UNIVERSITÄT LEIPZIG

Enterprise Computing
Einführung in das Betriebssystem z/OS

Prof. Dr. Martin Bogdan
Prof. Dr.-Ing. Wilhelm G. Spruth

WS2012/2013

Input/Output Teil 2

SCSI und FICON

Moderne Plattenspeicher-Anschlussarten

Moderne Plattenspeicher werden heute fast ausschließlich über ein serielles Protokoll angeschlossen. Die wichtigsten Protokolle sind:

- **SATA (serial ATA)** — **Nachfolger für parallell ATA (anderer Name ist IDE)**
- **SAS (serial attached SCSI)** — **Nachfolger für parallell SCSI**
- **iSCSI** — **Internet SCSI (benutzt Ethernet, von Mainframes nicht unterstützt)**
- **FC-SCSI** — **Fibre Channel SCSI**

SATA domiminiert beim PC und anderen Arbeitsplatzrechnern. Die verschiedenen SCSI Arten dominieren bei Servern. Fibre Channel SCSI kann sowohl für Punkt–zu–Punkt Verbindungen als auch als FC-AL Version (Fibre Channel Arbitrated Loop) eingesetzt werden.

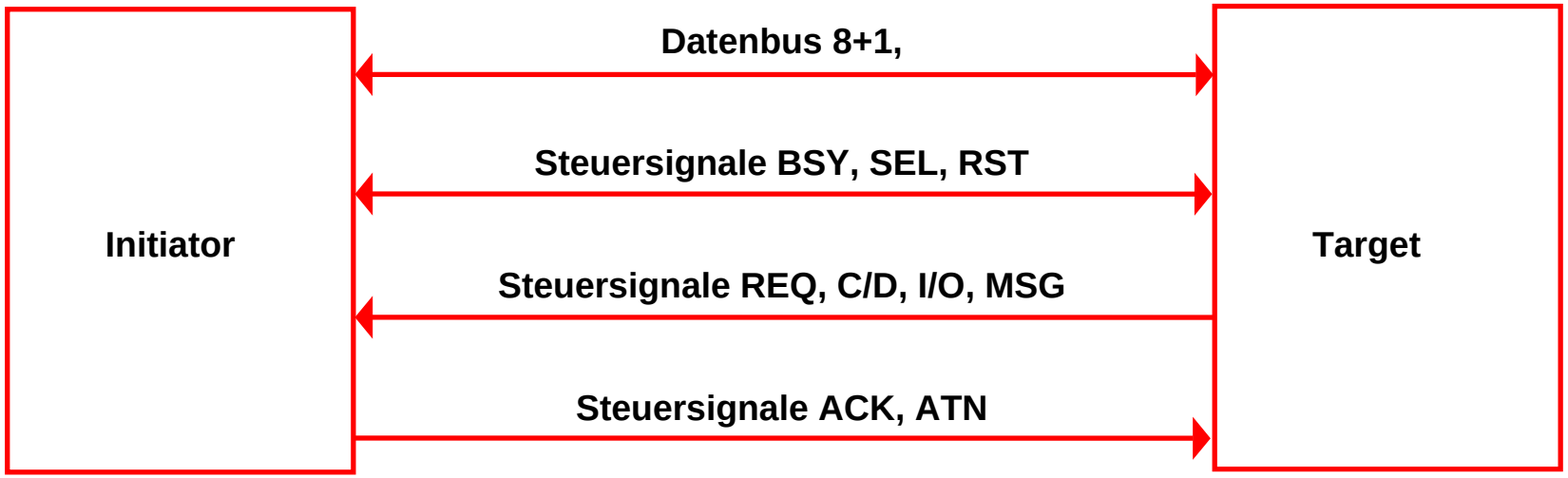
Seagate z.B. bietet z.B. (2010) die Barracuda Familie von Plattenspeichern mit der SATA und der SAS Schnittstelle an; Speicherkapazität bis zu 4,0 TByte. Für „Mission Critical Applications“ ist die Cheetah Familie von Plattenspeichern mit einer 4-Gb/s Fibre Channel interface verfügbar; Speicherkapazität bis zu 600 GByte. Cheetah Plattenspeicher sind laut Seagate für Anwendungen vorgesehen, „where system availability and reliability is of utmost importance“.

Letzteres ist vor allem bei Mainframes gegeben. Dafür wird die deutlich geringere Speicherkapazität in Kauf genommen. In anderen Worten: Plattenspeicher in Ihrem PC haben pro Einheit in der Regel eine deutlich höhere Speicherkapazität als die Plattenspeicher eines Mainframes.

Historische Entwicklung S/360 Channel und SCSI

Für den Anschluss von I/O Geräten System führte das Sytem/360 im Jahre 1964 den „Selektor“ und 1970 den „Block Multiplex“ Kanal ein. IBM veröffentlichte diese I/O Schnittstelle unter dem Namen OEMI (Original Equipment Manufacturer Interface) , was dazu führte, dass viele unabhängige Hersteller I/O Geräte für den Anschluss an die damaligen Mainframes entwickelten und installierten. Das ist auch heute noch der Fall.

Ab 1882 wurde auf Initiative der Firmen Shugart und NCR eine Modifikation des OEMI Standards durch das ANSI (American National Standards Institute) unter dem Namen SCSI (Small Computer System Interface) für den Einsatz in kleineren Systemen veröffentlicht. Hierbei wurde die auf extreme Zuverlässigkeit ausgelegte OEMI Verkabelung (elektrische Interface) vereinfacht. Die logische Interface wurde weitestgehend beibehalten.

