IBM zSeries Fault Tolerant Design
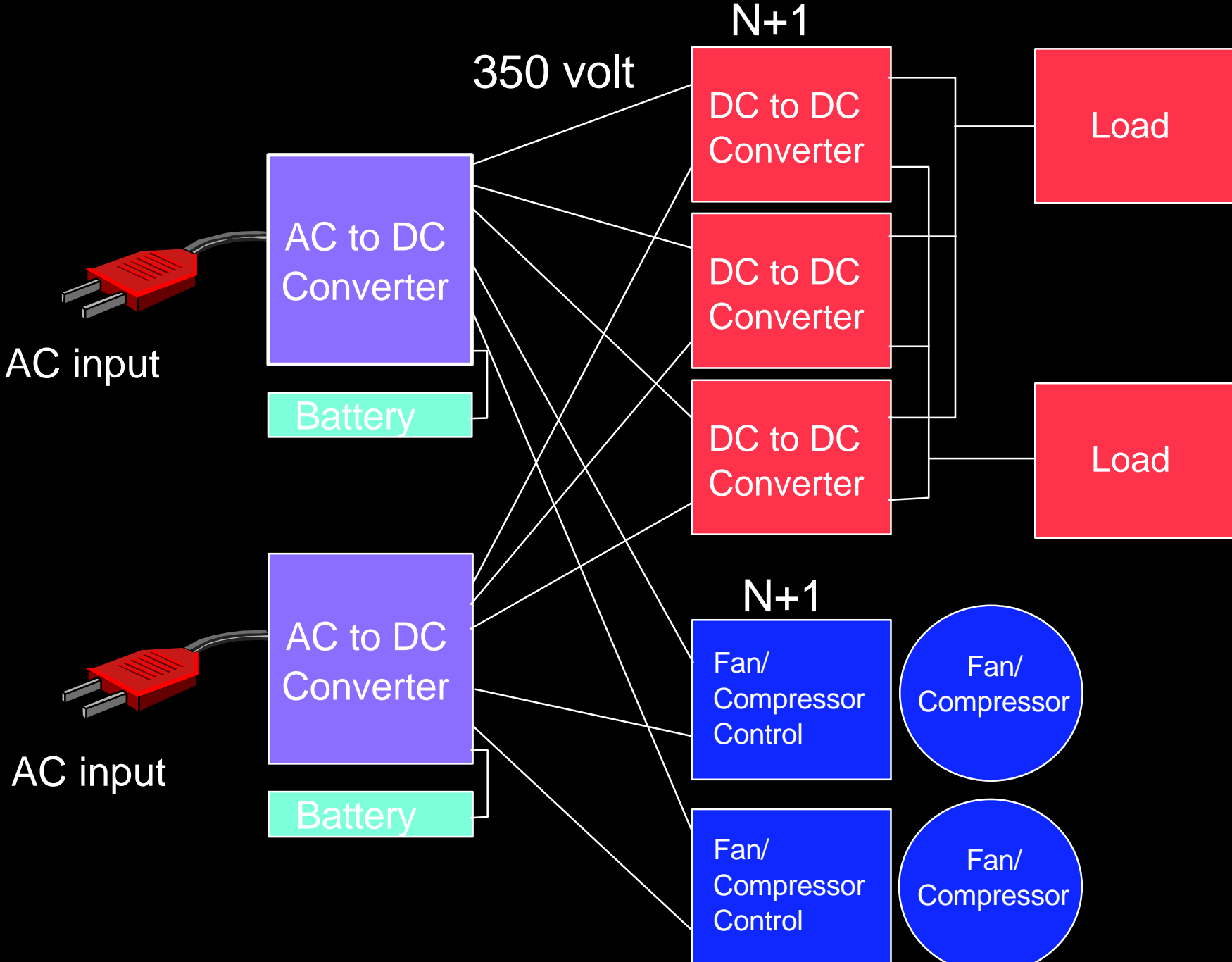
