

13 Das Zeitverhalten

Dieses Buch behandelt in erster Linie den Weg vom Algorithmus, d. h. vom Anwendungsproblem, zum Silizium. Hierbei in diesem Text sind Ziele wie die Erreichung höchster Schaltgeschwindigkeiten weitgehend ausgeklammert, da diese größtenteils eher zum Bereich der Elektrotechnik gehören, denn zur Informatik. Deshalb werden in diesem Kapitel nur sehr einfache Modelle eingeführt, die allenfalls eine grobe Abschätzung erlauben. In bestimmten Anwendungen (wie beispielsweise Mikroprozessoren und Signalprozessoren) ist höchster Durchsatz, also bestmögliches Zweitverhalten, jedoch ein primäres Ziel, wofür hier ein erheblicher Anteil des Entwurfsaufwandes aufgebracht werden muß. Wir können hier die folgenden Gebiete unterscheiden: (1) *Charakterisierung* des Zeitverhaltens von Bauelementen, Schaltungen, Subsystemen und Systemen (durch Berechnungen, Abschätzungen, Simulation oder Messungen, s. Abschnitt 13.1); (2) *konstruktive Maßnahmen* zur Durchsatz-Verbesserung (Abschnitt 13.5 bis 13.6).

Letztere Kategorie zerfällt in schaltungstechnische Maßnahmen (schnellere Schaltstufen, gekürzte kritische Leitungen etc., vgl. Abschn. 13.5.1, 13.5.2, 13.3.2), eine günstigere Gehäuse-Technik (Abschnitt 13.3.3), oder eine bessere (schnellere) Technologie: entweder eine neuere Version einer bereits vorhandenen Technologie (Skalierung: beispielsweise von 1,2 µm zu einer 0,8-µm-Technologie, s. Abschnitt 13.2), oder die Umstellung auf eine andersartige Technologie (beispielsweise BiCMOS, bipolar, oder Gallium-Arsenid statt MOS, s. u. a. [25] [42]). Die Darstellungen dieses Kapitels berühren hauptsächlich die Schaltkreis-Ebene (s. Bild 13.1).

13.1 Das Zeitverhalten aktiver Bauelemente

Die Beschreibung der elektrischen Eigenschaften einer Schaltung und des durch diese beeinflussten Zeitverhaltens wird als "Charakterisierung" bezeichnet. Die Schaltzeiten der einzelnen Transistoren beeinflussen ganz erheblich das Verhalten einer gesamten Schaltung. (Lediglich

13.1 Das Zeitverhalten aktiver Bauelemente.....	269
13.1.1 Das Zeitverhalten von MOSFETs	271
13.2 Skalierung von Layout.....	273
13.3 Das Zeitverhalten von Interkonnekt	275
13.3.1 Schätzung der Kapazität des Interkonnekt	276
13.3.2 Interkonnekt-Modellierung.....	276
13.3.3 Multi-Chip-Module (MCMs).....	277
13.4 Das Zeitverhalten von Schaltkreisen	279
13.4.1 Anwendung von Schaltkreis-Simulatoren	280
13.5 Maßnahmen zur Verbesserung des Zeitverhaltens	281
13.5.1 Der "Super-Buffer".....	284
13.5.2 Pad-Treiber	284
13.6 Schnelle Integrierte Schaltungen	285
13.6.2 Schnelle Bipolar-Schaltungen	286
13.6.1 BiCMOS-Schaltungen	285
13.7 Literatur	286

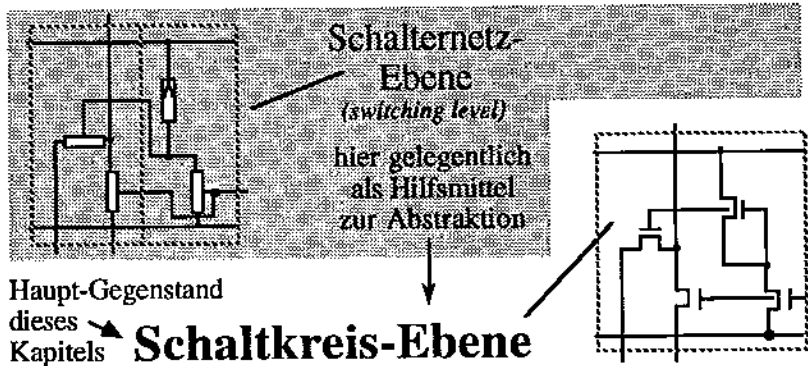


Bild 13.1: Durch dieses Kapitel berührte Abstraktionsebene(n).

bei sehr schnellen Schaltungen (s. Abschnitt 13.6) können die Eigenschaften nicht-aktiver Elemente (des Interkonnekt, s. Abschnitt 13.3) dominieren beim Zustandekommen des Verhaltens der Gesamtschaltung.) Wir beschränken uns hier zur Modellierung des Schaltverhaltens aktiver Elemente auf die Anwendung naive Modelle, zumal dieses Buch eher den VLSI-System-Implementierer, weniger jedoch den reinen Elektrotechniker anspricht. Genauere Modelle sind in der Literatur zu finden, wie beispielsweise in [1] [13] [27] [33] [41] [42] [46], für Bipolar-Transistoren u. a. in [10].

Wir modellieren das Transistor-Verhalten über RC-Glieder (s.a.[31]) aus Widerstand (R) und parasitärer Kapazität (C). Bild 13.3 ist eine Art Kompakt-Repetitorium über das RC-Glied. Das RC-Glied-Modell eignet sich insbesondere für die grobe Modellierung von MOS-Transistoren [29]. Die Widerstände, die hauptsächlich als Transistor-Eigenschaften auftreten, wurden bereits in Kapitel 6 behandelt. Nun sollen die parasitären Kapazitäten behandelt werden.

In Bild 13.2 a ist ein Teil einer Schaltung zu sehen. Logisch gesehen handelt es sich dabei um einen Inverter (der Eingang ist x), der einen Transistor ansteuert. In Bild 13.2 b ist die gleiche Schaltung mit Widerständen und Kapazitäten dargestellt. R_{pu} ist der Widerstand des Pullup-Transistors, R_{pd} der des Pulldown-Transistors. Die angedeutete Kapazität C ist die parasitäre Kapazität der Leitung und des Gatters der angesteuerten Schaltungsstufe. Diese Kapazität muß umgeladen werden, wenn der Eingang des Inverters seinen Wert ändert. Wird x von logisch "0" auf "1" gesetzt, so wird C über den Widerstand R_{pd} des pulldown-Transistors

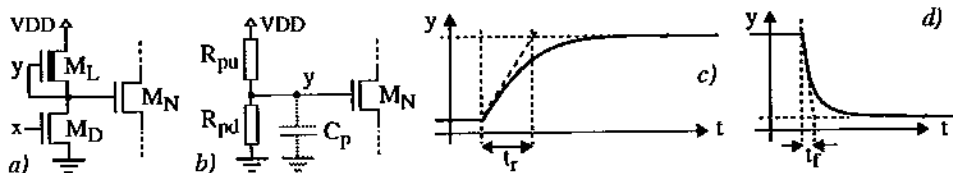


Bild 13.2: Charakterisierung des Zeitverlaufs einer Inverterschaltung: a) Transistorschaltung, b) Ersatzschaltbild: MOS-Schaltung und ihr Ersatzschaltbild, c) Anstiegszeit, d) Abfallzeit.

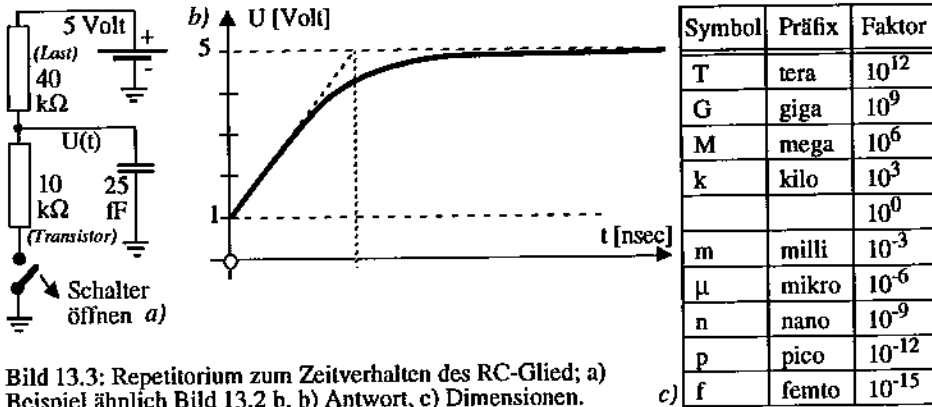


Bild 13.3: Repetitorium zum Zeitverhalten des RC-Glied; a) Beispiel ähnlich Bild 13.2 b, b) Antwort, c) Dimensionen.

entladen, im umgekehrten Fall über den Widerstand R_{pu} des Pullup-Transistors aufgeladen. Die Schaltzeit t errechnet sich aus $R \cdot C$, d.h.:

$$t_{\text{entladen}} = R_{pd} \cdot C \quad \text{und} \quad t_{\text{aufladen}} = R_{pu} \cdot C$$

Da der Widerstand des Pullup-Transistors größer ist als der des Pulldowns, dauert es länger die Kapazität aufzuladen, als sie zu entladen. Der Faktor, um den die Anstiegszeit länger ist als die Abfallzeit, entspricht dem Widerstandsverhältnis k (Bild 13.2 c, d).

Hiermit läßt sich auch die Zeitverzögerung eines **Inverterpaares** charakterisieren, d.h. also die Schaltzeitverzögerung die zwei hintereinander geschaltete Inverter verursachen. Diese Inverterpaar-Verzögerung beträgt $t + kt = (k+1)t$, da bei einem der Inverter die Kapazität entladen (Schaltzeit t) und beim anderen aufgeladen (Schaltzeit $k t$) werden muß. Die gesamte Schaltzeitverzögerung ergibt sich dann aus der Addition der beiden Einzelverzögerungen (Bild 13.4).

