IBM S/390 G5 Microprocessor

Authors

Timothy J. Slegel (presenter)
Robert M. Averill III
Mark A. Check
Bruce C. Giamei
Barry W. Krumm
Christopher A. Krygowski
Wen H. Li
John S. Liptay
John D. MacDougall
Thomas J. McPherson
Jennifer A. Navarro
Eric M. Schwarz
Kevin Shum
Charles F. Webb

IBM Corporation
Poughkeepsie, NY

IBM

IBM S/390 G5 Microprocessor

- Newest member of IBM's CMOS mainframe family
- Announced May 7, 1998
- Microprocessor design is based on previous generation G4 but with numerous enhancements:
  - Faster cycle time
  - Improvements to CPI
  - New architectural features
  - Improvements in Reliability-Availability-Serviceability (RAS)

IBM

Recent History of IBM's Mainframes

- Compatible enhancements from S/360 and S/370 architectures
- 1990 - S/390 architecture. Multiple 2GB address spaces
- 1993 - Last bipolar mainframe announced
- 1994 - IBM begins transition to CMOS technology for future systems
- 1995 - G2 system announced. Second generation of CMOS for S/390
- 1996 - G3 announced. Significant performance improvements over G2
- 1997 - G4 announced. Performance comparable to IBM's bipolar mainframe
- 1998 - G5 announced. 2X performance over any prior IBM mainframe and fastest S/390 in the industry

IBM

Chip Technology

- IBM CMOS6X technology
  - .25µm drawn, .15µm L_{eff} (nFET)
  - 6 levels of metal
  - 1.9V (at circuit)
- 14.6mm x 14.7mm
- 25 million transistors
- 25W power
- Dataflow logic is full-custom
- Most control logic is synthesized with some hand tweaking of the schematic and layout

IBM

“Speeds and Feeds”

- Frequency:
  - 500 MHZ (shipping product)
  - 600 MHZ (laboratory environment)
  - Other chips in the system run at ½ microprocessor frequency
- 256 KB L1 cache (4-way associative) contains instruction, operand and millicode data. Cache is 2-way interleaved.
- Cache line size is 256 bytes
- 1024 entry Translation-Lookaside Buffer (4-way associative)
- 2048 entry Branch Target Buffer (2-way associative)
- 32 KB writeable control-store contains routines for 64 of the most commonly used instructions executed by millicode
- Performance:
  - 150 S/390 MIPS uni-processor
  - 1040 S/390 MIPS 10-way
  - 2X performance of G4 and previous IBM bipolar systems

IBM

Processor Implementation

- Single issue in-order execution
- Pipeline length is 7 stages from I-fetch to writing of results
  - 3 additional stages used for Recovery-Unit (R-Unit) checkpointing of results. These can not cause pipeline stalls
- Dynamic Address Translation (and Access Register Translation) done in hardware for all S/390 addressing modes
- Instruction and operand Absolute Address History Tables used to predict virtual→real bits when accessing the cache
- Continuation fetches allow multiple cache accesses per cycle
- Fixed-point execution unit contains: 64-bit binary adder, 64-bit BLU/AIM, 8-digit decimal adder
- R-unit holds checkpointed copy of entire micro-architected state of the processor in 256 registers, each 32/64 bits
- Local shadow copies of GPRs, FPRs, and certain R-unit registers distributed throughout I and E units
- "Slow-mode" when processor detects exception conditions