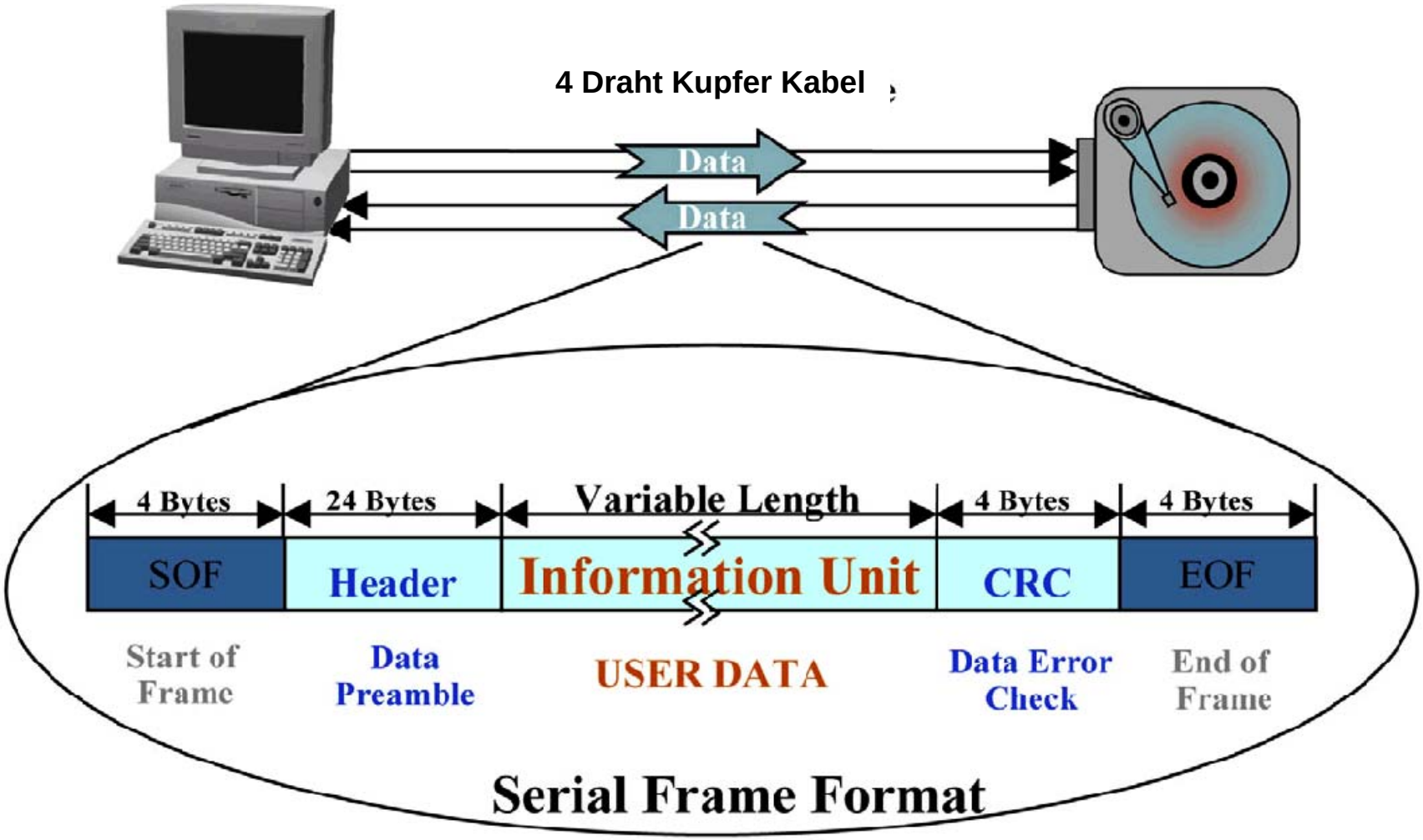


OEMI und SCSI BUS

Die logische OEMI und SCSI Interface bestand aus einem 8 Bit (plus Parity) Datenbus sowie einer Reihe von teils unidirektionalen, teils bidirektionalen Steuersignalen. Die OEMI Schnittstelle wird heute von IBM als „Parallel Channel“ bezeichnet. Ein Parallel Channel hat eine Datenrate von max. 4.5 MByte/s. und überbrückt Distanzen bis zu 130 m. Der Parallel Channel benutzt zwei Kupfer Kabel: *Bus* und *Tag*. . Ein Bus Kabel überträgt Information (ein Byte in jeder Richtung). Daten auf dem Tag Kabel definieren die Bedeutung der Information auf dem Bus Kabel.

Später entwickelten sich die OEMI und SCSI Schnittstellen unabhängig voneinander weiter. Der ursprünglich 8 Bit breite Datenbus der SCSI-1 Schnittstelle wurde auf 16 und später 32 Bit verbreitet. Anschließend entstanden serielle Versionen, die bei IBM zu den Glasfaser-gestützten ESCON und dann den FICON Kanälen führten. Bei SCSI entstanden die Serial SCSI und die Glasfaser FC-SCSI Versionen.

Wegen der höheren Anforderungen im Großrechnerbereich hatten die IBM Kanäle immer einen deutlich höheren Funktionsumfang als die SCSI Schnittstellen. Auch heute kann der mit einem FICON Netzwerk erzielbare Durchsatz durch ein SCSI Netzwerk nicht erreicht werden.



Die ursprünglich parallel übertragenen OEMI oder SCSI Daten werden bei SAS (serial attached SCSI) in einen Rahmen (Frame) gepackt und über ein Kupferkabel seriell übertragen. Die Fibre Channel SCSI (FC-SCSI) Version verwendet statt dessen zwei Glasfaserkabel (für Hin- und Rückleitung.

Das aus dem Parallel Channel (OEMI) entwickelte serielle Fibre Channel FICON Protokoll benutzt ebenfalls Glasfasern. Es hat im Vergleich zu FC-SCSI einen wesentlich höheren Funktionsumfang und damit eine bessere I/O Leistung.

Fibre Channel (FC)

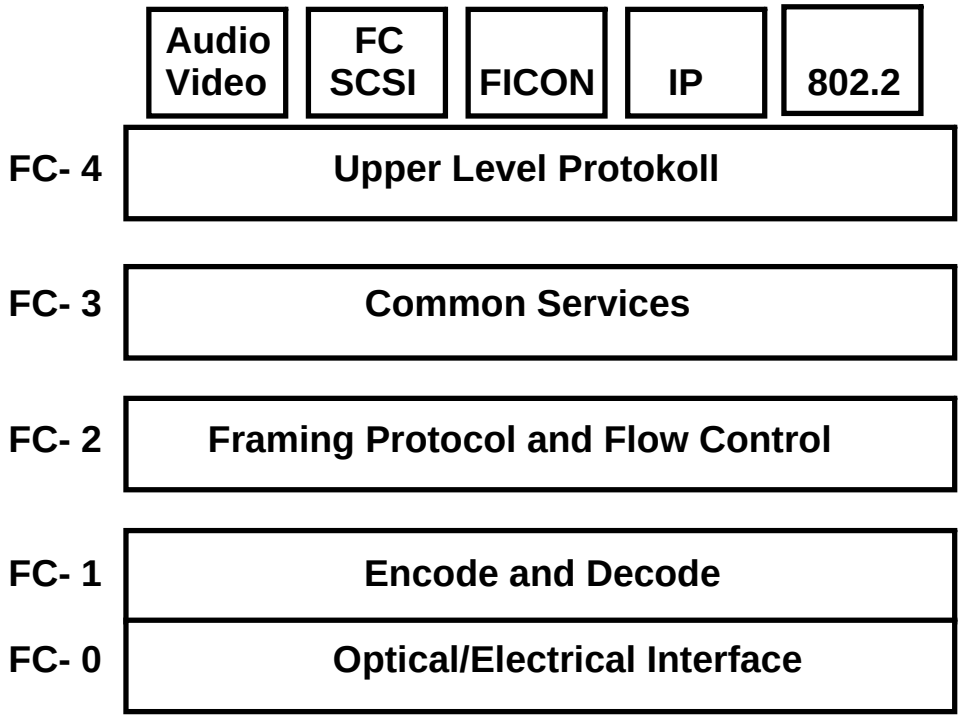
Fibre Channel ist für serielle, kontinuierliche Hochgeschwindigkeitsübertragung großer Datenmengen konzipiert worden. Die erreichten Datenübertragungsraten liegen heute bei 8 Gbit/s, was im Vollduplex-Betrieb für Datentransferraten von 800 MB/s ausreicht. Als Übertragungsmedium findet man gelegentlich Kupferkabel (hauptsächlich innerhalb von Storage-Systemen; überbrückt bis zu 30 m), meistens aber Glasfaserkabel. Letzteres wird vor allem zur Verbindung von Rechnern mit Storage-Systemen oder aber von Storage-Systemen untereinander eingesetzt. Hierbei werden Entfernungen bis zu 10 km überbrückt. Der Zugriff auf die Festplatten erfolgt blockbasiert.

Es können generell zwei Arten von Fibre-Channel-Implementierungen unterschieden werden, die Switched Fabric, die meist als Fibre Channel Switched Fabric (FC-SW) bezeichnet wird und die Arbitrated Loop, kurz als FC-AL bekannt.

Bei der Fibre Channel-Switched Fabric werden Punkt-zu-Punkt-Verbindungen (Point To Point) zwischen den Endgeräten geschaltet. Beim Fibre Channel-Arbitrated Loop handelt es sich um einen logischen Bus, bei dem sich alle Endgeräte sich die gemeinsame Datenübertragungsrate teilen.

Das Fibre Channel-Switched Fabric ist die leistungsfähigste und ausfallsicherste Implementierung von Fibre Channel. In den meisten Fällen ist Switched Fabric gemeint, wenn nur von Fibre Channel gesprochen wird. Im Zentrum der Switched Fabric steht der Fibre Channel Switch (von IBM als „Director“ bezeichnet). Über dieses Gerät werden alle anderen Geräte miteinander verbunden, so dass es über den Fibre Channel Switch möglich wird, direkte logische Punkt-zu-Punkt-Verbindungen zwischen je zwei beliebigen angeschlossenen Geräten zu schalten.