13.1.1 Das Zeitverhalten von MOSFETs

Durch die Steuerung des Kanalstromes über ein elektrisches Feld können MOSFETs leistungslos gesteuert werden, d.h. es muß nach Beendigung des Schaltvorganges kein Strom I_{gate} fließen, um einen FET durchgeschaltet zu halten. Der MOSFET ist also spannungsgesteuert (vgl. Übersicht in Bild 6.4). Gate und Substrat bilden mit der isolierenden (hier allerdings dickeren) Oxidschicht einen Kondensator, der aufgeladen keinen Stromfluß über das Gate mehr zuläßt. Zum Aufbau des Feldes wird allerdings Energie benötigt, d.h. beim Schaltvorgang selbst fließt zunächst ein Ladestrom. Neben der Kapazität des Transistors in der Praxis der Anwendung kommt noch die Kapazität der Signalleitung gegen Masse hinzu, da sie ja ebenfalls durch eine Oxidschicht isoliert über dem Substrat verläuft und somit ebenfalls einen langen, dünnen Kondensator darstellt. Zu C_{gate} kommt noch die Interkonnekt-Kapazität C_{Int} hinzu, was zusammen die Last-Kapazität C_L ergibt.

Diese Überlegung ergibt ein einfaches Modell [29]: Der Strom I_{ds} von Source zu Drain ergibt sich aus der zum Kanal induzierten Ladung dividiert durch die Übergangszeit τ , oder:

$$\text{Strom } I_{ds} = \frac{\text{induzierte Ladung}}{\text{durchschnittliche Übergangszeit für ein Elektron von S zu D}} \quad (13.1)$$

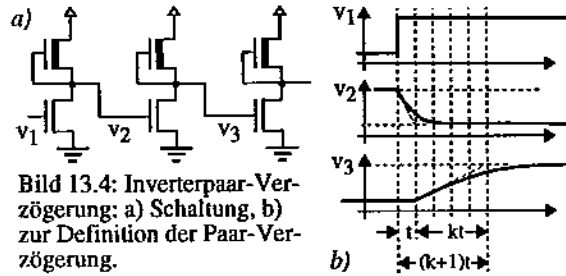


Bild 13.4: Inverterpaar-Verzögerung: a) Schaltung, b) zur Definition der Paar-Verzögerung.

oder:

$$\text{Übergangszeit } \tau = \frac{\text{Entfernung } L}{\text{Geschwindigkeit } v} \quad (13.2)$$

Etwas formaler ausgedrückt ergibt dies:

$$\tau = \frac{L}{v} = \frac{L}{\mu \cdot E} = \frac{L^2}{\mu \cdot V_{DS}} \quad (13.3)$$

wobei μ die Mobilität der Ladungsträger ist (hier: Elektronen) unter dem Einfluß eines elektrischen Feldes E in der Kanal-Gegend; genauer gesagt, die Mobilität je Feldstärke-Einheit ($\text{cm}^2/\text{Volt-Sekunden}$). Die im Übergang befindliche Ladungsmenge Q ist gerade die Gate-Kapazität C_{gate} mal der Gate-Source-Spannung jenseits der Schwellenspannung V_{th} , also:

$$Q = C_{\text{gate}} (V_{\text{gs}} - V_{\text{th}}) = \epsilon \frac{WL}{D} (V_{\text{gs}} - V_{\text{th}}) \quad (13.4)$$

wobei ϵ die Dielektrizitätskonstante des Gate-Oxids ist, $A=WL$ die Fläche der planparallelen "Kondensator-Platten", und D die Dicke des Gate-Oxids. (Dabei ist ϵ definiert als die Kapazität in Farad/cm, wenn $A = 1 \text{ cm}^2$ und $D = 1 \text{ cm}$ ist.) Deshalb ist $C_{\text{gate}} = \epsilon WL/D$. Daraus folgt:

$$I_{\text{ds}} = \epsilon \mu \frac{W}{DL} (V_{\text{gs}} - V_{\text{th}}) \cdot V_{\text{ds}} \quad (13.5)$$

Man beachte, daß für kleine V_{ds} (dort, wo der eigentliche Schaltvorgang stattfindet) I_{ds} proportional zu $(V_{\text{ds}} - V_{\text{th}})$ ist - ein Verhalten, das dem eines Widerstandes R gleicht mit:

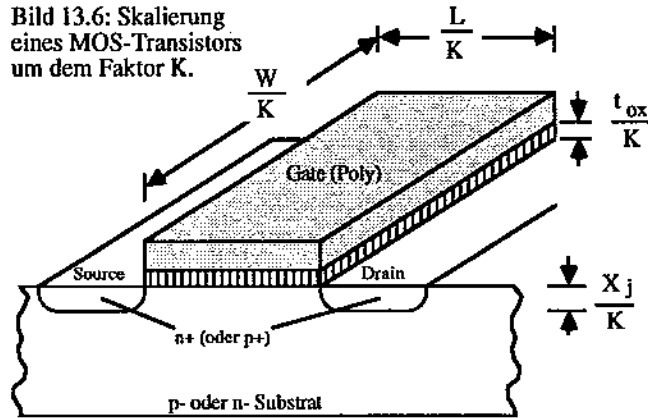
$$\frac{V_{\text{ds}}}{I_{\text{ds}}} = R = \frac{L^2}{\mu C_{\text{gate}} (V_{\text{gs}} - V_{\text{th}})} \quad (13.6)$$

Parameter	≈1980	≈1990
minimale Linienbreite (minimum feature size) in μm	≈5	≈0,5
τ in ns	≈0,5	≈0,05
System-Takt in ns	≈25	≈2,5

Bild 13.5: Übergangszeiten τ für eine modernere Technologie und eine ältere Technologie.



Bild 13.6: Skalierung eines MOS-Transistors um dem Faktor K .



Wenn wir hiermit einen weit offenen Transistor treiben für den die Gate-Kapazität C_{gate} vorliegt, so ergibt sich für den Schaltvorgang die folgende Zeitkonstante RC mit:

$$RC_{\text{gate}} = \frac{L^2}{\mu (V_{\text{gs}} - V_{\text{th}})} \quad (13.7)$$

Diese Übergangszeit $\tau = RC_{\text{gate}}$ kann als die elementare Zeiteinheit zur Geschwindigkeits-Abschätzung für jegliches als IC entworfene nMOS-System verwendet werden. Die schnellste mögliche Operation ist die ideale Übertragung eines Signals vom Gate eines kleinsten MOS-Transistors zum Gate eines anderen kleinsten Transistors, d. h. die Übertragung über idealen, also verzögerungsfreien Interkonnekt. Die Tabelle in Bild 13.5 gibt überschlägliche Werte von τ an für eine $5\text{-}\mu\text{-}$ Technologie und eine $0,5\text{-}\mu\text{-}$ Technologie. Genauere Ergebnisse sind erzielbar durch die Anwendung eines Schaltkreis-Simulators erzielt werden ([37] [49], Abschn. 13.4.1).

13.2 Skalierung von Layout

Ein sehr wichtiger Faktor in der raschen Weiterentwicklung der Technologie integrierter Schaltungen [9] [26] war die zunehmende Verkleinerung der Dimensionen der einzelnen Elemente des Layouts. Es soll hier keine detaillierte Analyse der durch diese Skalierung hervorgerufenen Effekte durchgeführt werden. Details zum Thema Skalierbarkeit sind u.A. in [42] nachzulesen. Dieses Kapitel soll vielmehr einen davon Eindruck vermitteln, welchen Nutzen die Skalierung in den verschiedenen Bereichen bringt und welcher Aufwand dazu nötig ist.

Hierzu wird das wohl gebräuchlichste Modell zugrundegelegt, welches davon ausgeht, daß durch eine Skalierung die in der Schaltung vorhandenen Felder unverändert bleiben. Dies bedeutet, daß alle Dimensionen (Breite, Länge, Dicke) in linear verändert werden müssen. Im Einzelnen bedeutet dies, daß die Länge und Breite jedes Schaltungselements (Transistor-Kanal, Kontaktloch (cut), Leiterbahn) durch einen festen Faktor K dividiert werden (Bild 13.6). Dies erhöht die Zahl der auf einem Chip integrierbaren Elemente um den Faktor K^2 . Die Reduzierung der Betriebsspannungen um den selben Faktor K vermindert die in der Schaltung vorhan-

Material	minimaler Wert	maximaler Wert
Gate (Poly über Kanal)	$4,0 \cdot 10^{-1}$	$5,0 \cdot 10^{-1}$
Poly über Dickoxid	$0,4 \cdot 10^{-1}$	$0,6 \cdot 10^{-1}$
Metall über Dickoxid	$0,4 \cdot 10^{-1}$	$0,6 \cdot 10^{-1}$
Metall über "field"	$0,15 \cdot 10^{-1}$	$0,3 \cdot 10^{-1}$
Metall über Diff	$0,8 \cdot 10^{-1}$	$1,0 \cdot 10^{-1}$
n-Diff	$0,8 \cdot 10^{-1}$	$1,0 \cdot 10^{-1}$
n-Kanal	$0,8 \cdot 10^{-1}$	$1,0 \cdot 10^{-1}$
p Diff	$7,0 \cdot 10^{-1}$	$9,0 \cdot 10^{-1}$
p-Kanal	$6,0 \cdot 10^{-1}$	$8,0 \cdot 10^{-1}$

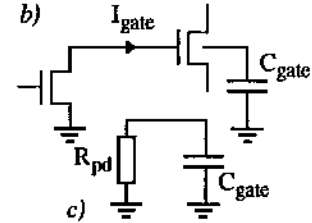


Bild 13.7: typische Kapazitätswerte ($\approx 5\text{-}\mu\text{-Proz.}$) in $\text{fF} / \mu\text{m}^2$ (a), b) einf. MOS-Transistor-Modell, c) Ableitung von (b) - Widerstandswerte s. .

denen Ströme um K und verringert die Leistungsabgabe der beteiligten Elemente um K^2 , wodurch die Leistungsaufnahme des gesamten Chips konstant bleibt.

