UNIVERSITÄT LEIPZIG

Enterprise Computing
Einführung in das Betriebssystem z/OS

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

WS 2012/13

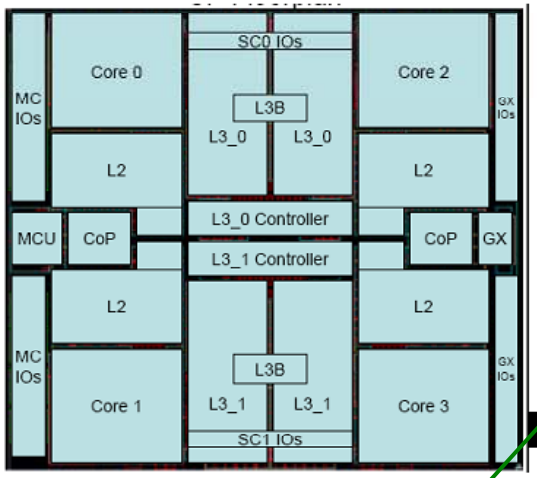
System z/Hardware Teil 3

Book und System Frame

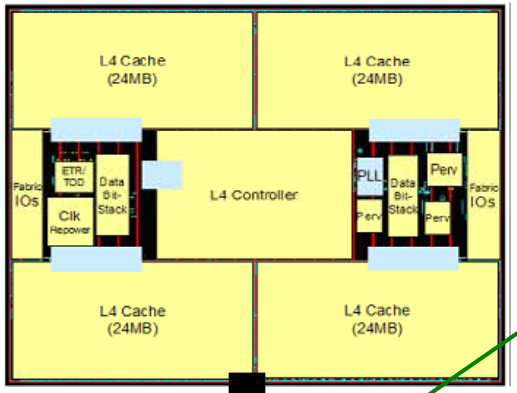
© copyright W. G. Spruth,

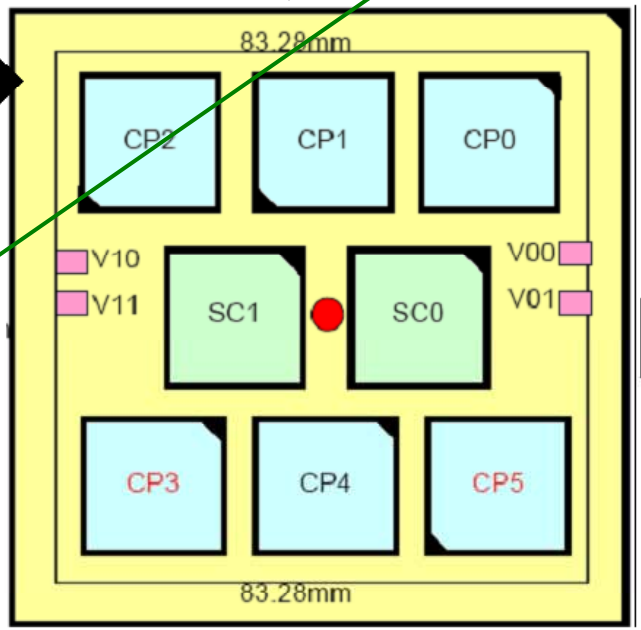
z196 PU chip, SC chip and MCM

z196 Quad Core PU CHIP



96 MB SC CHIP





MCM

Book

Book

Front View Fanouts

Gezeigt ist ein CPU Chip (CP) und ein L4 Cache Chip (SC Chip). 6 + 2 dieser Chips befinden sich auf auf einem Multichip Module (MCM).

Das MCM ist Bestandteil eines Books.

Book

Das MCM ist Bestandteil einer als „Book“ bezeichneten Baugruppe, welche neben dem MCM noch Steckplätze für 30 Hauptspeicher DIMMs (Dual Inline Memory Module) sowie „Fan Out Adapter“ Karten enthält. Fan Outs Cards werden auch als Host Connector Adapter (HCA) Cards bezeichnet. Sie erfüllen die gleiche Funktion wie die I/O Adapter Cards der Sun oder Hewlett Packard System Boards.

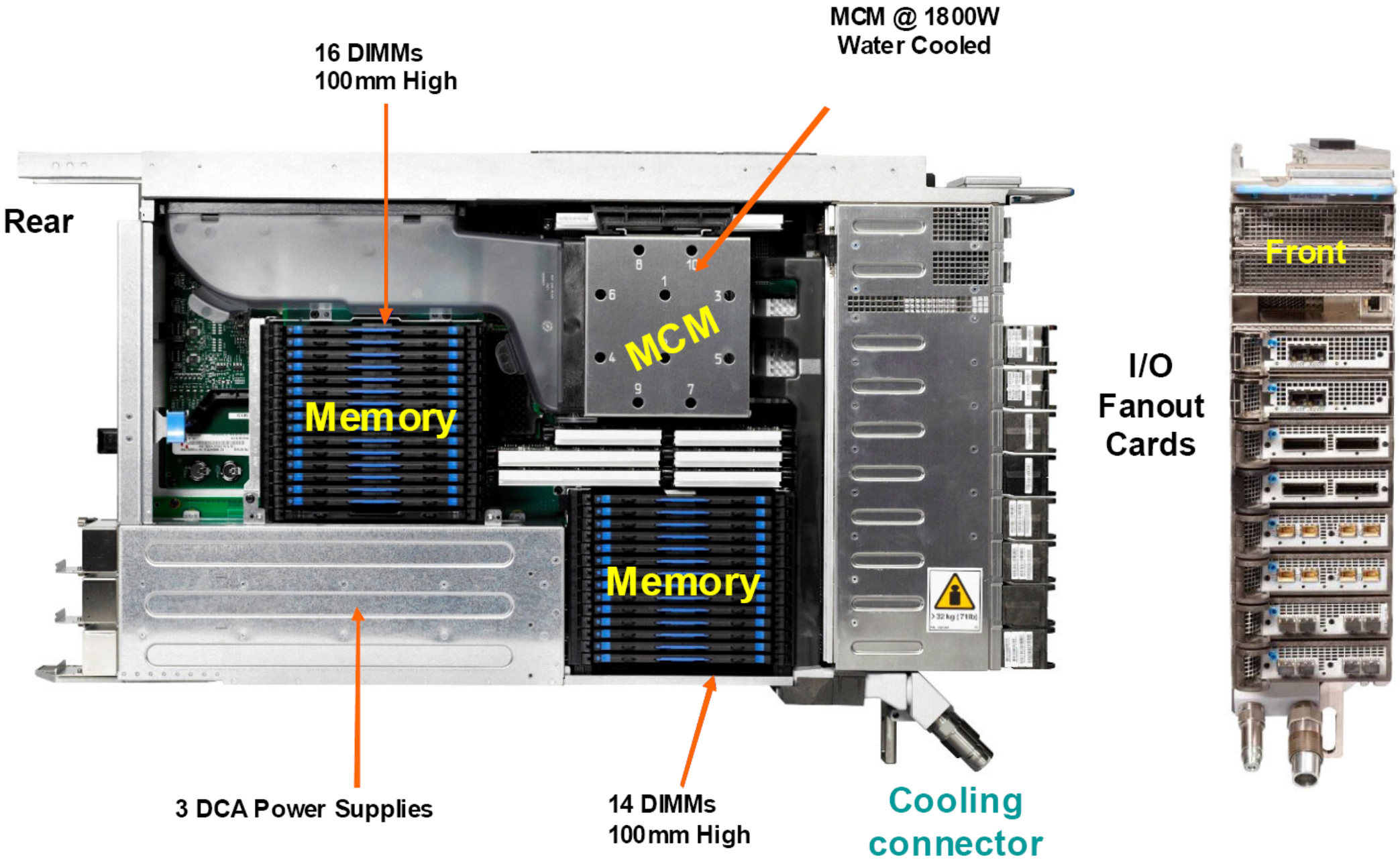
Hauptspeicher DIMMs haben eine Kapacität von je 4 GByte, 16 GByte oder 32 GByte. Die maximale physische Hauptspeicherkapazität beträgt somit 960 GByte, von denen nach Abzug für Fehlerkorrektur und Redundanz 768 GByte verfügbar sind.

