltage scalability from 0.25–1.2V and performance scalability from 20kHz–200MHz“. In: *Proceedings of the European Solid-State Circuits Conference*. 2008, S. 282–285. doi: 10.1109/ESSCIRC.2008.4681847.
- [140] M. E. Sinangil, N. Verma und A. P. Chandrakasan. „A Reconfigurable 8T Ultra-Dynamic Voltage Scalable (U-DVS) SRAM in 65 nm CMOS“. In: *IEEE Journal of Solid-State Circuits* 44.11 (2009), S. 3163–3173. doi: 10.1109/JSSC.2009.2032493.
- [141] A. K. Singh. *Digital VLSI Design*. Prentice-Hall of India Pvt.Ltd., 2010. ISBN: 978-8120341876.
- [142] C. Subramanian, J. Hayden, W. Taylor, M. Orlowski und T. McNelly. „Reverse short channel effect and channel length dependence of boron penetration in PMOSFETs“. In: *Proceedings of the International Electron Devices Meeting*. 1995, S. 423–426. doi: 10.1109/IEDM.1995.499229.
- [143] Y. Suzuki, K. Odagawa und T. Abe. „Clocked CMOS calculator circuitry“. In: *IEEE Journal of Solid-State Circuits* 8.6 (Dez. 1973), S. 462–469. ISSN: 0018-9200. doi: 10.1109/JSSC.1973.1050440.
- [144] R. M. Swanson und J. D. Meindl. „Ion-implanted complementary MOS transistors in low-voltage circuits“. In: *IEEE Journal of Solid-State Circuits* 7.2 (1972), S. 146–153. doi: 10.1109/JSSC.1972.1050260.
- [145] R. Swanson. *Complementary MOS transistors in micropower circuits*. Defense Technical Information Center, 1974. URL: <http://books.google.de/books?id=IEgFAAAAIAAJ>.
- [146] V. Sze und A. Chandrakasan. „A 0.4-V UWB baseband processor“. In: *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*. Aug. 2007, S. 262–267. doi: 10.1145/1283780.1283837.
- [147] Texas Instruments: SN5474, SN54LS74A, SN54S74, SN7474, SN74LS74A, SN74S74, *Dual D-Type Positive-Edge-Triggered Flip-Flops with Preset and Clear*. März 1988.

- [148] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis u. a. „Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. Bd. 1. 2002, S. 422–478. DOI: 10.1109/ISSCC.2002.993112.
- [149] J. Tschanz, S. Narendra, R. Nair und V. De. „Effectiveness of adaptive supply voltage and body bias for reducing impact of parameter variations in low power and high performance microprocessors“. In: *Proceedings of the IEEE Symposium VLSI Circuits*. 2002, S. 310–311. DOI: 10.1109/VLSIC.2002.1015112.
- [150] M.-H. Tu, J.-Y. Lin, M.-C. Tsai, C.-Y. Lu, Y.-J. Lin u. a. „A Single-Ended Disturb-Free 9T Subthreshold SRAM With Cross-Point Data-Aware Write Word-Line Structure, Negative Bit-Line, and Adaptive Read Operation Timing Tracing“. In: *IEEE Journal of Solid-State Circuits* 47.6 (Juni 2012), S. 1469–1482. ISSN: 0018-9200. DOI: 10.1109/JSSC.2012.2187474.
- [151] N. Verma und A. P. Chandrakasan. „A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy“. In: *IEEE Journal of Solid-State Circuits* 43.1 (2008), S. 141–149. DOI: 10.1109/JSSC.2007.908005.
- [152] N. Verma und A. P. Chandrakasan. „A High-Density 45 nm SRAM Using Small-Signal Non-Strobed Regenerative Sensing“. In: *IEEE Journal of Solid-State Circuits* 44.1 (2009), S. 163–173. DOI: 10.1109/JSSC.2008.2006428.
- [153] N. Verma und A. P. Chandrakasan. „A High-Density 45nm SRAM Using Small-Signal Non-Strobed Regenerative Sensing“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2008, S. 380–621. DOI: 10.1109/ISSCC.2008.4523216.
- [154] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag u. a. „A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System“. In: *IEEE Journal of Solid-State Circuits* 45.4 (2010), S. 804–816. DOI: 10.1109/JSSC.2010.2042245.
- [155] N. Verma und A. P. Chandrakasan. „A 65nm 8T Sub-Vt SRAM Employing Sense-Amplifier Redundancy“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2007, S. 328–606. DOI: 10.1109/ISSCC.2007.373427.
- [156] A. Viterbi. „Error bounds for convolutional codes and an asymptotically optimum decoding algorithm“. In: *IEEE Transactions on Information Theory* 13.2 (Apr. 1967), S. 260–269. ISSN: 0018-9448. DOI: 10.1109/TIT.1967.1054010.

