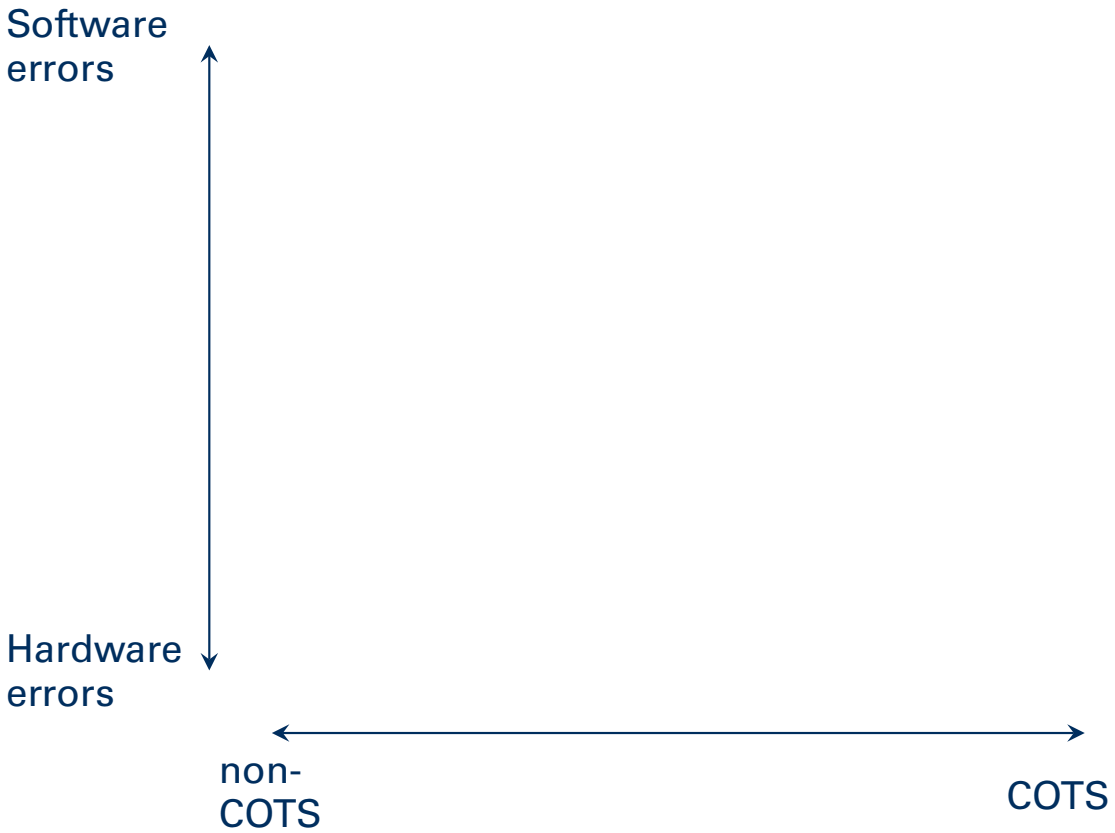
TECHNISCHE UNIVERSITÄT DRESDEN

OPERATING SYSTEM SUPPORT FOR REDUNDANT MULTITHREADING

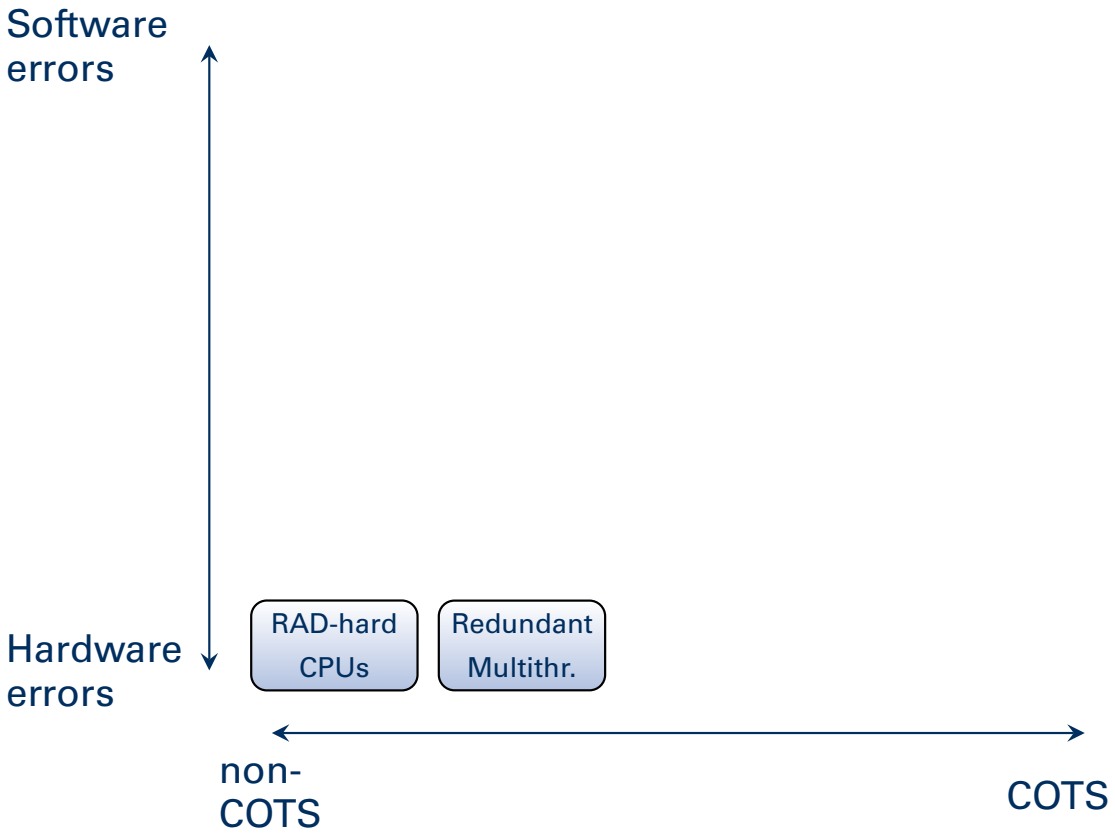
Björn Döbel (TU Dresden)
Hermann Härtig (TU Dresden)
Michael Engel (TU Dortmund)