Die Fanout Cards führen zu Steckkontakten auf der Vorderseite (Front View) des Books. Es existieren 8 Steckkontakte , die mit jeweils 2 Kabeln eine Verbindung zu einem „I/O Cage“ aufnehmen, in denen Steckkarten für I/O Anschlüsse untergebracht sind. Zwei weitere Steckkontakte (FSP) werden zur Verbindung zu zwei „Service Elementen“ verwendet die später diskutiert werden.

Es existiert eine zweistufige Stromversorgung. Die primäre Stromversorgung (Bulk Supply) versorgt einen ganzen Rechner, und ist aus Zuverlässigkeitsgründen doppelt vorhanden. In jedem Book befindet sich eine zusätzliche sekundäre Stromversorgung (Distributed Converter Assembly, DCA) , welche kurzfristige Versorgungs-schwankungen innerhalb eines Books ausgleicht.

Die folgende Abbildung zeigt die Seitenansicht eines (geöffneten) Books, in der das Multichip Modue (MCM), die Hauptspeicher DIMMs, die (secundäre) Stromversorgung und (hinter der Abdeckung verborgen) die Fanout Cards zu sehen sind.

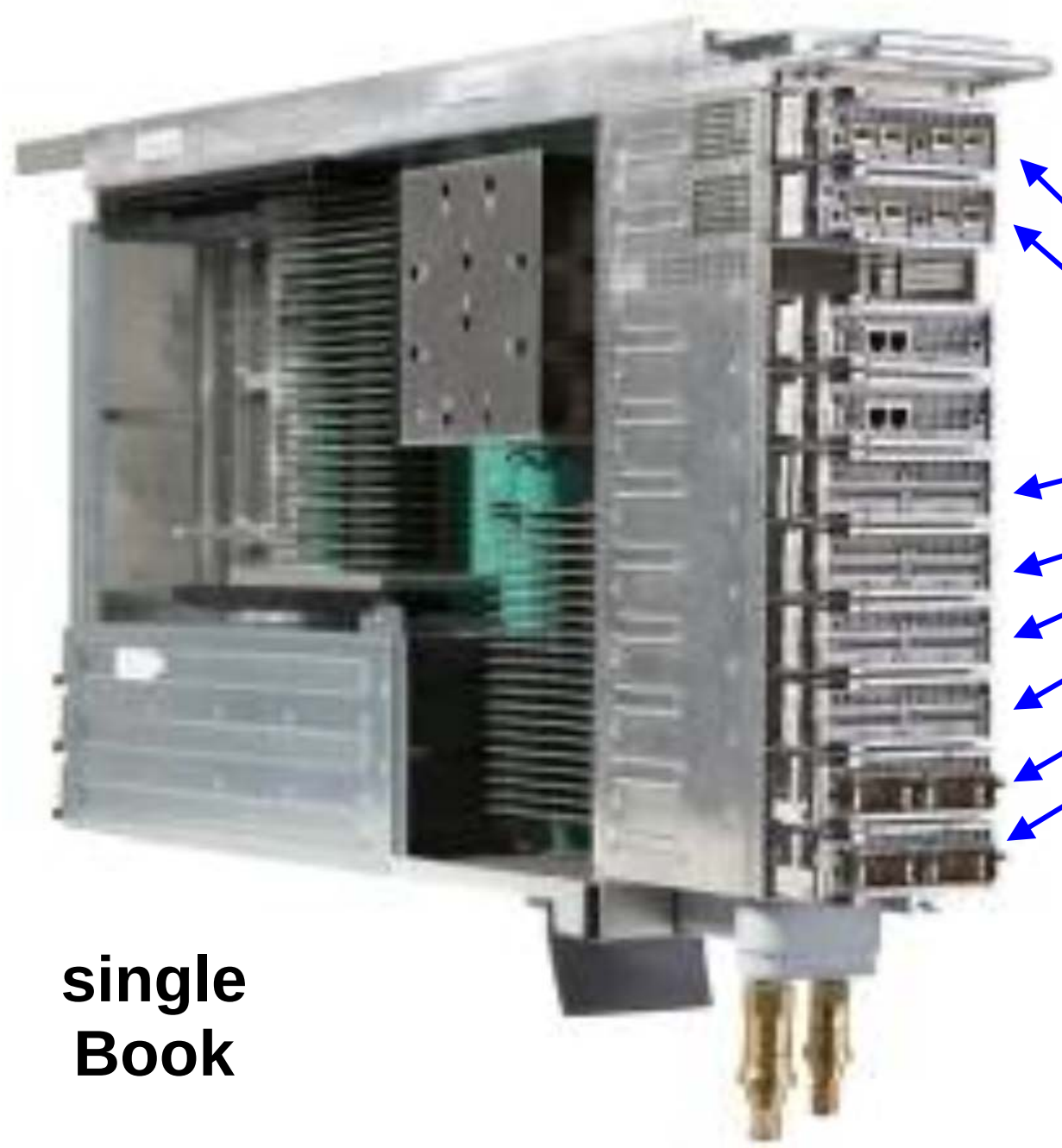
Voltage Tansformation Modules (VTM) bewirken Power Conversion in dem Book und benutzen Triple Redundancy. Die Redundanz schützt die Processoren vor einem Spannungsverlust als Folge des Versagens einer VTM Card. Triple redundancy wird auch für die Humidity (Luftfeuchtigkeit) Sensoren zur Verbesserung der Zuverlässigkeit eingesetzt.