Lisa Spainhower
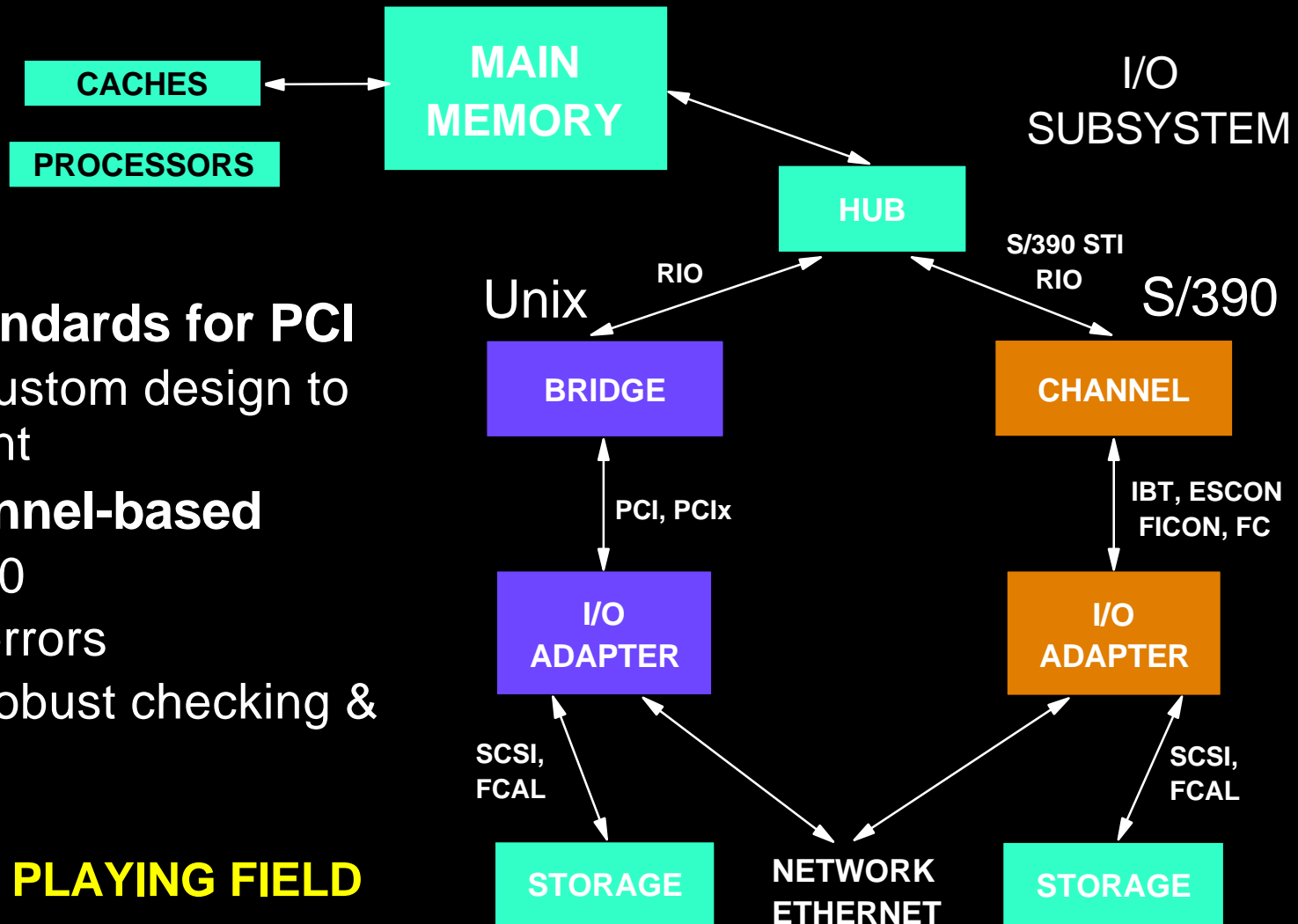
m   ^   Technology