Der wichtigste Effekt für einen einzelnen Transistor besteht (bis ca. $K = 0,1 \mu\text{m}$) in der Beschleunigung seiner Schaltzeiten um den Faktor K durch die Verringerung seiner Kanal-Länge um K . Unterhalb etwa $0,1 \mu\text{m}$ bis $0,2 \mu\text{m}$ nimmt die Bedeutung anderer Effekte zu (Rand-Effekten, Tunnelung von Elektronen, u. a.), auf die jedoch hier nicht weiter eingegangen wird.

Zunehmende Dominanz des Interkonnekt-Verhaltens [32] [48]. Ein anderer wichtiger Punkt ist die Skalierung von Verbindungen. Der Leiter-Querschnittsfläche verringert sich um den Faktor K^2 . Da gleichzeitig die Länge des Leiters um K abnimmt, erhöht sich dessen Widerstand um K . Die Kapazität des Leiters verringert sich um K , wodurch die RC-Konstante (Maß für die Signallaufzeit) unverändert bleibt. Die Leitungsverzögerung erhöht sich relativ zur Schaltzeit um den Faktor K . Ab ca. 1μ wird das Interkonnektverhalten dominant (vgl. Bild 13.5 und Beispiel in Bild 13.8). Der Vorteil von lokaler Kommunikation zwischen Teilen eines Chips tritt deshalb gegenüber langen Verbindungen globaler Verdrahtung in den Vordergrund.

Auch der Fertigungsprozeß wird von der Skalierung unmittelbar betroffen. Die Größe der Fläche, bei der ein Defekt im Kristall wahrscheinlich ohne Auswirkungen bleibt, verringert sich und damit vergrößert sich die Zahl der von einem Defekt betroffenen Elemente. Will man die

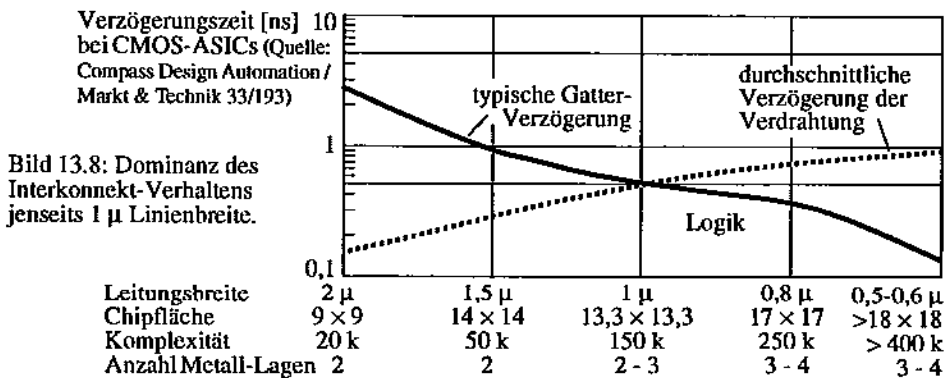


Bild 13.8: Dominanz des Interkonnekt-Verhaltens jenseits 1μ Linienbreite.

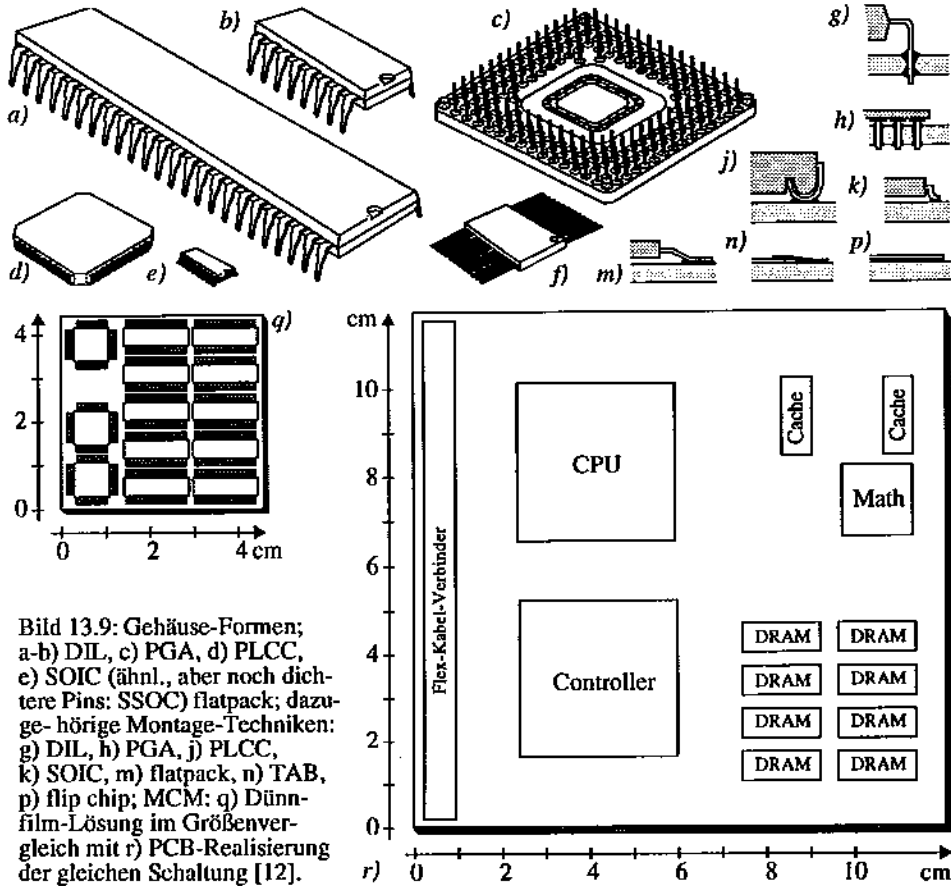


Bild 13.9: Gehäuse-Formen; a-b) DIL, c) PGA, d) PLCC, e) SOIC (ähnlich, aber noch dichtere Pins: SSOC) flatpack; dazugehörige Montage-Techniken: g) DIL, h) PGA, j) PLCC, k) SOIC, m) flatpack, n) TAB, p) flip chip; MCM: q) Dünnschicht-Lösung im Größenvergleich mit r) PCB-Realisierung der gleichen Schaltung [12].

Prozentzahl der Ausbeute beibehalten, sind zusätzliche Maßnahmen in Bezug auf den Fertigungsprozeß und den Reinheitsgrad erforderlich. Ein anderer Ausweg ist die zunehmende Verwendung fehlertoleranter Entwürfe, die ebenfalls zu einer erhöhten Ausbeute führen, aber nicht zum Thema dieses Kapitels gehören.

13.3 Das Zeitverhalten von Interkonnect

Mit abnehmender minimaler Linienbreite wird der Einfluß des Interkonnect-Verhaltens auf das Gesamt-Verhalten einer Schaltung immer bedeutender [2] [51]. Bei konservativen Schaltungs-Entwürfen genügt oft die Berücksichtigung der durch den Interkonnect eingebrachten parasitären Kapazitäten, etwa mittels eines einfachen RC-Glied-Modells (Abschnitt 13.3.1) oder über die Anwendung von Schaltkreis-Simulatoren (13.4.1). Bei Technologien unterhalb von ca. $1 \mu\text{m}$ wird das Interkonnect-Verhalten sogar dominant [51]. Insbesondere bei hohen Durch-

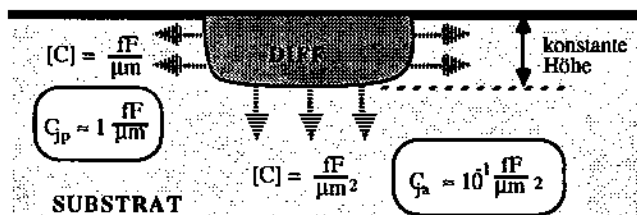


Bild 13.10: Kapazitätsberechnung bei Aktivbereichen (Diffusion).

satz-Anforderungen müssen dann anspruchsvollere Verfahren für die Verhaltens-Modellierung des Interkonnect angewandt werden (13.3.2).

13.3.1 Schätzung der Kapazität des Interkonnect

Die Schaltzeit t eines Transistors beträgt unter Belastung durch die eines getriebenen Transistors bei einer Technologie von $5 \mu m$ ca. $0,5 \text{ nsec}$ (vgl. Tabelle in Bild 13.5). Zwischen die beiden Inverter sei ein Transfer-Transistor geschaltet: d.h. $k=8$ und somit $(1+k)=10$. Als Zwischenergebnis ergibt sich: $(1+k)t=10 \cdot 0,5 \text{ nsec} \approx 5 \text{ nsec}$. Sonstige Lasten vergrößern die Schaltzeit vielleicht nochmal um das 2- bis 5-fache. Dies ergäbe also eine ungefähre Schaltzeit des Inverterpaares von 10 bis 25 nsec.

Im obigen Beispiel haben wir nur grob geschätzt, wie parasitäre Kapazitäten in die Schaltzeit mit eingehen ohne deren genaue Größe zu kennen. Es stellt sich also die Frage, wie Kapazitäten abgeschätzt bzw. berechnet werden. Dabei muß man einen kleinen Unterschied machen zwischen der Berechnung der Kapazität von Diffusion und der anderen Layer. Bei Diffusion muß die Kapazität der Fläche und des Umfanges gegenüber dem Substrat abgeschätzt werden, während es bei allen anderen Layern nur auf die Fläche ankommt.

Der Grund dafür ist, daß die Diffusion in das Substrat hineinreicht (Bild 13.10) und somit außer mit der Grundfläche, auch noch mit dem Rand eine Kapazität gegen das Substrat bildet. Die Tabelle in Bild 13.7 gibt einige typische kapazitive Werte für die Layer an. Die Werte sind alle in Femto-Farad (10^{-15} F) pro mm^2 angegeben (vgl. Tabelle in Bild 13.3 c).