16 DIMMs
100mm High
MCM @ 1800W
Water Cooled
Rear
Memory
MCM
Memory
Front
I/O
Fanout
Cards
3 DCA Power Supplies
14 DIMMs
100mm High
Cooling
connector

zEC12 Book

Jedes zEC12 Book enthält die folgenden Komponenten:

- **Ein Multi-Chip-Modul (MCM) mit sechs Hex-Core-Mikroprozessor-Chips, und zwei Storage Control Chips mit 384 MByte Level 4 Cache.**
- **Speicher DIMMs in 30 verfügbaren Slots. Dies ermöglicht von 60 GByte bis 960 GByte physischen Speicher pro Book.**
- **Eine Kombination von bis zu acht Host Channel Adapter (HCA) oder PCIe Fanout-Karten.**
- **PCIe Fanouts werden für 8Gbit/s Links zu den PCIe I/O-Karten verwendet. Die HCA-Optical Fanouts verbinden zu externen coupling links (CPC).**
- **Drei verteilte Wandler-Baugruppen (DCAs), die das Book mit Strom versorgen. Bei Verlust einer DCA ist genügend Power vorhanden (n +1 Redundanz), um den Strombedarf des Books zu befriedigen. Die DCAs können während des laufenden Betriebes gewartet werden.**
- **Zwei „Flexible Service-Prozessor“ (FSP)-Karten für die Systemsteuerung.**



single Book

8 Fanout Card Anschüsse

2 Anschlüsse pro Fanout Card

Fanout Cards werden auch als „Host Channel Adapter“ (HCA) oder bei Sun und HP als I/O Cards bezeichnet. Pro Fanout Card sind 2 nebeneinander-liegende Steckkontakte für Kabel-verbindungen mit einem I/O Cage vorhanden.

Book

Das MCM ist Bestandteil einer als „Book“ bezeichneten Baugruppe, welche neben dem MCM noch Steckplätze für 30 Hauptspeicher DIMMs (Dual Inline Memory Module) sowie „Fan Out Adapter“ Karten enthält. Fan Outs Cards werden auch als Host Connector Adapter (HCA) Cards bezeichnet. Sie erfüllen die gleiche Funktion wie die I/O Adapter Cards der Sun oder Hewlett Packard System Boards.

Hauptspeicher DIMMs haben eine Kapacität von je 4 GByte, 16 GByte oder 32 GByte. Die maximale physische Hauptspeicherkapazität beträgt somit 960 GByte, von denen nach Abzug für Fehlerkorrektur und Redundanz 768 GByte verfügbar sind.

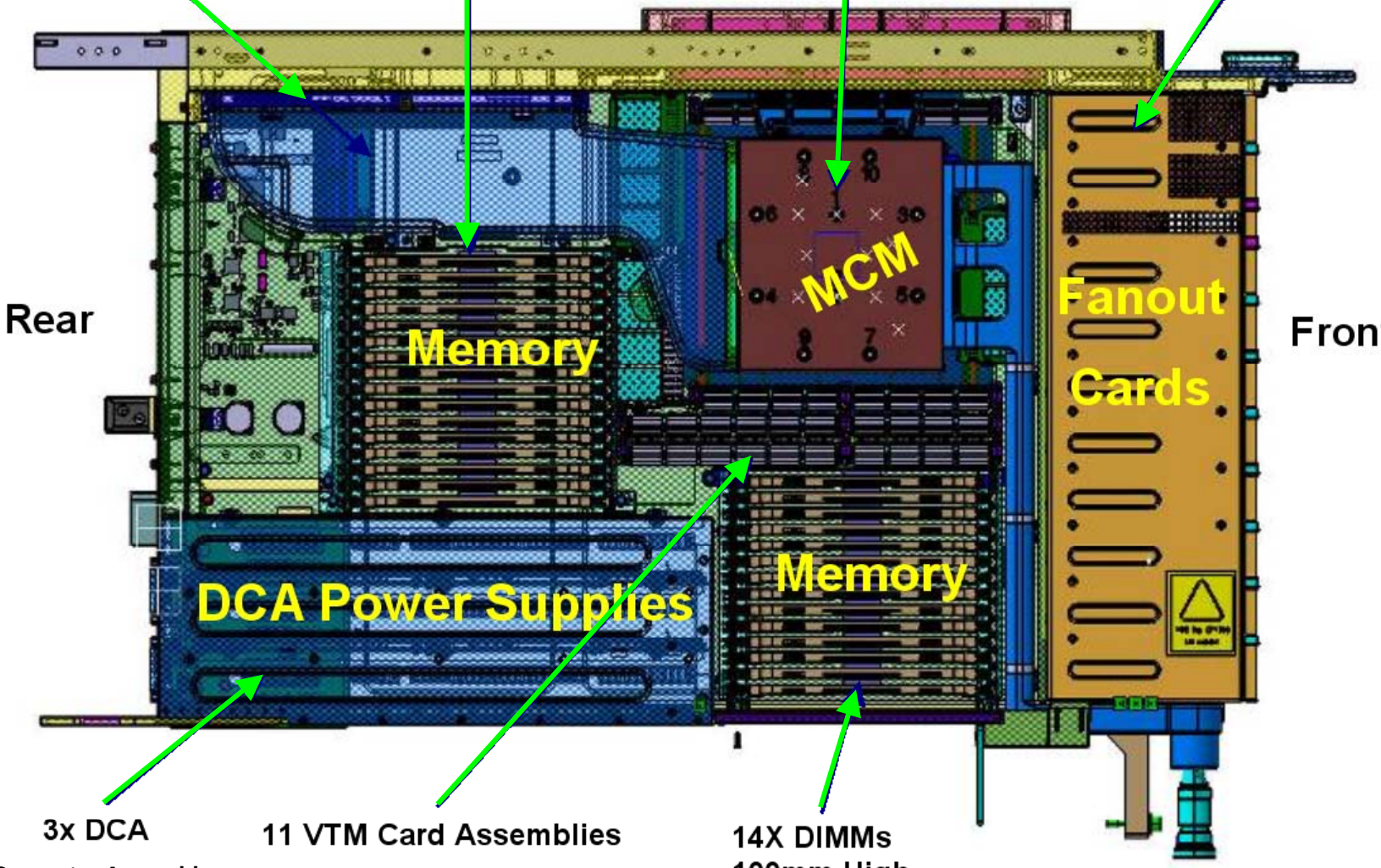
Die Fanout Cards führen zu Steckkontakten auf der Vorderseite (Front View) des Books. Es existieren 8 Steckkontakte , die mit jeweils 2 Kabeln eine Verbindung zu einem „I/O Cage“ aufnehmen, in denen Steckkarten für I/O Anschlüsse untergebracht sind. Zwei weitere Steckkontakte (FSP) werden zur Verbindung zu zwei „Service Elementen“ verwendet die später diskutiert werden.