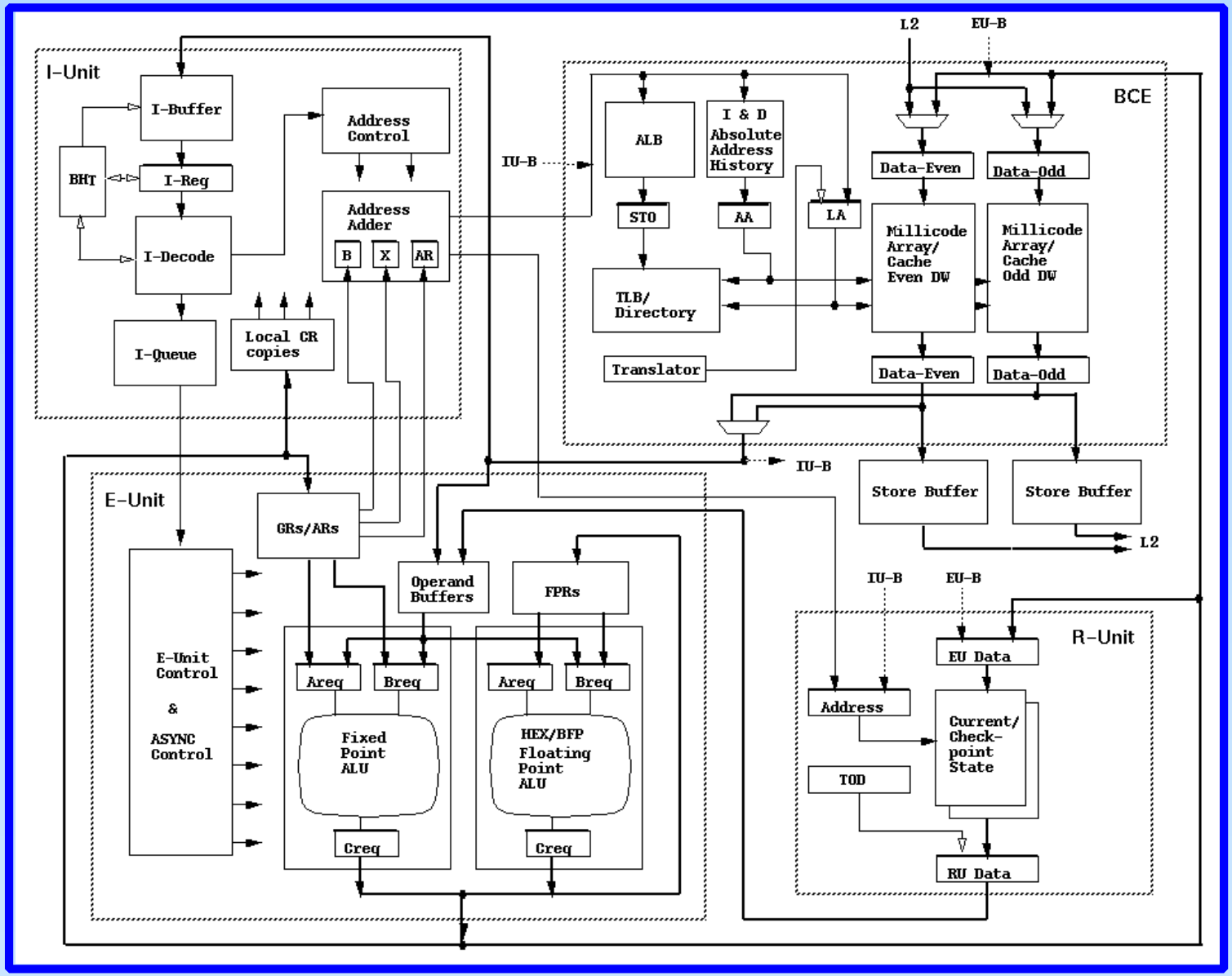
IBM

Processor Implementation

- Hardware support for two levels of virtual machine guests
- Millicode:
  - Vertical microcode
  - Used to implement complex S/390 instructions and exception processing. Also handles various service functions
  - Consists of S/390 hard-wired instructions plus about 100 instructions only usable by millicode
  - Executing a complex S/390 instruction is like executing a hardwired subroutine call instruction. When millicode completes execution, it is like a hardwired subroutine return
  - Has complete read/write access to all R-unit registers that contain control/state information not normally accessible by S/390 programs
  - Has its own set of GPRs and control registers