13.3.2 Interkonnect-Modellierung

Bei Technologien mit einer minimalen Linienbreite (feature size) etwa unterhalb $1 \mu m$ beginnt das Verhalten des Interkonnect das Gesamtverhalten der Schaltung zu dominieren, wohingegen der Einfluß der aktiven Devices zurückgeht oder gar ganz vernachlässigt werden kann. Mit anderen Worten ausgedrückt bedeutet dies: die Transistoren sind so schnell, daß die durch diese hervorgerufenen Verzögerungen gering sind oder gar vernachlässigt werden können. Dadurch kommt der Charakterisierung des Interkonnect eine größere oder gar ausschließliche Bedeutung zu. Bei sehr schnellen Devices wird sogar der relative Einfluß der Induktivität von Leitungen so bedeutend, daß reine RC-Modellierung (Abschnitt 13.3.1) nicht mehr ausreicht. Bei noch schnelleren Schaltungen muß sogar auf Methoden der Mikrowellen-Technik (evtl. Hochfrequenz-Streifenleiter oder Mikrostreifenleiter) zurückgegriffen werden. Es werden für sehr schnelle Schaltungen also erheblich anspruchsvollere Verfahren der Modellierung benö-

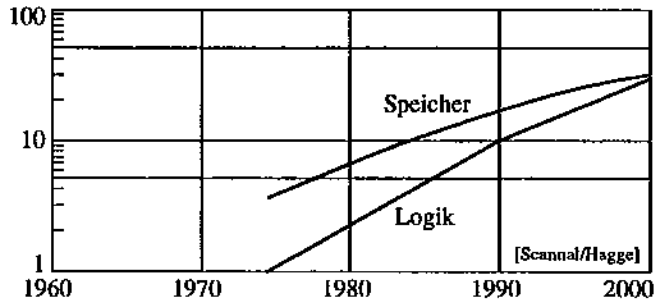


Bild 13.11: Wachsende maximale Chip-Dimension (Millimeter).

tigt, womit die Interkonnect-Modellierung eine eigenständige Disziplin geworden ist [2] [34] [35] [39] [51].

13.3.3 Multi-Chip-Module (MCMs)

Multi-Chip-Module sind Produkte moderner Montage-Technik (packaging technique) in der Mikroelektronik [4] [5] [6]. Montage-Technik für die Mikroelektronik [3] [32] [45] [47] ist übrigens ein wichtiger Markt-Faktor. So wird beispielsweise bis zum Jahre 2001 eine Verdoppelung des Weltmarktes für gedruckte Schaltungen erwartet von gegenwärtig 8,2 Mrd. US-\$ auf insgesamt 19,4 Mrd. \$ [36]. Diese Techniken sind beispielsweise auch das größte und wichtigste Fachgebiet auf der jährlich im Juni stattfindenden "SMT/ASIC/Hybrid", einer Fachmesse in Nürnberg mit 400 Direkt-Ausstellern und 21000 m² Ausstellungsfläche [36]. Einen wichti-

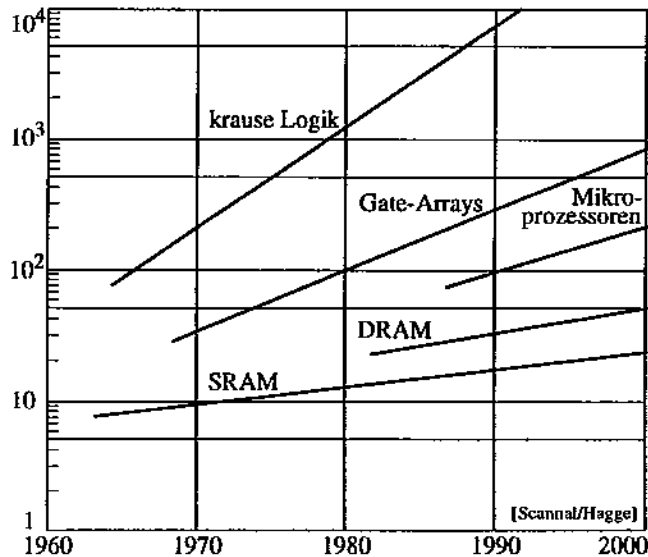


Bild 13.12: Steigender Bedarf an Anschluß-Pins (I/O terminals).

Einführung	Art	Pin-Abstand [mm]
mit Bohrung:		
1963	DIL	2,54
Oberflächen-montiert:		
1979	EIAJ-SOIC	1,27
1982	PLCC	0,635
1884	EIAJ-SSOP	0,635
1989	SOIC	0,5
1993	beherrschbar ¹	0,4
	Studie ²	0,3

1) OEMs und Betreiber von Bestückungs-Automaten
 2) Hersteller von Bestückungs-Automaten [21]

Bild 13.13: Zur Entwicklung der Gehäuse-Formen für integrierte Digitalschaltungen.

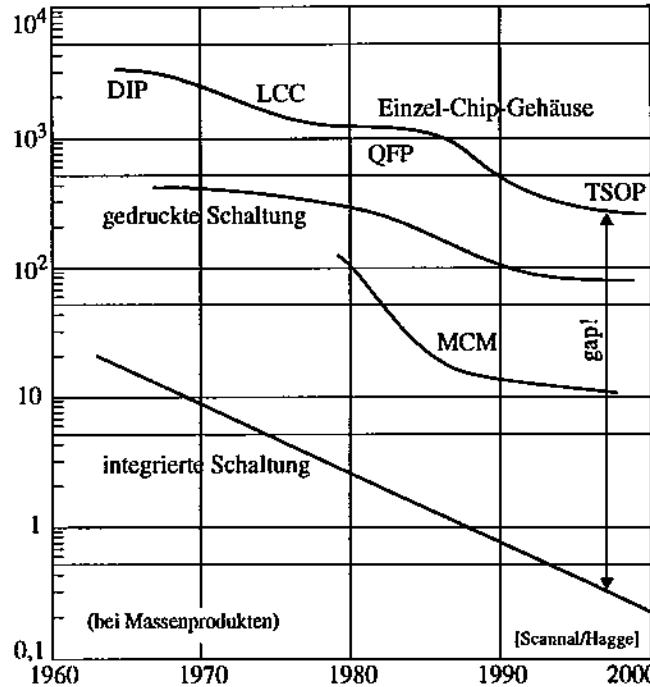


Bild 13.14: Entwicklung der minimalen Leitungsbreite in μm .

gen Teilmarkt der Montage-Technik bildet die damit zusammenhängende Fertigungstechnik, zu welcher beispielsweise Bestückungsmaschinen gehören (Marktvolumen in Europa: 0,5 Mrd. \$). Leistungen von Bestückungsmaschinen reichen übrigens bis zu 12 500 Bauteile pro Stunde. Dies sind 7 Bauteile in 2 Sekunden, bei einer Genauigkeit von 0,4 mm je Pin [36] (vgl. auch Bild 13.13).

Die traditionelle Gehäuse-Technik schließt ICs in ein Einzelchip-Gehäuse ein (Bild 13.9a-f), montiert auf einer gedruckten Schaltung (PCB). Kritisch ist der Einfluß dieser Technik auf das Zeitverhalten: eine Kombination aus Anstiegszeiten der Signale aus den ICs und den Verzögerungen durch Signal-Laufzeiten außerhalb der ICs. Ein Ausweg ist die Minimierung der Abstände zwischen den ICs, sowie der Widerstände und parasitären Kapazitäten der Leitungen. Das Problem wurde (trotz verringerter Gehäusegrößen: Bild 13.9 a-f und Bild 13.13) gravierender durch die zunehmende Geometrie-Lücke zwischen ICs und PCBs (vgl. Bild 13.14).

Neben verbesserten PCB-Techniken (Bild 13.16) wie SMT (surface mount technique) und FPP (fine-pitch PCBs: Bild 13.13 zeigt die Tendenz zur Verringerung der Pin-Abstände) bieten sich MCMs (Multi-Chip-Module) als neue Gehäuse-Technik an [16]: mehrere Chips werden in einem gemeinsamen Gehäuse montiert und verdrahtet. Der über Einzelchip-Gehäuse oft nicht realisierbare hohe Pin-Bedarf bei VLSI (Bild 13.12) wegen wachsender Chip-Größe (Bild 13.11) und Integrationsdichte (Bild 1.10) kann über MCM-Technik gedeckt werden, indem ein Großteil der Pins nur noch intern von Chip zu Chip geht und daher praktisch Fanout-frei ist.



DIP	dual-in-line package	P/I	packaging / interconnect
DSR	double-sided rigid PWB	PLCC	plastic chip carrier package
EIAJ	Electronic Industry Association of Japan	PMCM	programmable MCM
FCP	few chip packaging	PQFP	plastic QFP
FP-MCM	fully programmable MCM	PWB	printed wiring board
FPP	fine-pitch PCB	PTH	print-through hole (gelocht)
HDI	high density interconnection	QFP	quad flat package
HIC	hybrid (integrated) circuit	QTAI	quick turn-around interconnection
IC	integrated circuit	SCP	single chip packaging
JEDEC	Joint Electron Device Engineering Council	SMT	surface mount technology
KGD	known good die	SOIC	small outline IC
LCC	lead chip carriers	SOP	small outline package
MCM	multi chip module	SP-MCM	semi programmable MCM
MCM-C	MCM, ceramic	SPPD	signal processor packaging design
MCM-D	MCM, deposited	SQFP	shrink QFP
MCM-L	MCM, laminated	SSOP	shrink SOP
MLB	multi-layer board	SWAP	size, weight, and power
MLB-SM	surface-mount MLB	TAB	tape-automated bonding
MLB-TH	thru-hole MLB	TFP	thin film on polyimide
MLC	multilayer ceramic	TSOP	thin small-outline package
PCB	printed circuit board, PWB	TSSOP	thin shrink SOP
		WSI	wafer scale integration

Bild 13.15: Abkürzungen aus dem Bereich der Gehäuse-Technik.