Tampere, 08.10.2012

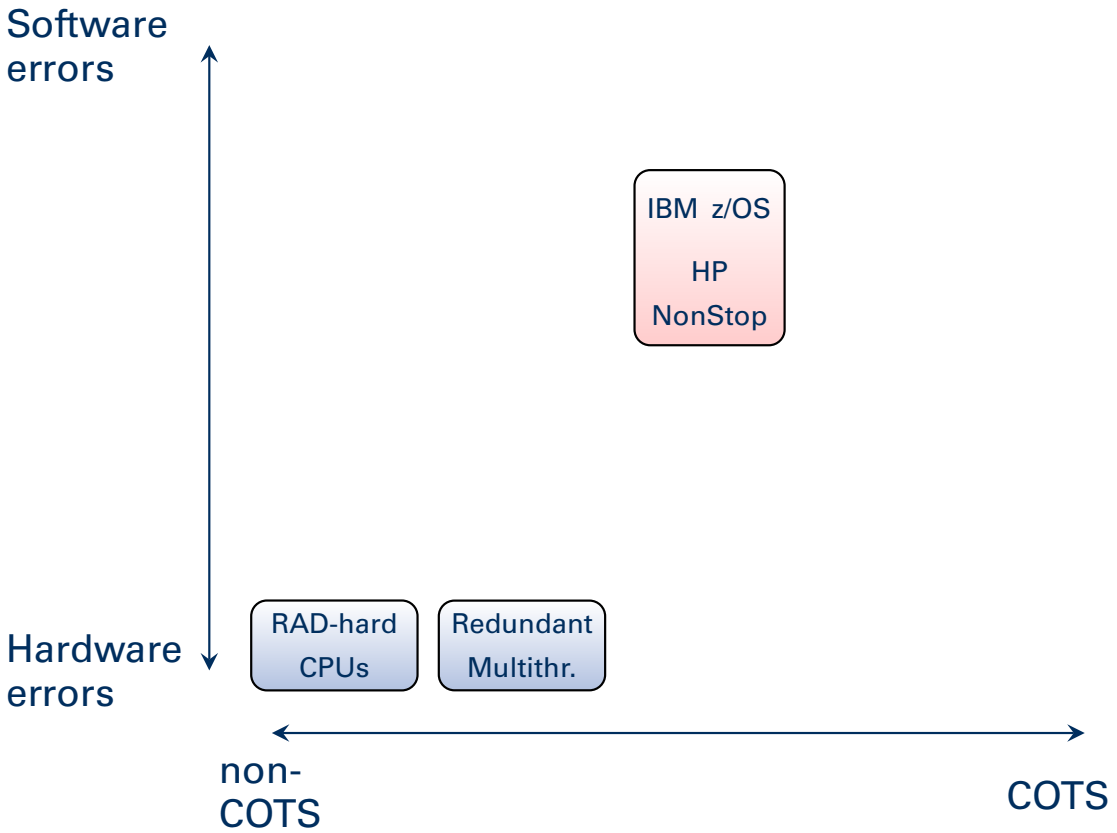
Fault Tolerance: State of the Union



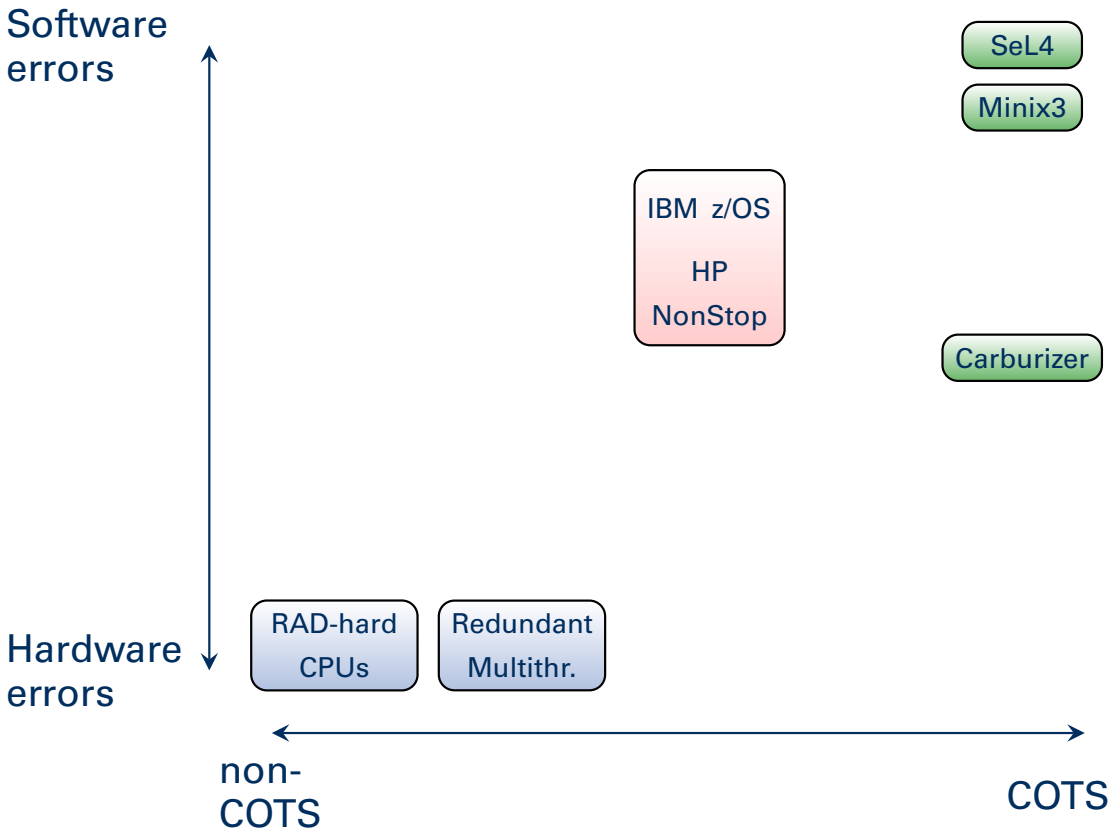
Fault Tolerance: State of the Union



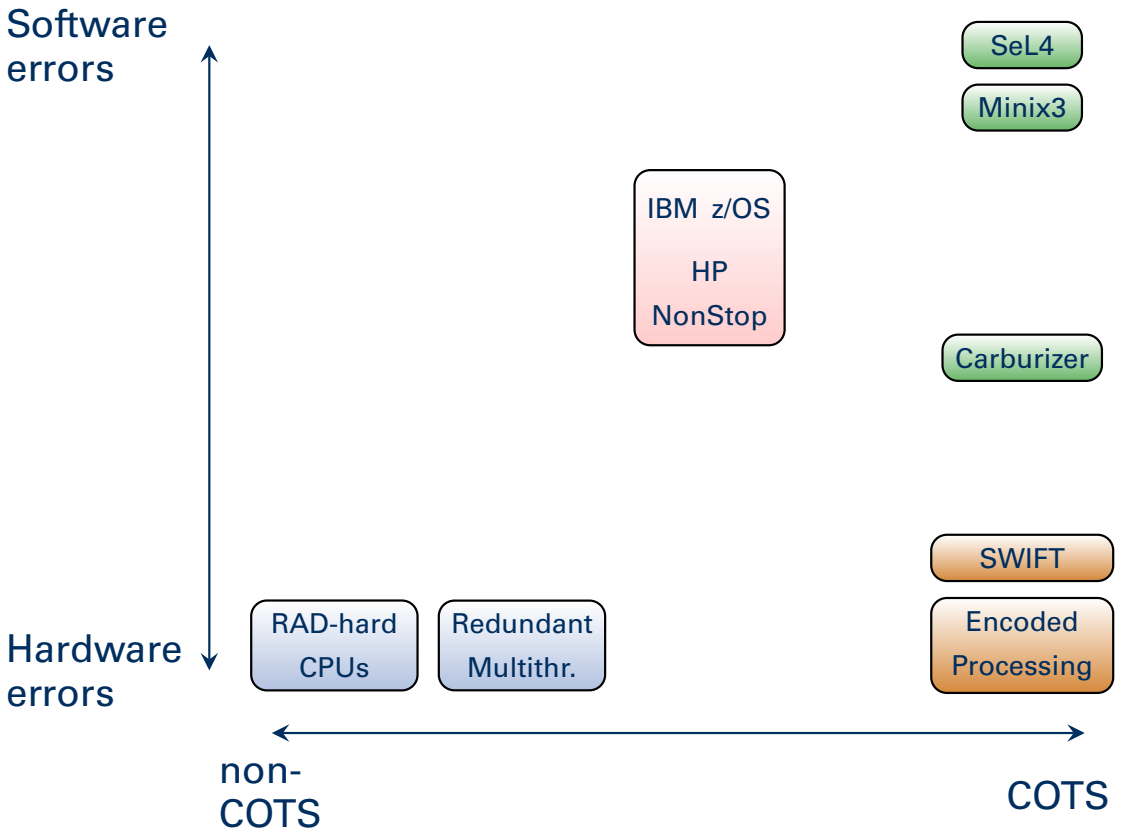
Fault Tolerance: State of the Union



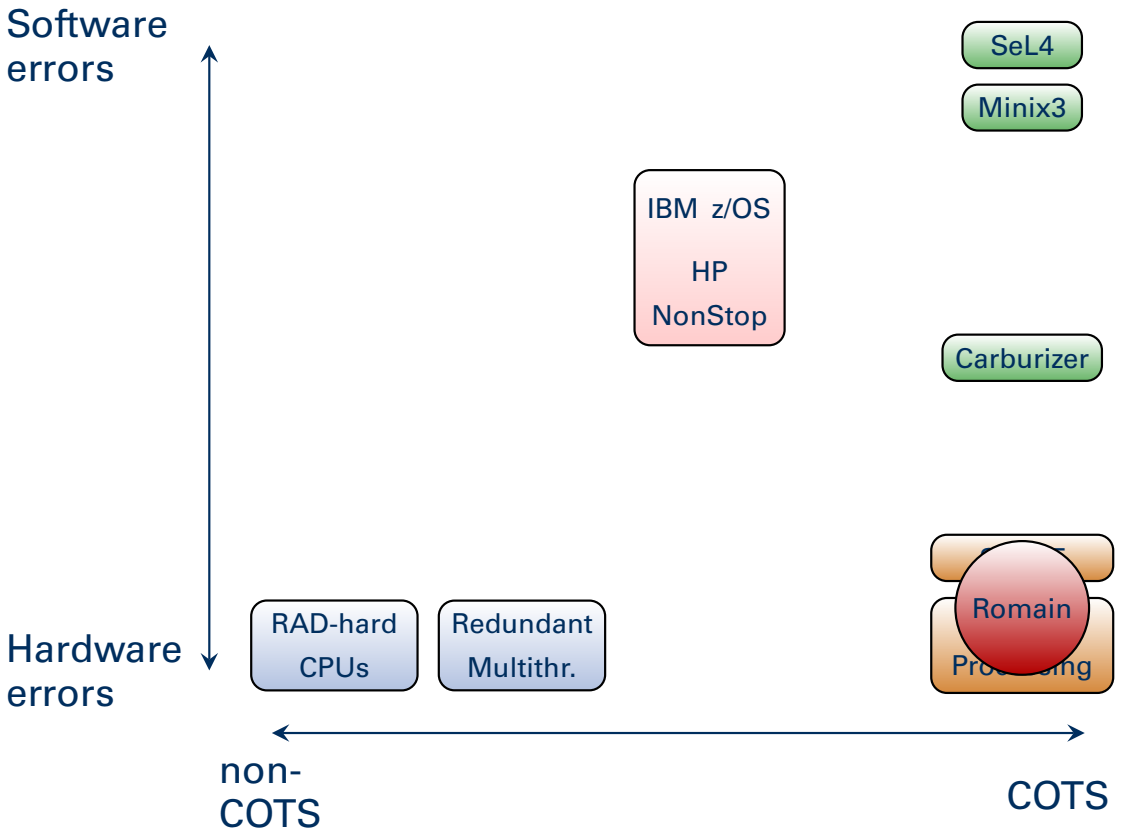
Fault Tolerance: State of the Union



Fault Tolerance: State of the Union



Fault Tolerance: State of the Union



Process-Level Redundancy [Shye 2007]

Binary recompilation

- Complex, unprotected compiler