September 20, 2001

Power/Cooling Fault Tolerance

N+1

350 volt

AC input

AC to DC Converter

Battery

AC input

AC to DC Converter

Battery

DC to DC Converter

DC to DC Converter

DC to DC Converter

Load

Load

N+1

Fan/ Compressor Control

Fan/ Compressor Control

Fan/ Compressor

Fan/ Compressor

I/O ED and Recovery
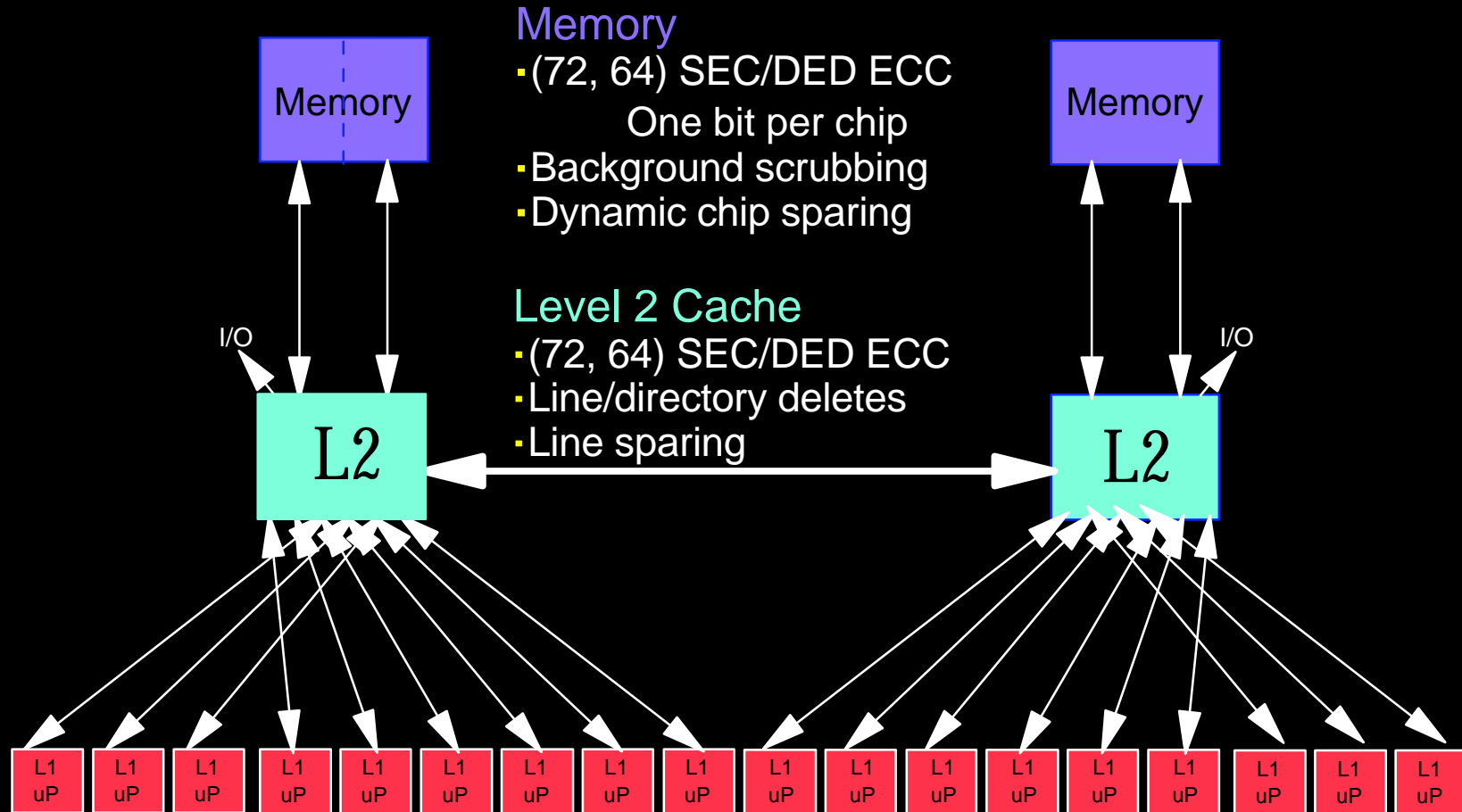
NO ED standards for PCI
- ▶ RS/AIX custom design to circumvent