Bei Multi-Chip-Moduln (MCMs) ist Interkonnekt ein Kostenfaktor [16] [22] [23] [30]. Mehrschicht-Dünnschicht-Substraten erlauben bei MCMs eine wesentlich höhere Verdrahtungs-Dichte (vgl. z. B. Bild 13.9 q und r), in dem eine Angleichung der Anschluß-Raster möglich wird (0,1 - 0,3 mm bei ICs, 0,6 - 2,5 mm bei PCBs [44]). Dies zusammen mit der Freiheit von Fanout hat erhebliche Vorteile nicht nur bezüglich der Geschwindigkeit (Bild 13.16). Bei der MCM-Version des SPPD-Projekt wurde im Vergleich mit einer 4 PCBs umfassenden älteren Version [44] beispielsweise die durchschnittliche Signalfadlänge von Chip zu Chip von 20 bis 40 cm auf 8 bis 20 cm verringert. Nicht nur das SPPD-Projekt zeigt, daß der MCM-Entwurf nebst der dazu nötigen Partitionierungs-Strategie einen erheblichen Anteil des gesamten Entwurfs-Aufwandes einnimmt.

13.4 Das Zeitverhalten von Schaltkreisen

Das gesamte Verhalten einer Schaltung wird beeinflußt vom Transistor-Verhalten und vom Interkonnekt-Verhalten. Das Transistor-Verhalten wird in Abschnitt 13.1 behandelt. Die Veränderung des Transistor-Verhaltens und des Gesamtverhaltens einer Schaltung durch Skalierung wird in Abschnitt 13.2 behandelt. Das Interkonnekt-Verhalten ist Gegenstand von Abschnitt 13.3. Das Gesamtverhalten einer Schaltung ist im hier vorliegenden Abschnitt dargestellt. Bei Technologien bis $\approx 1 \mu$ genügt meist die Modellierung einer Gesamt-Schaltung (beispielsweise eines Gatters) mit der Last-Kapazität C_L (Beispiel in Bild 13.18 a mit parasitären Kapazitäten

Eigenschaft (relativ)	gelochte PCBs	surface mount PCB	fine pitch PCB	MCM
Größe	10	4	2	1
Gewicht	6	3	1,5	1
Signal-Verzögerung	3	2	1,5	1
Temperatur-Vorteil	0	0	0	- 20°C
Zuverlässigkeit	0,25	0,2	0,25	1
IC-Pin-Raster (mm)	2,54	1,27	0,5	0,3

Bild 13.16: MCM im Vergleich mit anderen Gehäuse-Techniken (bei gleichem Chip-Satz [44]).

C_A , C_B , und C_{out}). Hierzu sei auf Abschnitt 13.1 und 13.3 verwiesen. Bei größeren Genauigkeits-Ansprüchen, insbesondere auf Detail-genaue Kurvenformen der Zeitfunktionen, ist die Verwendung eines Schaltkreis-Simulators zu empfehlen, eingeführt in Abschnitt 13.4.1 .

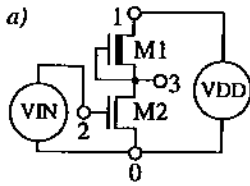
13.4.1 Anwendung von Schaltkreis-Simulatoren

Relativ genau läßt sich das Zeitverhalten von Transistor-Schaltkreisen durch die Anwendung von Schaltkreis-Simulatoren ermitteln (wenn die verwendeten Technologie-Parameter stimmen). Die Genauigkeit erstreckt sich auch auf die detaillierte Wiedergabe der Kurvenformen mit Überschwängern, Unterschwingern und Spikes etc. (vgl. V(3) in Bild 13.17 f). Es wird deutlich, daß die e-Funktion der RC-Glied-Modellierung eine grobe Näherung ist (beispielsweise in Bild 13.2 und Bild 13.3). Der bekannteste Schaltkreis-Simulator ist SPICE [37] [49]. SPICE steht für "Simulation Program with Integrated Circuit Emphasis". Diese Version von SPICE bietet je nach Genauigkeits-Ansprüchen drei verschiedene Ebenen ("Level 1" bis "Level 3").

Bild 13.17 zeigt ein einfaches Ein- und Ausgabe-Beispiel für eine SPICE-Simulation eines nMOS-Inverters (Schaltplan in Bild 13.17 a). Die in den Simulator einzugebende Beschreibung beginnt mit den technologischen Parametern: beim Beispiel in Bild 13.17 b ein Modell NENH für selbstsperrende nMOS-Transistoren NMOS, und ein Modell NDEP für selbstleitende nMOS-Transistoren NMOS). Je nach "Level" werden bis zu mehr als 50 Parameter benötigt [49] (eine Liste auch in [13], S. 271 - 280). Bei der meist üblichen Form der Zusammenarbeit von VLSI-System-Entwicklern mit Herstellern über *Silicon-Broker* werden diese Parameter als "black box" quasi unbesehen übernommen und in die SPICE-Eingabe eingesetzt.

Bild 13.17 c zeigt die hinter die Parameter einzusetzende Netzliste, d. h. die textuelle Form von Bild a (der Stimuli-Spannungs-Generator VIN gehört eigentlich nicht mehr zur Schaltung). Das Format sei am Beispiel der ersten Zeile erläutert: M1 ist der Transistor-Name, die vier Nummern (hier: 1 3 3 0) zeigen, an welche Netz-Knoten der Transistor angeschlossen ist (vgl. Knoten-Nummern 1, 3, 0 in Bild a, die vierte Nummer gibt an, wo das Substrat des Transistors angeschlossen ist). NDEP verweist auf das Modell (vgl. Bild b). W = 5U und L = 20U geben Breite (width) und Länge (length) des Kanal in μm an (U, da μ kein ASCII-Zeichen ist). VDD ist die Spannungsversorgung.

Bild 13.17 d gibt die Simulations-Kommandos an. VIN beschreibt einen Stimuli-Impuls-Spannungs-Generator (Impulsform V(2) in Bild f). Kommando .TRAN 1NS 40NS gibt an, daß eine Transienten-Analyse von der 0. Nanosekunde bis zur 40. Nanosekunde mit einem Zeitschritt von 1 ns durchgeführt werden soll. Das Kommando .PRINT TRAN V(2) V(3) besagt, daß der Verlauf der Spannungen V(2) und V(3), also an Knoten 2 und 3 aus dieser Transienten-Analyse

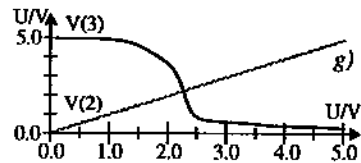
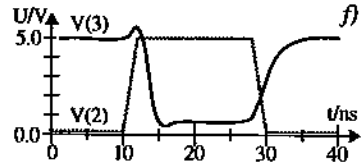


b) `.MODEL NENH NMOS`
`....(technologische Parameter).....`
`.MODEL NDEP NMOS`
`....(technologische Parameter).....`

`VIN 2 0 PULSE(0 5 10NS 2NS 2NS 20NS 40NS)`
`.TRAN INS 40NS`
`.PRINT TRAN V(2) V(3)`
`.END`

`M1 1 3 3 0 NDEP W=5U L=20U`
`M2 3 2 0 0 NENH W=10U L=5U`
`VDD 1 0 5`

`VIN 2 0 DC 5`
`.DC VIN 0 5 0.25`
`.PRINT DC V(2) V(3)`
`.END`



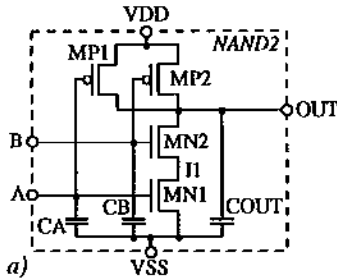
ausgedrückt werden soll (Ergebnis ist der Plot in Bild 13.17 f). Bild 13.17 e und g zeigen als Beispiel Simulations-Kommandos und Ergebnis-Plot für eine Gleichspannungs-Analyse (DC steht für *direct current*).

Bild 13.18 zeigt die Netzliste (Bild b) einer anderen Schaltung (CMOS-NAND-Schaltung; Bild a) mit den Transistor-Modellen NFET (nMOS-Transistor) und PFET (pMOS-Transistor) und den parasitären Kapazitäten CA, CB, und COUT. Solche Kapazitäten sind wichtig für die Ermittlung des Zeitverhaltens einer Schaltung durch eine SPICE-Simulation. Das Beispiel in Bild 13.18 zeigt die Fähigkeit von SPICE, Hierarchien darzustellen. Die Zeilen `.SUBCKT` (für "subcircuit") und `.ENDS` (für "end subcircuit") als *header* und *tail* rahmen die Netzliste ein, sodaß eine Zelle NAND2 aus einer höheren Ebene referenziert werden kann. VDD, VSS, A, B, OUT (vgl. *header*, d. h. erste Zeile, in Bild 13.18 b) sind Ports der Außenansicht (external view) der Zelle NAND2 (vgl. Rähmchen in Bild 13.18 a).

