

# 25 Jahre DKRZ

**Vom Institutsrechenzentrum  
zum  
Hochleistungsrechenzentrum**

# Acknowledgement

Viele Diagramme und Bilder sind mir von ehemaligen Kollegen aus dem DKRZ und aus Rechnerfirmen überlassen worden.

Ich danke Ihnen dafür vielmals.

# Themen

- Organisatorischer Rahmen
- Das GRZ als Vorläufer des DKRZ
- Das DKRZ zu Zeiten der CRAY-Rechner
- Das DKRZ zu Zeiten des HLRE

# Organisatorischer Rahmen

- Planungsphase des DKRZ
- DKRZ Gesellschafter
- Organe des DKRZ
- Personalplan
- Unterbringung
- Umstrukturierung

# Planungsphase des DKRZ (1985 -1987)

- Anlass: Erhöhter Rechenbedarf des MPIM
- Außerdem geeignete Infrastruktur für die Klimamodellierung in Deutschland
- Vorbild: NCAR oder GFDL in den USA
- Arbeitsgruppe aus BMFT, FHH, GKSS, MPG und MPIM zur Vorbereitung der Gründung eines „Klimarechenzentrums“

# Ergebnisse der Arbeitsgruppe

- Gemeinnützige GmbH als Rechtsform
- Etwa 50 Personalstellen im Endausbau
- Vektorrechner der CRAY-Klasse
- Leistungsgerechter Universalrechner
- Gesellschafter tragen einen großen Anteil der Betriebskosten (etwa 2/3)
- BMFT übernimmt Investitionskosten voll und trägt die restlichen Betriebskosten, alles im Wege der Projektfinanzierung

# DKRZ-Gründung

- Unterzeichnung der Gründungsvereinbarung durch BMFT, FHH, GKSS und MPG im Juli 1987
- Gründung des DKRZ als GmbH am 11.11.1987 in Hamburg
- Das DKRZ hat seinen Betrieb am 1.1.1988 aufgenommen

# DKRZ Gesellschafter

- Gründungsgesellschafter waren
  - Die Max-Planck-Gesellschaft mit 6 Geschäftsanteilen
  - Die Freie und Hansestadt Hamburg mit 3 Geschäftsanteilen
  - Die GKSS GmbH mit 1 Geschäftsanteil
- Zum 1.1.1991 trat die Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung als vierter Gesellschafter mit 1 Geschäftsanteil bei

# Organe des DKRZ

- Notwendige Organe einer GmbH sind
  - Gesellschafterversammlung
  - Geschäftsführung
- Ein Aufsichtsrat (AR) ist ein fakultatives Organ einer GmbH
  - Über den Aufsichtsrat sollte der BMFT/BMBF wegen seiner starken finanziellen Beteiligung Einfluss auf die Entwicklung des DKRZ nehmen können
  - Der AR bestand von März 1988 bis Dezember 1999

# Personalentwicklung

- Im Laufe des Jahres 1988 wurden die Abteilungen Systemsoftware, Anwendungssoftware, Kommunikation, Operating und Verwaltung mit insgesamt 39 Mitarbeitern aufgebaut
- Zum 1.1.1991 kam die Abteilung Modelle und Daten hinzu
- 1995 hatte das DKRZ zusammen mit Personal aus Drittmittelfinanzierung knapp 60 Mitarbeiter
- Nach der Umstrukturierung beschäftigte das DKRZ noch 20 Mitarbeiter

# Organisatorischer Rahmen

- Planungsphase des DKRZ
- DKRZ Gesellschafter
- Organe des DKRZ
- Personalplan
- **Unterbringung**
- Umstrukturierung

