

# 1 Einleitung

Das exponentielle Wachstum der Schaltungskomplexität<sup>1</sup> ist die Triebfeder für den wirtschaftlichen Erfolg der Halbleiterindustrie. Es wird durch das *Moore'sche Gesetz* [Moo65] beschrieben, das eine Verdopplung alle 24 Monate vorhersagt. Das Moore'schen Gesetz gilt seit etwa 35 Jahren, für die Zukunft ist es allerdings durch folgende Hindernisse beschränkt oder zumindest gefährdet:

- **Physikalische Grenzen:** Bei Strukturen in der Größenordnung von einzelnen Atomen ist keine weitere Verkleinerung möglich. Bauelemente in atomaren Dimensionen haben außerdem vollständig andere Eigenschaften als klassische Halbleiterbauelemente, die die Basis heutiger integrierter Schaltungen bilden.
- **Technologische Grenzen:** Es werden Fertigungstechnologien benötigt, die letztendlich Strukturen in der Größenordnung von einzelnen Molekülen oder sogar Atomen herstellen oder bearbeiten können. Die Fertigung von Schaltungen mit Strukturgrößen von etwa 35 nm scheint jedoch gesichert.
- **Entwurfskomplexität:** Die Steigerung der Produktivität bei der Schaltungsentwicklung hält nicht mit der Entwicklung der Fertigungstechnologie mit (*Design Productivity Gap*). Dadurch kann die technologisch mögliche Anzahl von Bauelementen einer integrierten Schaltung nicht sinnvoll genutzt werden. Durch Innovationsschübe bei der Entwicklungsmethodik kann diese Lücke zeitweilig geschlossen werden. In jüngerer Vergangenheit hat sich beispielsweise die Wiederverwendung von größeren Teilschaltungen, so genannten *IP-Cores* (Intellectual Property Cores), zur Integration eines extrem komplexen Systems in einer einzigen integrierten Schaltung, einem *Systems-on-Chip*, durchgesetzt.
- **Testkosten:** Die Testkosten von integrierten Schaltungen beginnen, die Herstellungskosten zu dominieren, so dass deren Fertigung unwirtschaftlich werden kann. Die Testkosten entstehen im Wesentlichen durch die Investitionen für automatische Testsysteme und für die Testentwicklung. Die Ursachen für Kostensteigerungen sind einerseits mit höheren Taktraten steigende Ansprüche an die Genauigkeit der automatischen Testsysteme. Andererseits werden immer größere Testdatenmengen benötigt, um eine ausreichende Testqualität immer komplexerer Schaltungen zu erzielen, so dass die Testdauer für eine einzelne Schaltung kritisch wird.

---

<sup>1</sup> Die Anzahl der aktiven Bauelemente einer integrierten Schaltung.