Es existiert eine zweistufige Stromversorgung. Die primäre Stromversorgung (Bulk Supply) versorgt einen ganzen Rechner, und ist aus Zuverlässigkeitsgründen doppelt vorhanden. In jedem Book befindet sich eine zusätzliche sekundäre Stromversorgung (Distributed Converter Assembly, DCA) , welche kurzfristige Versorgungs-schwankungen innerhalb eines Books ausgleicht.

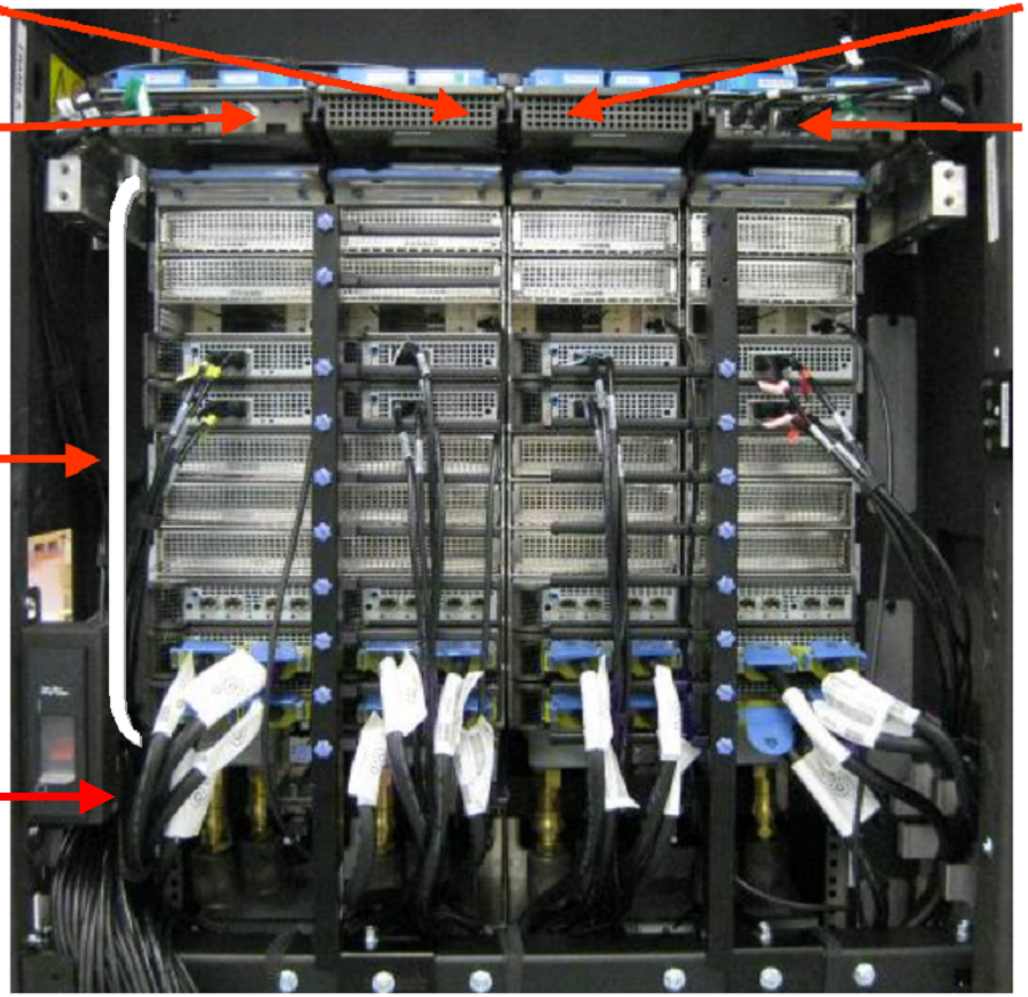
Die folgende Abbildung zeigt die Seitenansicht eines (geöffneten) Books, in der das Multichip Modue (MCM), die Hauptspeicher DIMMs, die (secundäre) Stromversorgung und (hinter der Abdeckung verborgen) die Fanout Cards zu sehen sind.

Voltage Tansformation Modules (VTM) bewirken Power Conversion in dem Book und benutzen Triple Redundancy. Die Redundanz schützt die Processoren vor einem Spannungsverlust als Folge des Versagens einer VTM Card. Triple redundancy wird auch für die Humidity (Luftfeuchtigkeit) Sensoren zur Verbesserung der Zuverlässigkeit eingesetzt.