13.5 Maßnahmen zur Verbesserung des Zeitverhaltens

Die Zeitverzögerung t ist also proportional zu R und C , d.h. daß bei großen parasitären Kapazitäten (bei sehr langen Leitungen), sowie großen Widerständen die Schaltzeiten sehr groß werden können. Dies ist besonders der Fall bei Signalen, die nach außerhalb des Chips gelegt werden. Die parasitären Kapazitäten die dann auftreten sind sehr viel größer als die innerhalb des Chips. Es gibt aber auch Fälle, in denen eine Zelle, möglicherweise eine wiederholt vorkommende Zelle, ein Durchsatz-Engpaß ist. Auch dies ist eine Motivation für einen besonders "schnellen" Design. Wir können die Gründe für einen Bedarf zur Verbesserung des Zeitverhaltens wie folgt zusammenfassen:

- externe Leitungen, zu treiben über die Chip-Peripherie (enorme kapazitive Last)
- kritische Pfade auf dem Chip (Durchsatz-Engpaß oder besonders lange Verzögerungen)
- interne Ausgangsleitungen mit besonders hohem Fanout (starke kapazitive Last)
- kritische Zellen (Durchsatz-Engpaß, oft vorkommend, oder beides)



```

.SUBCKT NAND2 VDD VSS A B OUT
MN1 II A VSS VSS NFET W=8U L=4U AD=64P AS=64P
MN2 OUT B II VSS VSS NFET W=8U L=4U AD=64P AS=64P
MP1 VDD A OUT VDD VDD PFET W=16U L=4U AD=128P AS=128P
MP2 VDD B OUT VDD VDD PFET W=16U L=4U AD=128P AS=128P
CA A VSS 50fF
CB B VSS 50fF
COUT OUT VSS 100fF
.ENDS

```

Bild 13.18: SPICE-Beschreibung [28]: a) Schaltplan, b) NDL-Netzliste; (Anm.: Transistor-Anschlußfolge: d g s b).

Um in all diesen Fällen noch akzeptable Schaltzeiten zu bekommen, kommt eine der folgenden Maßnahmen zur Verbesserung des Schaltverhaltens in Betracht:

- spezielle Pad-Treiber-Schaltungen (Verstärker-Zellen mit integriertem Bonding-Pad),
- Minimierung kritischer Pfade, wozu Zellen neu platziert werden müssen
- Verwendung von Treiber-Schaltungen, wie beispielsweise Super-Buffer (Bild 13.20)
- Die Anwendung von schnellen (z. B. Precharge-)Schaltungstechniken (u. a. Kap. 16)
- Flächen- und Interkonnekt-Optimierung durch strukturierten Entwurf (Kap. 19 u. 20)
- Architekturmaßnahmen: durch effizientere Algorithmen oder Basisstruktur-Konzepte
- Einführung einer anderen Gehäuse-Technik (kürzere Leitungen..., s. Abschn. 13.3.3)
- Anwendung einer neueren, schnelleren Version der gegebenen Technologie
- Anwendung einer anderen Technologie (bipolar, BiCMOS, GaAs..., s. Abschn. 13.6)

Für Ausgänge aus dem Chip heraus in seine Anwendungs-Umgebung hinein muß man das Signal durch sogenannte Treiber (hier "Pad-Treiber" genannt, s. auch Abschnitt 13.5.2) verstärken, das hier die parasitären Kapazitäten um Größenordnungen höher sind als auf dem Chip selbst. Meist sind das Inverter, deren Transistoren größer und niederohmiger sind als die der ihn treibenden Schaltungsstufe (z.B. Bild 13.20, Bild 13.21). Der Faktor, um den die Schaltungsstufe größer ist, als die vorausgegangene, nennt man f . In Bild 13.19 ist eine Kurve zu sehen, die das Verhältnis von Vergrößerungsfaktor f zur relativen Zeitverzögerung angibt.

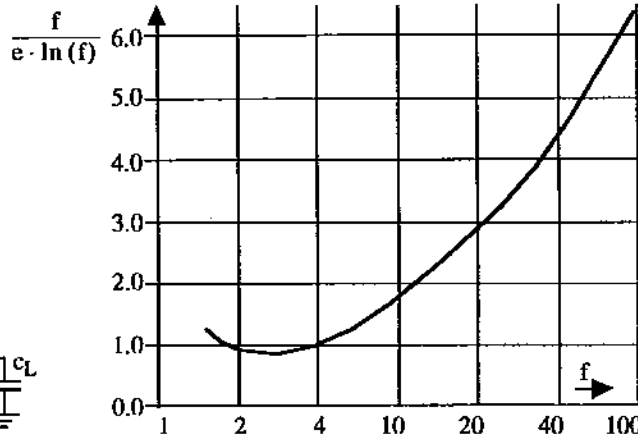
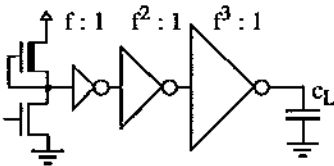
Spezielle Treiber-Schaltungen sind oft auch für Pfade innerhalb des Chips notwendig, etwa bei hohem Fanout-Faktor (etwa bei Takt- und Steuerungs-Eingängen in größere Registerfelder). Bekannte solche Treiber sind beispielsweise die sogenannten *Super-Buffer* (Bild 13.20). Diese helfen bei nMOS-Schaltungen auch, die durch das Pullup/Pulldown-Verhältnis gegebene Asymmetrie des Zeitverhaltens invertierender Stufen (Bild 13.4) abzubauen. Solche Super-Buffer verwenden anstelle des üblichen Last-Transistors mit $U_{gs} = 0$ einen gesteuerten Transistor (ebenfalls selbstleitend, Tr4 in Bild 13.20). Nähere Einzelheiten behandelt Abschn. 13.5.1 .

Für die Ermittlung kritischer Pfade auf dem Chip gibt es Analyse-Algorithmen, welche durch Extraktion aus dem Layout gewonnene Daten bearbeiten (über Extraktoren oder Extraktion s. S. 291 in Kapitel 14). Oft sind solche Analyse-Programme zur Ermittlung kritischer Pfade Bestandteil von Software-Paketen für Routing & Placement ([20], s. a. S. 294 in Kapitel 14).

Die Anwendung von Precharge-Schaltungstechniken (u. a. Kap. 16) ist ebenfalls zur Erzielung höherer Schaltgeschwindigkeiten geeignet. Bei solchen Schaltungstechniken wird die bei nor-



Bild 13.19: Treiben starker kapazitiver Lasten: a) Treiber-Kaskade, b) relative "Bestrafung" im Verhältnis zum Vergrößerungsfaktor f .



maler nMOS-Verhältnis-Logik auftretende Asymmetrie des Zeitverhaltens (s. u. a. in Bild 13.4) vermieden indem die Aufwärtsflanke durch Precharge-Technik beschleunigt wird.

Bei Flächen- und Interkonnekt-Minimierung durch strukturierten Entwurf (Kap. 19 u. 20) können separate Verdrahtungsflächen oft durch *wiring by abutment* vermieden werden (vgl. Bild 19.12). Die bei strukturiertem Entwurf ebenfalls angestrebte Verminderung der Anzahl von Zellentypen erlaubt auch verminderten Design-Aufwand bei der Entwicklung auf minimale Schaltzeit optimierter Zellen. Eigentlich ist strukturierter Entwurf schon fast eine Architektur-Entwicklung ("Chip-Architektur"). Weitere mögliche Maßnahmen zur Verbesserung des Zeitverhaltens sind echte Architekturmaßnahmen: durch die Wahl effizienterer Algorithmen oder Basisstruktur-Konzepte. Hier sei an den Begriff des Schlüssel-Algorithmus (key algorithm) erinnert, s. S. 406 in Abschnitt 20.2). Manchmal können sogar ganze Sätze von Schlüssel-Algorithmen automatisch generiert werden [14] [15] [24], sodaß der am besten geeignete davon ausgewählt werden kann. Ein typisches Architekturproblem ist die Kommunikations-Lücke zwischen Halbleiterspeicher (DRAM) und Mikroprozessor (Bild 13.23) mit der Folge, daß wegen der zu langsamen Speicher-Schnittstelle die Leistungsfähigkeit des Prozessors nicht ausgenutzt werden kann. Bild 13.23 zeigt die fortschreitende Entwicklung dieser Durchsatzlücke. Typische Architekturkonzepte mit dem Ziel der Abhilfe sind hierzu der Cache-Speicher (ein schneller Zwischenpuffer, s. beispielsweise Abschnitt 2.4 in [46]), verschränkte Speicherzugriffe über eine Aufteilung des Speichers in viele "Speicher-Bänke" (z. B. Abschnitt 3.1.4 in [18]), die RAMbus-Schnittstelle [11] u. a. [7] [8].

Eine ganz andere Gruppe von Maßnahmen zur Verbesserung der Schaltgeschwindigkeit ist die technologischer Natur. Dazu gehört die Einführung einer anderen Gehäuse-Technologie (kürzere Leitungen, geringere parasitäre Kapazitäten etc). Abschnitt 13.3.3 befaßt sich ausführlicher mit Gehäuse-Formen, insbesondere mit der besonders Durchsatz-fördernden MCM-Technologie. Eine andere mögliche Maßnahme ist die Anwendung einer neueren, schnelleren Version der gegebenen Technologie (s. Abschnitt 13.6) zur Ausnutzung des raschen Technologie-Fortschritts (vgl. Bild 1.10 oder die Kurve "integrierte Schaltungen" in Bild 13.14). Meist auf Kosten der Integrationsdichte und einer höheren Verlustleistung kann die Anwen-

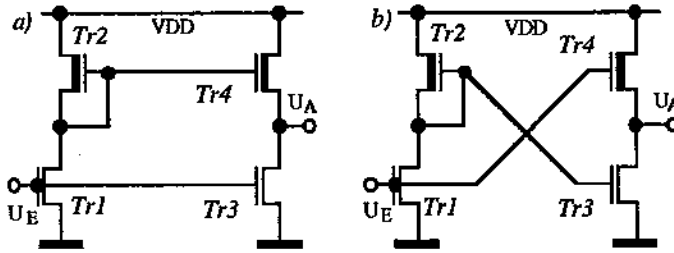


Bild 13.20: Super Buffer, a) invertierend, b) nicht invertierend.

dung einer anderen, schnelleren Technologie oft dramatische Durchsatz-Verbesserungen ermöglichen (Technologie mit Bipolar-Transistoren [10] (Abschnitt 13.6.2), BiCMOS-Technologie (Abschnitt 13.6.1), oder Gallium-Arsenid-Technologie [25] [42] u. a. [9] [26]).