IBM

Diagram of G5 Microprocessor



Processor CPI enhancements for G5

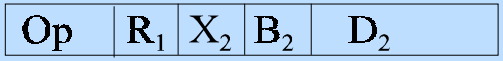
- Branch Target Buffer:
  - Active for normal S/390 and millicode instruction branch prediction
  - For conditional branches, records taken-once or more-than-once
- 6 instruction buffers (each 32 bytes)
- Decimal instruction performance improvements: totally implemented in hardware or addition of millicode assists
- Program Status Word modifying instructions in hardware
- Fixed-point multiply and divide improvements
- Quiesce enhancements for instructions that modify TLB entries:
  - Processor executing instruction sends quiesce request to all other processors
  - Receiving processors can now resume executing instructions, subject to limitations, after completing their TLB changes

IBM

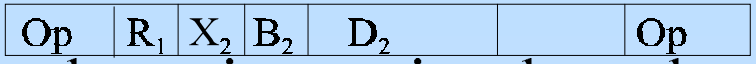
Floating-point Unit

- Continues to support traditional S/390 hex floating-point architecture
- Now provides IEEE 754 compliant execution on the S/390 platform
- 121 new opcodes (87 binary, 26 hex, 8 support) in addition to the 54 traditional opcodes. New opcode format:

  RX format (traditional):



  RXE format (new):



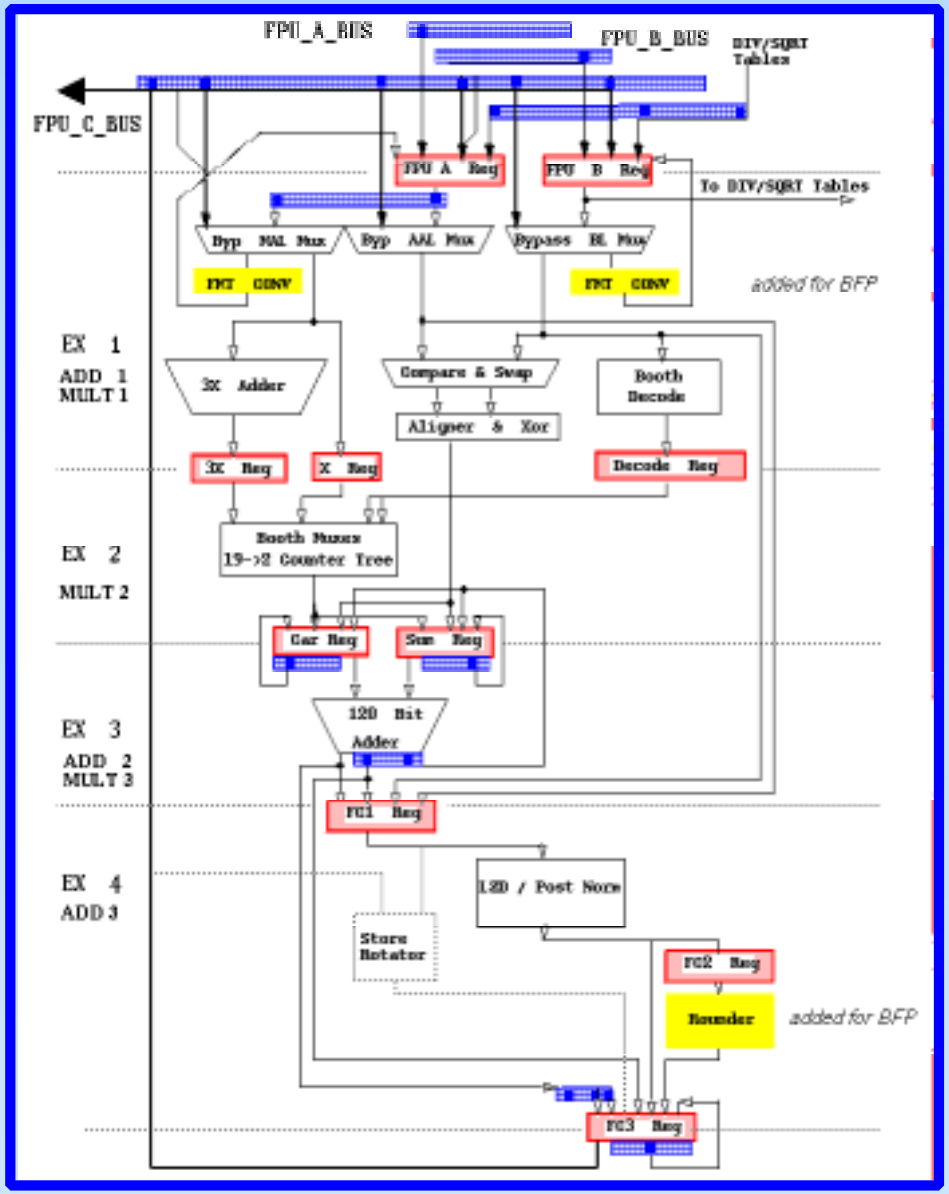
- Hex - 3 cycles latency / 1 cycle per instruction throughput
- Binary - 5 cycles latency / 2 cycles per instruction throughput
- Floating-point registers: increased from 4 to 16