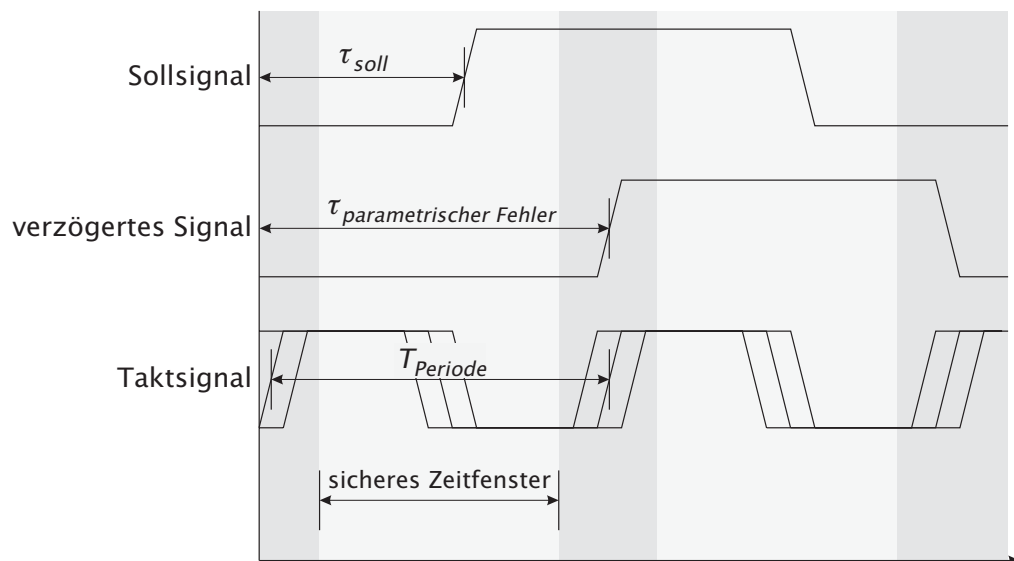


Abbildung 1.1: Verzögerungsfehler

Insbesondere die Hindernisse Entwurfskomplexität und Testkosten haben die Architektur integrierter Schaltungen in den vergangenen Jahren bereits stark geprägt und führen zu charakteristischen Eigenschaften, die in den folgenden Abschnitten kurz erläutert werden. Diese Eigenschaften führen auf das spezielle Problem, einen Selbsttest für Systeme von Leitungen zu entwickeln, die Teilschaltungen eines System-on-Chip miteinander verbinden. Die Lösung dieses Testproblems ist das Thema dieser Arbeit.

Mit dem Moore'schen Gesetz ist eine exponentielle Steigerung der Leistungsfähigkeit mikroelektronischer Systeme verbunden. Diese Steigerung wird durch die Integration von immer mehr und immer komplexeren Funktionen, insbesondere aber durch höhere Taktraten erzielt. Diese haben steilere Signalflanken zur Folge und bedeuten daher einen zunehmenden Hochfrequenzanteil bei den übertragenen Signalen.

Insgesamt werden die zulässigen Toleranzen für das zeitliche Verhalten einer integrierten Schaltung immer geringer. Demgegenüber steigt aber die Unsicherheit bei der Schaltungsentwicklung, weil mit hohen Signalfrequenzen der Einfluss von parasitären, d.h. ungewollten, Leitungsinduktivitäten und -kapazitäten auf den Signalverlauf innerhalb eines mikroelektronischen Systems zunimmt. Störsignale durch elektromagnetische Leitungskopplung, Reflexion an Leitungsenden und Dispersion, die in der Vergangenheit keine Rolle gespielt haben, können zu Fehlfunktionen führen. In digitalen Schaltungen wirken sich diese Störungen beispielsweise als eine Signalverzögerung aus.

Abbildung 1.1 zeigt, wie eine Signalverzögerung zu einem logischen Fehler führen kann. Die obere Zeile zeigt den nominellen Zeitverlauf eines Signals mit einer Verzögerung  $\tau_{soll}$  gegenüber dem Taktsignal am Eingang eines flankengesteuerten Flipflops. Die mittlere Zeile zeigt das gleiche Signal, das aber durch einen nun angenommenen Fehler oder durch elektromagnetische Kopplung zweier Leitungen eine größere Verzögerung  $\tau_{parametrischer\ Fehler}$  als das nominelle Signal in der oberen Zeile aufweist. Die untere Zeile zeigt das Taktsignal der Periode  $T_{Periode}$  mit möglichen zeitlichen Schwankungen (so genannte *Jitter*). Durch diese Schwankungen wird der Zeitbereich verbreitert, in dem das Eingangssignal für ein sicheres Speichern des Signalwerts stabil bleiben muss. Dieser Bereich ist in Abbildung 1.1 dunkel unterlegt. Umgekehrt wird der Zeitbereich, in dem ein Signalwechsel erlaubt ist, schmaler. Durch die zusätzliche Verzögerung des fehlerhaften Signals verschieben sich dessen Signalwechsel in die dunkel unterlegten Bereiche, so dass das Flipflop den Signalwert nicht sicher speichern kann. Eine logische Fehlfunktion ist die Folge.

Kritische parametrische Abweichungen durch Schwankungen im Fertigungsprozess bedeuten im Allgemeinen eine zusätzliche Verringerung des Zeitfensters für einen sicheren Signalwechsel innerhalb eines Taktzyklus. Das hat zur Folge, dass bei bestimmten extremen Signalverläufen, die ohne die Abweichungen nicht zu Fehlern führen, ein Signalwechsel nicht mehr in das reduzierte Zeitfenster fällt, so dass ein logischer Fehler auftritt. Extreme Signalverläufe sind Signalübergänge an einem oder mehreren Schaltungseingängen, die beispielsweise zu Übersprechen von Signalen auf benachbarten Leitungen oder zu kurzzeitigen Einbrüchen der Versorgungsspannung digitaler Schaltungen führen. Um die entsprechenden Fehler zu testen, müssen diese extremen Signalverläufe stimuliert werden. In dieser Arbeit wird untersucht, welche Eigenschaften diese Tests für lange, überwiegend parallel verlaufende Leitungen haben müssen.