- Architecture-dependent

System calls for replica synchronization

Virtual memory fault isolation

- Restricted to Linux user-level programs

Process-Level Redundancy [Shye 2007]

Binary recompilation

- Complex, unprotected compiler
- Architecture-dependent

Reuse OS mechanisms

System calls for replica synchronization

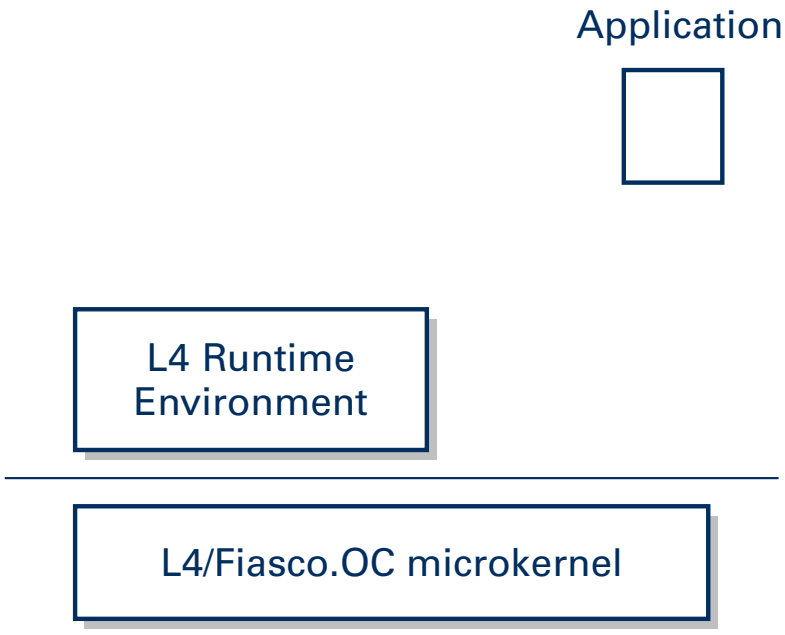
Additional synchronization events

Virtual memory fault isolation

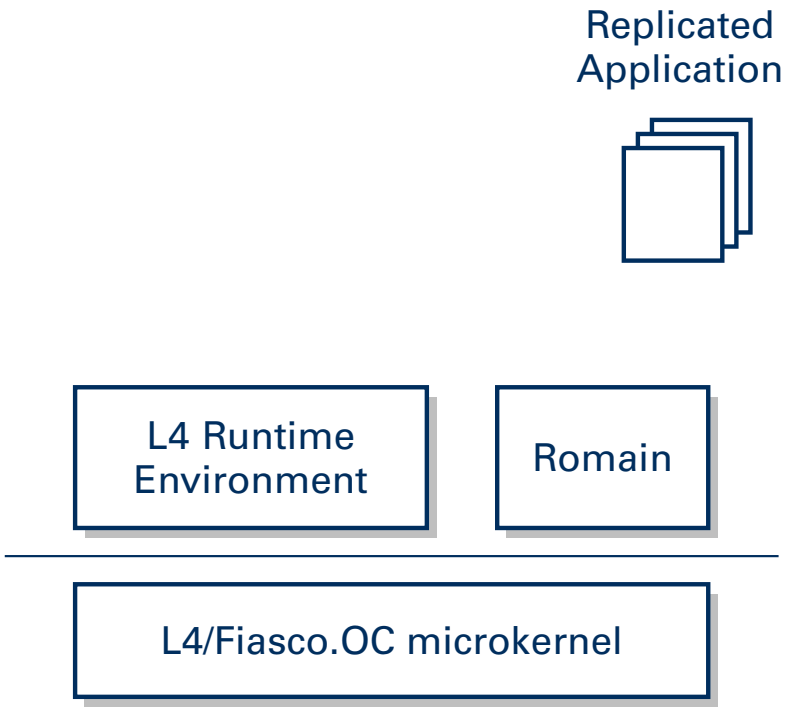
- Restricted to Linux user-level programs