S/390 Floating-point formats

Format	Sign	Exp bits	Exp bias	Fraction	Total width
Hex short	1	7	64	24	32
Hex long	1	7	64	56	64
Hex extend	1	7	64	112	128
Bin single	1	8	127	24	32
Bin double	1	11	1023	53	64
Bin Double Ext	1	15	16383	113	128

IBM

Diagram of Floating-point Unit



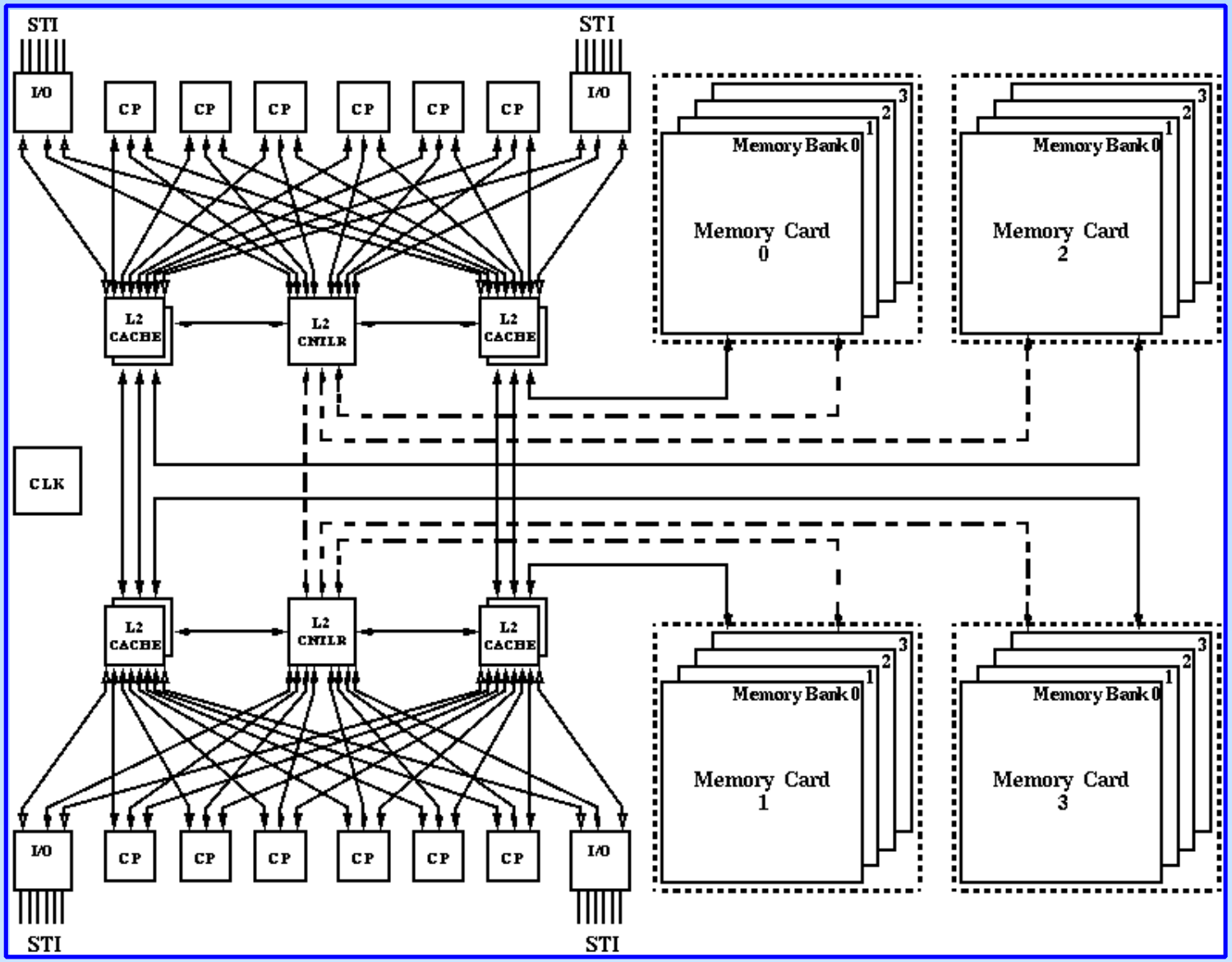
IBM

System Design

- MCM contains:
  - 12 processor chips (maximum of 10 available for customer use)
  - 2 L2 cache control chips - L2 directory, configuration array, system control functions
  - 8 L2 cache dataflow/array chips - 1 MB of L2 cache per chip
  - 4 I/O interface chips - each with 6 STI ports that connect to channel subsystem. Each STI port runs at 333 MBytes/sec in each direction and all 24 can be running simultaneously
  - 2 cryptographic co-processor chips - hardware RSA and DES
  - 1 clock/service chip - oscillator distribution, contains interface to the Service Element (laptop running OS/2)
- 8 MB of L2 cache in a bi-nodal configuration
- 4 memory cards, each with 4 banks. Total system memory up to 24GB.
- All main data busses in the system are 128 bits wide

IBM

Diagram of G5 System Structure



IBM

Reliability-Availability-Serviceability (RAS)

- Continues mainframe tradition of state-of-the-art error detection and recovery
- Processor has essentially 100% error detection and recoverability from any transient error
- R-Unit contains checkpointed micro-architected state and is protected with ECC
- Instruction and Execution Units are completely duplicated with both copies performing identical functions every cycle
- The two I/E copies results are cross-checked every cycle in the R-Unit and L1 cache
- If an error is detected (implemented totally in hardware):
  - Clear all processor arrays - L1, TLB, BTB
  - Hardware reset of critical latches
  - Shadow register copies of micro-architected state are reloaded from R-Unit checkpointed copy
  - Restart I-fetching and execution
- Array delete and soft repair of arrays at power-on reset

IBM

Recovery Mechanisms for Processor Checkstops

- Processor hardware recovery is always attempted first
- If that fails due to a solid error:
  - Concurrent CP Sparing allows a spare processor in the system to be brought on-line in a running system with customer intervention
  - Processor Availability Facility moves S/390 architected state to another processor in cooperation with the Operating System
  - SAP Sparing (for I/O processors) occurs dynamically without OS or customer intervention