13.5.1 Der "Super-Buffer"

Das Problem der geringer werdenden Flankensteilheit liegt nicht zuletzt am Konzept der n-MOS-Verhältnislogik. Zwecks ausreichender *High/Low*-Differenz wird der selbstleitende Transistor meist so dimensioniert, daß sein Widerstand viermal so groß ist wie der größtmögliche Widerstand des Schaltnetzes im durchgeschalteten Zustand. Dadurch ist der Ladestrom, der ja durch den *Pullup* fließen muß, relativ gering. Wird der Ausgang hingegen auf Masse gezogen, können die Kapazitäten der angesteuerten Lasten über den dort viermal geringeren Widerstand deutlich schneller entladen werden. Abfallenden Flanken sind deutlich steiler.

Um dieser Asymmetrie in der Fähigkeit, kapazitive Lasten zu treiben, entgegenzuwirken, stehen sogenannte *Super Buffer* zur Verfügung, die sowohl die Abstiegs-Flanke wie auch die Anstiegs-Flanke eines Signales trotz anzusteuender Lasten nur geringfügig verfälschen. Sie erfüllen quasi die Aufgabe eines Signalverstärkers. Bild 13.20 zeigt die Realisierung solcher Super Buffer, die es invertierend und nicht-invertierend gibt.

Es handelt sich um zwei gekoppelte Standard-Inverter mit dem üblichen Widerstandsverhältnis 4:1. Allerdings ist der selbstleitende Transistor des Ausgangsinverters nicht mit dem Gate an den Ausgang gekoppelt, sondern an ein zum *Pulldown*-Transistor inverses Signal. Wenn nun der Ausgang von *Low* auf *High* gehen soll, so geht die Spannung am Gate des *Pullup*-Transistors schnell auf VDD und es werden zusätzliche Elektronen in die Implantat-Schicht gezogen, während in der normalen Betriebsart ja überhaupt keine Spannung zwischen Gate und Source anliegt und nur die implantierten Ionen zur Leitung verwendet werden. Durch die so erhöhte Leitfähigkeit sind nun aber höhere Ladeströme für die zu treibenden kapazitiven Lasten möglich, so daß nunmehr ansteigende und abfallende Flanken nahezu unverfälscht und symmetrisch sind.

13.5.2 Pad-Treiber

Ein Pad-Treiber ist eine Verstärker-Kaskade (vgl. Bild 13.19 a) zum Treiben eines Bonding-Pad nebst extern daran angeschlossenen Leitungsteilen (Bonding-Draht, Anschlußfahnen des Chip-Gehäuse, Leitung auf der gedruckten Schaltung, und daran angeschlossene Schaltungseingänge, evtl. noch Kabel zu anderen gedruckten Schaltungen), die eine enorme kapazitive Last einbringen. Ein Pad-Treiber ist üblicherweise mit dem Bonding-Pad zu einer Zelle inte-

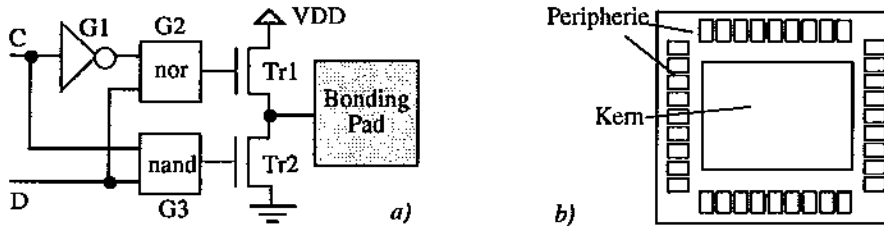


Bild 13.21: Pad-Treiber: a) Tri-State-Treiber-Schaltung, b) Pad-Zellen als Chip-Peripherie.

griert. Die Gesamtheit aller Pad-Zellen eines Chip wird als *Chip-Peripherie* bezeichnet. Die Bezeichnung *Peripherie* deutet nicht nur die Ein/Ausgabe-Funktion an, sondern auch die Tatsache, daß wegen Anforderungen seitens der Bonding-Technik alle Pad-Zellen jeweils in nur einer einzigen Reihe am Chip-Rand angeordnet werden müssen (Bild 13.21 b, Bild 20.34 b).

Bild 13.21 a zeigt den Schaltplan eines Tri-State Pad-Treibers [41]. Die Transistoren Tr1 und Tr2 sind sehr groß, um eine große kapazitive Last gut treiben zu können (vgl. auch Bild 13.24). Die Größe der Transistoren von G1 über G2 und G3 bis zu Tr1 und Tr2 sind optimal gestuft gemäß Bild 13.19 a, bzw. der Kurve in Bild 13.19 b. Bild 13.24 zeigt ein anderes Beispiel: Bild b das Layout und Bild a den dazu isomorphen Schaltplan. Layout nebst Dokumentation von Pad-Zellen ist u. a. zu finden in [29] [40] (nMOS) und [38] (CMOS).

13.6 Schnelle Integrierte Schaltungen

Allein schon durch den raschen Fortschritt der Technologie werden integrierte Schaltungen immer schneller, denn mit abnehmender minimaler Linienbreite werden die Transistoren nicht nur kleiner, sondern auch schneller. Zur Lösung eines Durchsatzproblems reicht manchmal schon das Warten auf eine schnellere Version der gegebenen Technologie. Wenn dies nicht oder nicht-technologische Maßnahme nicht ausreicht, ist der Übergang zu einer anderen Technologie zu überlegen: BiCMOS ist unter Umständen schneller (Abschnitt 13.6.1). Mit Bipolar-Transistoren lassen sich viel schnellere Schaltungen realisieren als mit MOSFET-Technologien [10] (s. Abschnitt 13.6.2), allerdings auf Kosten der Integrationsdichte und unter hoher Verlustleistung. Außergewöhnlich schnell sind die Devices bei Gallium-Arsenid-Technologie wegen der viel höheren Ladungsträger-Beweglichkeit (Mobilität). Die Schaltungstechnik hat ein wenig Ähnlichkeit mit der von nMOS- oder CMOS-Schaltungstechnik. Allerdings gibt es nur wenige Hersteller. Zudem ist viel Erfahrung nötig zur erfolgreichen Anwendung. Eine Einführung in GaAs-Technologie und-Schaltungstechnik ist zu finden: u. a. in [25] [42].

13.6.1 BiCMOS-Schaltungen

Modulare Technologie erlauben die Kombination verschiedener Technologien auf dem gleichen Chip. Die BiCMOS-Technologie ist ein Beispiel dafür (*Bi* steht für "Bipolar-"): eine Kombination aus CMOS und Bipolar-Technologie [41] [43]. Kapitel 12.1 (insbesondere Abschnitt 12.3) und Bild 10.9 versuchen dies zu veranschaulichen. Bild 13.22 b zeigt den Schaltplan eines BiCMOS-Inverters ([41], Bild 12.30 b zeigt das Layout). BiCMOS-Schaltungen eignen sich wegen der sehr niederohmigen Bipolar-Transistoren am Ausgang insbesondere für

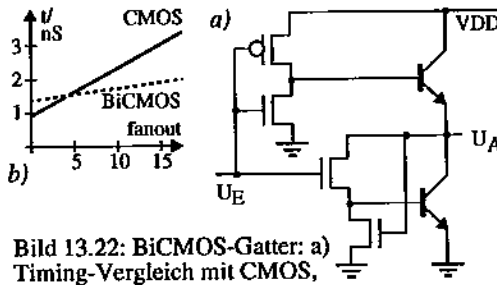


Bild 13.22: BiCMOS-Gatter: a) Timing-Vergleich mit CMOS, b) Beispiel (Inverter).

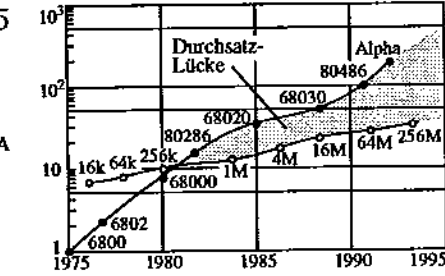


Bild 13.23: DRAM/Proz'or-Durchsatzlücke.

sehr großen Fanout (vgl. Bild 13.22 a), bringen bei sehr geringem Fanout aber keinen Gewinn, oder gar eine Verschlechterung (vgl. Bild 13.22 a).

13.6.2 Schnelle Bipolar-Schaltungen

BiCMOS ist ein Kompromiß zwischen CMOS- und Bipolar-Technik insofern, als Bipolar-Transistoren nur als Ausgangs-Transistoren von Gattern verwendet werden (siehe Beispiel in Bild 13.22 b). Dafür sind nur unter bestimmten Umständen (z. B. große Fanout-Faktoren) relativ bescheidene Verbesserungen zu erwarten (vgl. Diagramm in Bild 13.22 a). Drastisch größere Verbesserungen sind zu erwarten, wenn gesamte Schaltungen oder Subschaltungen vollständig mit Bipolar-Transistoren entwickelt werden. Der Bipolar-Transistor als Schalter hat im Sättigungszustand nur noch einen Restwiderstand in der Größenordnung eines Ohm (wohingegen ein MOSFET einen on-Widerstand in der Größenordnung 10 kΩ hat). Wir erhalten daraus bei einfacher RC-Glied-Modellierung (vgl. Bild 13.2, Bild 13.3) als grobe Näherung eine um etwa vier Größenordnungen verbesserte Zeitkonstante. Allerdings ist hierbei die beeinträchtigende Speicherzeit des gesättigten Transistors noch zu berücksichtigen. Es ist jedoch eine ganze Reihe von Schaltungstechniken bekannt (beispielsweise Emitter-coupled Logic (ECL)), die solche negativen Effekte wie die Speicherzeit vermeidet oder zumindest reduziert. Stark vereinfachte Einführungen in verschiedene Aspekte der Anwendung von Techniken Bipolar-Transistoren sind zu finden in den Kapiteln 6, 10, und 12. Für tiefergehende ausführlichere Darstellungen sei auf andere Literatur verwiesen, wie z. B. auf [10] [46].