IBT is channel-based
- ▶ Like S/390
- ▶ Defined errors
- ▶ Defined robust checking & isolation

LEVEL THE PLAYING FIELD

CACHES

PROCESSORS

MAIN MEMORY

I/O SUBSYSTEM

HUB

Unix

RIO

S/390 STI RIO

S/390

BRIDGE

CHANNEL

PCI, PCIx

IBT, ESCON FICON, FC

I/O ADAPTER

I/O ADAPTER

SCSI, FCAL

SCSI, FCAL

STORAGE

NETWORK ETHERNET

STORAGE

Memory Hierarchy Fault Tolerance

Memory
- (72, 64) SEC/DED ECC
    One bit per chip
- Background scrubbing
- Dynamic chip sparing

Level 2 Cache
- (72, 64) SEC/DED ECC
- Line/directory deletes
- Line sparing

Level 1 Cache
- Parity Protected
- Store-through to L2
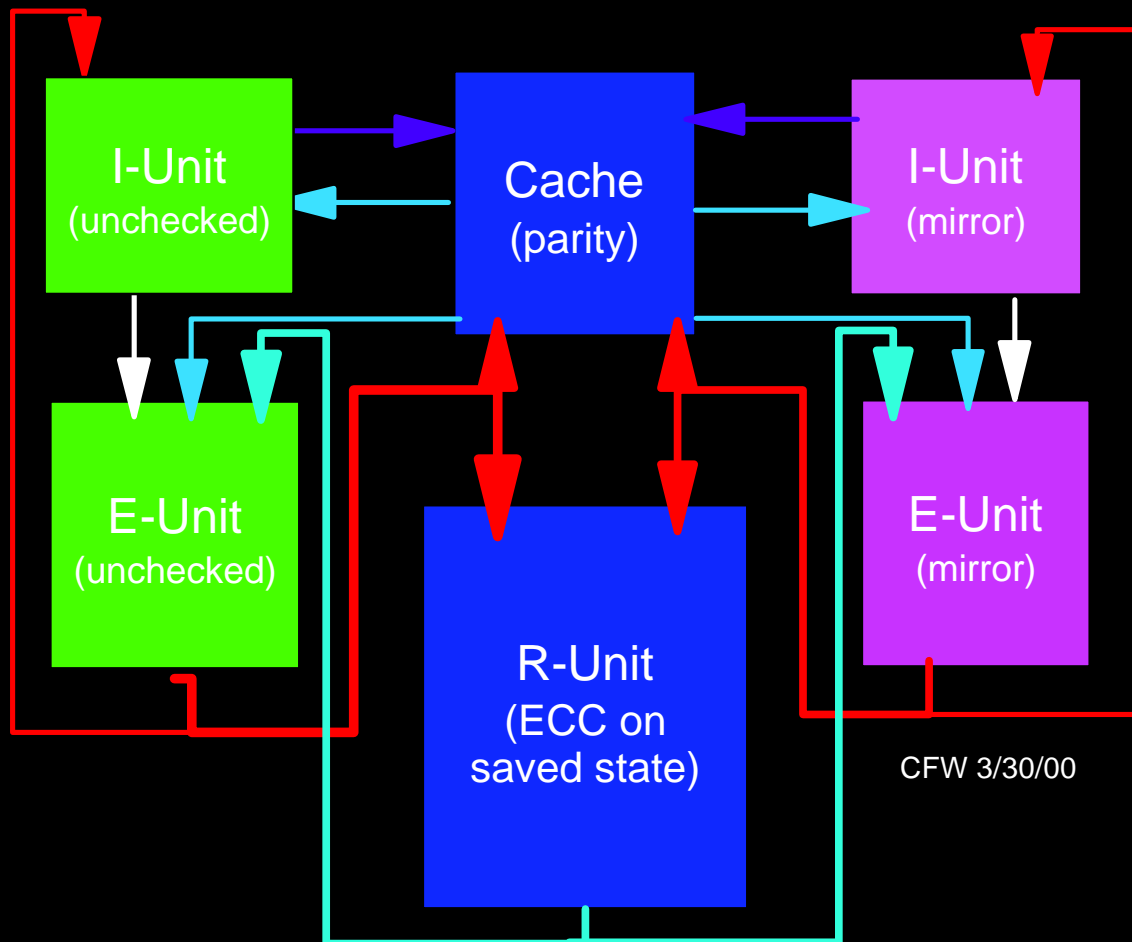- ECC'd Store Buffer on uP
- Line delete/sparing
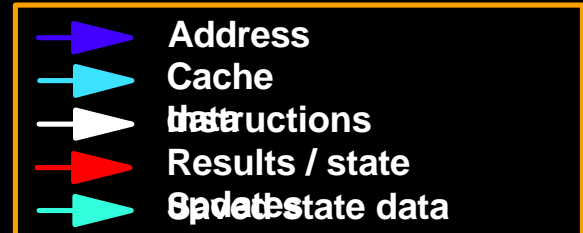
CP Error Detection & Recovery