FC-AL erlaubt es, bis zu 127 Geräte an einem logischen Bus zu betreiben. Dabei teilen sich alle Geräte die verfügbare Datenübertragungsrate (bis 8 GBit/s). Die Verkabelung kann sternförmig über einen Fibre Channel Hub erfolgen. Es ist auch möglich, die Geräte in einer Schleife (Loop) hintereinander zu schalten (Daisy Chain), da viele Fibre-Channel-Geräte über zwei Ein- bzw. Ausgänge verfügen. Dies ist z.B. beim IBM DSS 8700 Enterprise Storage Server der Fall.



Fibre Channel Schichtenmodell

Fibre Channel ist ähnlich wie TCP/IP ein Schichten-Protokoll und besteht aus 5 Lagen:

- **·FC0 Die physical layer beschreibt Kabel Fiber Optics, Konnektoren, Pinouts usw.**
- **·FC1 Die data link layer implementiert das the 8b/10b Encoding und Decoding der Signale.**
- **·FC2 Die network layer, definiert durch den FC-PI-2 standard, ist das Kern Element des Fibre Channel.**
- **·FC3 Die common services layer, ist für Erweiterungen vorgesehen, und könnte in Zukunft Funktionen wie Encryption oder RAID implementieren.**
- **·FC4 Die Protocol Mapping layer.**

In der Fibre Channel Schicht 4 (Protocol Mapping Layer) werden Protokolle wie FC-SCSI oder FICON abgebildet.

Fibre Channel Architektur

Die Fibre Channel Architektur verwendet ein Schichtenmodell, vergleichbar mit (aber unabhängig von) den TCP/IP oder OSI Schichtenmodellen. Die unterste Schicht verwendet in den meisten Fällen optische Kabel. Wichtig ist besonders die oberste Schicht FC4. Hierüber ist es möglich, unterschiedliche Protokolle zu betreiben.

FC-SCSI ist eine serielle Form des SCSI Protokolls, die über Fibre Channel Verbindungen erfolgt. (Das als „Serial SCSI“ bezeichnete Protokoll benutzt keinen Fibre Channel).

FICON ist das universell von Mainframes eingesetzte Protokoll, um Rechner miteinander und mit I/O Geräten zu verbinden.

Ein spezieller Fibre Channel Adapter wird für die Echtzeitübertragung von Fernsehprogrammen benutzt.

„IP over Fibre Channel“ und „Ethernet over Fibre Channel“ wurde standardisiert, wird aber nur wenig benutzt.

Literatur:
http://www.answers.com/topic/fibre-channel?cat=technology

Fibre Channel ATA
FATA

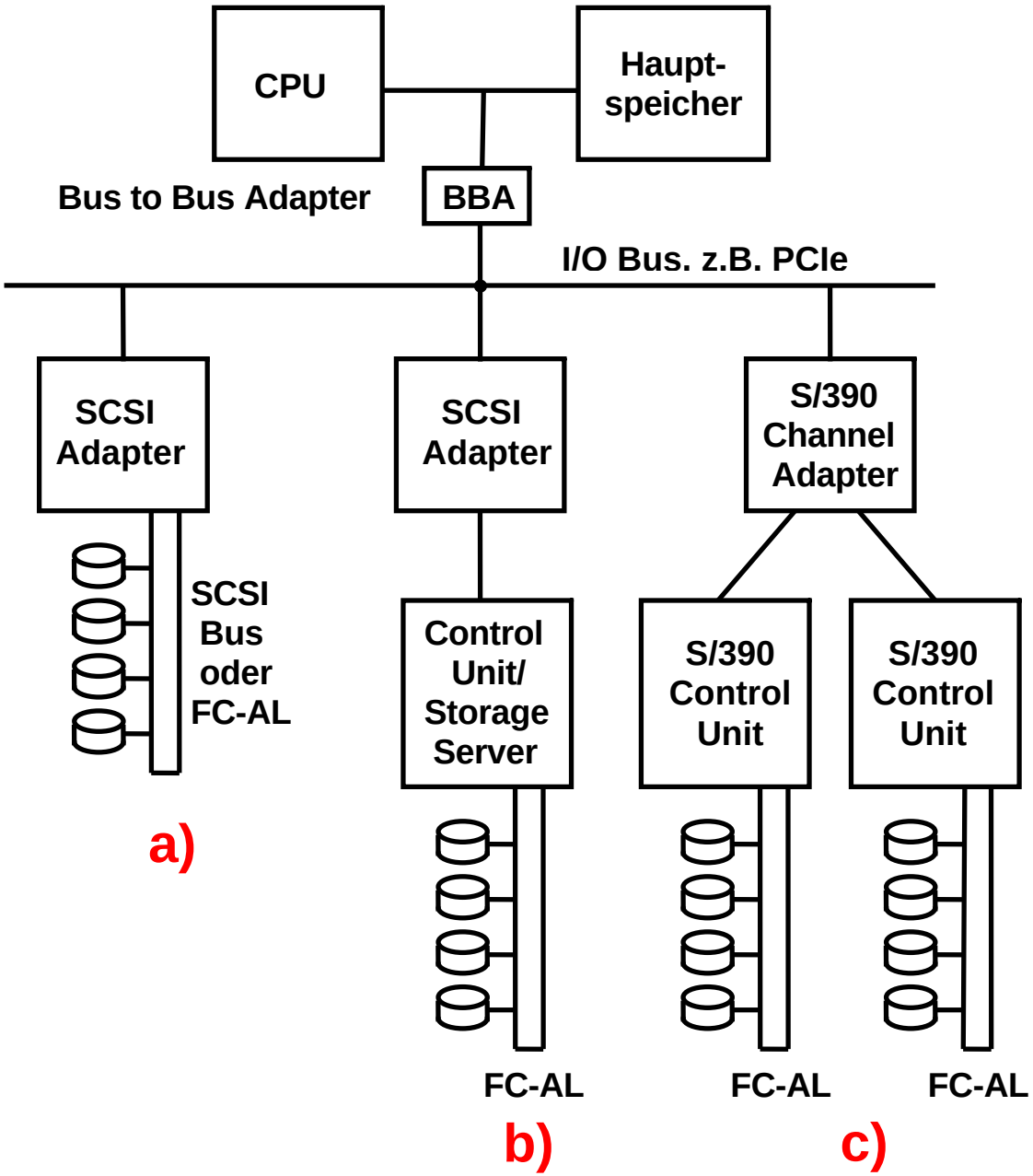
In der Praxis verwenden SCSI Platten bessere mechanische und elektronische Komponenten als SATA Platten. Dies bewirkt:

- **schnellere Zugriffszeiten**
- **höhere Zuverlässigkeit**
- **deutlich höhere Kosten**

Aus Zuverlässigkeitsgründen verwenden SCSI Platten selten die neueste Plattenspeichertechnologie. Dies ist einer der Gründe, warum für einen Personal Computer Plattenspeicher mit einer höheren Speicherkapazität erhältlich sind als dies im Mainframe Bereich der Fall ist.

FATA-Laufwerke sind SATA-Plattenlaufwerke mit einemr Fibre Channel (FC) Anschluss. Sie arbeiten mit den mechanischen Komponenten von ATA-Festplatten, jedoch mit vorgeschalteter FC-Schnittstelle. In anderen Worten, es sind SATA Platten, bei denen die elektrische serielle ATA Schnittstelle durch eine Fibre Channel Schnittstelle ersetzt wurde. Die FATA-Technologie hat den Vorteil, dass sie in einer Mischung mit anderen FC-Laufwerken betrieben werden können.

FATA Platten werden auch als „Nearline-Platten“ bezeichnet. Sie werden im Großrechnerbereich dann eingesetzt, wenn Zuverlässigkeit und Zugriffszeit weniger wichtig sind, z.B. um Bilddateien (Images) zu archivieren.