# GRZ-Räume 1978 – 1984

## 15. OG Geomatikum



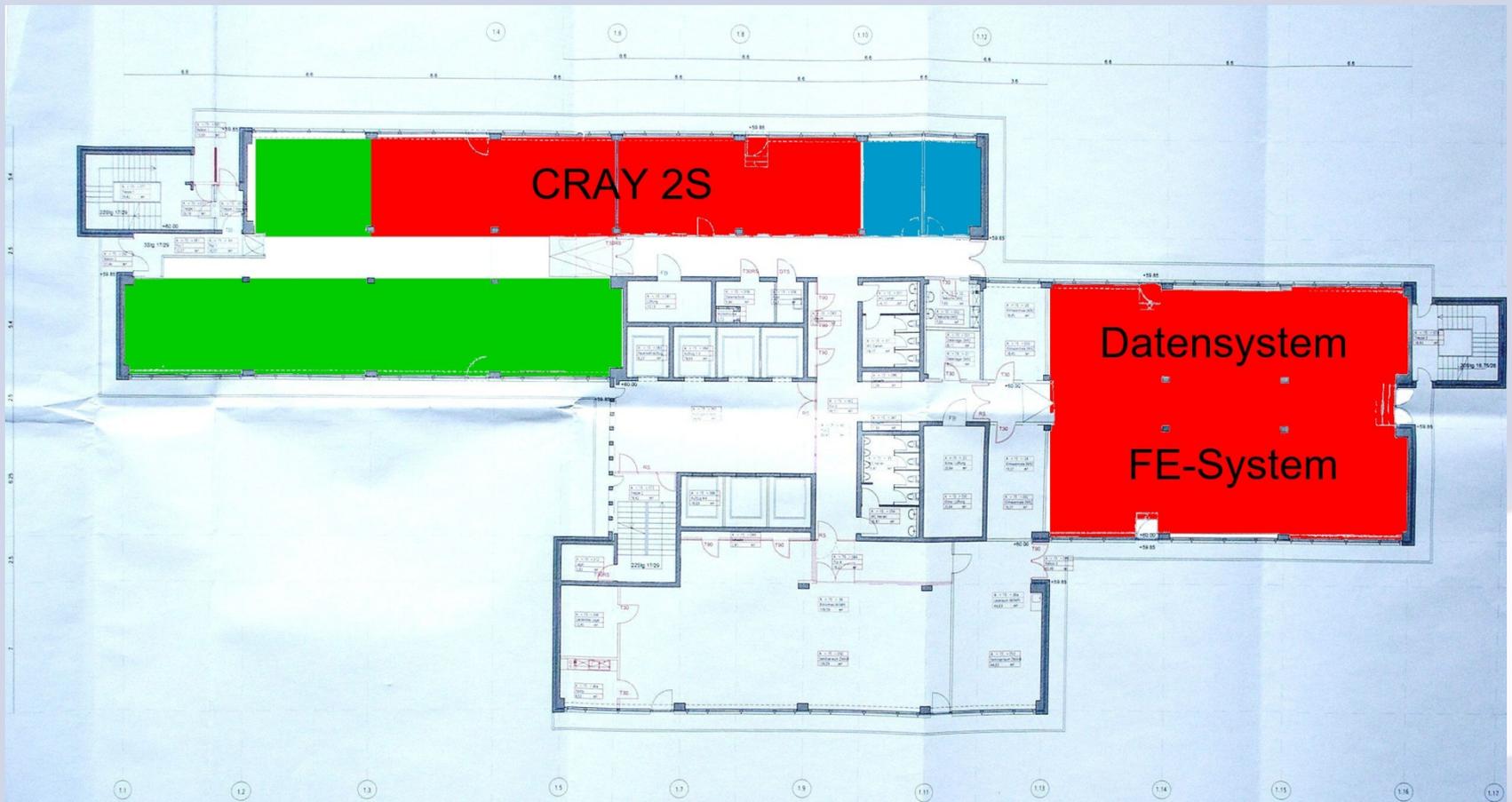
# GRZ-Räume 1984 – 1988

## 15. OG Geomatikum



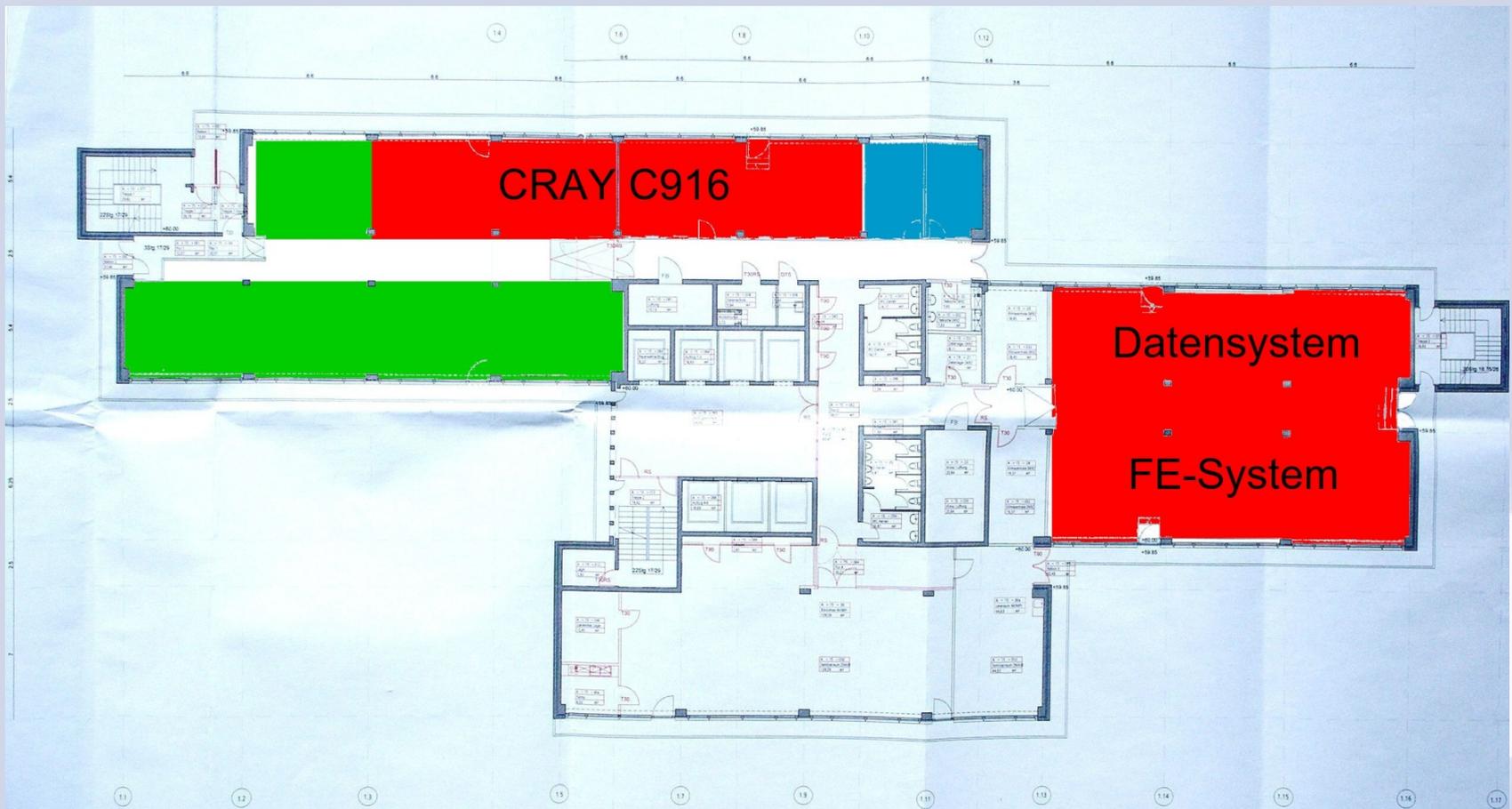
# DKRZ-Räume 1988 -1995

## 15. und 17. OG Geomatikum



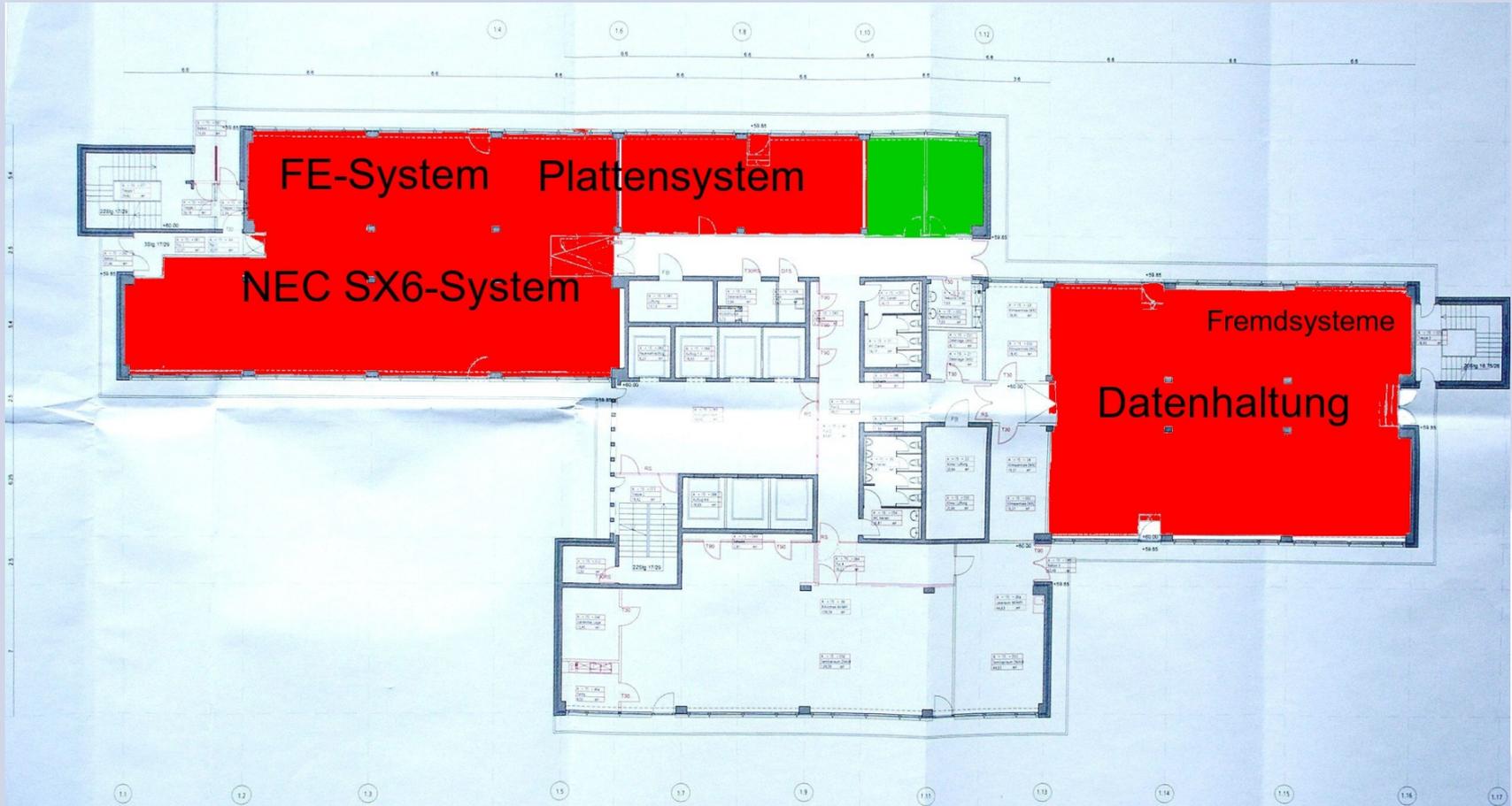
# DKRZ-Räume 1995 – 2001

## 15. und 17. OG Geomatikum



# DKRZ-Räume 2001 – 2004

## 12., 15., 17. OG Geomatikum, Pavillon



# Maschinenraumplanung für HLRE 2 gesamtes 15. OG Geomatikum



# Organisatorischer Rahmen

- Planungsphase des DKRZ
- DKRZ Gesellschafter
- Organe des DKRZ
- Personalplan
- Unterbringung
- Umstrukturierung

# Umstrukturierung des DKRZ

## Teil 1

- Im Mai 1995 wechselte das BMBF den Aufsichtsratsvorsitzenden aus
- 1996 teilte das BMBF seine Absicht mit, sich aus der Finanzierung des DKRZs zurückziehen zu wollen
- Gemeinnützigkeitsprüfung des DKRZ 1997
- 1998 Pralle Kommission spricht sich für Erhalt eines umstrukturierten DKRZs aus
- Anschließend wird ein neues Finanzierungsmodell zwischen BMBF und Gesellschaftern beraten

# Umstrukturierung des DKRZ

## Teil 2

- Zum 1.1.2000 wurde das bisherige DKRZ mit 55 Mitarbeitern in einen reinen Rechenzentrumsteil mit 20 Mitarbeitern und die Gruppe Modelle & Daten mit 15 Mitarbeitern zerlegt
- Die DKRZ Rechner standen zu 50 % den Gesellschafter-Einrichtungen und zu 50 % Drittmittelgeförderten Projekten zur Verfügung
- Das DKRZ bot den Gesellschaftern nicht mehr Dienstleistungen als Institutsrechenzentrum an

# Umstrukturierung des DKRZ

## Teil 3

- Der BMBF stellte nach der Berufung von Professor Brasseur zum MPIM-Direktor als Nachfolger von Professor Hasselmann 120 Mio DM für Investitionen bereit, beteiligte sich aber nicht mehr an Betriebskosten
- Die Betriebskosten müssen die Gesellschafter seit 2000 allein tragen

# Umstrukturierung des DKRZ

## Teil 4

- 70 Mio DM der zugesagten Investitionsmittel wurden zur Beschaffung des HLRE eingesetzt
- Die restlichen Investitionsmittel wurden für die HLRE2-Beschaffung übertragen
- Am 1.1.2010 kehrte die Gruppe Modelle und Daten wieder ins DKRZ zurück

# Themen

- Organisatorischer Rahmen
- **Das GRZ als Vorläufer des DKRZ**
- Das DKRZ zu Zeiten der CRAY-Rechner
- Das DKRZ zu Zeiten des HLRE

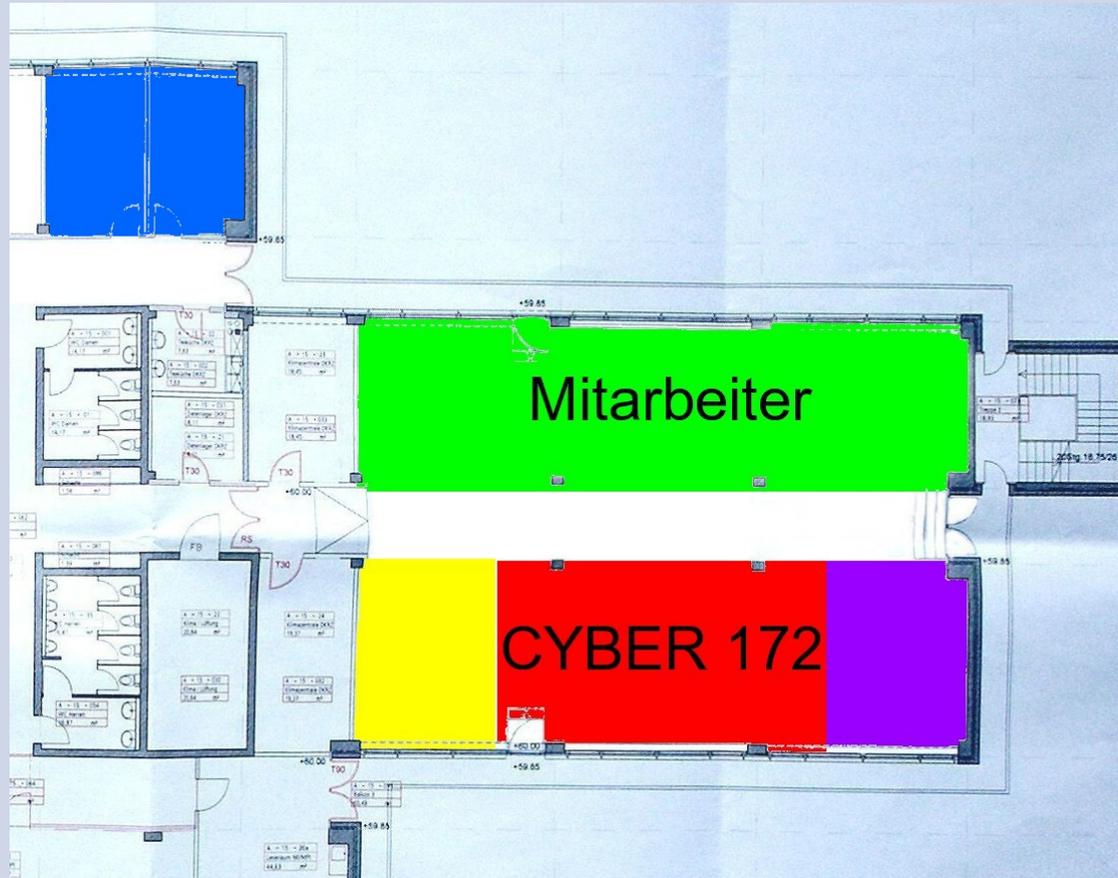
# Das GRZ als Vorläufer des DKRZ

- Inbetriebnahme 1978
- Gemeinsames Rechenzentrum (GRZ) des MPIM und der Universität Hamburg (Meteorologie, Geophysik, Meereskunde)
- Aufgaben des GRZ
- Räumlichkeiten des GRZ
- CYBER 172 Phase 1978 - 1984
- CYBER 205 Phase 1984 - 1988