Microkernel-based

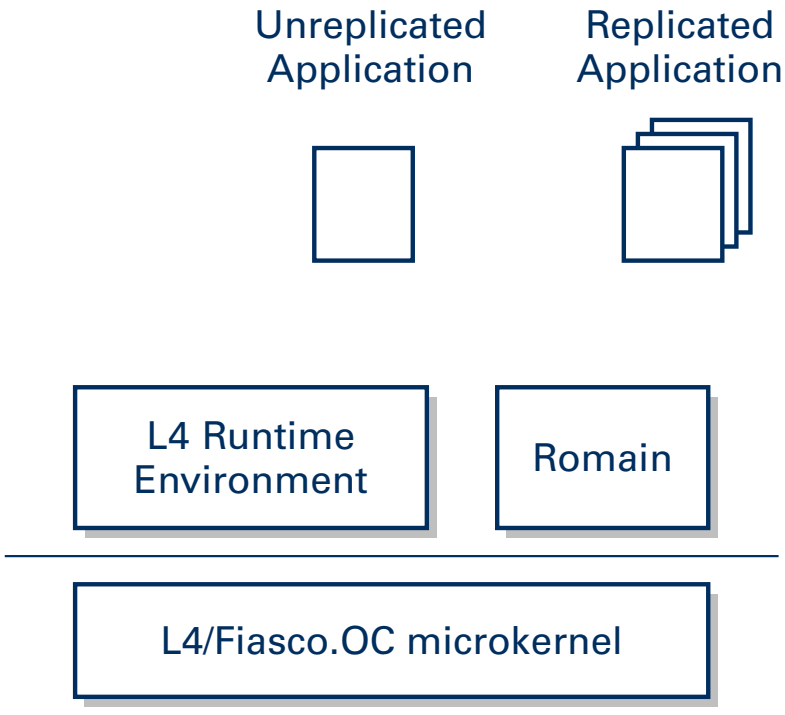
Transparent Replication as OS Service



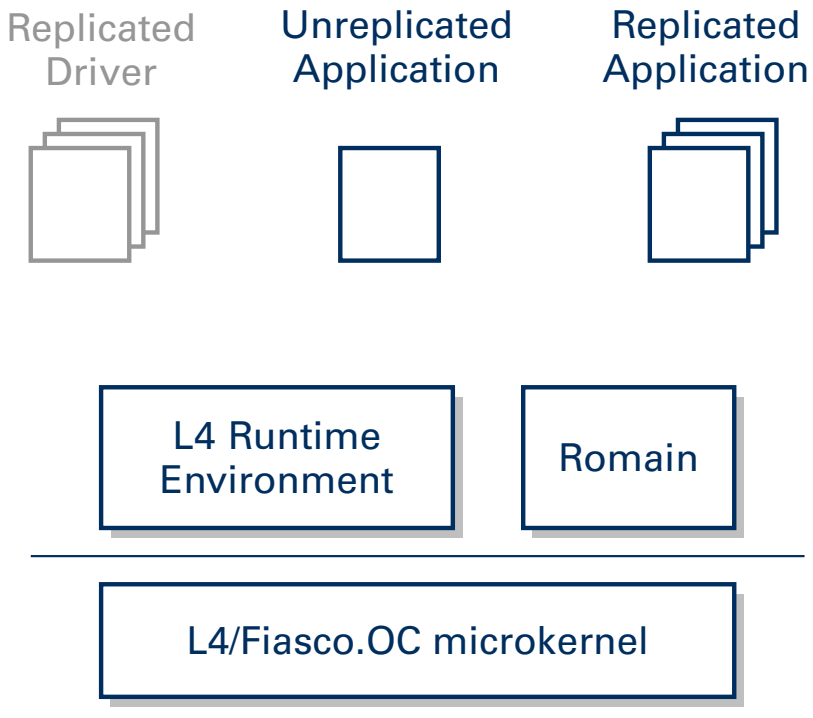
Transparent Replication as OS Service



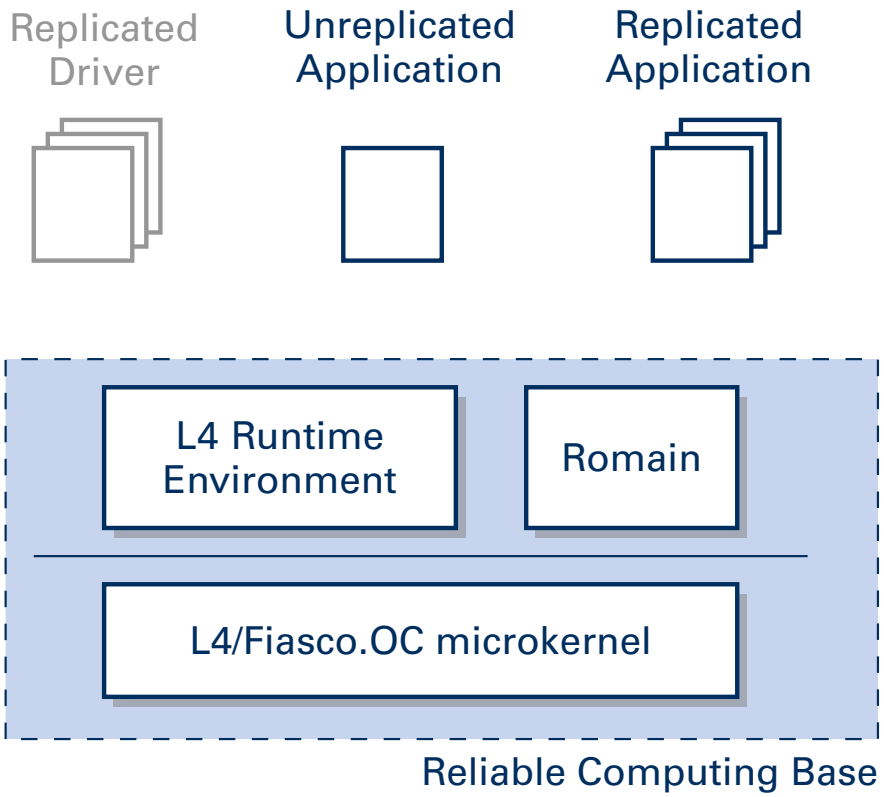
Transparent Replication as OS Service



Transparent Replication as OS Service