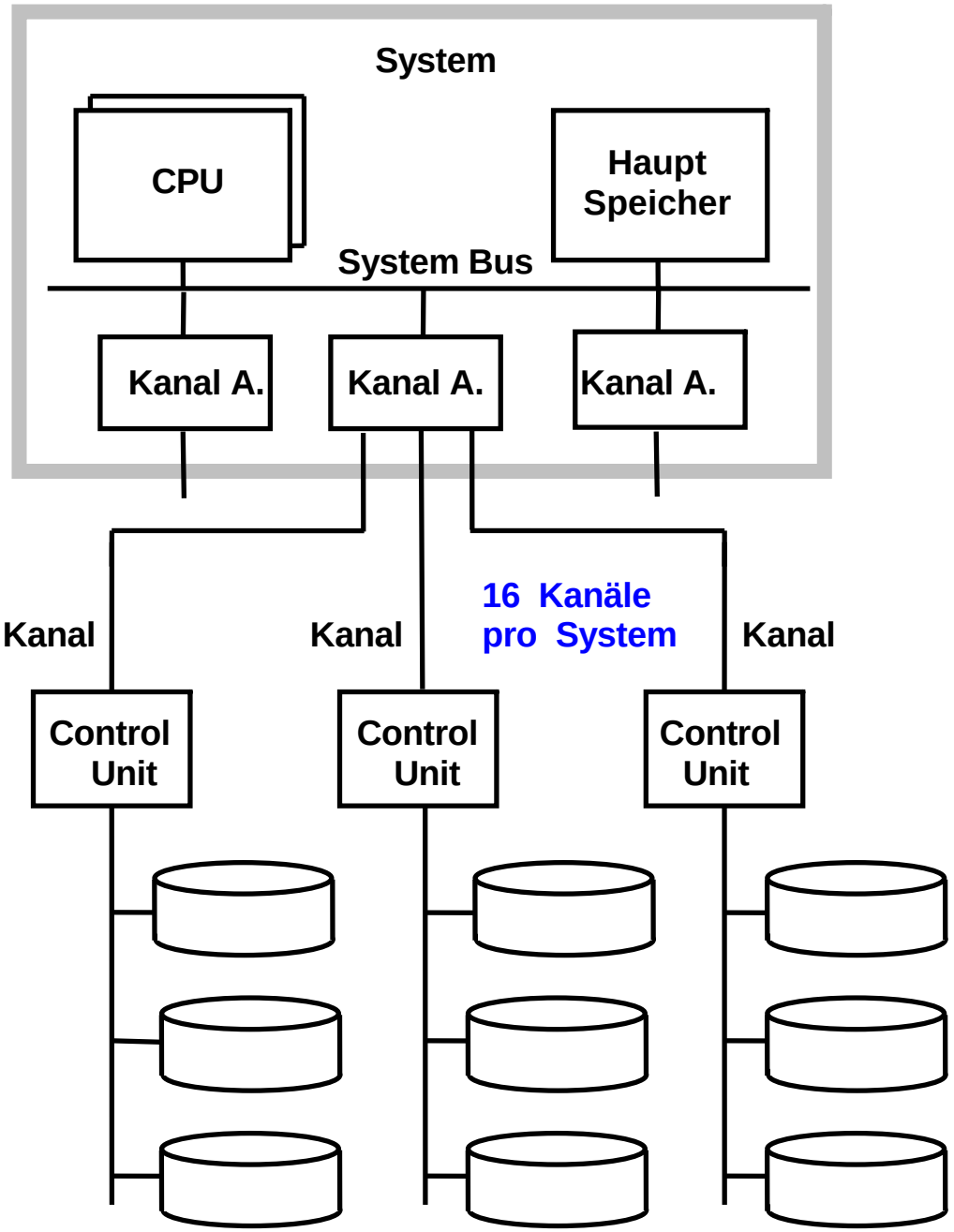


Plattenspeicher Anschluss Alternativen

a) Ein SCSI Adapter kann eine oder mehrere SCSI Platten mit dem I/O Bus eines Rechners verbinden.

b) Bei größeren Mengen an Plattenspeichern sind diese über eine Control Unit oder einen Storage Server mit dem SCSI Adapter verbunden.

c) Mainframes ersetzen den SCSI Adapter durch einen Channel Adapter. Plattenspeicher sind grundsätzlich über Control Units mit dem Channel Adapter verbunden .



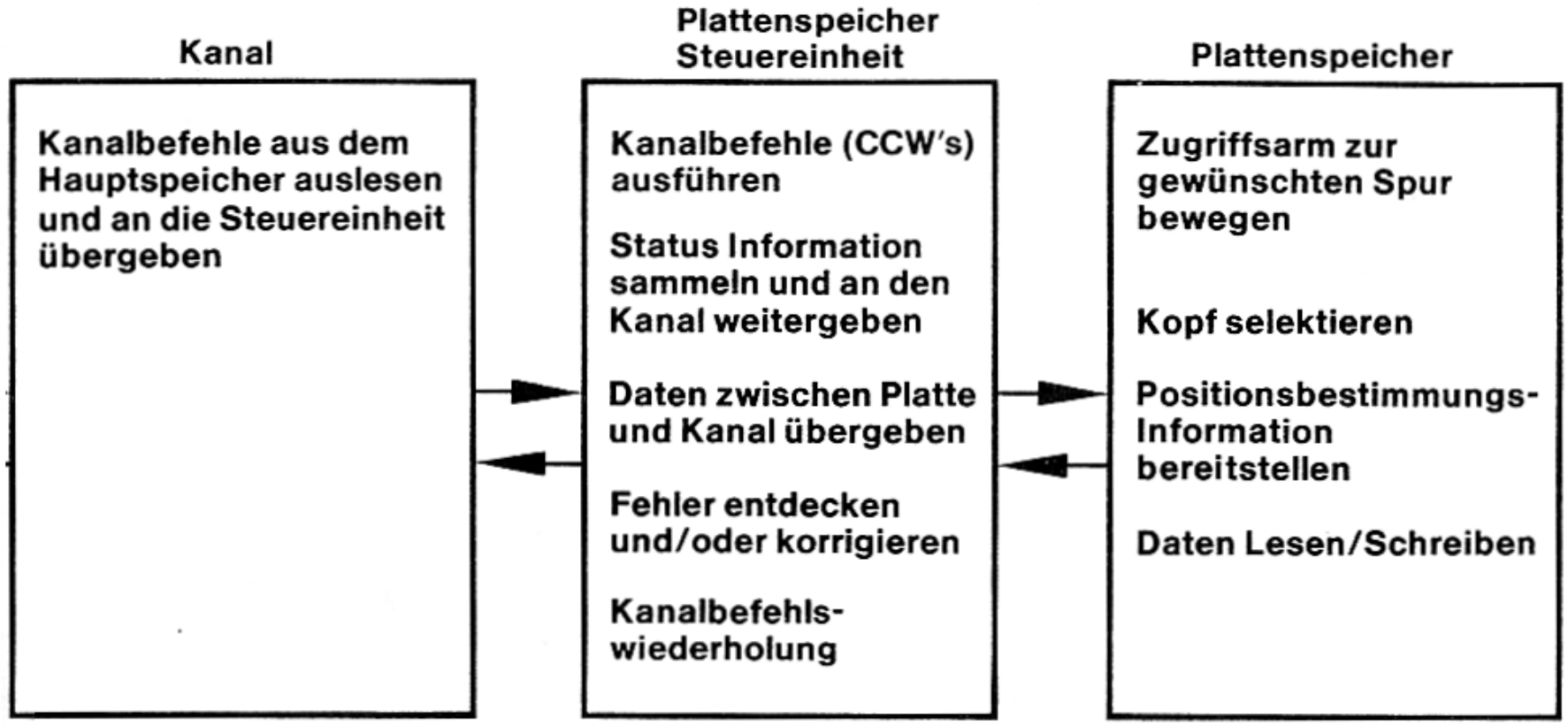
S/360 I/O Konfiguration

Dargestellt ist die ursprüngliche S/360 I/O Konfiguration.

Plattenspeicher sind über Control Units, Kanal-Verbindungskabel (Channel Cables) und Kanal-Adapter mit dem Hauptspeicher des Systems verbunden. Die Verbindungskabel des „Parallel Channels“ waren bis zu 400 Fuß (130 m) lang. Die Kanal-Adapter konnten mittels DMA direkt auf den Hauptspeicher des Systems zugreifen.