# Das GRZ im 15. OG des Geomatikum



# Arbeitsräume des GRZ im 15. OG des Geomatikum



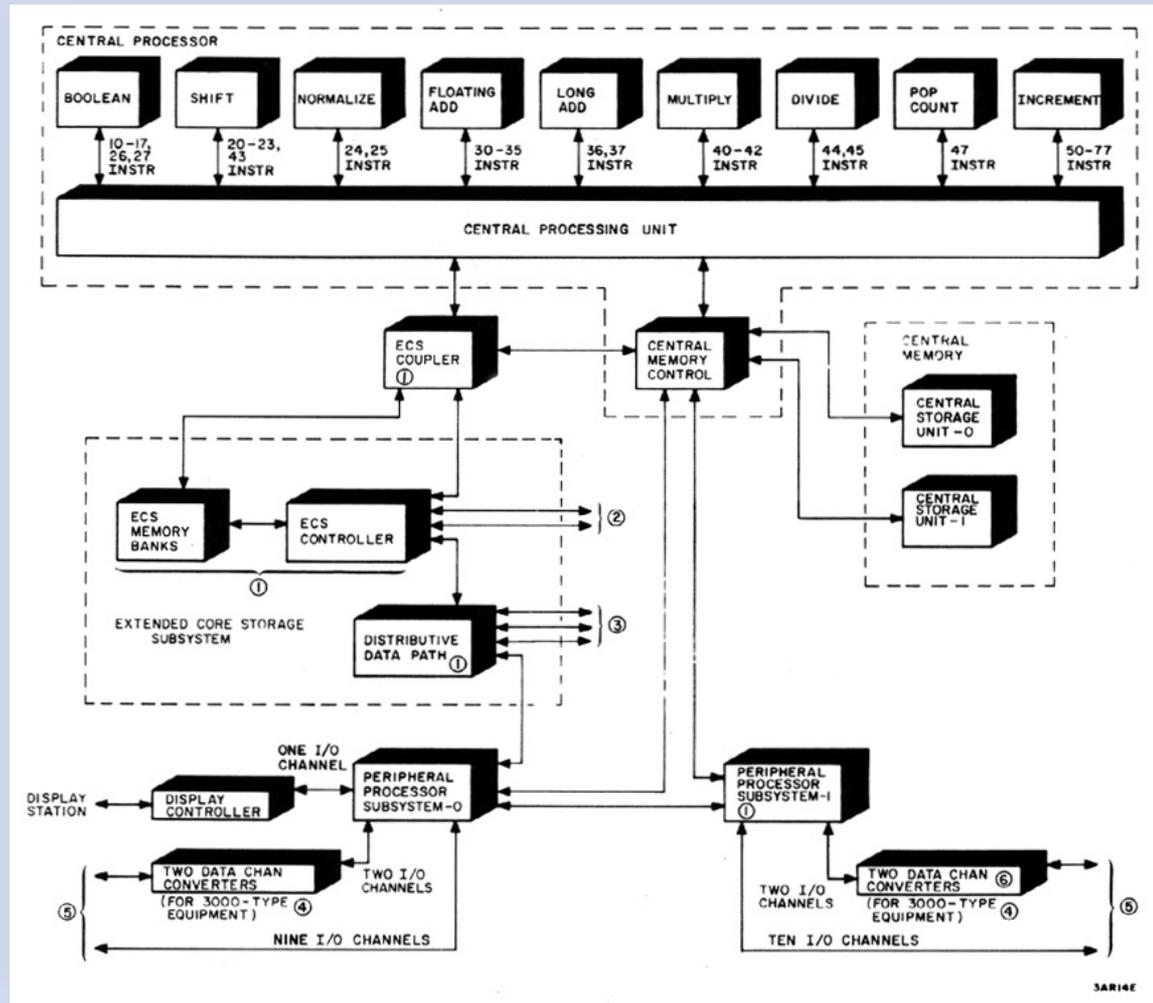
# Prozessrechner CYBER 18-17



# Universal-Rechnersystem CYBER 172



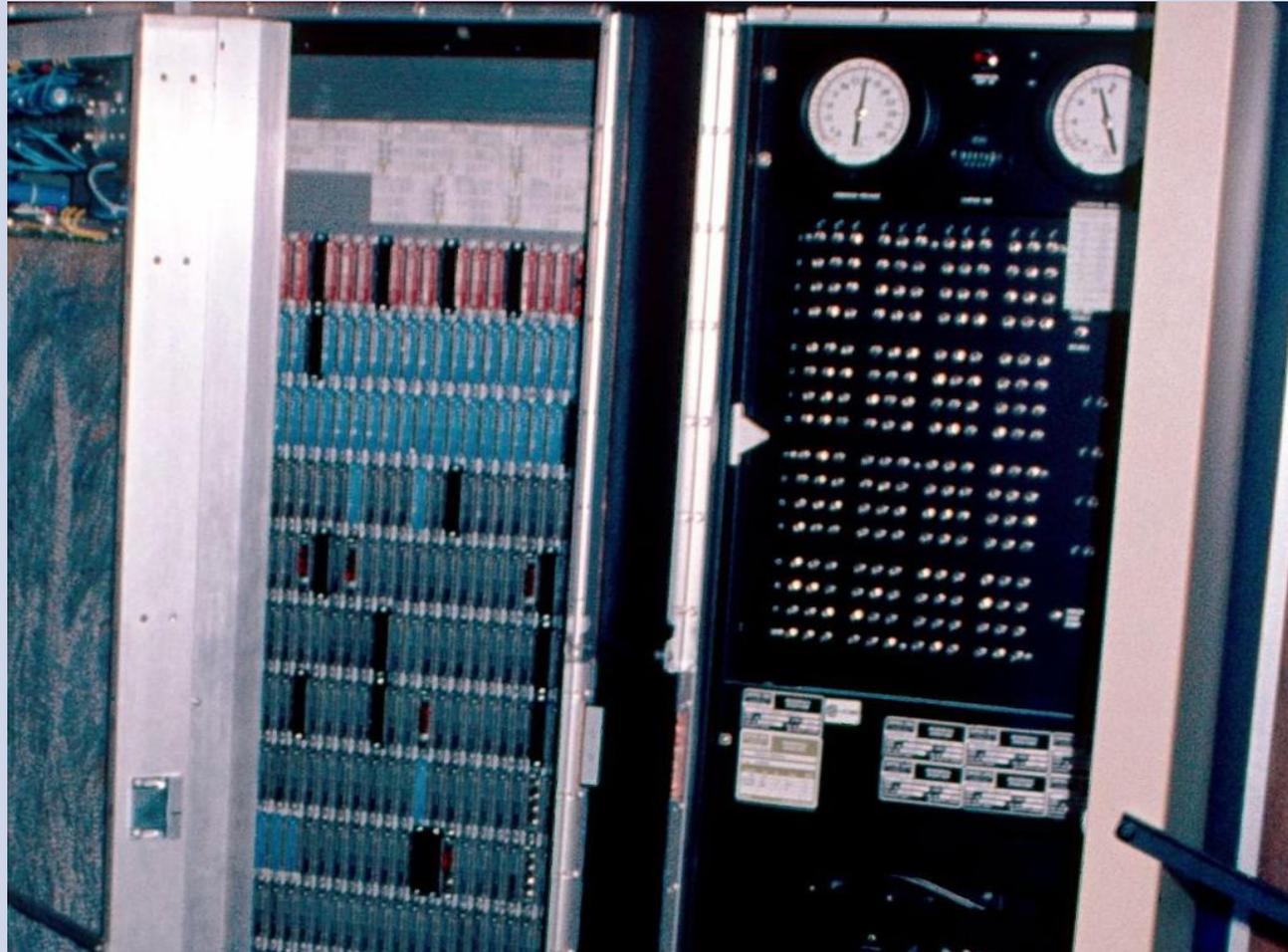
# CYBER 170 Systemarchitektur



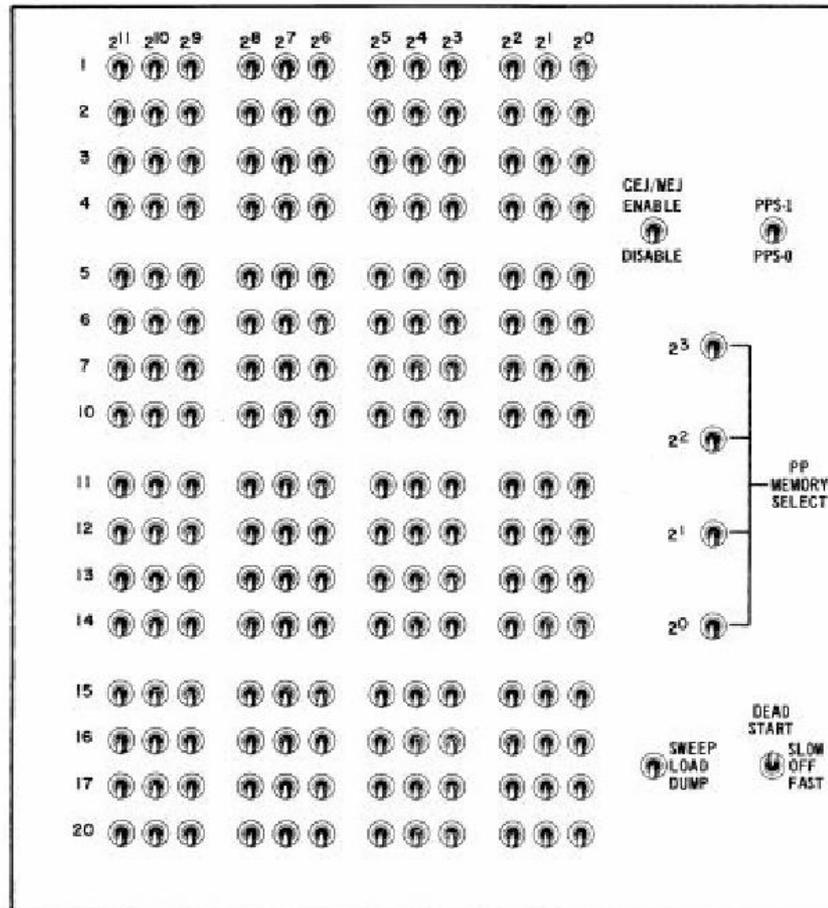
# CYBER 172 Innenleben



# CYBER 172 Deadstart-Panel



# CYBER 172 Deadstart-Panel



NOTE:

# CYBER 172 Papierperipherie



# GRZ Terminalraum



# Das GRZ als Vorläufer des DKRZ

- Inbetriebnahme 1978
- Gemeinsames Rechenzentrum (GRZ) des MPIM und der Universität Hamburg (Meteorologie, Geophysik, Meereskunde)
- Aufgaben und Räumlichkeiten des GRZ
- CYBER 172 Phase 1978 - 1984
- **CYBER 205 Phase 1984 – 1988**

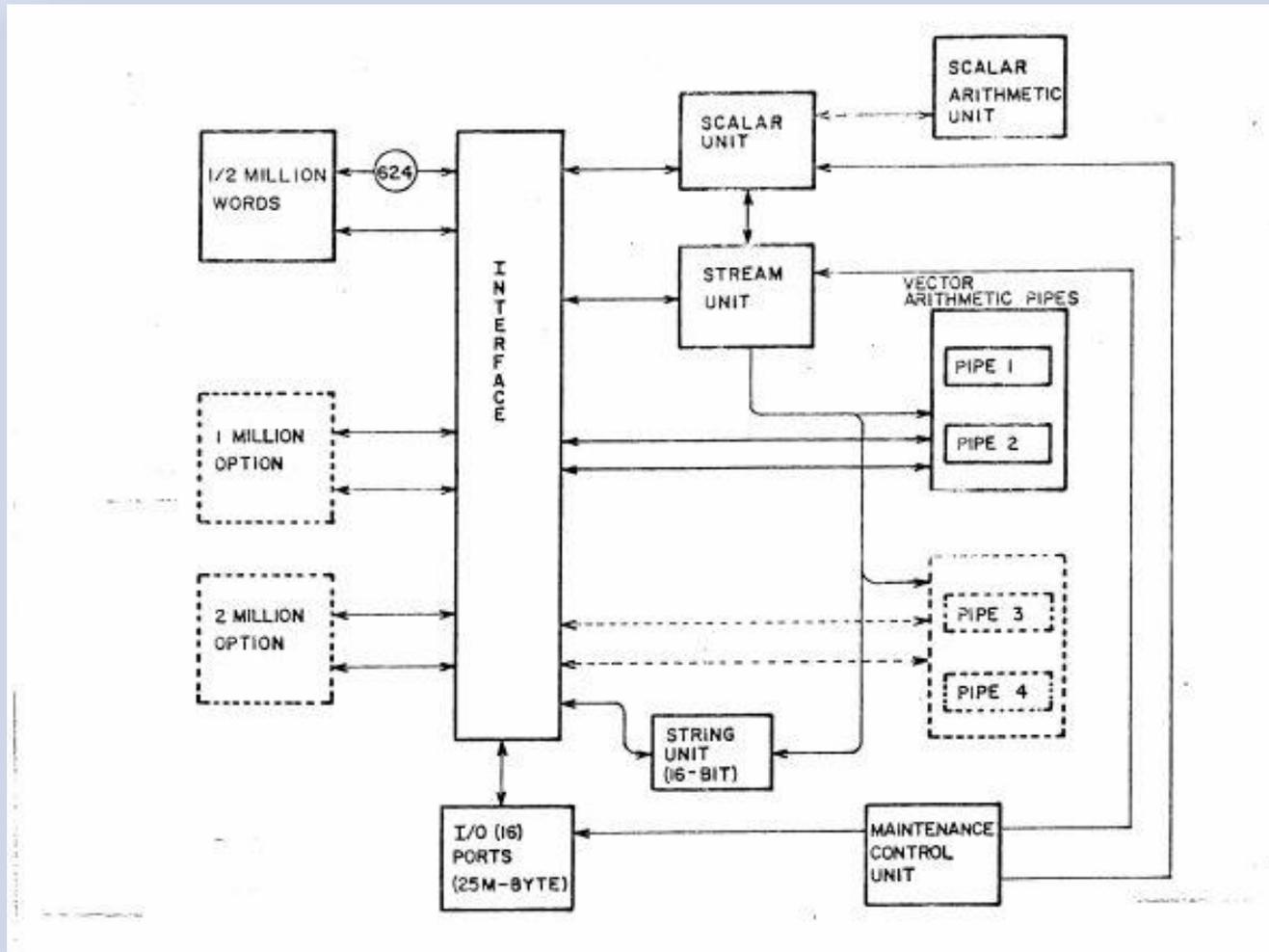
# Anlass und Alternativen

- Änderung der Arbeitsrichtung des MPIM führt zu stark erhöhtem Rechenbedarf
- Im Rahmen der zur Verfügung stehenden Mittel kommen ein Vektorrechner der Firma CRAY Research oder der Firma Control Data in Betracht
- Unterschiede der Systemarchitektur
- Entscheidung für CYBER 205