Duplicated:
- **Complex controls**
- **Arithmetic dataflow**

Shared:
- **Cache controls**
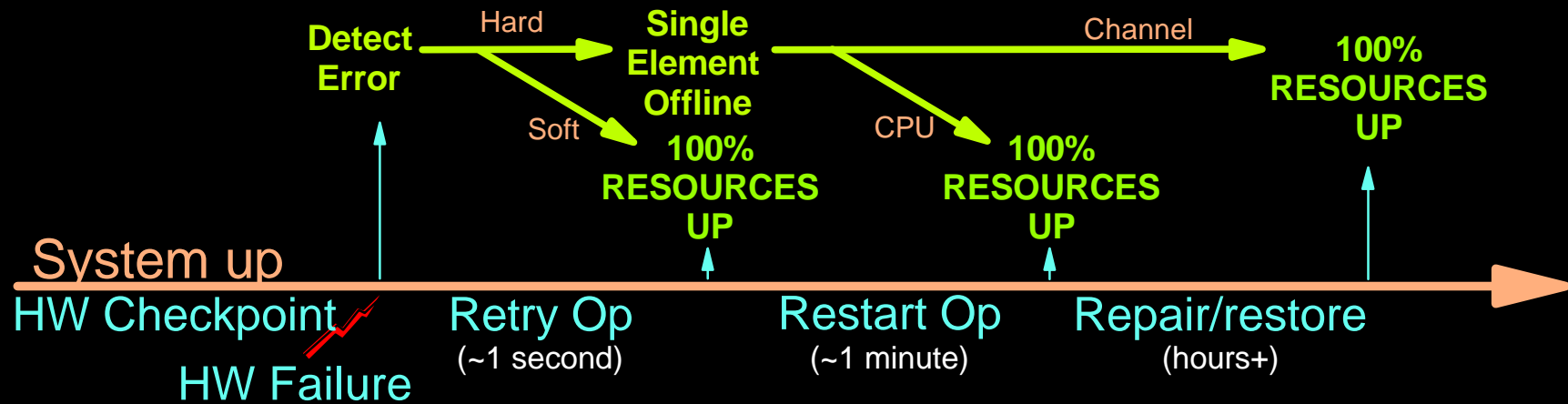- **Cache data/address flow**
- **R-Unit**