-
- [157] E. Vittoz und J. Fellrath. „CMOS analog integrated circuits based on weak inversion operations“. In: *IEEE Journal of Solid-State Circuits* 12.3 (1977), S. 224–231. doi: 10.1109/JSSC.1977.1050882.
- [158] A. Wang, A. P. Chandrakasan und S. V. Kosonocky. „Optimal supply and threshold scaling for subthreshold CMOS circuits“. In: *Proceedings of the IEEE Computer Society Annual Symposium*. 2002, S. 5–9. doi: 10.1109/ISVLSI.2002.1016866.
- [159] A. Wang und A. Chandrakasan. „A 180-mV subthreshold FFT processor using a minimum energy design methodology“. In: *IEEE Journal of Solid-State Circuits* 40.1 (2005), S. 310–319. doi: 10.1109/JSSC.2004.837945.
- [160] A. Wang und A. Chandrakasan. „A 180mV FFT processor using subthreshold circuit techniques“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2004, S. 292–529. doi: 10.1109/ISSCC.2004.1332709.
- [161] A. Wang, B. H. Calhoun und A. P. Chandrakasan. *Sub-threshold Design for Ultra Low-Power Systems (Series on Integrated Circuits and Systems)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN: 0387335153.
- [162] J.-S. Wang, P.-Y. Chang, T.-S. Tang, J.-W. Chen und J.-I. Guo. „Design of Subthreshold SRAMs for Energy-Efficient Quality-Scalable Video Applications“. In: *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on* 1.2 (Juni 2011), S. 183–192. ISSN: 2156-3357. doi: 10.1109/JETCAS.2011.2158345.
- [163] R. P. Weicker. „Dhrystone: a synthetic systems programming benchmark“. In: *Communications ACM* 27.10 (Okt. 1984), S. 1013–1030. ISSN: 0001-0782. doi: 10.1145/358274.358283. URL: <http://doi.acm.org/10.1145/358274.358283>.
- [164] R. Weinstein. „RFID: a technical overview and its application to the enterprise“. In: *IT Professional* 7.3 (Mai 2005), S. 27–33. ISSN: 1520-9202. doi: 10.1109/MITP.2005.69.
- [165] Xilinx: *Spartan-6 Family Overview, DS160 (v1.4)*. März 2010.
- [166] Xilinx: *Spartan-6 FPGA Clocking Resources, User Guide, UG382 (v1.3.1)*. Apr. 2010.
- [167] Xilinx: *Spartan-6 FPGA Data Sheet: DC and Switching Characteristics, DS162 (v1.4)*. März 2010.
- [168] Xilinx: *Spartan-6 FPGA SelectIO Resources, User Guide, UG381 (v1.3)*. März 2010.
- [169] Xilinx: *Virtex-5 Family Overview, DS100 (v5.0)*. Feb. 2009.