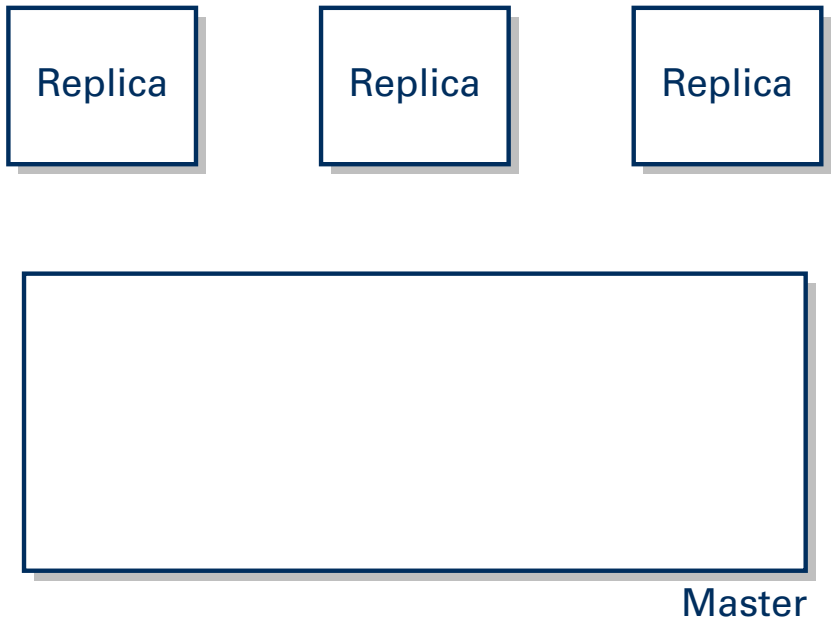


Transparent Replication as OS Service

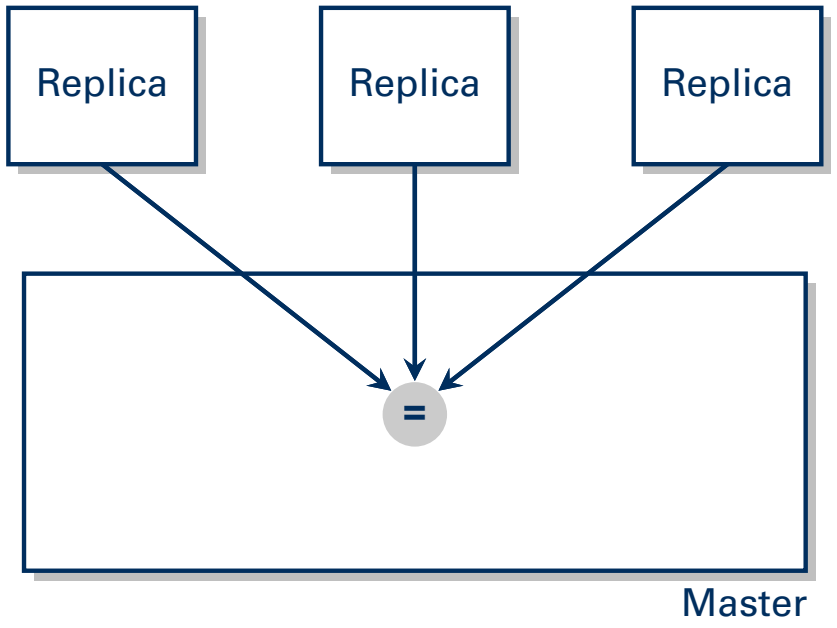


Romain: Structure

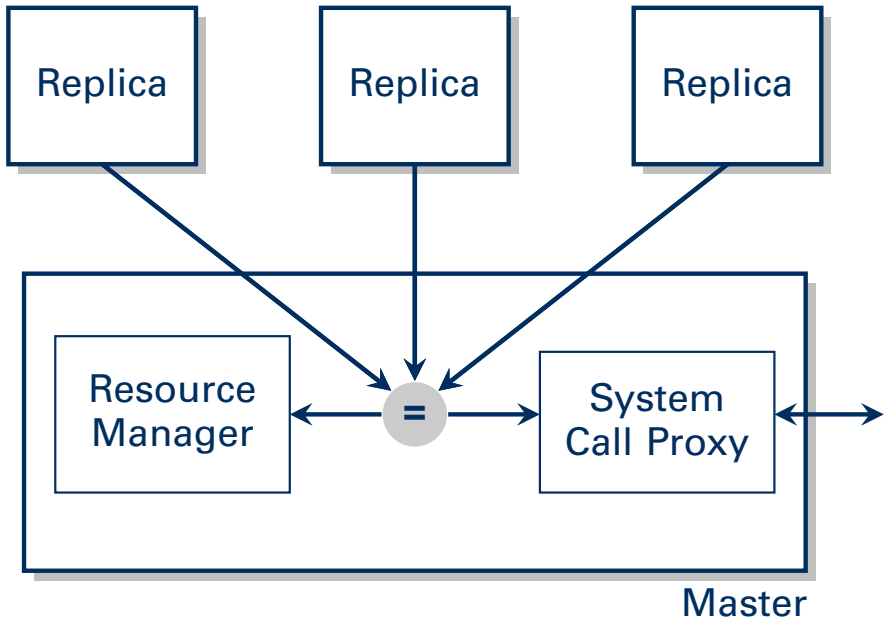
Romain: Structure



Romain: Structure



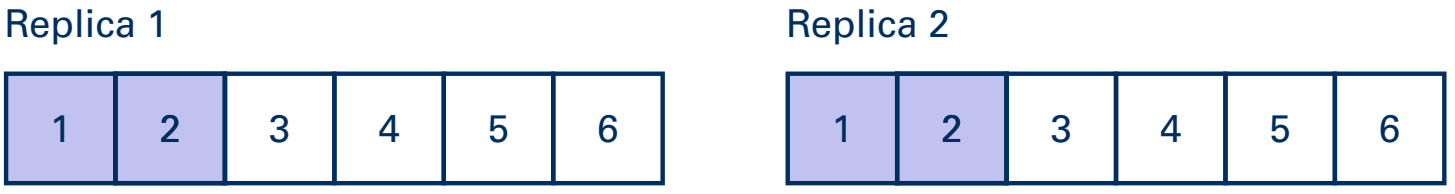
Romain: Structure



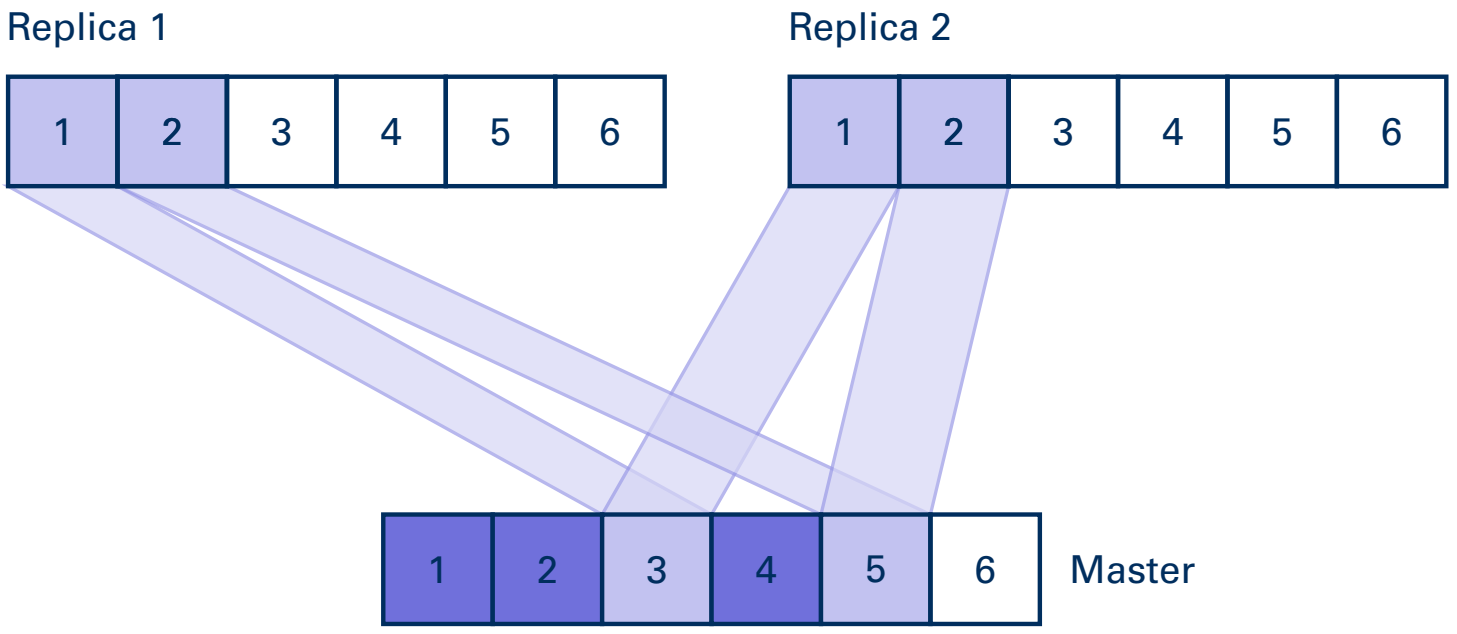