Die Control Unit führte Befehle aus, die vom Kanal-Adapter aus dem Hauptspeicher ausgelesen und zwecks Ausführung an die Control Unit übergeben wurden.

Magnetbandgeräte und Drucker werden ähnlich wie Plattenspeicher über Control Units an den Mainframe Rechner angeschlossen.



Aufgaben der Plattenspeicher-Steuereinheit

Dargestellt ist die Aufgabenaufteilung zwischen Kanal-Adapter, Steuereinheit (Control Unit) und der Plattenspeicher-Elektronik.

Der Kanal (die Kanal-Adapter Karte) ist nur dazu da, um stellvertretend für die Steuereinheit per DMA Daten und Kanalbefehle (CCWs) aus dem Hauptspeicher auszulesen und an die entfernte Steuereinheit weiter zu geben.

Der Plattenspeicher selbst enthält umfangreiche elektronische und logische Komponenten.

Serielle und parallele Kanäle

Der Parallel Channel dominierte die Mainframe I/O Konfigurationen bis zum Anfang dieses Jahrhunderts. Er hat viel Ähnlichkeit mit der parallelen SCSI Interface. Heutige Mainframes unterstützen den Parallel Channel nicht mehr.

Serielle Channels haben den älteren parallel Channels abgelöst. Es existieren zwei Serial Channel Typen:

- **Der (ältere) „ESCON Channel erlaubt Datenraten von 17 MByte/s. Er wurde 1990 eingeführt, und wird in zukünftigen Mainframe Modellen nicht mehr verfügbar sein.**
- **Ein FICON Channel erlaubt Datenraten bis zu 800 MByte/s. Das FICON Protokoll wurde 1997 eingeführt.**

Die über 10 Jahre alte „Shark“ Plattenspeichereinheit unseres eigenen Mainframe Rechners jedi.informatik.uni-leipzig.de wurde über ESCON Verbindungen angeschlossen. Unsere neuere DS6800 Plattenspeichereinheit verwendet FICON Verbindungen.

Serielle Channel verwenden Glasfaser Kabel an Stelle von Kupferverbindungen. Es können Entfernungen bis zu 100 km überbrückt werden. Außerdem verfügen sie über eine erweiterte I/O Adressierung.

Formal wird ein Channel als „Channel Path“ bezeichnet, und durch einen 8 Bit CHPID (Channel Path Identifier) gekennzeichnet. In der Umgangssprache werden Channel Path nach wie vor als Kanäle (Channels) bezeichnet, und auch Experten kennen den Unterschied nicht genau.

Was ist Firmware ?

Komponenten, die früher mittes hart verdrahteter Transistorlogik erstellt wurden, verwenden heute häufig statt dessen einen dedizierten Mikroprozesser mit speziellem Code. Dieser Code hat die Eigenschaft, dass ein normaler Benutzer nicht darauf zugreifen, ihn ändern oder erweitern kann. Derartiger Code wird wahlweise als Microcode oder als Firmware bezeichnet ist.

Firmware Code eines Mainframes wird von Prozessoren mit der System z Architektur ausgeführt. Microcode eines Mainframes wird von nicht-System z Prozessoren ausgeführt, z.B. von PowerPC Prozessoren. Auf den weiter unten erwähnten SAPs (System Assist Prozessor) läuft Firmware; auf der Channel Adapter Card befindet sich ein PowerPC Prozessor, der Microcode ausführt.

Außerhalb der Mainframes Welt ist der Unterschied zwischen Firmware und Microcode weniger sauber definiert. Firmware/Microcode läuft z.B. auf dem Prozessor, der einen WLAN Access Point, einem Mobiltelefon, eine Geschirrspülmaschine oder den elektrischen Fensterheber Ihres Mercedes S-Klasse Autos steuert.

Das auf der folgenden Abbildung erwähnte „Channel Subsystem“ besteht aus „System Assist Processoren (SAP)“ plus Firmware. Control Units und Enterprise Storage Server verwenden sehr leisungsfähige Prozessoren aber keine Firmware, da ihr Code nicht auf System z Prozessoren läuft.

System Assist Processor

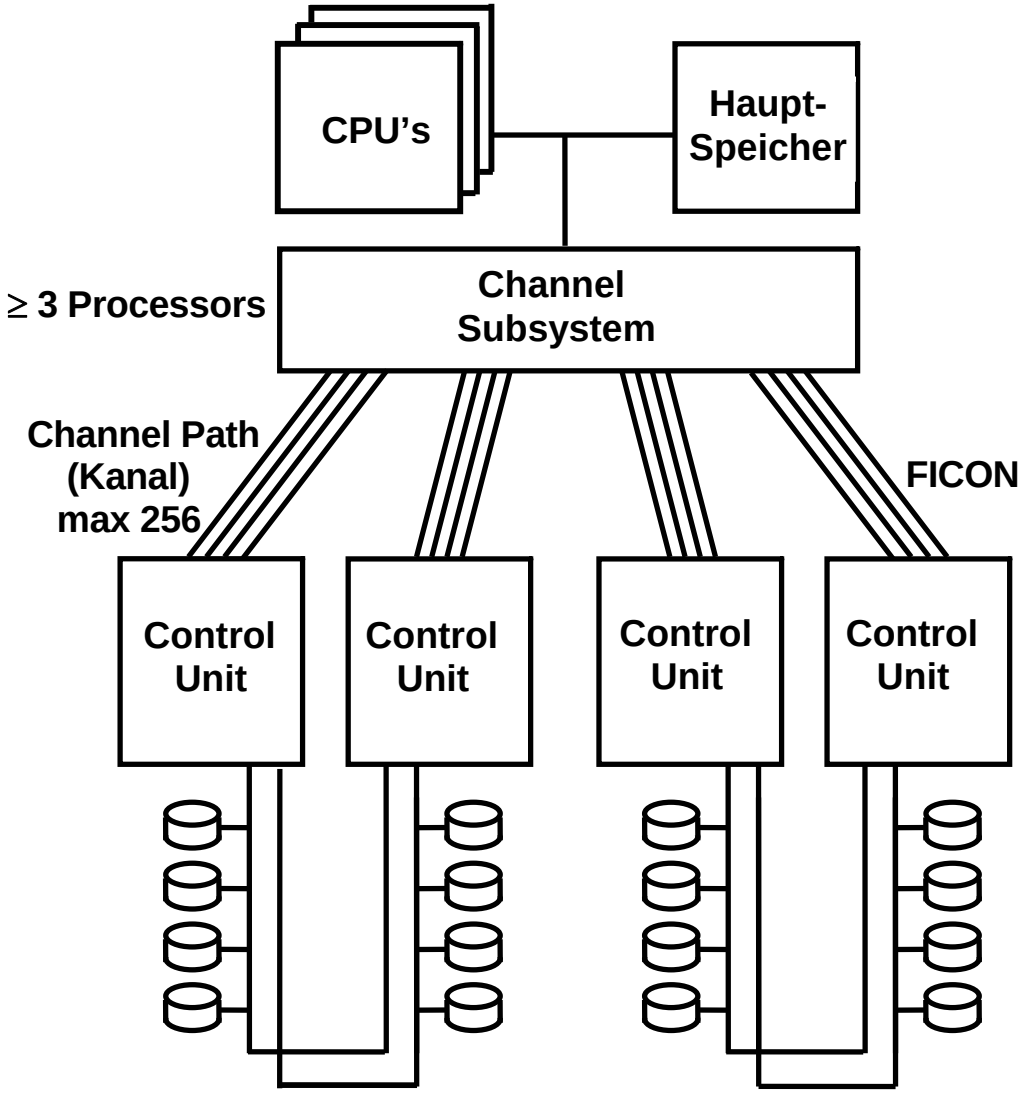
Ein z196 Rechner hat bis zu 4 x 24 = 96 Prozessoren (Cores). Von diesen wird nur der größere Teil als CPUs eingesetzt. Mehrere Prozessoren, als „System Assist Processoren“ (SAP) bezeichnet und vorkonfiguriert, führen ausschließlich Firmware Code aus. Bei einem maximal hochgerüsteten z196 Rechner sind dies mindestens 14 SAPs. Die SAPs weisen die gleiche Hardware Architektur auf wie die CPUs; auf ihnen läuft aber Firmware und kein z/OS. Bei der Installation eines neuen System z Rechners wird über eine Konfigurationsdatei eingestellt, wie viele der vorhandenen Prozessoren als CPUs bzw. SAPs eingesetzt werden.