z196 Book Layout



30 Dual Inline Memory Module (DIMM) insgesamt



FSP: Flexible Service Processor

STP Server Time Protocol

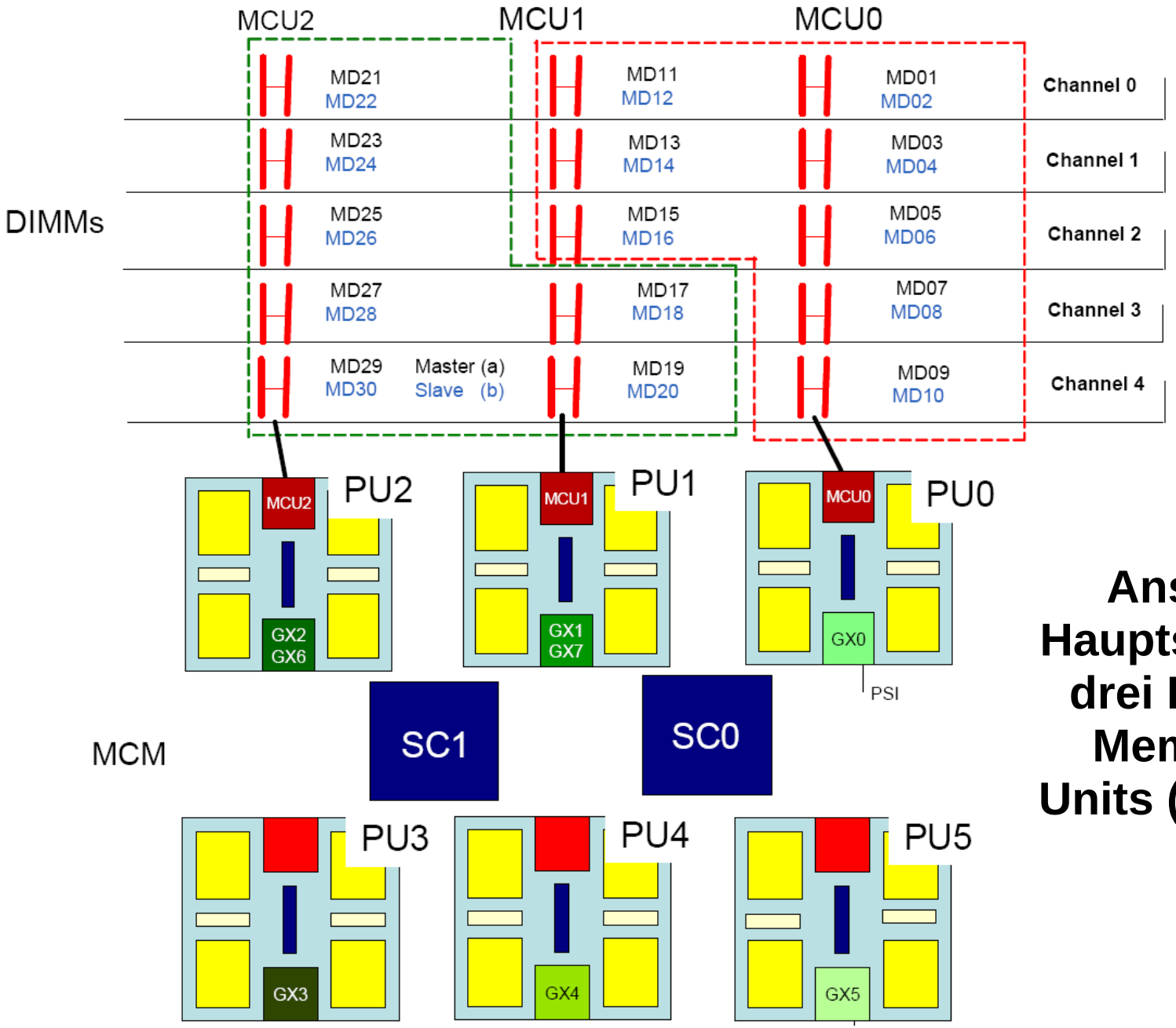
PCIe I/O Infrastructur mit einer Bandbreite von 8 Gbit/s

InfiniBand I/O Infrastructure mit einer Bandbreite von 6 Gbit/s

Die obige Abbildung zeigt die Vorderseite(n) von 4 nebeneinander stehende Books. Zu sehen sind die Steckkontakte, über die mit Infiniband oder PCIe Kabeln die Verbindung zu einem (max 3) I/O Cage hergestellt werden, welcher I/O Adapter Karten (z.B Plattenspeicher Anschlüsse) aufnimmt, die eine Verbindung zur Außenwelt übernehmen.

Zwei weitere Steckkontakte (FSP) werden zur Verbindung zu zwei „Service Elementen“ verwendet die später diskutiert werden. Zwei Oscillator Karten stellen Clock Signale zur Verfügung und stellen eine Synchronisation aller CPUs mittels des Server Time Protocols (STP) sicher.

Hier entfernt ein Entwicklungsingenieur ein Book aus einem z9 Rechner.



Anschluss des Hauptspeichers über drei Busse an drei Memory Control Units (MCU) an jedes MCM.

Verbindung MCM - Hauptspeicher

Die folgende Abbildung zeigt den Hauptspeicheranschluss.