## 1.1 Systems-on-Chip

Die technologische Entwicklung stellt durch steigende Integrationsdichte dem Schaltungsentwickler immer größere Ressourcen bezüglich der Anzahl integrierbarer Bauelemente zu Verfügung. Unter dem Druck immer kürzerer Entwicklungszyklen können diese Ressourcen nur durch die Verwendung von vorab entworfenen und verifizierten Teilschaltungen, den *IP-Cores*, ausgenutzt werden. Cores sind strukturell und funktionell eng verknüpfte Schaltungsteile, die über definierte Schnittstellen mit anderen Cores kommunizieren. Sie realisieren Funktionen wie Mikroprozessoren (z.B. eine RISC-CPU), Speicher (z.B. DRAM, SRAM oder ROM), Standardschnittstellen (z.B. PCI, UART oder DMA) oder auch spezialisierte Prozessoren für die Bearbeitung von Audio- oder Bilddaten (z.B. MPEG). Cores sind die Bausteine eines

komplexen integrierten Systems, nur noch geringe Teile werden für ein System neu entwickelt (*User Defined Logic*, UDL). Abbildung 1.2 zeigt ein derartiges System-on-Chip (SOC).

Dieser neue Entwurfsansatz, der auf der schwerpunktmäßigen Verwendung von *IP-Cores* beruht, unterscheidet sich vom „klassischen“ Entwurf dadurch, dass der Entwickler nur noch eine Verhaltensbeschreibung der verwendeten Teilschaltungen erhält, nicht aber detaillierte Informationen über die schaltungstechnische Realisierung.

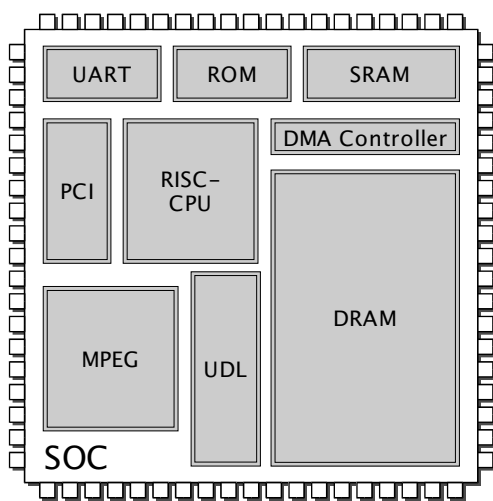


Abbildung 1.2: System-on-Chip (SOC) mit *IP-Cores*

Für einen erfolgreichen Entwurf ist es daher notwendig, dass sich die einzelnen Komponenten an kompatible Standards für die Interaktion halten. In der Regel interagieren die Cores eines solchen Systems über standardisierte Busse wie den PCI-Bus oder den AMBA-Bus von Arm Ltd. [Arm99]. Diese Systeme werden auch als Bus-basierte Systeme beschrieben. Die *Inter-Core*-Busse sind dabei von gleicher Bedeutung für die Funktion und Leistungsfähigkeit eines SOC's wie die Cores selbst.

Diese Busse zeichnen sich dadurch aus, dass ihre Länge in der gleichen Größenordnung liegen kann wie die Kantenlänge der integrierten Schaltung. Die Busse werden daher auch als global bezeichnet, weil sie einerseits die größte Länge besitzen und andererseits funktional gesehen Module auf der obersten logischen Hierarchieebene miteinander verknüpfen. Demgegenüber werden Leitungen als lokal bezeichnet, deren Länge in der Größenordnung einiger Dutzend Mikrometer liegt und die auf tieferer Ebene der logischen Hierarchie Schaltungsteile eines Cores miteinander verbinden.

Die Testentwicklung muss durch den Anbieter eines Cores erfolgen, weil diese auf detaillierten strukturellen Kenntnissen des Moduls beruht. Mittlerweile stehen

umfangreiche Bibliotheken von Cores für den Chipentwurf zur Verfügung. Diese enthalten im Allgemeinen auch verifizierte Tests für die einzelnen Komponenten. Die Tests müssen zu den einzelnen Cores transferiert werden. Dies soll in Zukunft durch den Standard IEEE 1500 [IEE05] geleistet werden, der neben einer standardisierten Testbeschreibungssprache (*Core Test Language*, CTL) eine standardisierte Testhülle (*Core Test Wrapper*) als Schnittstelle für den Test definieren soll [MKL<sup>+</sup>02]. Der Standard IEEE 1500 wird in Abschnitt 2.3 kurz dargestellt.