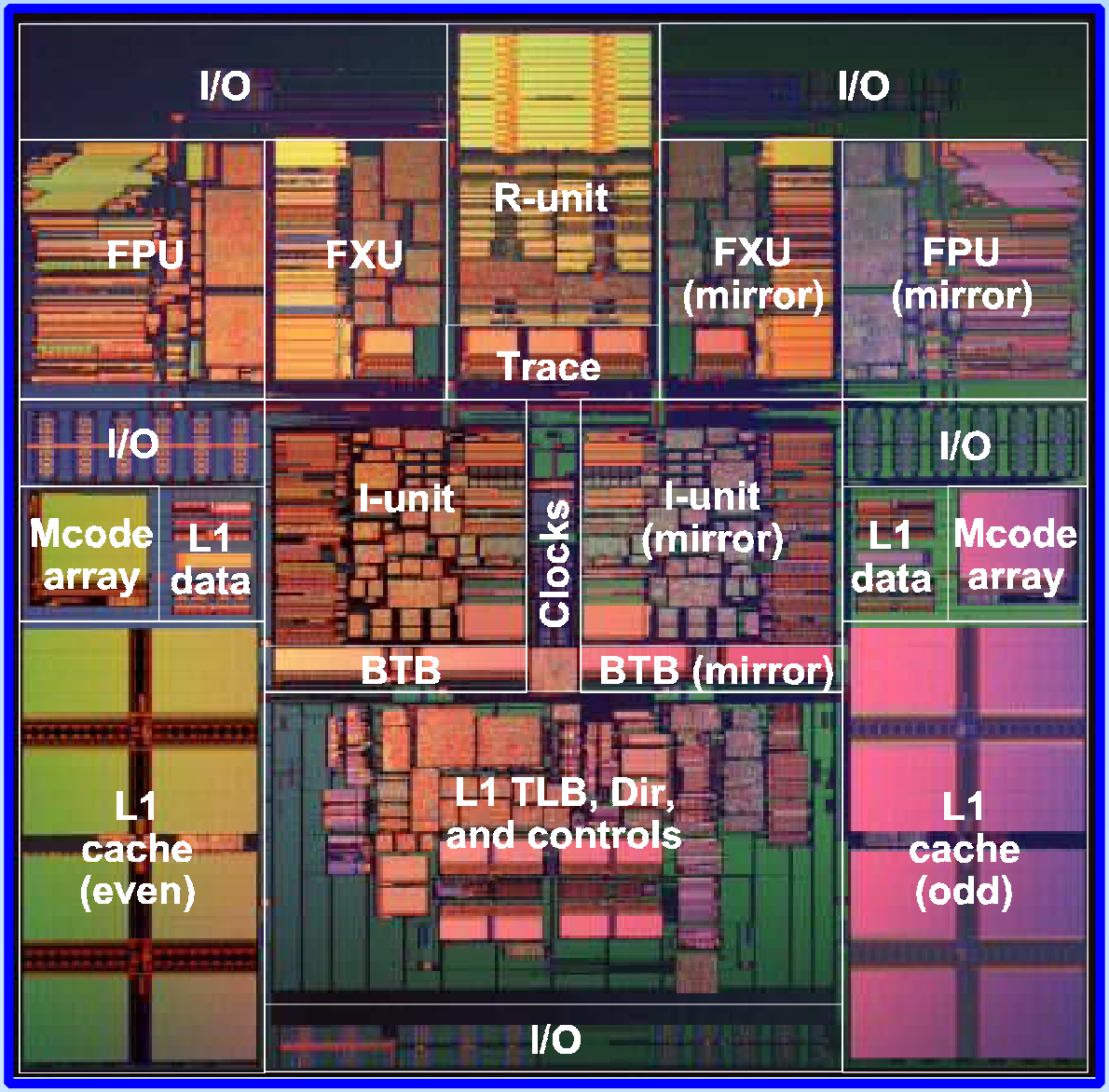
IBM

Transparent CP Sparing

- Processor encounters solid hardware-detected error and checkstops
- Service Element scans out all latches on failed processor
- SE extracts the micro-architected state from these latches
- SE sends micro-architected state back to system where it is placed in System Area memory
- Spare processor is notified
- Millicode on spare processor makes any required changes to the micro-architected state
- Millicode executes a hardware instruction that loads the state into the processor in one atomic operation
- At completion, I-fetching and execution resumes where it had stopped on the failed processor
- Operating System and the customer are not even aware this event has happened