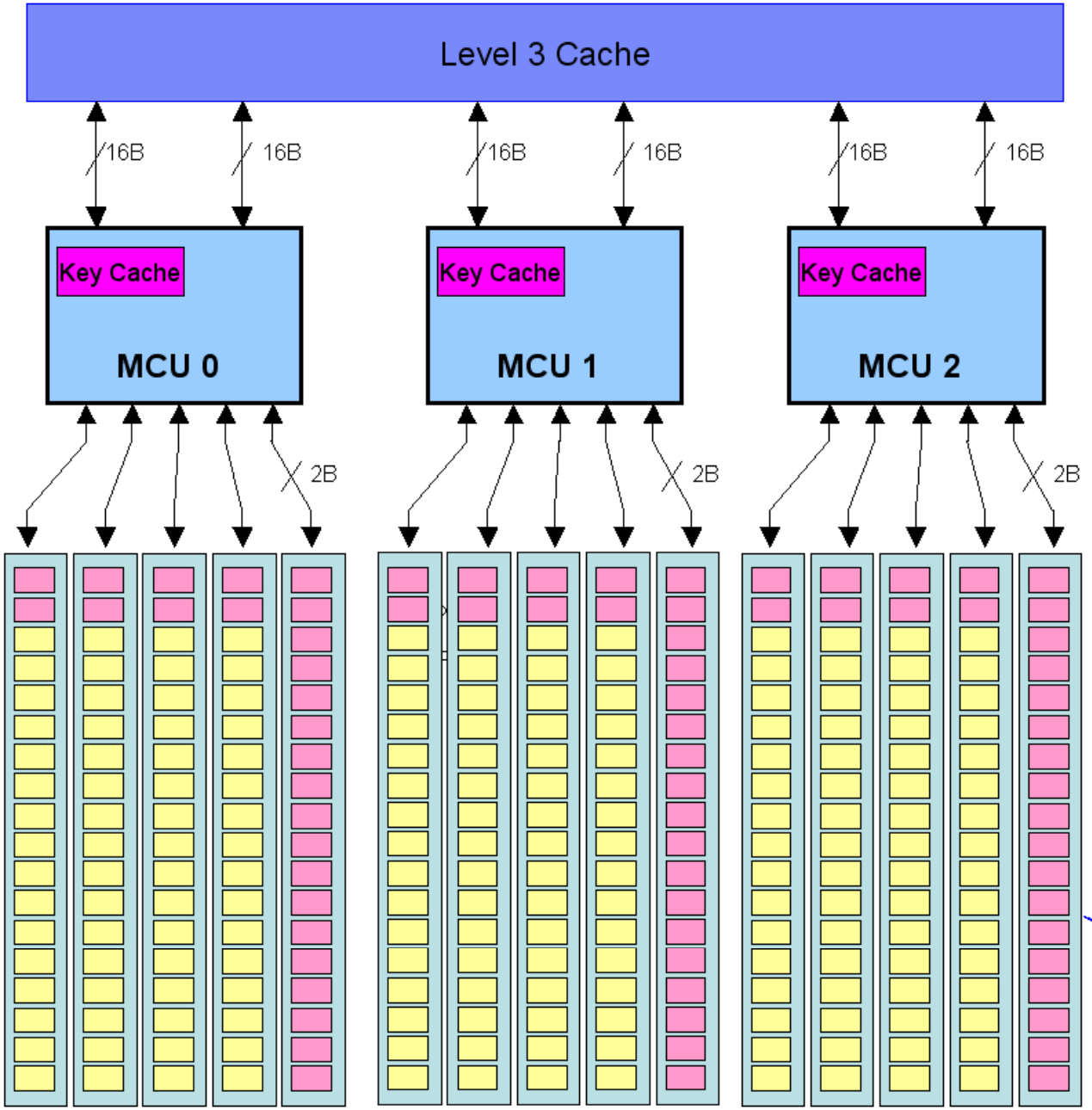
Dieser erfolgt mit Hilfe von drei „Memory Control Units“ (MCU), welche über zwei je 16 Byte (2 x 128 Bit) breite Busse mit den L3 Caches der einzelnen CPU Chips verbunden sind. Die Memory Control Units enthalten zusätzlich einen Cache, der die am häufigsten gebrauchten Storage Protection Keys (siehe Einführung Teil 3 dieser Vorlesung) enthält.

Der Hauptspeicher selbst enthält 4 „Channels“ (Colums) von Hauptspeicher Quad High DIMMs (Dual Inline Memory Module), vom Aussehen her ähnlich wie die DIMMs in Ihrem PC. Jeder Channel verwendet eine spezielle Version des ReedSolomon Codes an Stelle des sonst für Hauptspeicher üblichen Hamming Fehlerkorrektur Codes.

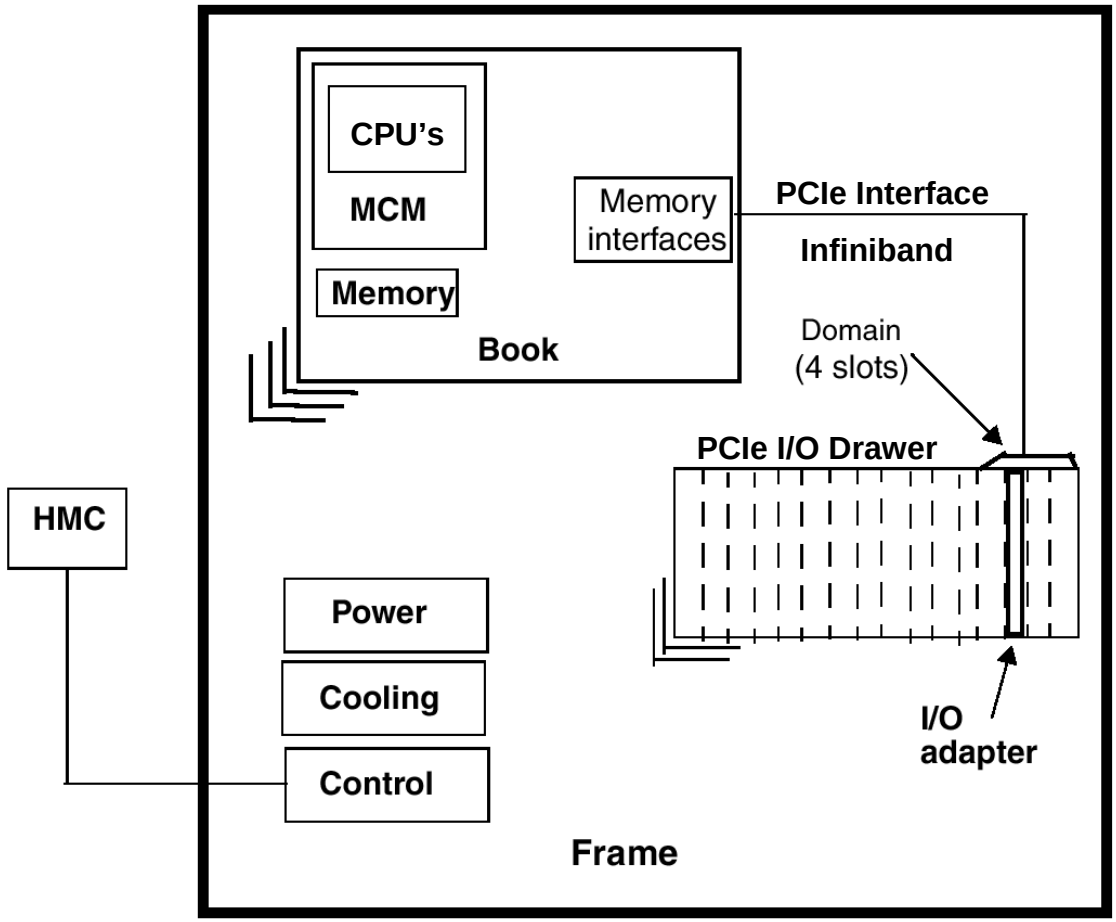
Als weitere Fehlercorrekturmaßnahme existiert ein 5. RAIM (Redundant Array od Independent Memory) Channel. RAIM benutzt das gleiche Verfahren wie RAID für Plattenspeicher. Wenn eine der 4 Colums ausfällt, kann der Hauptspeicherinhalt mit Hilfe der 5. Colum wieder hergestellt werden. RAIM wurde erstmalig mit der z196 für kommerziell erhältliche Rechner eingeführt.