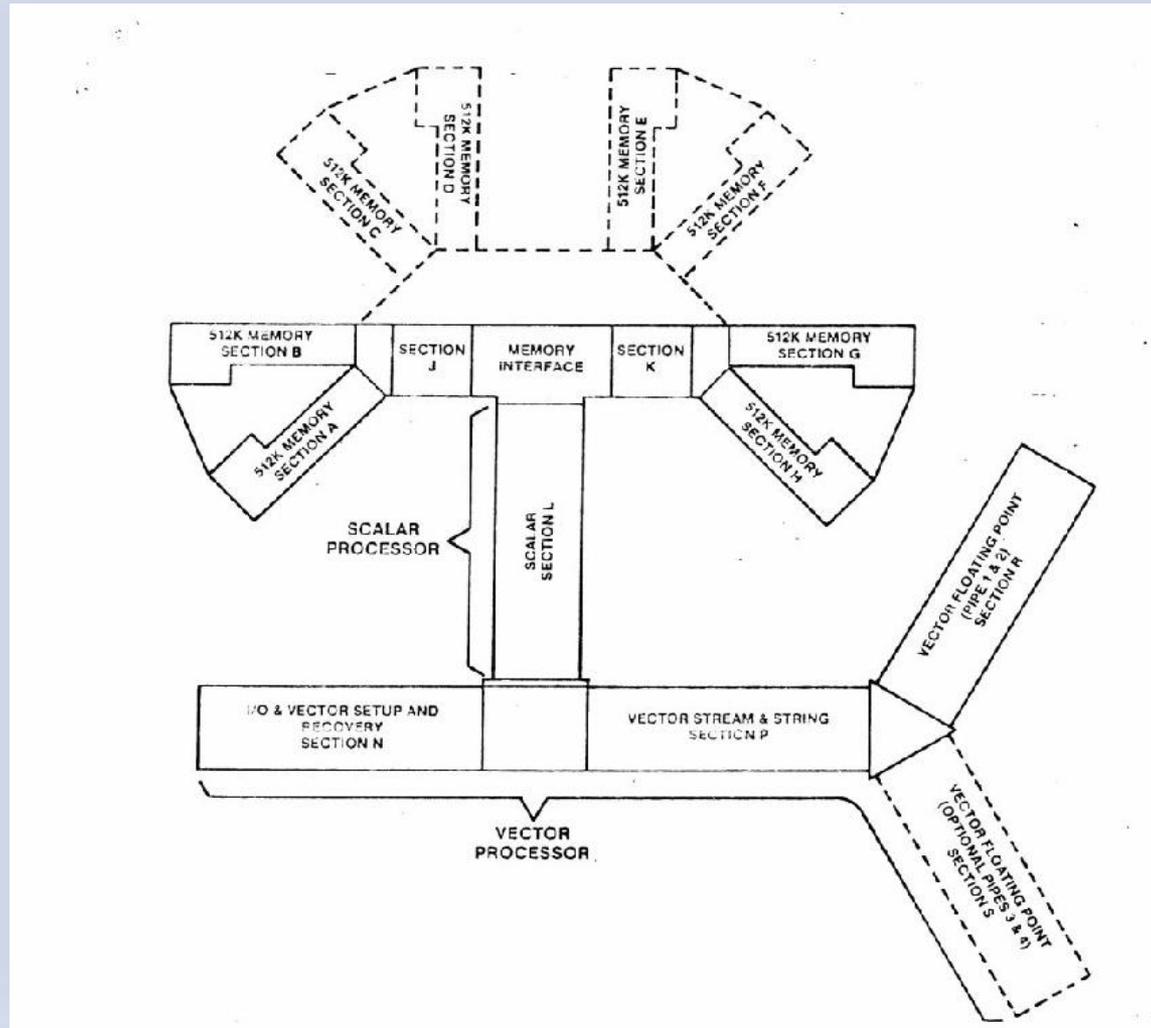
# CYBER 205 Specs

- Vektorrechner mit 2 oder 4 Pipes
- 32- oder 64-Bit Wortlänge + SECDED
- Rechenleistung 100 MFLOPS/Pipe bei 64 Bit-, 200 MFLOPS/Pipe bei 32-Bit-Arithmetik
- Maximale Vektorlänge 65.535 Elemente
- Virtueller Speicher mit 48 Bit Adressraum, d.h. O(TB)
- Realer Speicher mit 8, 16 oder 32 MB
- Schwierige Programmierung, 240 Q8-Befehle

# CYBER 205 Systemarchitektur



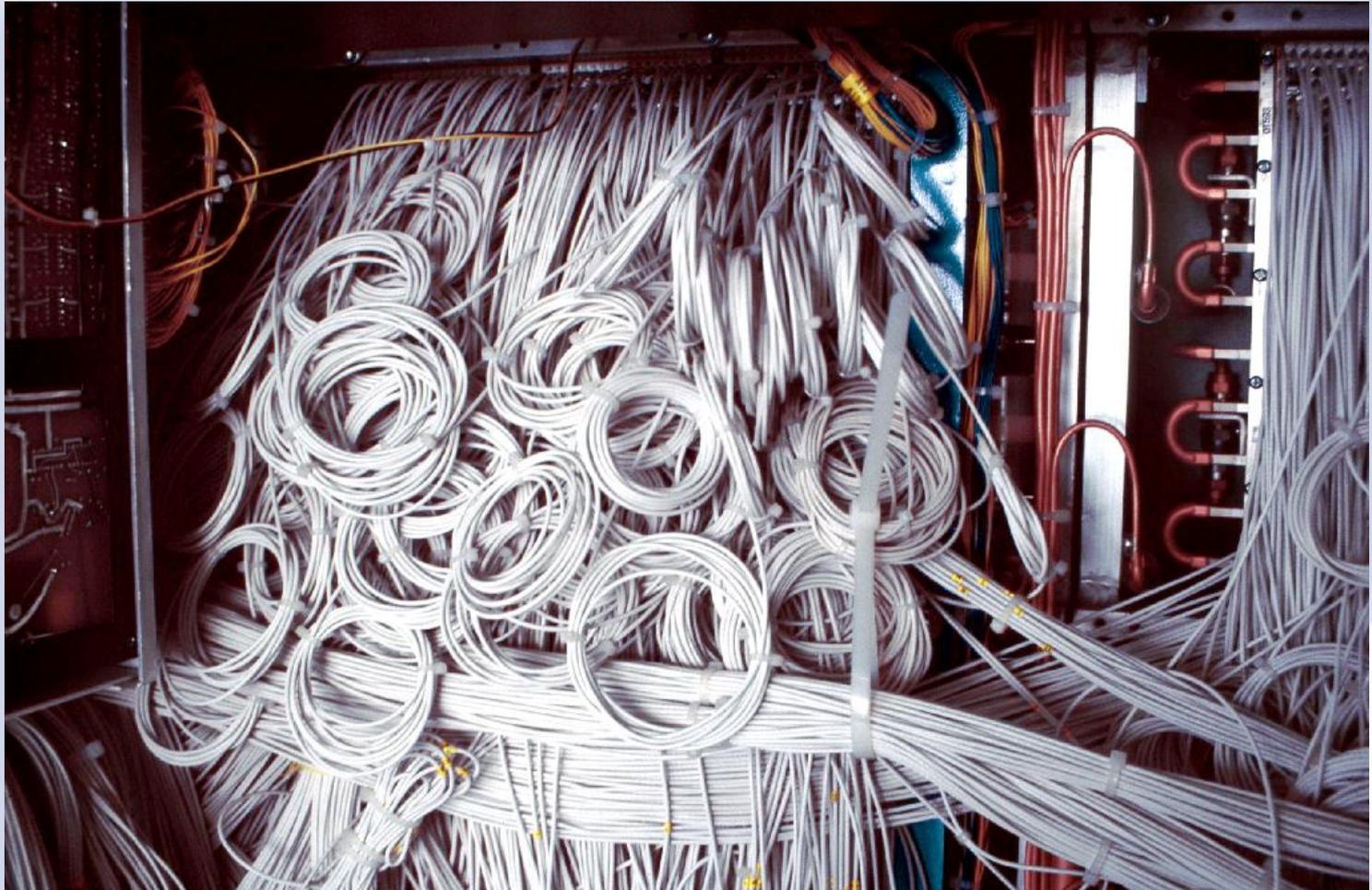
# CYBER 205 Grundriss des Vollausbaus



# CYBER 205 Draufsicht



# CYBER 205 Delaylines (Coax)



# CYBER 205 Programmierung

- Grundsätzlich FORTRAN für Anwendungsprogramme
- Jedoch: Kümmerliche Autovektorisierung, deshalb insgesamt 240 sog. Q8-Befehle, wie z.B.

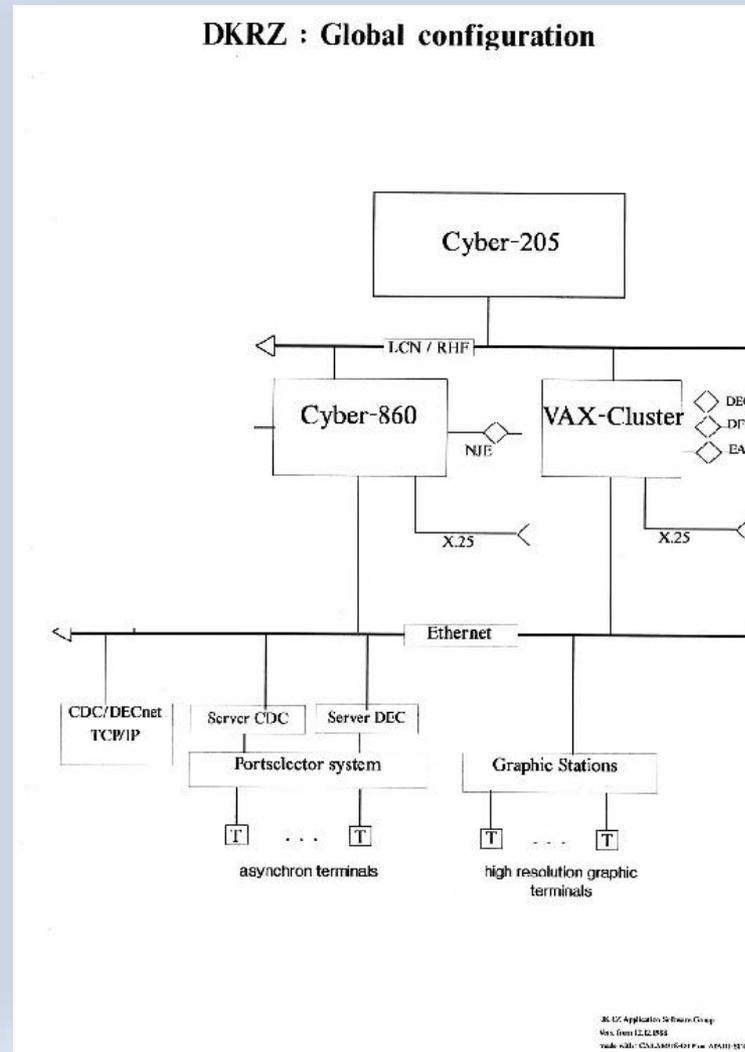
Q8AM011 (A(2),B(2),C(1),A(1), N-1) an Stelle des DO-Loops

```
DO 10 I=2,N
```