SAPs und ihr Firmware Code wird vor allem für drei Funktionen benötigt:

1. **Fehlerbehandlungs- und Recovery Funktionen**
2. **Channel Subsystem**
3. **PR/SM Hypervisor Software für LPAR virtuelle Maschinen (wird später diskutiert).**

Für Firmware (und dazugehörige Daten) sind in einem z196 Rechner 16 oder 32 GByte Speicherplatz vorgesehen. Von dem installierten physischen Speicher werden 16 oder 32 GByte abgeteilt und steht als „Hardware System Area“ (HSA) für Firmware Zwecke zur Verfügung. Der Rest kann als Hauptspeicher genutzt werden.

Bitte beachten: Den Begriff SAP (System Assist Processor) nicht verwechseln mit dem Namen der Firma SAP AG in Walldorf (Baden).



System z Plattenspeicher Anschluss

Ein Mainframe kann über mehrere (bis zu 8) Kanäle mit einer bestimmten Control Unit verbunden werden, und ein I/O Gerät kann an mehr als eine Control Unit angeschlossen werden.

Das Channel Subsystem bildet die logische I/O Konfiguration, wie sie das Betriebssystem sieht, auf die physische Konfiguration ab.

Heute werden mehrere Control Units und angeschlossene Plattenspeicher zu einem physischen „Enterprise Storage System“ zusammengefasst.

Ein Plattenspeicher ist typischerweise an zwei Control Units angeschlossen. Die Kommunikation mit der CPU erfolgt wahlweise über eine der beiden Control Units.

z/Architecture Principles of Operation, SA22-7201

Mehrfache Pfade zu einem I/O Gerät

System z I/O Geräte werden über Steuereinheiten (Control Units) an das Kanal Subsystem angeschlossen. Viele Funktionen, die auf Ihrem PC von der CPU als I/O Driver Code ausgeführt werden, sind bei einem Mainframe in den Control Units implementiert. Sie belasten die CPU's nur wenig, und ermöglichen die Ansteuerung einer sehr großen Anzahl von Festplatten.

Steuereinheiten können über mehr als einen Kanalpfad an das Kanalsubsystem angeschlossen werden und I/O Geräte (z.B. Plattenspeicher) können an mehr als eine Steuereinheit angeschlossen werden.

z/OS kann auf ein beliebiges I/O Gerät über bis zu 8 unterschiedliche Kanalpfade zugreifen (und umgekehrt).

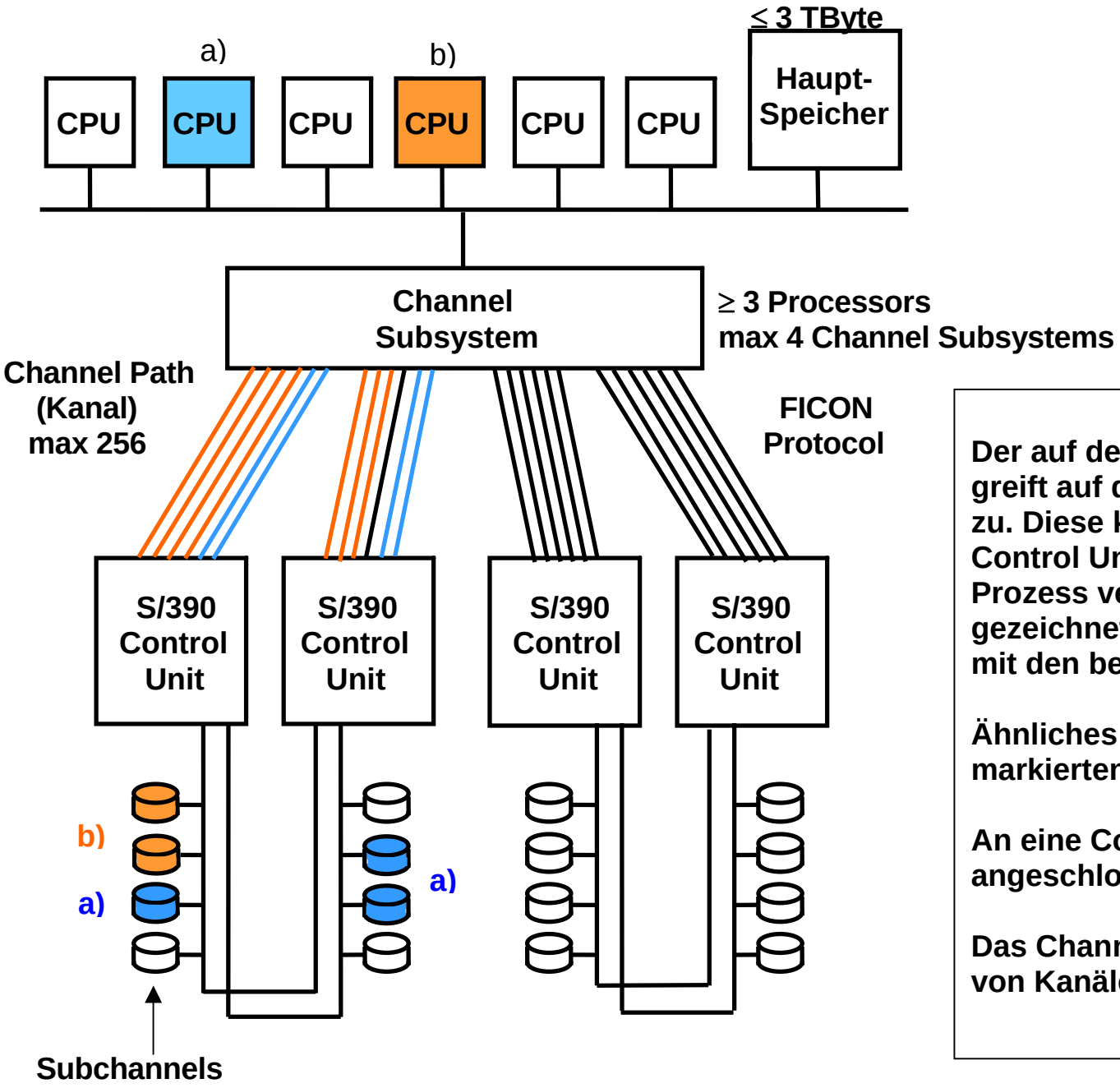
Der Zugriffsweg kann dynamisch geändert werden (DPS - Dynamic Path Selection). Eine I/O Operation muss nicht auf dem gleichen Weg abgeschlossen werden, auf dem sie gestartet wurde. Hiermit kann erreicht werden, dass sequentielle und zufallsbedingte Zugriffe sich nicht gegensätzlich beeinträchtigen.

Z.B. angenommen zwei Plattenspeicher, die an die gleiche Steuereinheit angeschlossen sind. Ein Plattenspeicher überträgt einen großen Block sequentieller Daten, während der zweite Plattenspeicher gleichzeitig viele kurze Datenpakete mit einem zufallsbedingten Zugriffsmuster überträgt. Kein Plattenspeicher soll die Nutzung einer Verbindung für einen längeren Zeitraum usurpieren.

In anderen Worten, es gibt mehrere Wege, auf denen Daten (und Steuerinformation) zwischen Platte und CPU übertragen werden können.

Diese dynamische Weg-Steuerung ist rechenaufwendig. Deswegen belastet man mit dieser Aufgabe nicht die CPUs und das Betriebssystem, sondern überträgt sie einer getrennten Verarbeitungseinheit, dem Channel Subsystem. Eine Beispiel-Konfiguration ist in der folgenden Abbildung wiedergegeben.