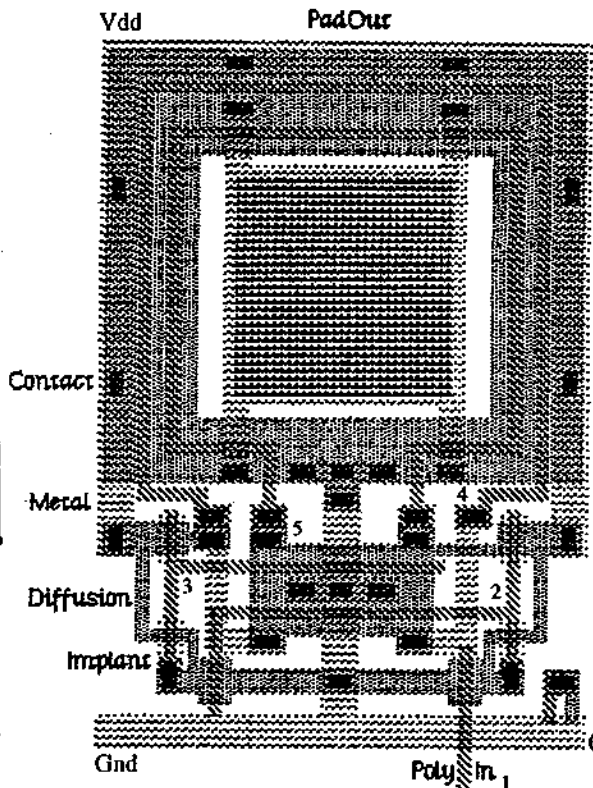
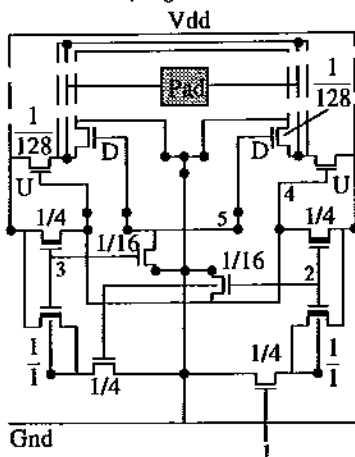
13.7 Literatur

- [1] M. Annaratone: Digital CMOS circuit design; Kluwer Academic Publishers, 1986
- [2] D. Auvergne, R. Hartenstein (editors.): Power and Timing Modeling for Performance of Integrated Circuits; IT Press, 1993
- [3] H. B. Bakoglu: Circuits, Interconnections and Packaging for VLSI; Addison-Wesley, 1990
- [4] J. W. Balde: Multichip Packaging and the Need for new Materials; J. Electron. Mater., vol. 18, no. 2 (Febr. 1989), auch in [19]
- [5] S. Burman, N. A. Sherwani: Programmable Multi-Chip Modules; IEEE Micro 4, 1993
- [6] D. H. Carey: Trends in Low-Cost, High-Performance Substrate Technology; IEEE Micro 13,2 (April 1993)
- [7] R. Comerford, G. F. Watson: Memory catches up; IEEE Spectrum, Oct. 1992
- [8] W. Daum et al.: Overlay High-Density Multichip Module Technology; Computer, 4, 1993



Bild 13.24: Beispiel eines Ausgangs-Pad-Treibers:
 a) Schaltplan mit Angabe der Transistor-Kanal-Größen (Verhältnis Länge / Breite), b) Layout (ICARUS-Plot [17]).

D: 2 Symbole für nur einen Transistor (Ring-Transistor)
 U: 2 Symbole für nur einen Transistor (Ring-Transistor)



- [9] B. El-Kareh, R. J. Bombard: Introduction to VLSI Silicon Devices, Physics, Technology and Characterization, Kluwer Academic Publ., 1986
- [10] M. I. Elmasry: Digital Bipolar Integrated Circuits; A Wiley & Sons, 1983
- [11] E. Farnwald, D. Mooring: A fast path to one memory; IEEE Spectrum, Oct 1992
- [12] P. Franzon, R. Evans: A Multichip Module Design Process for Notebook Computers; Computer 26,4 (April 1993)
- [13] R. L. Geiger, P. E. Allen, N. R. Strader: VLSI Design Techniques for Analog and Digital Circuits; McGraw-Hill, 1990
- [14] R. Hartenstein: Einführung in den VLSI-Entwurf; Skriptum, Univ. Kaiserslautern, 1992
- [15] R. Hartenstein, K. Lemmert: SYS³: A CHDL-based Systolic Synthesis System; in: (ed.: J. A. Darringer, F. J. Rammig) Computer Hardware Description Languages and their Applications; North Holland, Amsterdam 1990
- [16] D. J. Herrill: Addressing the Challenges of Advanced Packaging and Interconnections; IEEE Micro 13,2 (April 1993)
- [17] R. W. Hon, C. H. Sequin: A Guide to LSI Implementation, Xerox PARC, 1980
- [18] K. Hwang, F. Briggs: Computer Architecture & Parallel Processing; McGraw-Hill, 1984
- [19] R. W. Johnson, R. K. F. Teng, J. W. Balde: Multichip-Modules; IEEE Press, 1991

- [20] N. P. Jouppi: Timing Analysis and Performance Improvement of MOS VLSI Designs; IEEE Trans. CAD 6,4 (July 1987)
- [21] L. Katz: Kühlen Kopf bewahren - Gehäuse-Entwicklungen bei Logik; Elektronik 23/1992
- [22] W. H. Knausenberger, L. W. Schaper: Interconnection Costs of Various Substrates - The Myth of Cheap Wire; IEEE Trans. Components, Hybrids, Manuf. Technol., Sept 1984
- [23] D. P. LaPotin et al.: Early Packaging Analysis: Considerations and Case Study; Computer 23,4 (April 1993)
- [24] K. Lemmert: SYS3 Systolic Synthesis System around KARL; Diss, Univ. Kaiserslautern 1989
- [25] S. Long, S. Butner: Gallium Arsenid Digital Integrated Circuit Design; McGraw-Hill 1990
- [26] W. Maly: Atlas of IC-Technologies; Benjamin/Cummings, Menlo Park, CA, USA, 1987
- [27] J. Mavor, M. Jack, P. Denyer: Introduction to MOS LSI design; Addison-Wesley, 1983
- [28] P. Marwedel: Synthese und Simulation von VLSI-Systemen; Hanser-Verlag 1992
- [29] C. Mead, L. Conway: Introduction to VLSI Systems, Addison-Wesley, USA, 1980
- [30] G. Messner: Cost-Density Analysis of Interconnections; IEEE Trans. on Components, Hybrids, Manuf. Technol., vol. CHMT-10 (Juni 1987), auch in [19]
- [31] J. Millman, H. Taub: Impuls- und Digital-Schaltungen; Berliner Union, Stuttgart 1963
- [32] D. Misunas: Guest Editor's Introduction: Advanced Packaging and Interconnection Technology; IEEE Micro 13,2 (April 1993)
- [33] A. Mukherjee: Introduction to nMOS & CMOS VLSI systems design; Prentice-Hall, 1986
- [34] P. R. Mukund, J. G. McDonald: MCM - The High-Performance Electronic Packaging Technology; Computer 23,4 (April 1993)
- [35] N. N.: BRA No. 3237 PATMOS final report; Univ. Kaiserslautern et al., 1992
- [36] N. N.: Markt & Technik 24, 1993 (11. Juni)
- [37] L. Nagel, D. Pederson: Simulation Program with Integrated Circuit Emphasis (SPICE); 16th Midwest Symp. On Circuit Theory, Waterloo, Ontario, 1973
- [38] C. Neugebauer, R. Carlson: Comparison of Wafer Scale Integration with VLSI Packaging Approaches; IEEE Tr. CHMT-10 Components, Hybrids, Manuf. Technol., Juni 1987
- [39] C. Neugebauer et al.: Multichip Module Designs for High Performance Applications; Proc. NEPCON West 1989, auch in [19]
- [40] J. Newkirk, R. Mathews: The VLSI Designer's Library; Addison-Wesley, 1983
- [41] H.-U. Post: Entwurf und Technologie hochintegrierter Schaltungen; Teubner, 1989
- [42] D. A. Pucknell, K. Eshraghian: Basic VLSI Design, Systems and Circuits, Silicon Systems Engineering Series; Prentice Hall, 1988
- [43] A. Rothermel: Digitale BiCMOS-Schaltungen; Oldenbourg, München, 1990
- [44] R. K. Scannel, J. K. Hagge: Development of a Multichip Module DSP; Computer 4, 1993
- [45] D. Seraphim, R. Laskey, C. Y. Li: Principles of Electronic Packaging; McGraw-Hill, 1989
- [46] U. Tietze, Ch. Schenck: Halbleiter-Schaltungstechnik, Springer-Verlag, 1989
- [47] R. R. Tummala, E. J. Rymaszewski: Microelectronics Packaging Handbook; Van Nostrand Reinhold, New York, 1988
- [48] R. P. Vidano et al.: Technology and Design for High Speed Digital Components in advanced Applications; Proc. 3rd Electron. Components Conf. (ECC) 1983, auch in [19]
- [49] A. Vladimirescu, S. Liu: The Simulation of MOS integrated circuits using SPICE2; Memorandum no. UCB/ERL M80/7, University of California, Berkeley, Febr. 1980
- [50] D. von Heinbuch: The CMOS Cell3 Library; Addison-Wesley, 1987
- [51] D. Zhou, F. Preparata, S. Kang: Interconnect Delay in Very High-Speed VLSI; IEEE Trans. on Circuits and Systems 38, p. 779-790 (Juli 1991)