Auch die Spezifikation des AMBA-Busses legt ein Verfahren für den Test der Cores fest. Dabei wird der AMBA-Bus als breitbandige Schnittstelle zur Übertragung von Testdaten von der Peripherie zu den integrierten Cores verwendet. Insbesondere beim AMBA-Bus ist die Funktionsfähigkeit der Inter-Core-Busse für den Test der Cores notwendig. Inter-Core-Busse müssen daher sogar als erste Komponente des Systems getestet werden. Mit beiden Standards werden keine spezielle Methoden für den Test von Inter-Core-Leitungssystemen zur Verfügung gestellt. Diese Leitungssysteme stellen aber in Zukunft einen Engpass bei der Leistungsfähigkeit von Core-basierten Systemen dar und müssen daher auch beim Test gezielt berücksichtigt werden.

## 1.2 Globale Verbindungsnetze bei Systems-on-Chip

Im vorausgehenden Abschnitt 1.1 wurde dargestellt, wie die Notwendigkeit, die Entwurfskomplexität zu beherrschen, zu Systemen führt, die aus Cores zusammengesetzt sind, die über globale Busse miteinander interagieren. Aber auch die Veränderung der elektrischen Parameter von Verbindungsleitungen mit den Technologiegenerationen führt zu einem busbasierten Ansatz für die Schaltungsarchitektur.

### 1.2.1 Die Dominanz von Leitungen in integrierten Schaltungen

Verbindungsleitungen von integrierten Schaltungen haben einen stetig wachsenden Einfluss auf die Gesamtpformance, weil sie im Gegensatz zu aktiven Bauelementen bezüglich ihrer Signalverzögerung nicht von der allgemeinen Strukturverkleinerung profitieren. Diese Entwicklung kann bereits mit dem sehr einfachen längshomogenen Leitungsmodell nach Abbildung 1.3 erklärt werden [BRK093, Boh95]. Der linke Teil der Abbildung stellt, stark vereinfacht, den Querschnitt zweier Leitungen auf einem Chip einer bestimmten Technologiegeneration dar. Der rechte Teil stellt die gleichen Leitungen in einer folgenden Technologiegeneration dar, wobei die lateralen Abmessungen um einen Faktor  $k < 1$  reduziert werden, während die

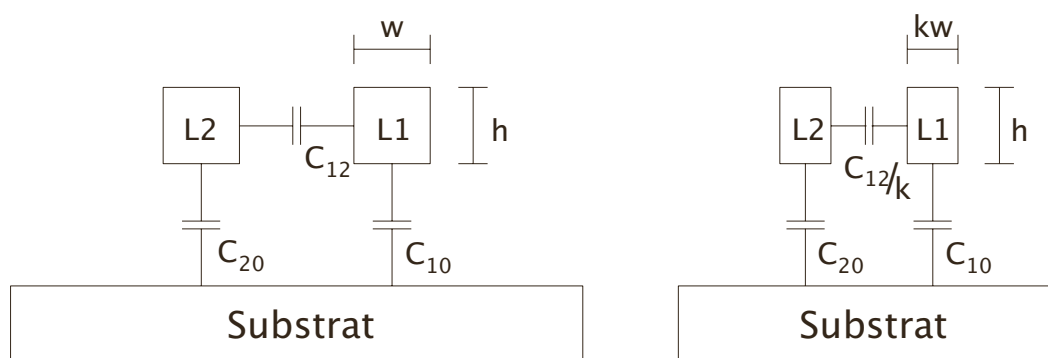


Abbildung 1.3: Einfaches kapazitives Leitungsmodell: nicht skaliert (links), laterale Dimensionen skaliert mit Faktor  $k = 0,66$  (rechts).

vertikalen Abmessungen in diesem einfachen Beispiel als konstant angenommen werden<sup>2</sup>.