System z Disk Storage Attachment

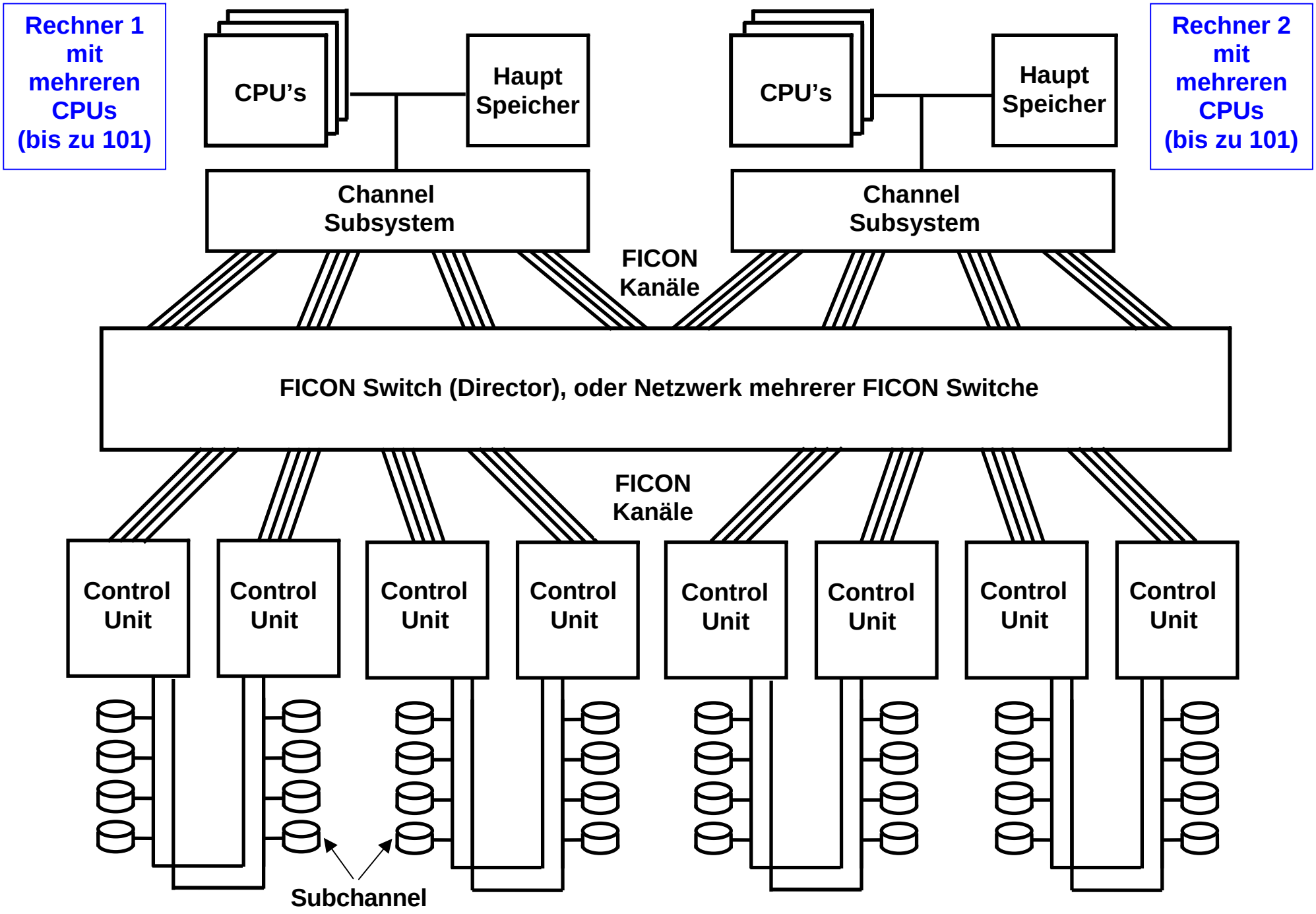


Der auf der blauen CPU a) laufende Prozess greift auf die drei blauen Plattenspeicher a) zu. Diese können über die ersten beiden Control Units erreicht werden. Der blaue Prozess verwendet hierzu vier blau gezeichnete Kanäle, von denen jeweils zwei mit den beiden Control Units verbunden sind.

Ähnliches gillt für den Prozess auf der orange markierten CPU b).

An eine Control Unit können bis zu 8 Kanäle angeschlossen sein.

Das Channel Subsystem kann die Zuordnung von Kanälen zu Prozessen dynamisch ändern.


Rechner 1 mit mehreren CPUs (bis zu 101)
CPU's
Haupt Speicher
Channel Subsystem
Rechner 2 mit mehreren CPUs (bis zu 101)
CPU's
Haupt Speicher
Channel Subsystem
FICON Kanäle
FICON Switch (Director), oder Netzwerk mehrerer FICON Switche
FICON Kanäle
Control Unit
Control Unit
Control Unit
Control Unit
Control Unit
Control Unit
Control Unit
Control Unit
Subchannel

FICON Director

Die obige Abbildung zeigt zwei Mainframe Rechner und zahlreiche Control Units mit ihren Plattenspeichern. Jeder zEC12 Mainframe Rechner kann bis zu 101 CPUs enthalten und verfügt über ein eigenes Channel Subsystem.

Es ist in der Regel notwendig, beide (bis zu 32) Mainframe Rechner mit allen Control Units und allen Plattenspeichern zu verbinden.

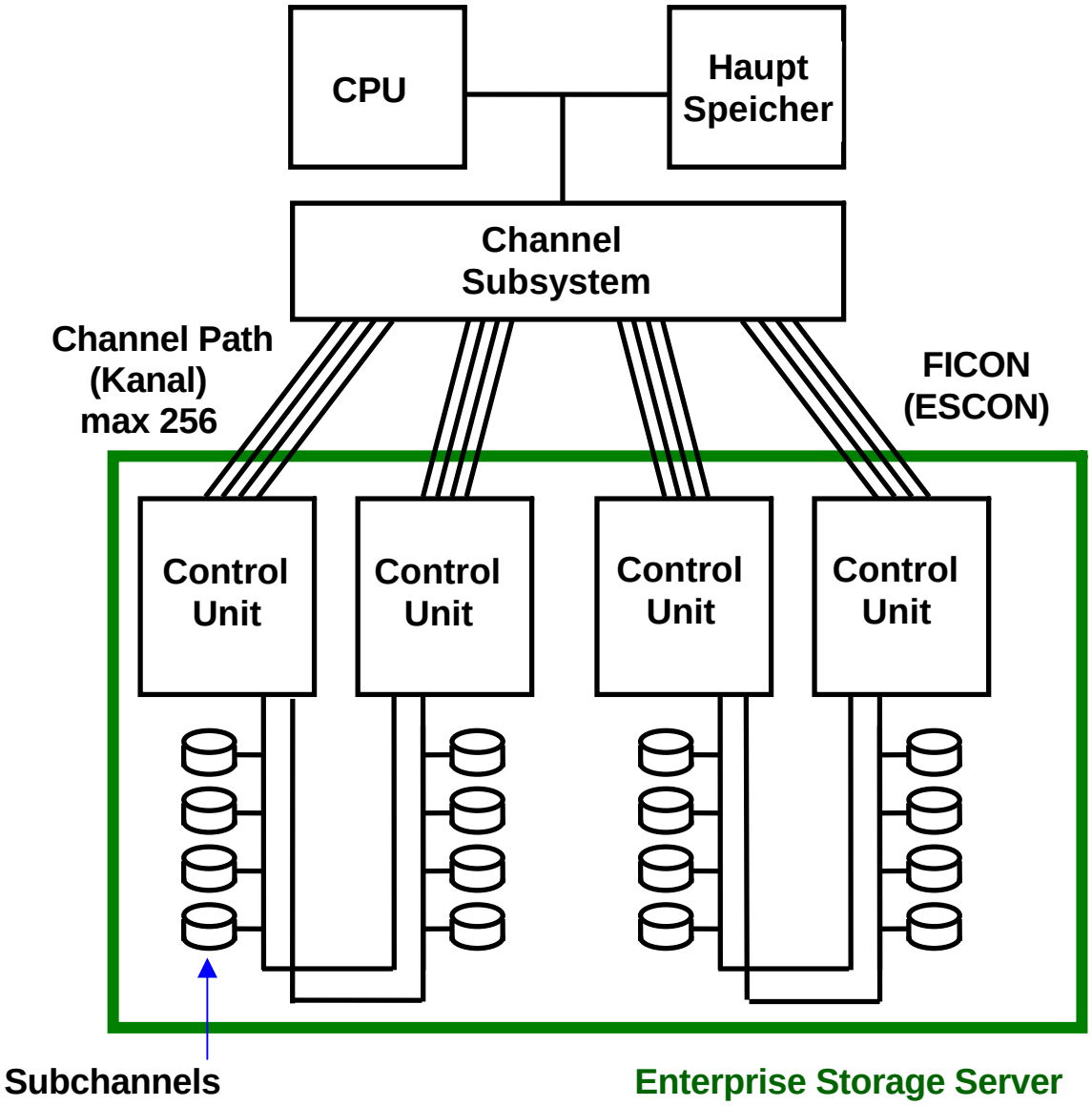
Da die Verkabelung sehr unübersichtlich ist, schaltet man einen FICON Switch (offizielle Bezeichnung „FICON Director“) zwischen die Channel Subsysteme und und die Control Units. Vielfach wird an Stelle eines einzelnen FICON Switches ein ganzes Netzwerk von FICON Switchen eingesetzt (nicht gezeigt).