Resource Management: Capabilities

Replica 1

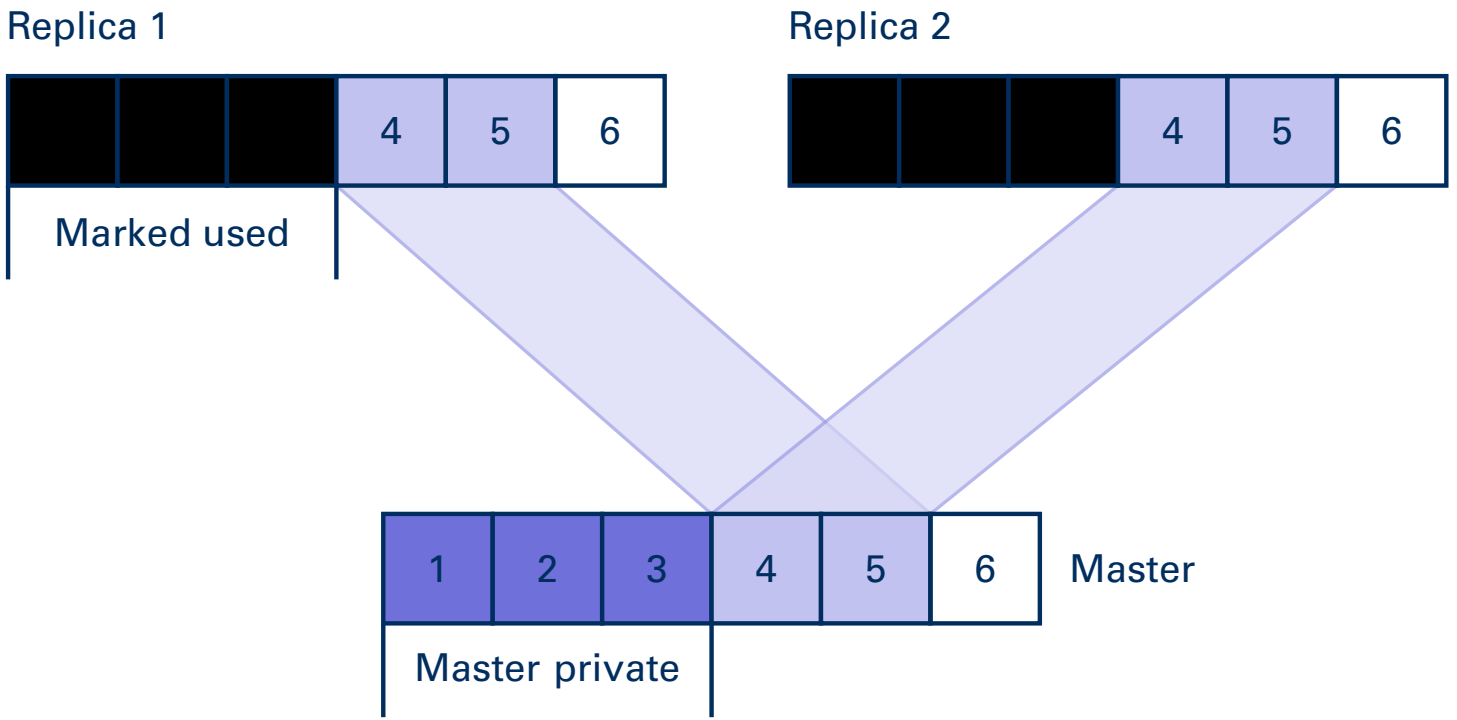
Resource Management: Capabilities



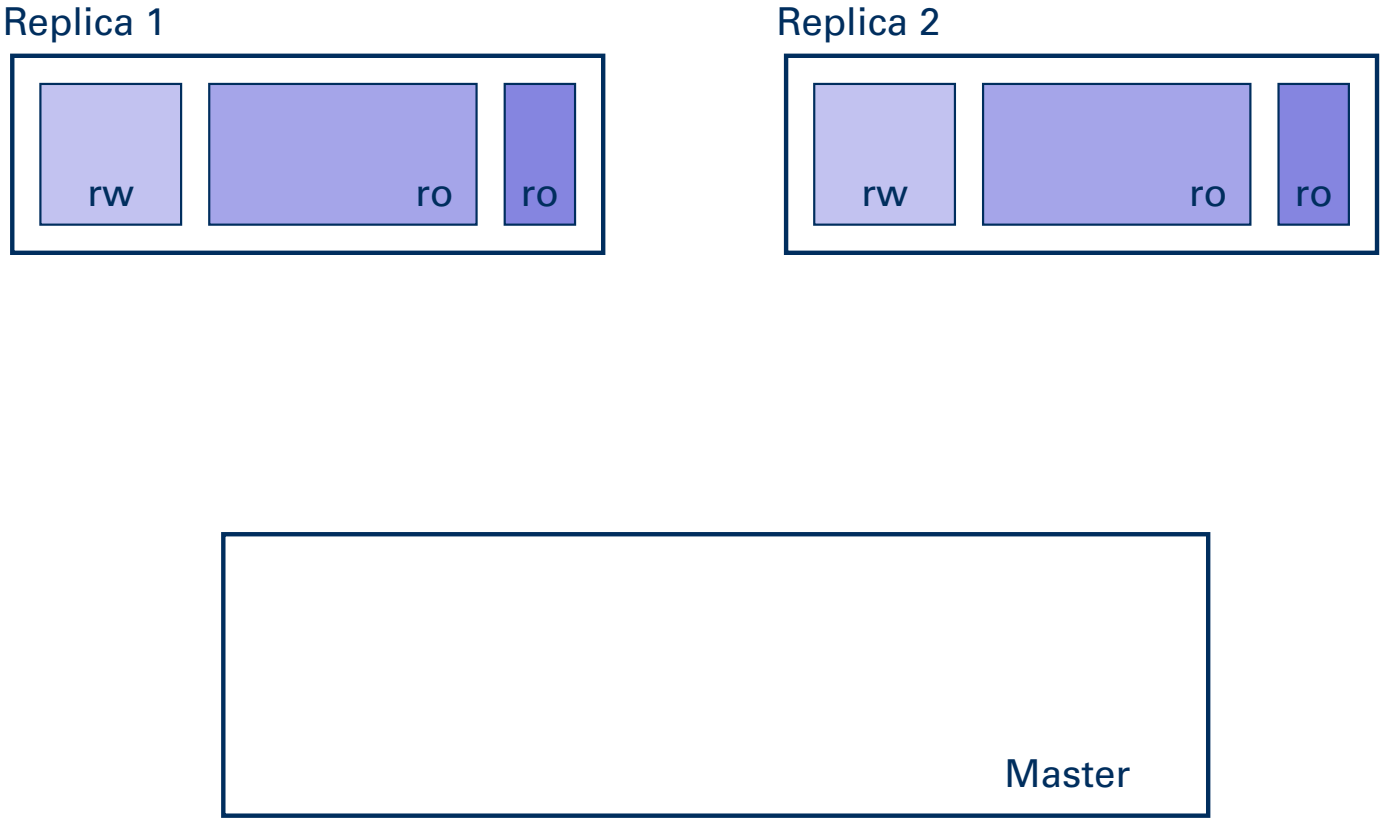
Resource Management: Capabilities



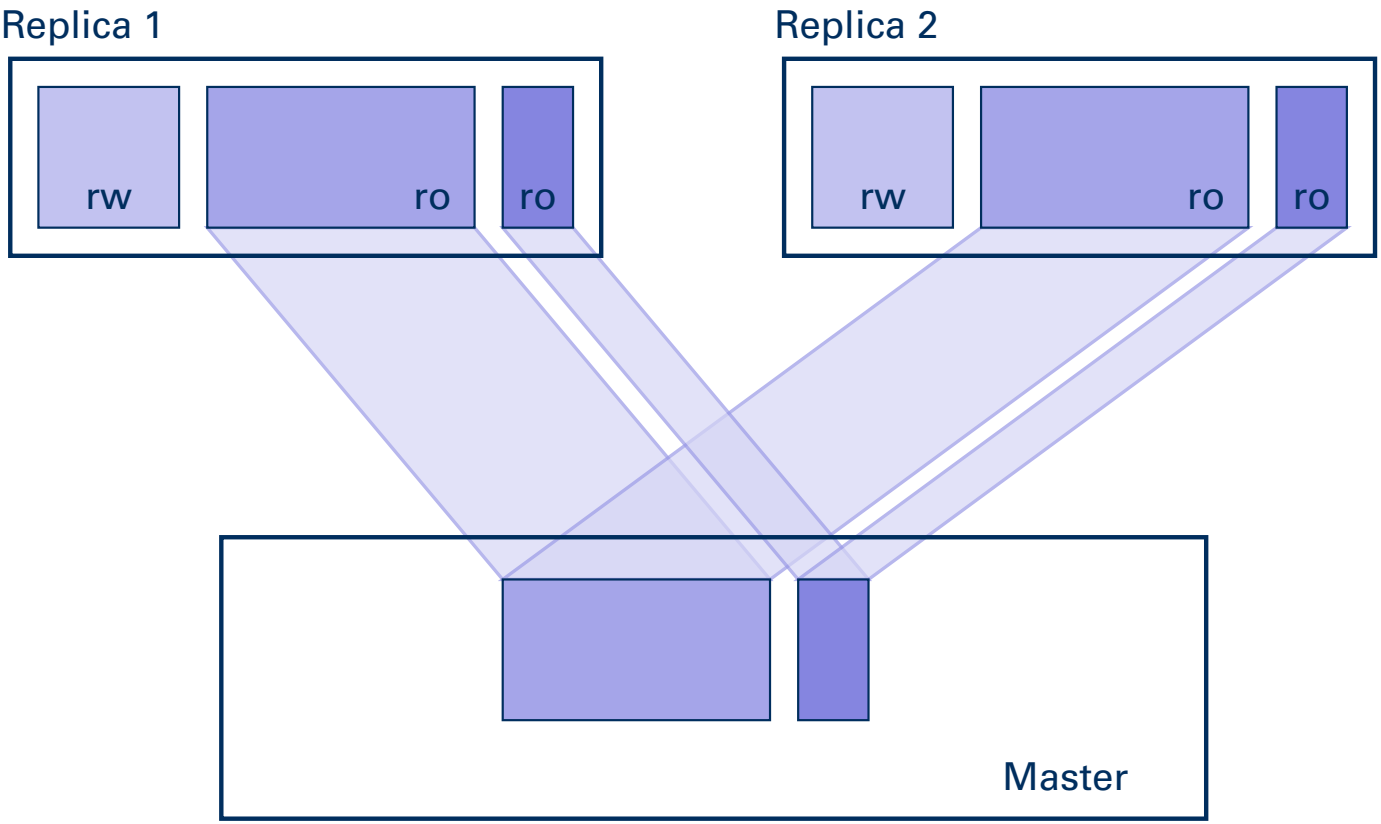
Partitioned Capability Tables