Die Signalverzögerung  $\tau$  einer Leitung ist in erster Näherung das Produkt ihres Widerstands  $R$  und ihrer Kapazität  $C$ , also  $\tau = RC$ . Widerstand  $R$  und Kapazität  $C$  skalieren mit einer Strukturverkleinerung. Bei längshomogenen Leitungen gibt man die Kapazität  $C$  und den Widerstand  $R$  bezogen auf die Leitungslänge  $L$  als Kapazitätsbelag  $C' = C/L$  und Widerstandsbelag  $R' = R/L$  an. Damit gilt für die Signalverzögerung  $\tau = R'C'L^2$ .

Eine Verringerung der Leitungsbreite  $w$  um einen Faktor  $k < 1$  bedeutet eine entsprechende Reduktion des Leitungsquerschnitts, also letztendlich eine Vergrößerung des Widerstandsbelags  $R'$  um den Faktor  $k$ , wenn man annimmt, dass die Leitungsdicke  $h$  konstant bleibt.

Die Kapazitätsbeläge der Leitungen können nach Abbildung 1.3 in die Anteile  $C'_{10}$  bzw.  $C'_{20}$  zum Substrat und in die Koppelkapazität  $C'_{12}$  zur benachbarten Leitung aufgeteilt werden. Unter Annahme eines einfachen Parallelplattenmodells für die Kapazität nehmen die Beläge  $C'_{10}$  und  $C'_{20}$  um den Faktor  $k$  ab, während der Koppelungsanteil  $C'_{12}$  um den Faktor  $\frac{1}{k}$  zunimmt, weil der Abstand von Leitungen wie die anderen lateralen Dimensionen um diesen Faktor abnimmt. In erster Näherung bleibt damit das Produkt von Widerstandsbelag und Kapazitätsbelag  $R'C'$  einer Leitung günstigstenfalls konstant, wenn die Kapazität zum Substrat dominiert, steigt aber sonst quadratisch mit der Abnahme der Strukturgrößen.

Bei der Entwicklung der Leitungslänge  $L$  gibt es zwei gegenläufige Tendenzen. Zum einen kann die Länge von lokalen Leitungen, die benachbarte logische Zellen verbin-

<sup>2</sup> Die Skalierung von einer Technologiegeneration zur nächsten ist in Wirklichkeit sehr viel komplexer. Es erfolgt insbesondere ebenfalls eine Verringerung der vertikalen Abmessungen, die aber mit den einzelnen Schichten variiert, z.B. von den unteren zu den oberen Metalllagen. Außerdem können sich auch die Materialien für Leiter und für Dielektrika verändern.

den, mit deren lateralen Abmessungen verkürzt werden. Die Verzögerung lokaler Leitungen nimmt also bei konstantem Widerstands- und Kapazitätsbelag proportional zu  $k$  ab, und stellt für die Gesamtverzögerung eines Schaltungspfads bestehend aus Gattern und lokalen Leitungen kein Problem dar.

Mit jeder Technologiegeneration nimmt jedoch die Fläche integrierter Schaltungen um einen Faktor  $l > 1$  zu, der von  $\frac{1}{k}$  verschieden ist. Die maximale mögliche Länge für eine globale Leitung, die zwei Zellen an gegenüberliegenden Stellen der Schaltung verbindet, nimmt damit um den Faktor  $\sqrt{l}$  zu. Die Verzögerung globaler Leitungen nimmt also wenigstens proportional zu  $l$  zu.

Während die Signalverzögerungen auf Leitungen im Zuge der Skalierung also stagnieren (oder sogar zunehmen), geht man bei der intrinsischen Gatterverzögerung<sup>3</sup> von einer Reduzierung um 15 % pro Jahr aus [ITR03]. Die Folge ist, dass die Signalverzögerung globaler Leitungen das zeitliche Verhalten einer integrierten Schaltung bei weiterer Skalierung dominiert [HMH01] und limitierend für die Schaltungsperformance wirkt. Damit kann ohne neue Leitungsarchitekturen die mit dem Wechsel der Technologiegenerationen verbundene Steigerung der Leistungsfähigkeit durch Erhöhung der Taktraten nicht weiter realisiert werden.