Große Mainframe Installationen können über 10 000 Glasfaserkabel aufweisen.

Die I/O Anforderungen der Prozesse auf den einzelnen CPUs müssen ihren Weg durch das Netzwerk zu dem gewünschten Plattenspeicher finden. Die Channel Subsysteme der einzelnen Rechner haben die Aufgabe, das Routing der I/O Anforderungen über das FICON Netzwerk zu übernehmen. Ähnlich wie im Internet kann sich der Weg zwischen CPU und Plattenspeicher während der Ausführung einer I/O Operation dynamisch ändern.

Ein derartiges Netzwerk wird als „Storage Area Netzwerk“ (SAN) bezeichnet. Ein SAN benutzt das Fibre Channel Protokoll an Stelle der in Communicationsnetzen üblichen TCP/IP oder SNA Protokolle.

Die in der obigen Abbildung gezeigten beiden Channel Subsysteme haben eine weitere Aufgabe: Sie verbergen die komplexe I/O Konfiguration vor den z/OS Betriebssystemen, die auf den beiden Rechnern laufen. Ein z/OS Betriebssystem arbeitet mit der Illusion, dass alle Plattenspeicher über individuelle Punkt-zu-Punkt Verbindungen direkt an seine CPUs angeschlossen sind. Diese Verbindungen werden als „Subchannels“ bezeichnet; je ein Subchannel pro Plattenspeicher. Es ist die Aufgabe des Channel Subsystems, logische Subchannels auf physische Kanalpfade und die Topologie des FICON Netzwerkes dynamisch abzubilden.



Enterprise Storage Server

Früher waren Control Units getrennte physische Einheiten in ihren eigenen Gehäusen.

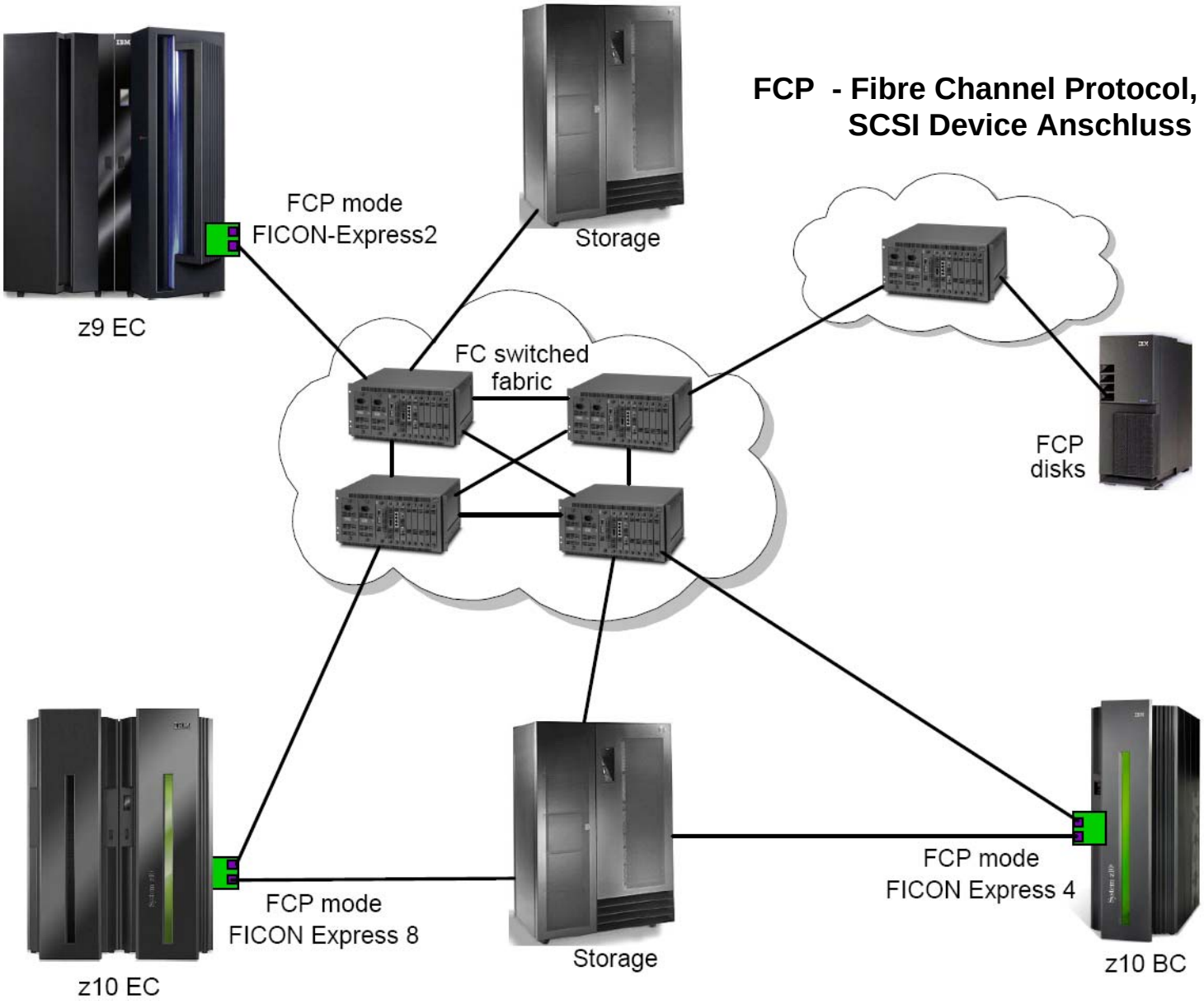
Heute werden mehrere Control Units und angeschlossene Plattenspeicher zu einem physischen „Enterprise Storage Server“ (ESS) zusammengefasst, der auch die angeschlossenen Plattenspeicher enthält.

Der ESS emuliert mehrere logische Control Units, hat Anschlüsse für zahlreiche FICON Kanäle und bringt zahlreiche Plattenspeicher im gleichen Gehäuse unter.

Es können beliebig viele ESS angeschlossen werden.

z/Architecture Principles of Operation, SA22-7201

FCP - Fibre Channel Protocol,
SCSI Device Anschluss



Heutige Mainframes können alternativ neben FICON Plattenspeichern auch Fibre Channel SCSI Plattenspeicher anschließen. Dies ist z.B. beim Einsatz von zLinux üblich. Der Großteil der Daten einer Mainframe Installation wird aber auf FICON Platten gespeichert.