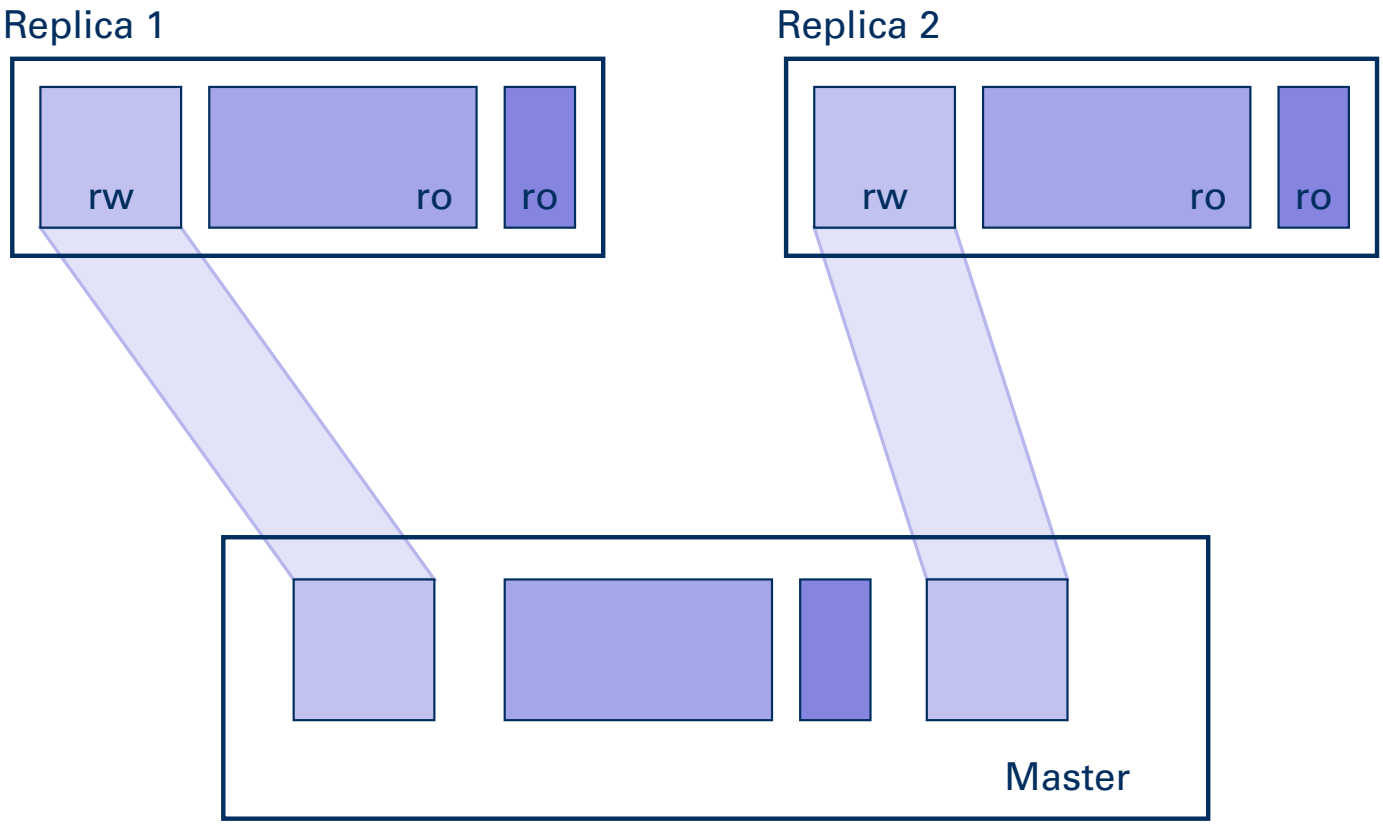
Replica Memory Management



Replica Memory Management



Replica Memory Management



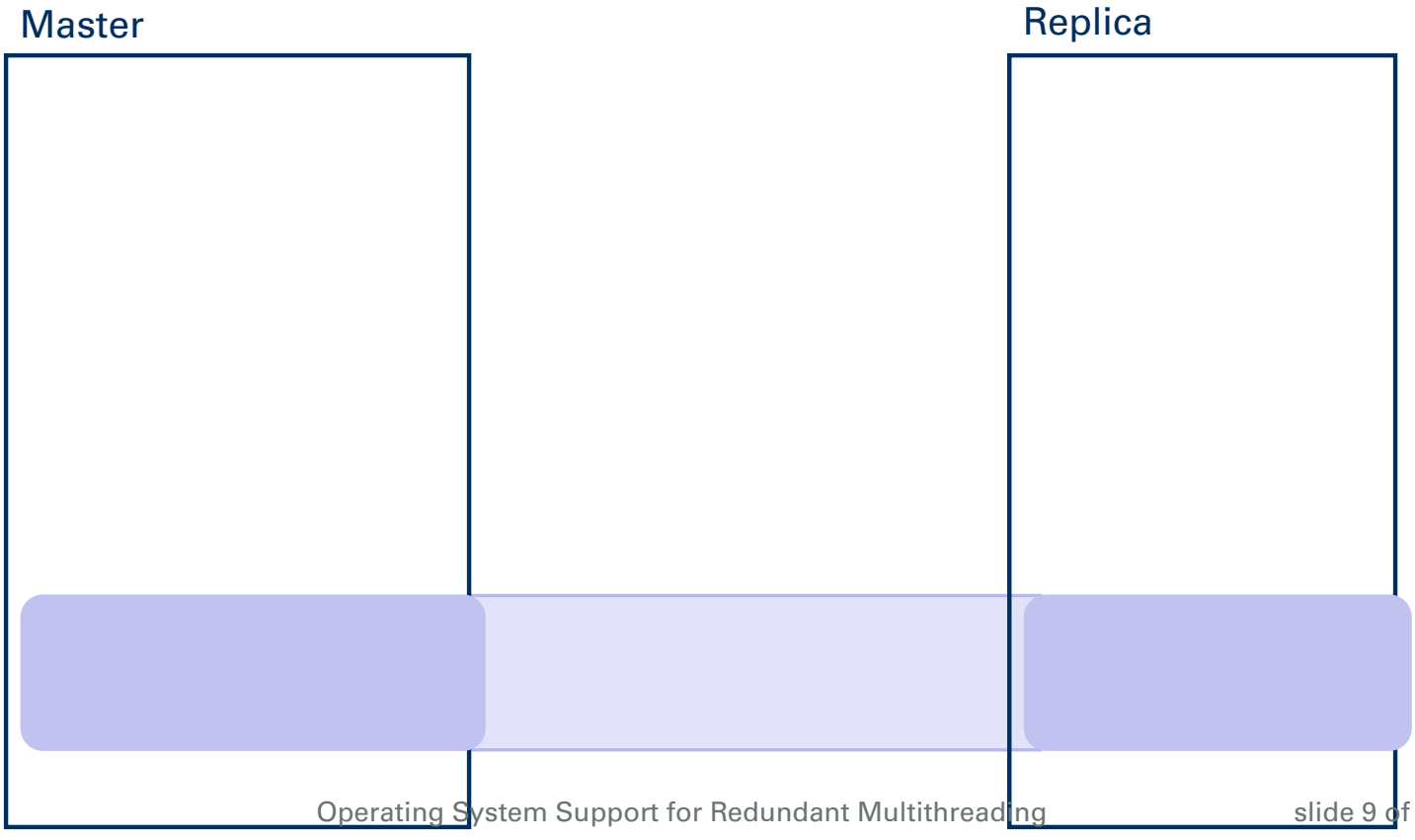
Shared Memory

- Not in complete control of master
- Standard technique: trap&emulate
  - Execution overhead (x100 - x1000)
  - Adds complexity to RCB