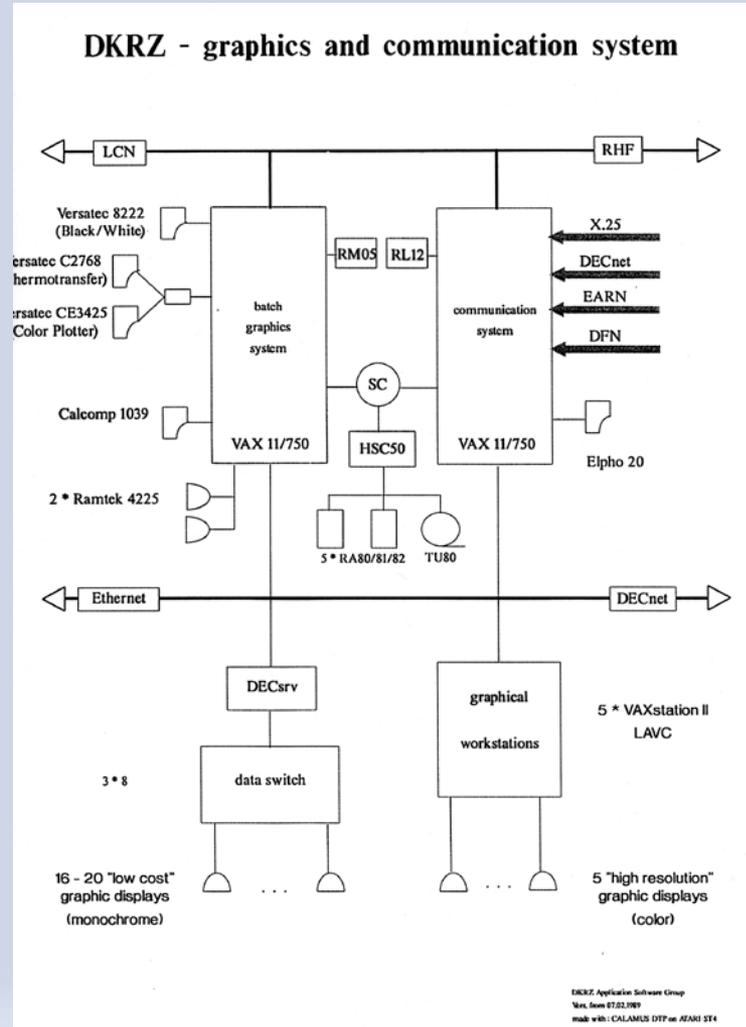
```
10 A(I)=B(I)+(C(I-1) * A(I-1))
```

zur Steigerung der effektiven Rechenleistung

# GRZ-Rechnersystem mit CYBER 205



# DEC VAX 750 Frontendrechner



# Themen

- Organisatorischer Rahmen
- Das GRZ als Vorläufer des DKRZ
- Das DKRZ zu Zeiten der CRAY-Rechner
- Das DKRZ zu Zeiten des HLRE

23

22

# Der erste DKRZ Compute Server

- Am 1.1.1988 wurde zunächst die CYBER 205 übernommen
- Auswahl des Nachfolgerechners:
  - Parallelrechner oder Vektorrechner
  - ETA
  - CRAY Research

# Parallelrechner

- 1988 eigentlich noch keine Produktionsrechner, mehr experimentell
- Suprenum wurde in Deutschland vom BMFT gefördert, wollte aber erst 1989 einen Prototyp ausliefern
- Suprenum-Förderung wurde 1990 aus Kostengründen eingestellt
- Andere Hersteller waren auch nicht attraktiv

# ETA-Rechner

- CDC's Nachfolgemodell zur CYBER 205 war die ETA 10
- 7 ns CPU Zykluszeit statt 20 ns
- CPU in CMOS- statt ECL-Technologie
- Deshalb Kühlung in Flüssigstickstoff
- 10 GFLOPS Nominalspitzenrechenleistung pro System bei 32 Bit-Arithmetik
- Erste Auslieferung Dezember 1987

# ETA-Rechner

- Vom Konzept war der ETA-Rechner interessanter als die CRAY Rechnermodelle, selbst als das neueste Modell CRAY YMP
- Aber CRAY hatte gut funktionierende Software und eine erprobte Entwicklungs-, Produktions- und Wartungsmannschaft
- Deshalb: Lieber den Spatz in der Hand als die Taube auf dem Dach, also einen CRAY-Rechner
- ETA stellte 1990 seinen Betrieb ein

# CRAY-Rechner-Auswahl

- Vom Leistungsvergleich konkurrierten ein großes YMP-Modell mit einem großen CRAY 2-Modell
- YMP war zu teuer
- CRAY 2 DRAM Modell mit 2 GB Hauptspeicher oder CRAY 2 SRAM Modell mit 1 GB Hauptspeicher finanziell gleichwertig
- CRAY 2S im Benchmark 30% schneller als CRAY 2D
- DKRZ erhielt CRAY 2S mit der Seriennummer 2012, die erste Maschine mit SRAM-Speicher - gebraucht



# CRAY 2 mit Primärkühlung



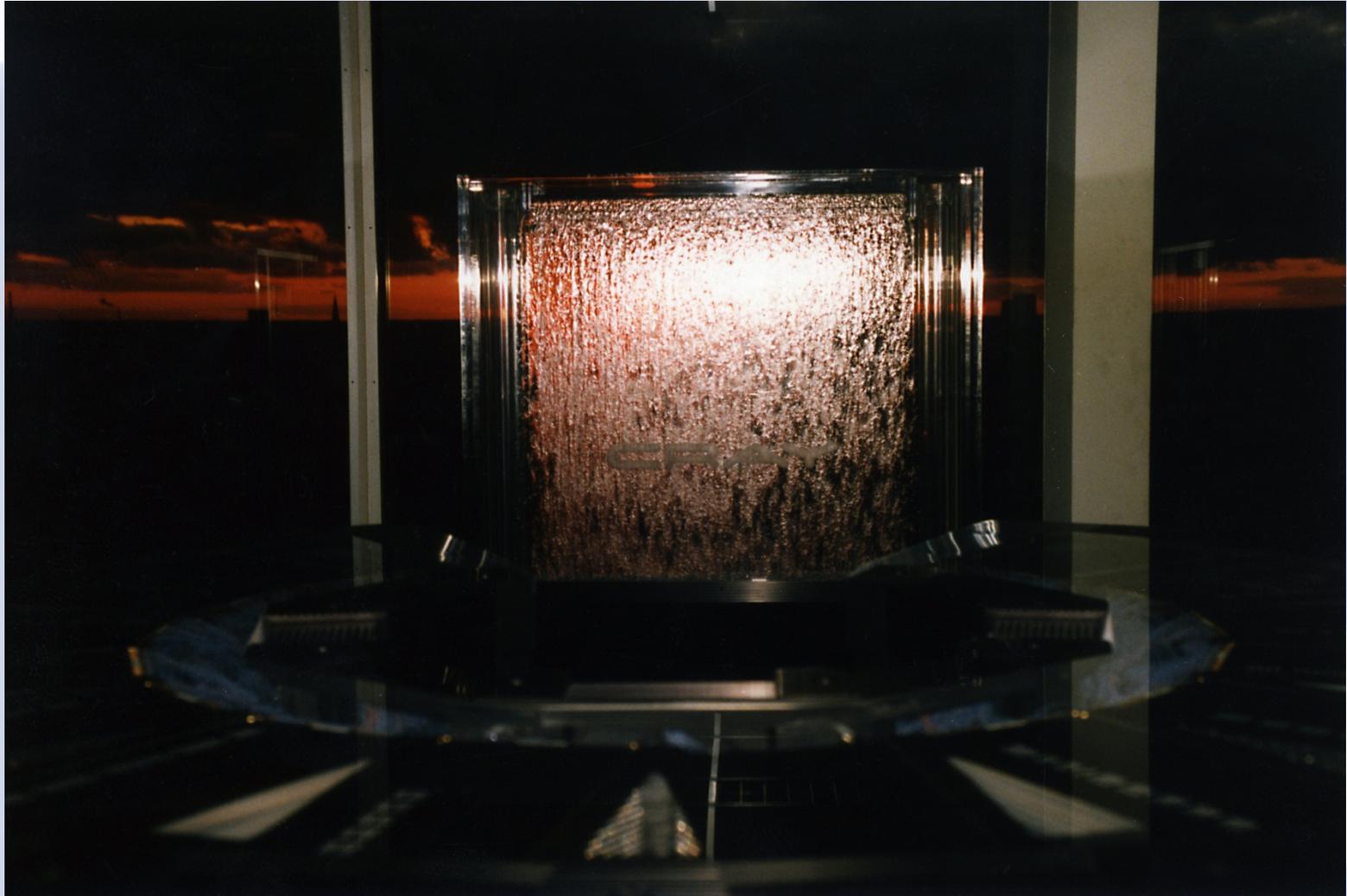
# CRAY 2 Interne Verkabelung (Twisted Pairs)



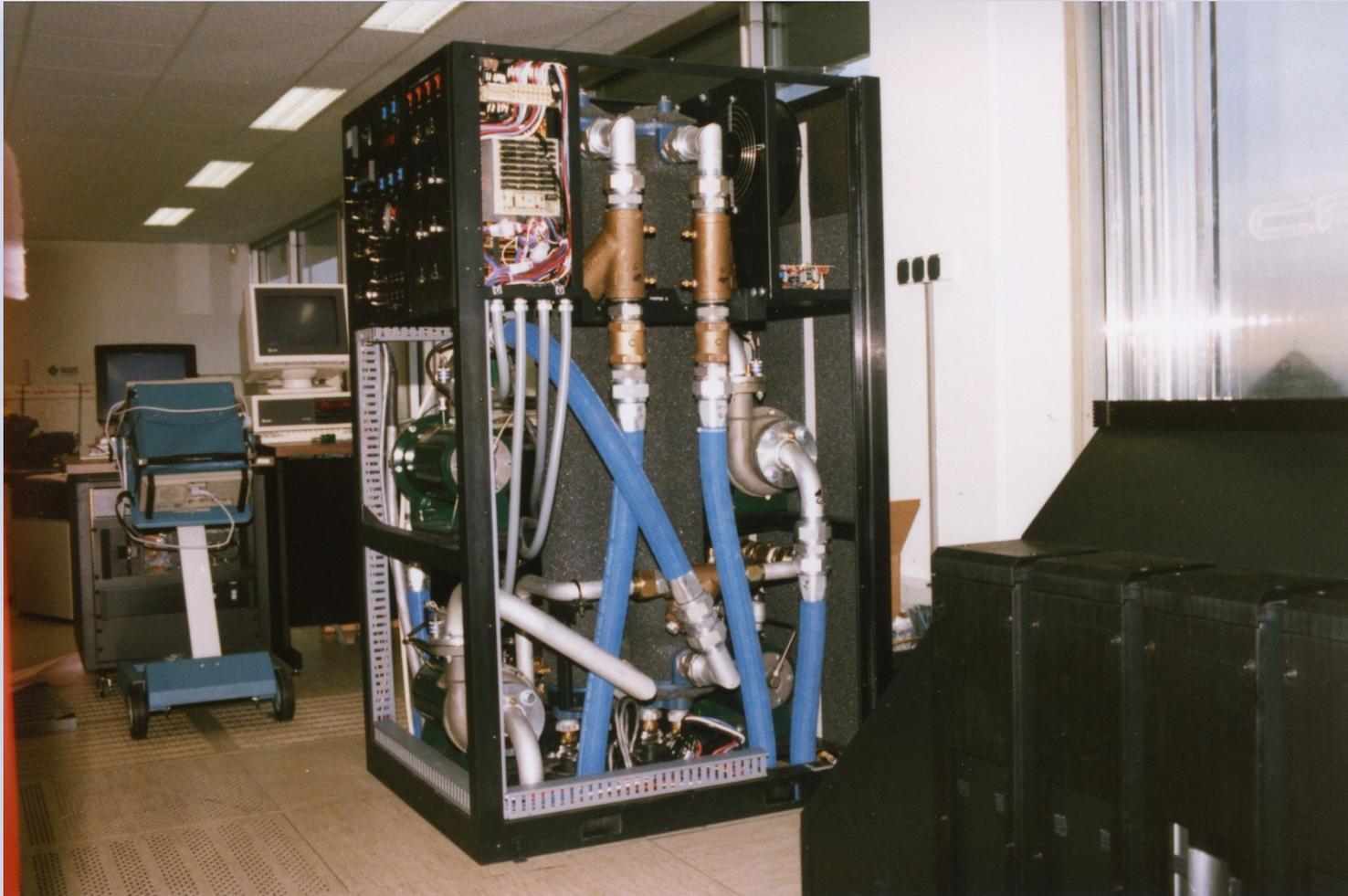
# CRAY 2 Specs

- 135 cm Durchmesser, 115 cm hoch
- 14 Segmente in 300° Zirkel
- Fluorinert-Kühlung
- 360 Module, 3 D mit 8 Schichten/Module
- Masse 2.500 kg, davon 1/3 Fluorinert
- 195 kW Anschlussleistung
- 400 Hz Spannungsversorgung