An jedem Channel hängen 2 DIMMs; jede einzelne MCU bedient 10 DIMMs. Jedes DIMM Module speichert 32 GByte. Jedes Book verfügt über 30 Dual In-Line Memory Modules (DIMMs). Jedes DIMM hat eine Speicherkapazität von 32 GByte, für eine maximale Speicherkapazität von 960 GByte pro Book, oder 3 480 GByte für ein 4 Book System.

Da die RAIM DIMMs 20 % der Speicherkapazität in Anspruch nehmen, stehen dem Benutzer lediglich 768 GByte pro Book, oder 3 072 GByte pro System zur Verfügung. Davon werden 16 GByte für die „Hardware System Area“ (HSA) benötigt. Die HSA speichert Firmware, und wird später erläutert.

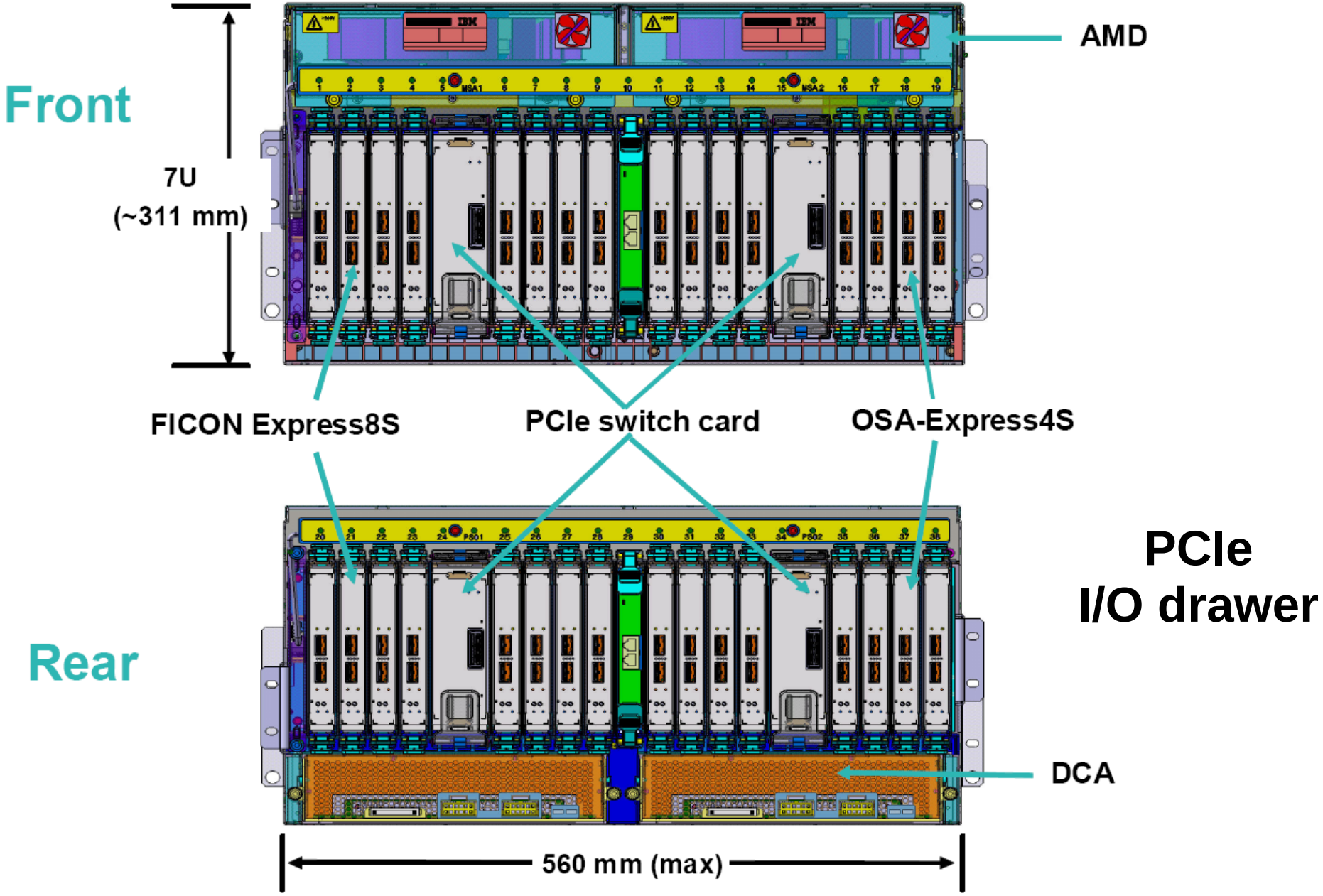


The parity of the four "data" DIMMs are stored in the DIMMs attached to the fifth memory channel.



Struktur eines zEC12 Mainframe Systems.

Bis zu 32 I/O Adapter Cards pro I/O Cage, für Verbindungen zu Plattenspeichern, Magnetbändern und anderen I/O Geräten.


Front
7U
(~311 mm)
AMD
FICON Express8S
PCIe switch card
OSA-Express4S
Rear
DCA
560 mm (max)
PCIe
I/O drawer

PCIe I/O drawer

Die Abbildung zeigt die interne Struktur eines z9, z10 z196 Mainframe oder zEC12 Systems. CPUs und Caches befinden sich auf einem Multi Chip Module (MCM). Ein einziges MCM, zusammen mit dem Hauptspeicher (main store) bilden eine als “Book” bezeichnete Baugruppe. Bis zu 4 Books sind möglich.

Die Hauptspeicherschnittstelle (memory interface) verbindet ein Book mit „I/O Adaptern“ (I/O Cards) die in einem „I/O Drawer” (auch als I/O Cage bezeichnet) untergebracht sind. Die Verbindung zwischen Book und I/O Drawer wird durch I/O Kabel hergestellt, welche das PCIe Protokoll implementieren, äquivalent zu dem PCIe Protokoll in einem PC. Die PCIe Drawer ist eine zweiseitige Drawer (I/O-Adapter auf beiden Seiten). Die Drawer enthält 32 I/O-Steckplätze, und kann bis zu 32 I/O Adapter Cards aufnehmen, für Verbindungen zu Plattenspeichern, Magnetbändern und anderen I/O Geräten. Letztere sind grundsätzlich in getrennten Gehäusen untergebracht.