Um die durch die Dominanz der Leitungen gegebene Grenze zu überschreiten, hat die Halbleiterindustrie neue Materialien und Leitungsarchitekturen eingeführt. Die Verwendung von Kupfer als ein Leitungsmaterial mit höherer Leitfähigkeit und von isolierenden Materialien mit niedriger Permittivität  $\epsilon$  sowie andere fertigungstechnische Fortschritte [HH00] verringern die Signalverzögerung auf Leitungen. Sie können den grundsätzlichen Trend aber nicht aufheben, sondern die Dominanz der Verzögerung langer Leitungen nur auf eine spätere Technologiegeneration verschieben. Der Grund für die Verwendung von Kupferleitungen ist daher eher in der größeren Widerstandsfähigkeit gegen Elektromigration zu sehen [BRKO93].

In dem dargestellten einfachen Verzögerungsmodell gilt für den Zusammenhang von Signalverzögerung  $\tau$  und Leitungslänge  $L$  der Zusammenhang  $\tau \sim L^2$ . Daher werden lange Leitungen in mehrere Abschnitte aufgeteilt, die jeweils von starken Leitungstreibern getrieben werden. Bei optimaler Auslegung solcher Leitungssysteme hinsichtlich der Länge der Abschnitte und der Stärke der Treiber kann die Gesamtverzögerung gegenüber einer langen Einzelleitung deutlich verringert werden [LCL96, CDBS01].

---

<sup>3</sup> Die intrinsische Gatterverzögerung ist die Signalverzögerung von einem Eingang eines logischen Gatters zu seinem Ausgang, die sich ergäbe, wenn keine Last am Ausgang vorhanden wäre.

### 1.2.2 Hierarchische Leitungssysteme

Neue Leitungsarchitekturen nutzen aus, dass die steigenden Signalverzögerungen nur die langen globalen Verbindungsleitungen betreffen. Dagegen skaliert die Verzögerung auf den kurzen lokalen Leitungen mit der Strukturverkleinerung, wenn auch nicht in dem gleichen Maße wie die Transistorschaltzeiten. Die lokalen Leitungen machen außerdem den Hauptanteil der Gesamtzahl aller Verbindungsleitungen aus, weil mit ihnen die Verbindungen innerhalb der strukturell eng verknüpften Cores realisiert werden. Daher werden lokalen Leitungen mit dem gleichen Faktor skaliert wie die aktiven Bauelemente, um die Dichte der Leitungen entsprechend der Dichte der Bauelemente zu vergrößern.

Für den geringen Anteil der langen globalen Leitungen wird eine inverse Skalierung durchgeführt. Das bedeutet, dass Leitungsdicke, -breite und -abstand in verschiedenem Maße vergrößert anstatt verringert werden, so dass Signalverzögerung auf einer Leitung und die elektromagnetische Kopplung zwischen mehreren Leitungen reduziert werden [KMSS98]. Inverse Skalierung bedeutet jedoch auch, dass für die globalen Verbindungsleitungen eine wesentlich größere Fläche benötigt wird. Da die benötigte Fläche direkt die Herstellungskosten beeinflusst, muss ein Kompromiss zwischen Leistungsperformance und Kosten gefunden werden [DM98, Mei03].

Insgesamt lässt sich die Signalverzögerung der globalen Leitungen durch inverse Skalierung, Optimierung von Leitungstreibern und Anwendung neuer Materialien soweit verringern, dass sie zumindest für die Technologiegenerationen der kommenden zehn Jahre kein Hemmnis für die Leistungsfähigkeit integrierter Schaltungen darstellen. Eher scheint die zur Verfügung stehende Verdrahtungsfläche ein Hindernis zu werden [The00].

Inverse Skalierung führt zu einer – im einfachsten Fall zweistufigen – Hierarchie der Leitungen einer integrierten Schaltung: Eine Ebene lokaler kurzer Leitungen, die mit den übrigen lateralen Dimensionen skalieren, und eine Ebene globaler, langer Leitungen, die überproportional große Leitungsbreiten und -abstände aufweisen, und deren Verzögerung wie oben beschrieben optimiert wird.

Diese Hierarchie der Leitungen passt zu der in Abschnitt 1.1 beschriebenen Entwurfsmethodik, die auf der Verwendung von Cores beruht. Die enge interne Verknüpfung innerhalb der Cores erfolgt auf der Ebene der lokalen Leitungen, die daher als Intra-Core-Leitungen bezeichnet werden und deren Verzögerung durch Skalierung ihrer Länge reduziert werden kann. Die Verbindung mit anderen Cores erfolgt über die Ebene globaler, langer Leitungen, die daher als Inter-Core-Leitungen bezeichnet werden.