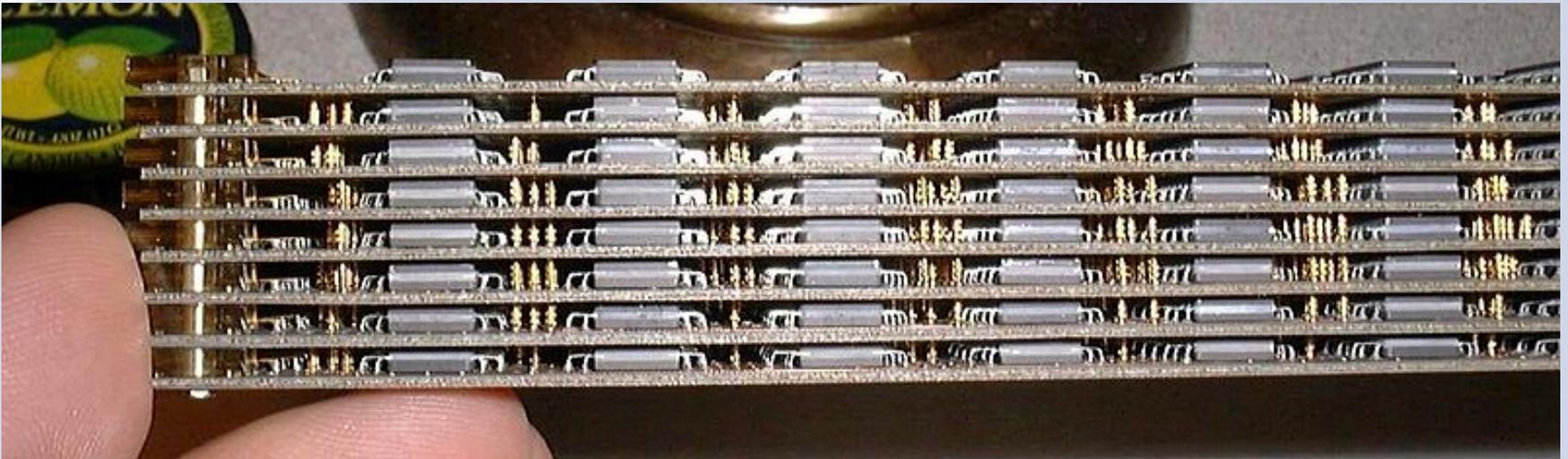
# CRAY 2 Wasserfall



# CRAY 2 Fluorinert-Rückkühleinheit



# CRAY 2 Modul

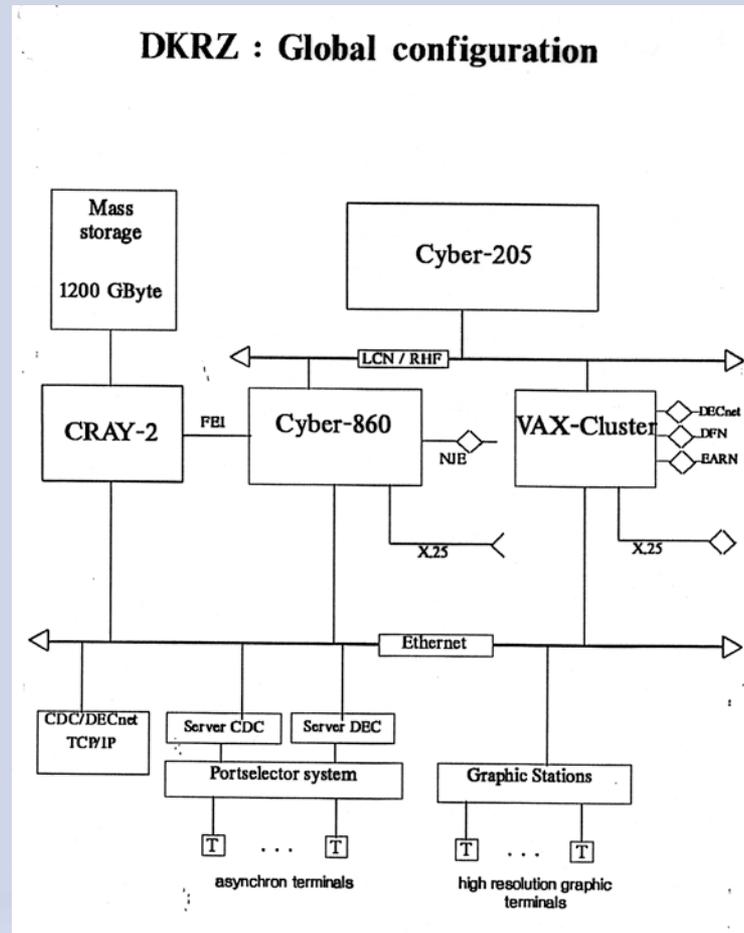


Ein CRAY 2 Modul enthält 8 übereinanderliegende Platinen mit jeweils 8 mal 12 Chips. Das Modul produziert etwa 500 Watt Wärme und wird zur Kühlung von Fluorinert umspült.

# CRAY 2 Specs

- 4 CPUs mit 4,1 ns CPU-Zykluszeit
- CPU-interne Taktrate 10 GHz
- 1,956 GFLOPS Nominalspitzenrechenleistung
- 1 GB SRAM Hauptspeicher unterteilt in 128 Speicherbänke
- Plattenspeicher
- Verbindung zum Frontendrechner CYBER 860
- Online-Anschluss eines Magnetband-Speichersilos

# DKRZ-Rechnerkonfiguration Dezember 1988



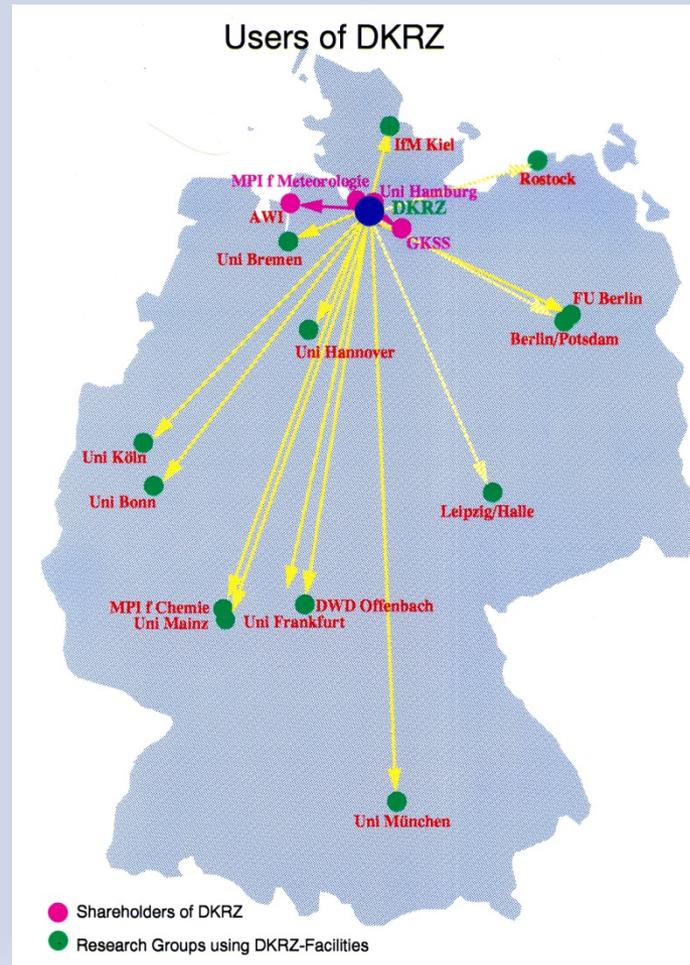
# CRAY 2 Nutzung

- Der CRAY 2 Rechner wurde im November 1988 in Betrieb genommen
- Steigerung der Rechenleistung etwa um den Faktor 10 gegenüber CYBER 205
- Gekoppelte Atmosphären/Ozean-Modelle über 100 Jahre Simulationszeitraum
- Während der CRAY2-Standzeit waren die Rechnungen 1990/1991 für die Vorbereitung der Rio Klimakonferenz 1992 das wichtigste Projekt

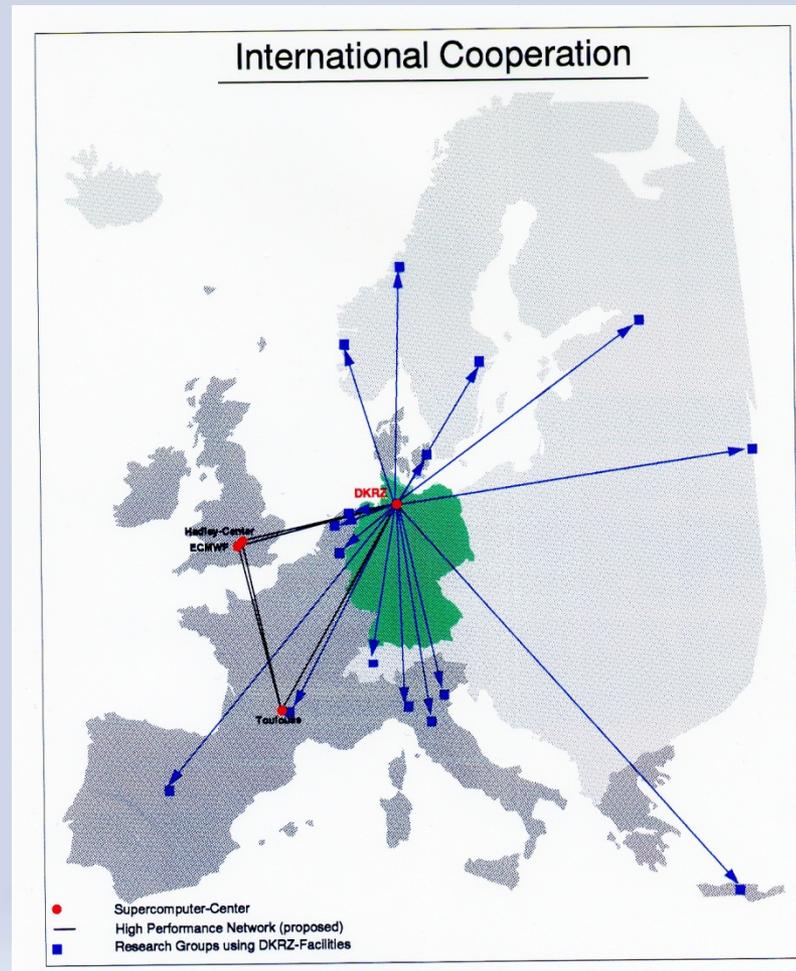
# Außenwirkung der CRAY 2

- Rechen- und Datenhaltungsmöglichkeiten machten das DKRZ ab 1990 auch für Nutzer außerhalb Hamburg interessant, da das DFN mit einer Bandbreite von 2 Mbps zur Verfügung stand
- 1992 wurde dann das ECCN angeboten

# DKRZ-Nutzung in Deutschland



# Verbund Europäischer Klimarechenzentren



# Forschungsminister Dr. Riesenhuber



Forschungsminister  
Dr. Heinz Riesenhuber (1991)