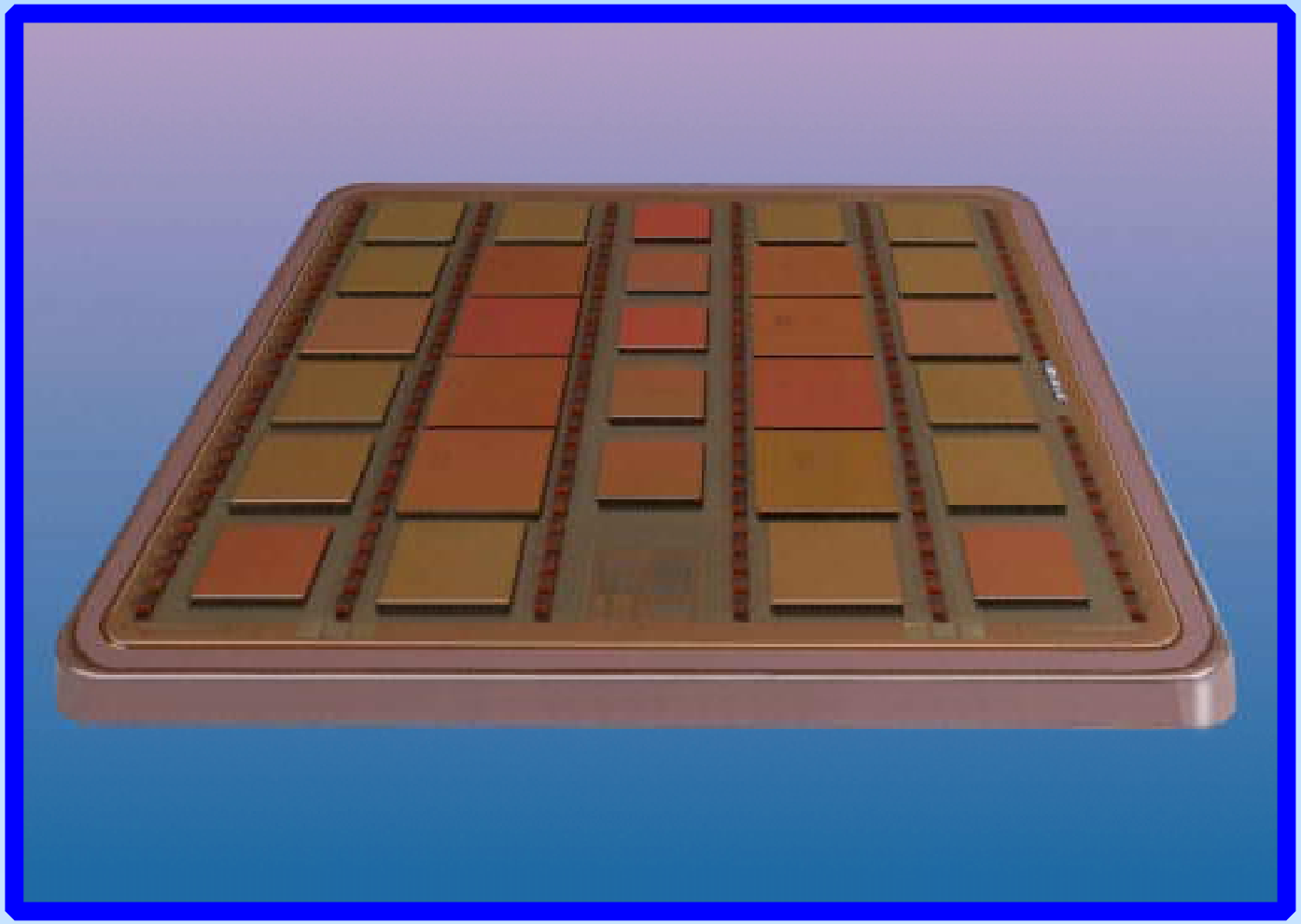
IBM

IBM S/390 G5 Microprocessor



IBM

IBM S/390 G5 MCM



IBM