- **Check all state updates**
- **Preserve known good state**
- **If error**
  1. Stop state updates
  2. Refresh from saved state
  3. Restart CPU
- **If error persists**
  1. Extract saved state (SE)
  2. Load into spare CPU
  3. Start spare CPU

I-Unit (unchecked)

Cache (parity)

I-Unit (mirror)

E-Unit (unchecked)

R-Unit (ECC on saved state)

E-Unit (mirror)

CFW 3/30/00

Address
Cache data
Instructions
Results / state updates
Saved state data

2Q01 zSeries Full Field Data

➡ **MTTHardware Repair = 8 months**
➡ **81-83% of repairs are concurrent**

TYPICAL REPAIR SCENARIO

Detect Error → Hard → **Single Element Offline**

Soft → **100% RESOURCES UP**

Channel → **100% RESOURCES UP**

CPU → **100% RESOURCES UP**

System up

HW Checkpoint

HW Failure

Retry Op (~1 second)

Restart Op (~1 minute)

Repair/restore (hours+)

➡ **13-15% of repairs are deferable**
➡ **2-6% of repairs are app loss: MTTAL = 24 years**

zSeries Error Reporting

~2 week interval "call home" recovery
data

Suppose CP hard logic (not array) fails caused app loss:
MTTAL from 24 yrs to 11 yrs

Suppose array (L1, L2, BHT) fails also caused app loss:
MTTAL from 11 yrs to 5 yrs
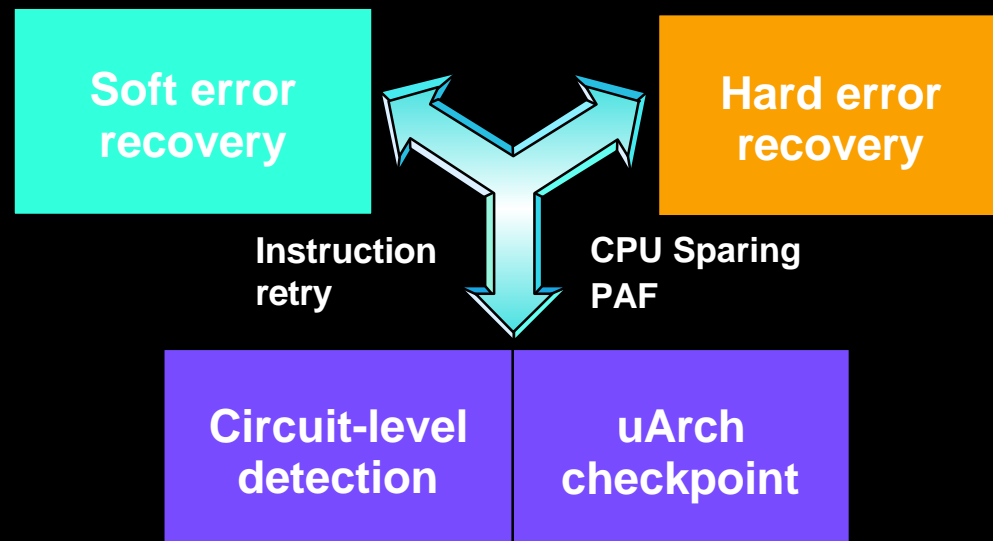
S/390 Evolution

S/390 uses same technology building blocks for soft and hard error recovery

- ► Enhanced over past 35 years

IT'S NOT THE ONLY OPTION

- ► Beginning afresh, might land elsewhere
- ► Need to be driven by current conditions
  - ✓ Technology
  - ✓ Workload

IT'S EFFICIENT & EFFECTIVE FOR S/390

Soft error recovery	Hard error recovery

Instruction retry   CPU Sparing PAF

Circuit-level detection	uArch checkpoint

Challenges for the 00s

➡ Increased importance of firmware

➡ Circuit failure mechanisms

➡ State encapsulation

➡ On-the-fly change

➡ Dynamic resource allocation

➡ Configuration validation