# Erweiterung des Rechnerparks 1990 - 1994

- Rechenleistung CRAY 2 allein unzureichend
- Deshalb Beschaffung von CRAY YMP 4E/364 als weiterer Compute-Server
- DKRZ erhielt die YMP-E mit der Seriennummer 1901, den ersten Rechner dieser Serie – praktisch neu !!!
- Außerdem File-Server Convex C220 und Convex C3860
- LAN wurde auf Ultranet mit 1 Gbps Übertragungsrate umgestellt

# CRAY YMP 4E



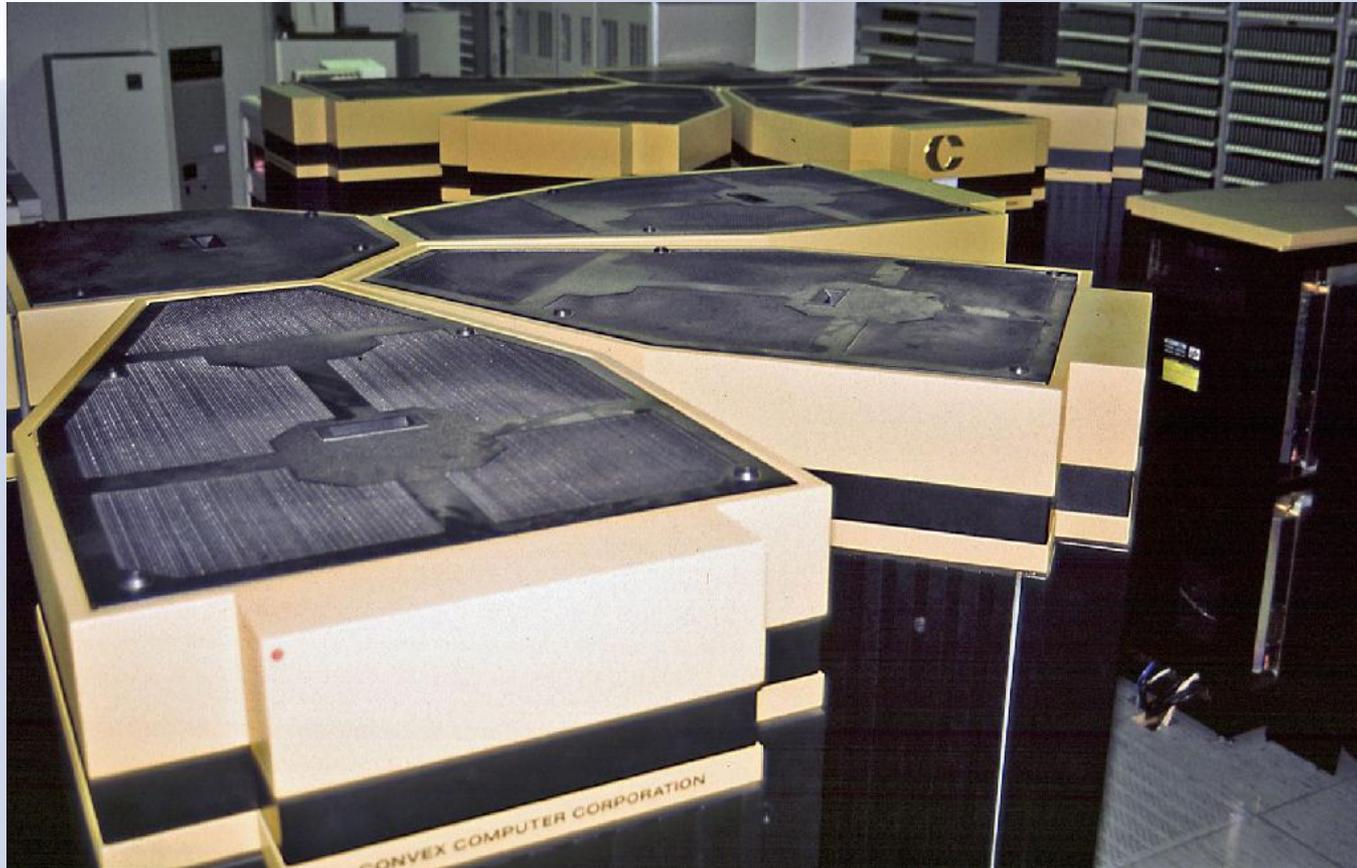
3 CPUs mit 1 GFLOPS, 64 MW und Ultranetanschluss

# Convex C220



Erster DKRZ File-Server

# Convex C38x0



Spätere DKRZ File-Server Convex C3860 und C3880

# Datenhaltung

- CRAY hatte mit der Lieferung der CRAY 2 auch den Anschluss eines Datensilos angeboten
- Seit 1987 lieferte StorageTek die tape library Powderhorn aus, die einen Operator-reduzierten Bandbetrieb ermöglichte
- 6000 Stellplätze pro Powderhorn-Silo (LSM = Library Storage Module genannt)
- Das erste Powderhorn-Silo wurde 1989 am DKRZ mit 1,2 TB Speicherkapazität installiert
- Ab 2002 bot das DKRZ in 3 LSM eine maximale Silo-Speicherkapazität von 3,6 PB an

# Datenhaltung

- Anfang der 90-er Jahre betrug die Speicherkapazität einzelner Medien um 200 MB
- Speicherbedarf wuchs am DKRZ zu schnell
- Deshalb übergangsweise METRUM Bibliothek mit 100 VHS-Kassetten als Archivspeicher
- Bis zu 14,5 GB Speicherkapazität/VHS-Kassette
- Ab 1995 SD3 Helical Scan Cartridges mit 10 GB/Cartridge für Powderhorn verfügbar
- Unzuverlässige VHS-Kassetten wurden abgeschafft
- Soviel zunächst zur Hardware, nun zur Software

# HSM und File-Residenz

- Software HSM = Hierarchical Storage Manager
- Was der Virtuelle Speicher für den Hauptspeicher des Rechners ist, ist der HSM für den Plattenspeicher
- Abhängig von der Nutzungshäufigkeit der Daten werden die Daten Platten-resident vorgehalten oder auf Tapes ausgelagert
- Der HSM alloziert die Residenz transparent für den Nutzer, d.h. der Filename allein erlaubt den Datenzugriff
- Der Nutzer sieht das File immer Online, die Zugriffszeit auf die Daten ist jedoch residenzabhängig

# Betrieb der Datensilos

- Tape Libraries sind von Vorteil, da Roboter die Magnetbänder automatisch auf den Bandeinheiten zur Verfügung stellen
- Besonders wertvoll werden die Datensilos, wenn man sie im Zusammenspiel mit einer HSM = Hierarchical Storage Manager Software einsetzt
- Zunächst das CRAY-Produkt DMF als HSM im Einsatz
- Nachteil: Compute-Server als File-Server missbraucht

# Engpass Plattenplatz

- Schnelle Rechner benötigen schnelle Platten
- Platten sind teuer, insbesondere schnelle
- Plattenplatz ist deshalb sehr begrenzt
- Abhilfe schafft ein HSM-System
- Zunächst das CRAY-eigene DMF auf den CRAY-Rechnern mangels geeigneter Alternativen
- Ab 1990/91 kommt Unitree auf eigenen File-Servern der Firma Convex zum Einsatz

# Unitree

- Ab 1991 kommt als HSM Unitree auf eigenen File-Servern der Firma Convex zum Einsatz
- Das HSM Unitree wurde am LLNL über viele Jahre entwickelt und 1989 auf der SC'89 vorgestellt
- 1990 hatte Unitree die bessere Funktionalität und lief vor allem auf unterschiedlichen Rechnerplattformen, auch solchen mit billigeren Platten
- Einfluss des DKRZ auf die Unitree-Entwicklung (Convex, Unitree Inc., Open Vision, ...)
- Unitree war bis zur Inbetriebnahme des HLRE2 am DKRZ im Einsatz

# Auswirkung des Unitree-Betriebes

- Die Begrenzung des Plattenspeicherplatzes wurde kaum noch empfunden
- Der Bestand der ausgelagerten Daten wuchs exponentiell wie die Rechenleistung der Computer-Server
- Wegen des riesigen Datenbestandes und des fast störungsfreien Unitree-Betriebes wurde weitgehend auf Sicherungskopien verzichtet
- Dies führte im Januar 2004 bei Systemarbeiten fast zum Totalverlust des Datenbestandes

# Auswirkung des Unitree-Betriebes

- Der erfolgreiche Einsatz von Unitree förderte einen Paradigmenwechsel im File- bzw. Datenzugriff
- Am Speicherplatz orientiert:
  - Benutzerverwaltete Magnetbänder
  - 1984 – 1988 durch Operator bereit gestellt
  - 1989 – 1991 durch Roboter bereit gestellt
- Filenamen orientiert
  - Seit 1992 transparent durch Unitree
  - 1995 um 1.000.000 Files unter HSM-Verwaltung
- Kontext orientiert
  - Seit 1998 durch die Klimadatenbank CERA

# Kontext-orientierte Datenhaltung

- 1993 Antrag beim BMFT für Mittel zum Aufbau „semantischer Klimadatenhaltung“
- M&D entwickelt CERA, eine Datenbank-basierte Software für „Climate and Environmental data Retrieval and Archiving“
- Ab 1998 am DKRZ verfügbar
- Beim HLRE-Betrieb im Routine-Einsatz
- M&D betreibt WDC on Climate am DKRZ
- 2005 die größte wissenschaftliche Datenbank mit 150 TB, Ende 2012 etwa 1 PB total

# Das Nachfolgesystem der CRAY 2

## Teil 1

- 1992/93 wurden Grenzen des CRAY 2 / YMP Rechnerkomplexes zur Bewältigung der Forschungsaufgaben erreicht
- Nachfolgekandidaten von CRAY Research, Cray Corporation, SSI und NEC standen zur Auswahl
- Die CRAY 3 von Cray Corporation konnte nicht überzeugen
- SSI musste 1993 den Betrieb einstellen
- NEC war in Europa unzureichend vertreten
- 37

8

# Das Nachfolgesystem der CRAY 2

## Teil 2

- CRAY Research konnte 1993 den Vektorrechner CRAY C90 oder das MPP-System T3D liefern
- Der Vektorrechner CRAY T90 wurde entwickelt und ab 1995 ausgeliefert, also zu spät
- Dasselbe galt für das MPP-System T3E
- MPP-Systeme erschienen 1993 für Klimacodes noch wenig geeignet
- Damit war die Vorentscheidung für den Vektorrechner CRAY C90 gefallen

# CRAY 2 Abtransport



# Das CRAY C90-System

- Am 2. Mai 1994 wurden die Compute-Server CRAY 2 und CRAY YMP 4-E abgeschaltet
- Am 25. Mai 1994 wurde die CRAY C916 (SN 4001-gebraucht) mit zunächst 12 CPUs, 2 GB Hauptspeicher und 4 GB SSD-Speicher zur Abnahme übergeben, die am 13. Juni 1994 erfolgreich beendet wurde
- Die feierliche Einweihung fand am 16. 09. 1994 statt
- Im Dezember 1994 wurde das System auf 16 CPUs mit einer Peak-Rechenleistung von 16 GFLOPS hochgerüstet

# CRAY C90 Anlieferung



# CRAY C90 Mainframe und SSD



# CRAY C90 Einweihung mit BM Dr. Krüger



# Erweiterung des CRAY Rechnerkomplexes

- Investitionen 1994/1995
  - J90 für Programmentwicklung
  - T3D Erfahrung mit dem Einsatz von MPP-Systemen bei Klimamodellcodes sammeln
- Danach erfolgte eine lange Durststrecke bis zur HLRE-Installation, da zunächst die Struktur- und Finanzierungsfrage für das DKRZ geklärt werden mussten

# Themen

- Organisatorischer Rahmen
- Das GRZ als Vorläufer des DKRZ
- Das DKRZ zu Zeiten der CRAY-Rechner
- **Das DKRZ zu Zeiten des HLRE**