- [170] *Xilinx: Virtex-II Platform FPGA User Guide, UG002 (v2.2)*. Nov. 2007.
- [171] B. Zhai, D. Blaauw, D. Sylvester und S. Hanson. „A Sub-200mV 6T SRAM in 0.13 μ m CMOS“. In: *Digest of Technical Papers of the IEEE International Solid-State Circuits Conference*. 2007, S. 332–606. DOI: 10.1109/ISSCC.2007.373429.
- [172] B. Zhai, R. G. Dreslinski, D. Blaauw, T. Mudge und D. Sylvester. „Energy efficient near-threshold chip multi-processing“. In: *Proceedings of the ACM/IEEE International Low Power Electronics and Design*. 2007, S. 32–37. DOI: 10.1145/1283780.1283789.
- [173] B. Zhai, S. Hanson, D. Blaauw und D. Sylvester. „A Variation-Tolerant Sub-200 mV 6-T Subthreshold SRAM“. In: *IEEE Journal of Solid-State Circuits* 43.10 (2008), S. 2338–2348. DOI: 10.1109/JSSC.2008.2001903.
- [174] B. Zhai, S. Hanson, D. Blaauw und D. Sylvester. „Analysis and mitigation of variability in subthreshold design“. In: *Proceedings of the International Symposium Low Power Electronics and Design*. 2005, S. 20–25. DOI: 10.1109/LPE.2005.195479.
- [175] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth u. a. „A 2.60pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency“. In: *Digest of Technical Papers of the IEEE VLSI Circuits Symposium*. 2006, S. 154–155. DOI: 10.1109/VLSIC.2006.1705356.
- [176] B. Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson u. a. „Energy-Efficient Subthreshold Processor Design“. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 17.8 (2009), S. 1127–1137. DOI: 10.1109/TVLSI.2008.2007564.
- [177] E. Zitzler und L. Thiele. „Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach“. In: *IEEE Transactions on Evolutionary Computation* 3.4 (1999), S. 257–271. DOI: 10.1109/4235.797969.

Eigene Veröffentlichungen

- [178] M. Blesken, S. Lütke-meier und U. Rückert. „Multiobjective optimization for transistor sizing of sub-threshold CMOS logic standard cells“. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*. 2010, S. 1480–1483. doi: 10.1109/ISCAS.2010.5537349.
- [179] T. Jungeblut, S. Lütke-meier, G. Sievers, M. Pormann und U. Rückert. „A modular design flow for very large design space explorations“. In: *Proceedings of the CDNLive! EMEA*. Mai 2010. URL: <http://pub.uni-bielefeld.de/publication/2286628>.
- [180] S. Lütke-meier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Pormann u. a. „A 65 nm 32 b Subthreshold Processor with 9T Multi-Vt SRAM and Adaptive Supply Voltage Control“. In: *IEEE Journal of Solid-State Circuits* 48 (1 Jan. 2013), S. 8–19. doi: 10.1109/JSSC.2012.2220671.
- [181] S. Lütke-meier, T. Jungeblut, M. Pormann und U. Rückert. „A 200mV 32b Subthreshold Processor with Adaptive Supply Voltage Control“. In: *Proceedings of the IEEE International Solid-State Circuits Conference*. 2012, S. 484–485.
- [182] S. Lütke-meier, T. Kaulmann und U. Rückert. „A Sub-200mV 32bit ALU with 0.45pJ/instruction in 90nm CMOS“. In: *Proceedings of the Semiconductor Conference Dresden*. Dresden, 2009.
- [183] S. Lütke-meier und U. Rückert. „A Subthreshold to Above-Threshold Level Shifter Comprising a Wilson Current Mirror“. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 57.9 (Sep. 2010), S. 721–724.

Betreute Arbeiten

- [184] M. Vohrmann. *Analyse und Optimierung der Ressourceneffizienz einer UMTS-Implementierung für einen eingebetteten Multiprozessor*. Diplomarbeit. Paderborn: Fachgruppe Schaltungstechnik, Universität Paderborn, Mai 2012.
- [185] M. Vohrmann. *Implementierung einer Performance and Power Management Unit*. Studienarbeit. Paderborn: Fachgruppe Schaltungstechnik, Universität Paderborn, Juni 2011.