Die PCIe I/O drawer nutzt PCIe als Infrastruktur. Die PCIe I/O-Bus-Infrastruktur Datenrate beträgt 8 Gbit/s. Bis zu 128 Kanäle (64PCIe I/O-Features) sind in einer I/O Drawer möglich. Die Drawer enthält 4 Switch Cards (zwei vorne, 2 hinten), und zwei DCAs für die redundante Stromversorgung.

Z Frame A Frame

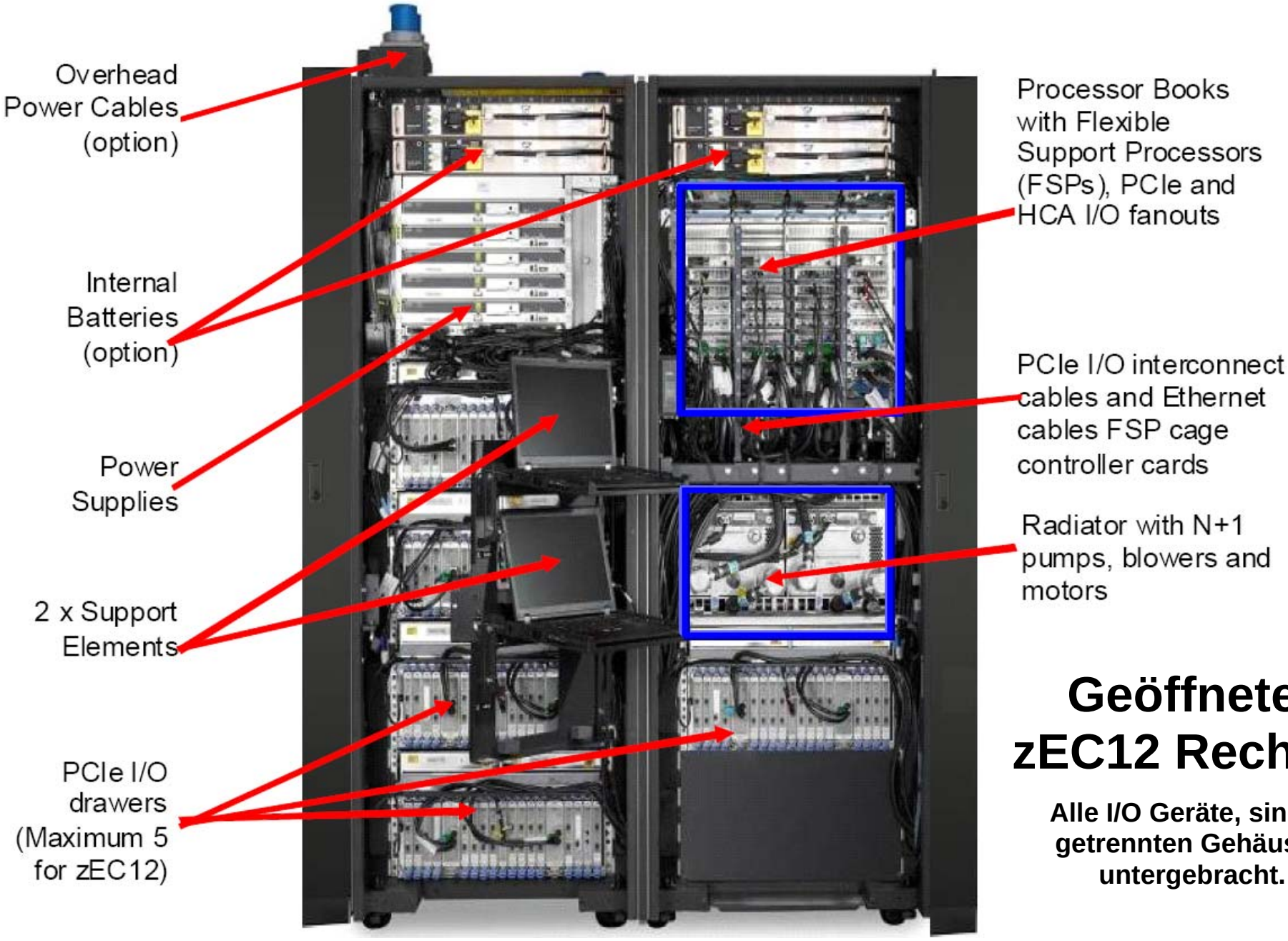
Ein z9, z10, z196 oder zEC12 Rechner besteht aus 2 meistens nebeneinander aufgestellten Rahmen, etwa 1,8 Meter hoch, welche von IBM als „Z Frame“ und „A Frame“ bezeichnet werden. Die Türen zu den beiden Frames sind künstlerisch geformed und enthalten viel leere Luft.

Ein z196 Rechner enthält bis zu 4 Books mit je 4 x 6 CPU Chips und 4 x 6 x 4 = 96 Prozessoren. Von diesen können 80 als CPUs eingesetzt werden, 16 arbeiten als „System Assist Prozessoren“ (SAP; erläutert später) und 2 dienen als Reseve (Spares), die aktiviert werden können, wenn ein anderer Processor ausfällt.

Es sind 786 GByte Hauptspeicher pro Book möglich, insgesamt also 3 TByte für ein System mit 4 Books.

Es kann ein Rechner mit 1, 2, 3 oder 4 Books ausgeliefert werden. Unser z9 Rechner hat nur 1 Book.

Beim Entwurf der Türen für die zEC12 Rahmen durften sich die künstlerisch motivierten Designer austoben.



Geöffneter zEC12 Rechner

Alle I/O Geräte, sind in getrennten Gehäusen untergebracht.

Geöffneter zEC12 Rechner

Gezeigt ist ein geöffneter zEC12 Rechner (ohne Türen). Rechts oben sind 4 Books zu sehen mit Anschluss Steckern auf der Vorderseite. Hinter den Steckern sitzen Host Connector Adapter (HCA) Cards, welche die Verbindung zu dem MCM herstellen. Diese nehmen entweder PCIe Bus Kabel für die Verbindung zu den I/O Drawers, Infiniband Kabel für die Verbindung zu anderen Rechnern, oder FSB Kabel (diskutiert weiter unten) auf.

Zu beachten ist: Alle I/O Geräte, besonders auch Plattenspeicher, sind in getrennten Gehäusen untergebracht.