# Vergabeverfahren für den HLRE

- Im Vorfeld der Vergabe hatten die Firmen Compaq, CRAY, Fujitsu-Siemens, Hitachi, HP, IBM, NEC, SGI und SUN Interesse zur Angebotsabgabe gezeigt
- Die Vergabe wurde EU-weit öffentlich ausgeschrieben
- Die Angebotsphase lief in der Zeit April bis Juni 2001
- Die Unterlagen wurden an 7 Rechnerhersteller und 8 andere Firmen versandt
- Nur die Firma NEC gab ein gültiges Angebot ab

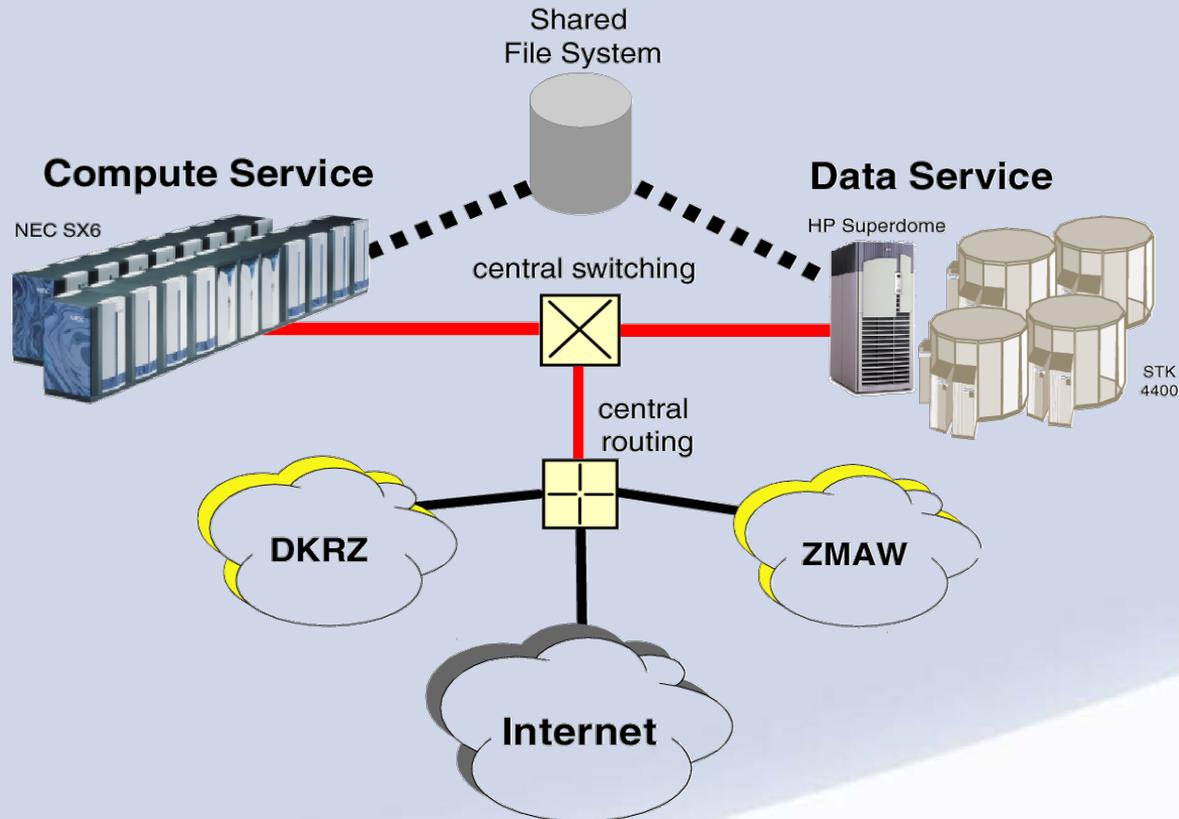
# HLRE Ausbauplan mit NEC-SX6

Datum	Feb 2002	Aug 2002	Apr 2003
Nodes	8	16	24
CPUs	64	128	196
Erwartete Sustained Performance [Gflops]	ca. 200	ca. 350	ca. 500
Erwartete Durchsatzsteigerung gegenüber CRAY C916	ca. 40	ca. 75	ca. 100
Hauptspeicher [Tbytes]	0.5	1.0	1.5
Festplatten-Kapazität [Tbytes]	ca. 30	ca. 50	ca. 60
Massenspeicher-Kapazität im Bandarchiv [Tbytes]	>720	>1400	>3400

# Gemessene FLOPS-Raten für HLRE

- 4 node performance > approx.100 GLFOPS  
( about 40 % Efficiency) for
  - ECHAM (70-75)
  - MOM
  - Radar Reflection on Sea Ice
- 24 node performance for Turbulence Code  
about 470 GFLOPS (30+ % Efficiency)

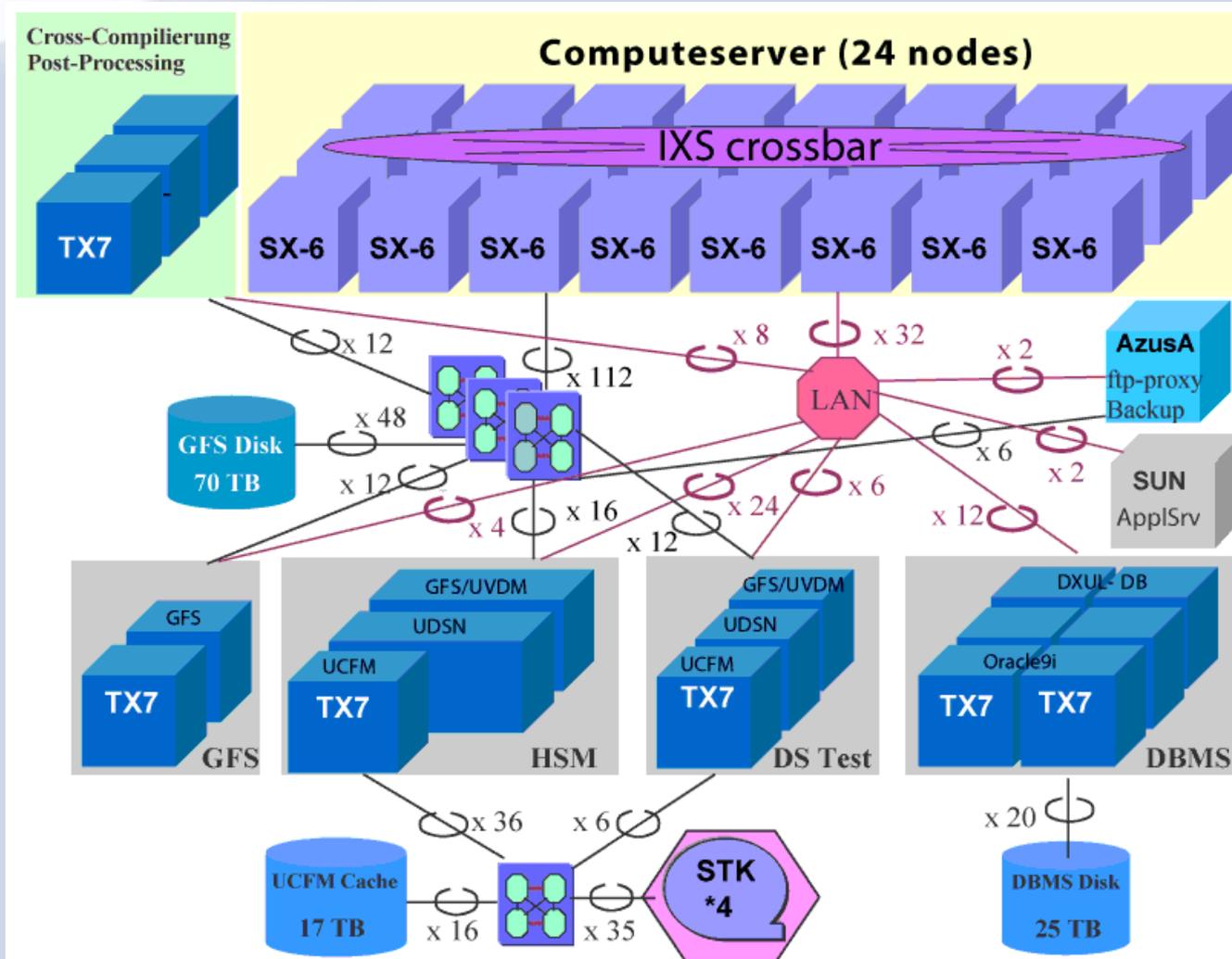
# HLRE System Architektur



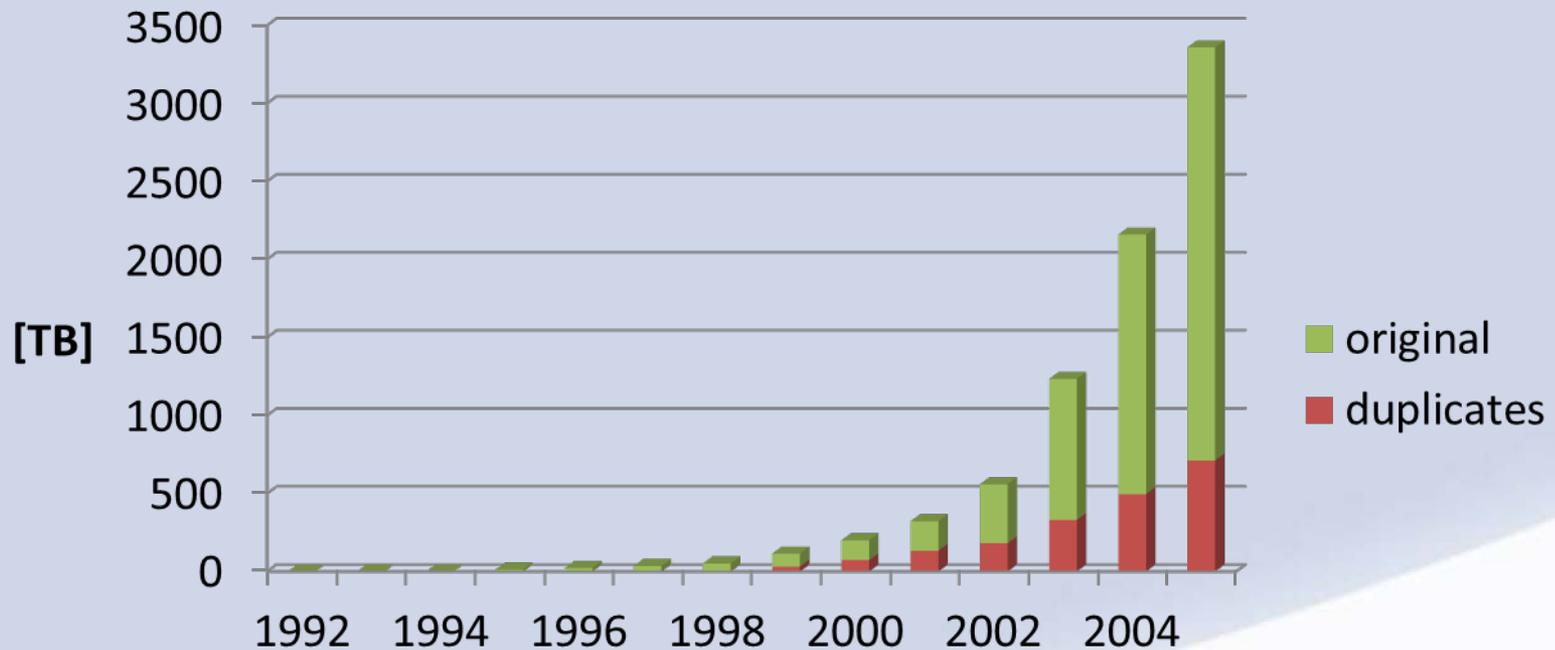
# 24 Knoten NEC SX6-Rechner



# HLRE-Konfiguration im Endausbau



# DS Archiv-Kapazität



# Dank

- Professor Hasselmann
- Dr. Krause, BMFT
- Herrn Fichtel, DKRZ
- Professor Graßl, MPIM
- Herrn Laubert, UniHH
- Mr. Al Kellie, NCAR