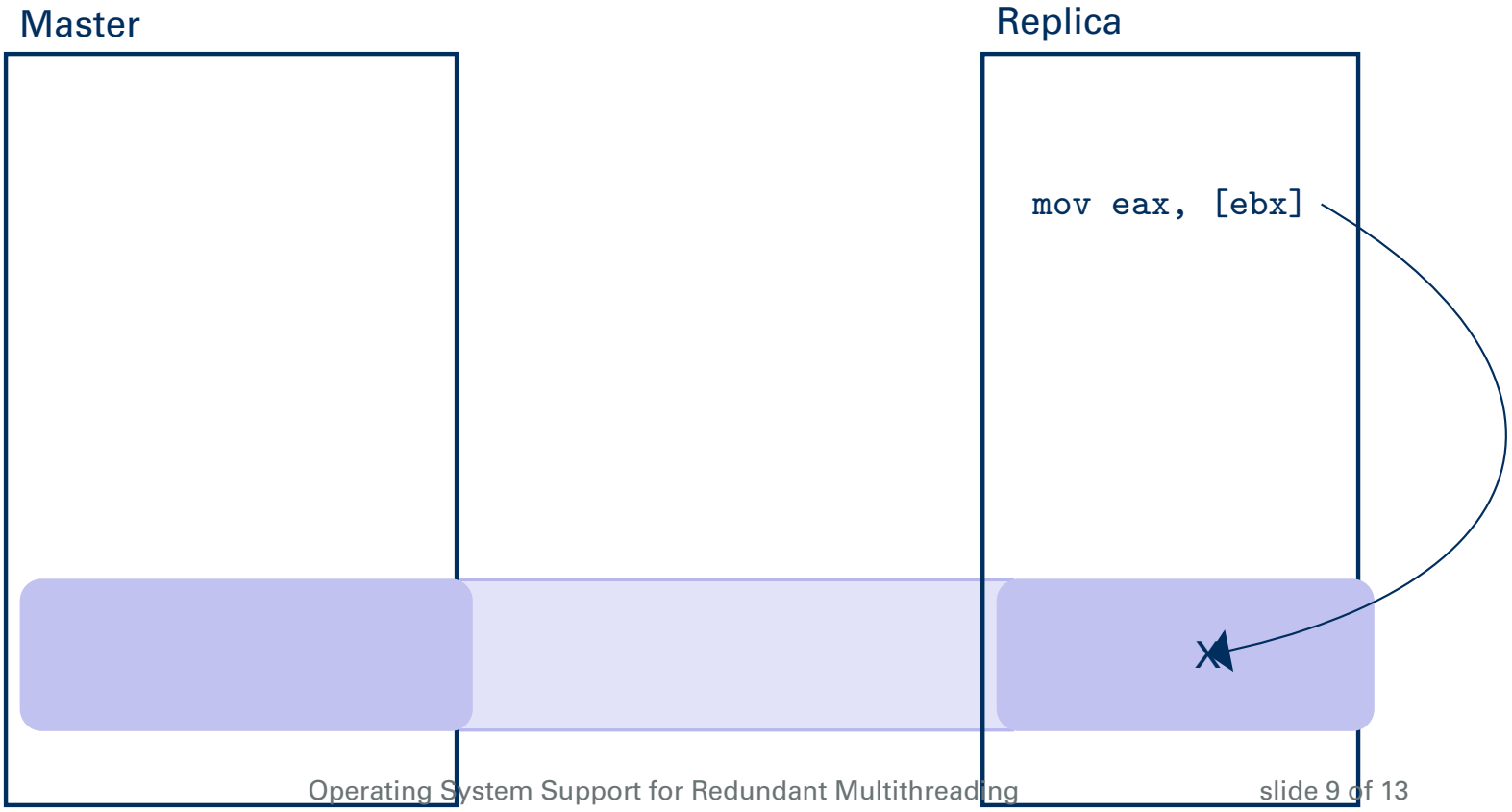
    Disassembler 6,000 LoC

    Tiny emulator 500 LoC
- Our implementation: copy & execute

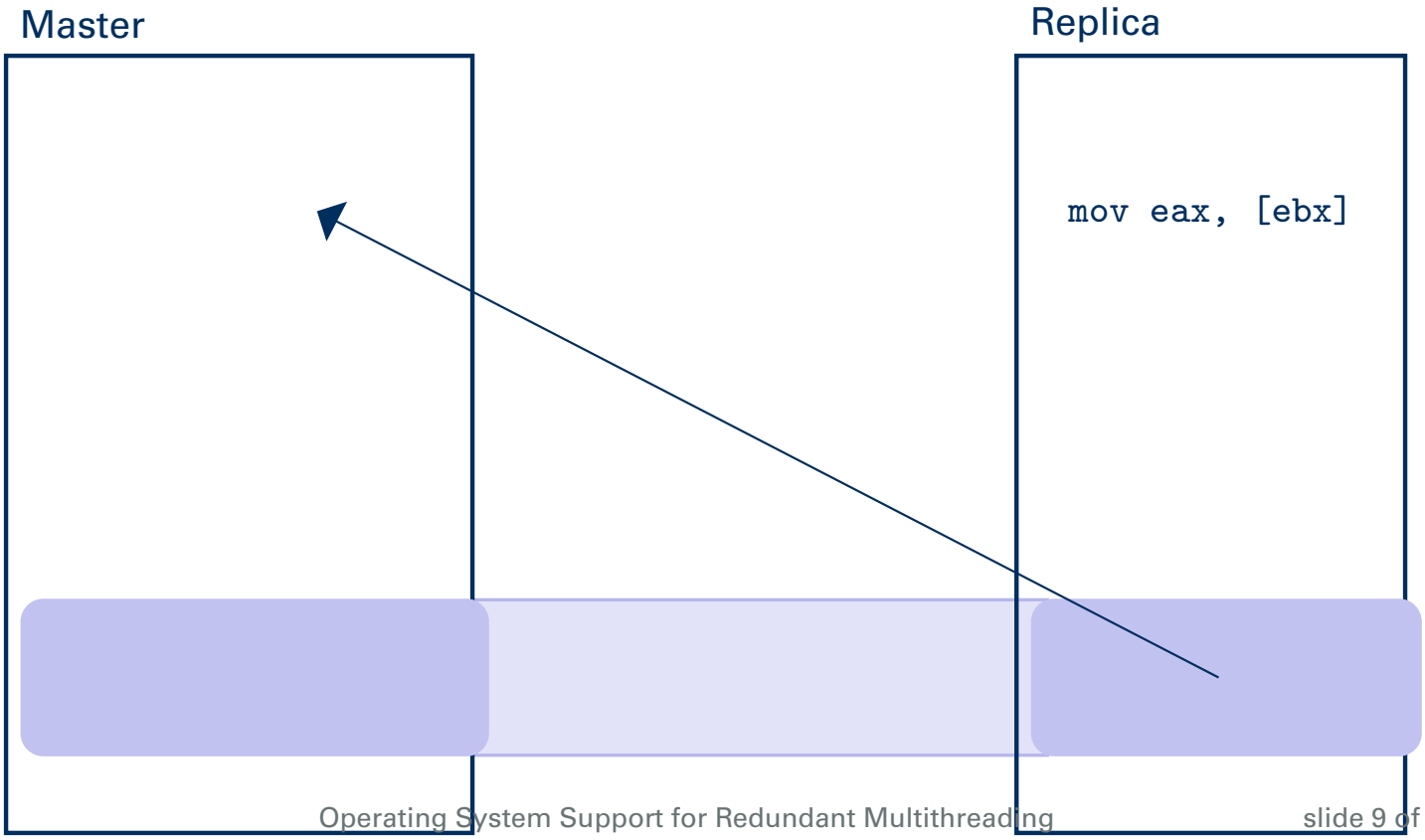
Copy&Execute



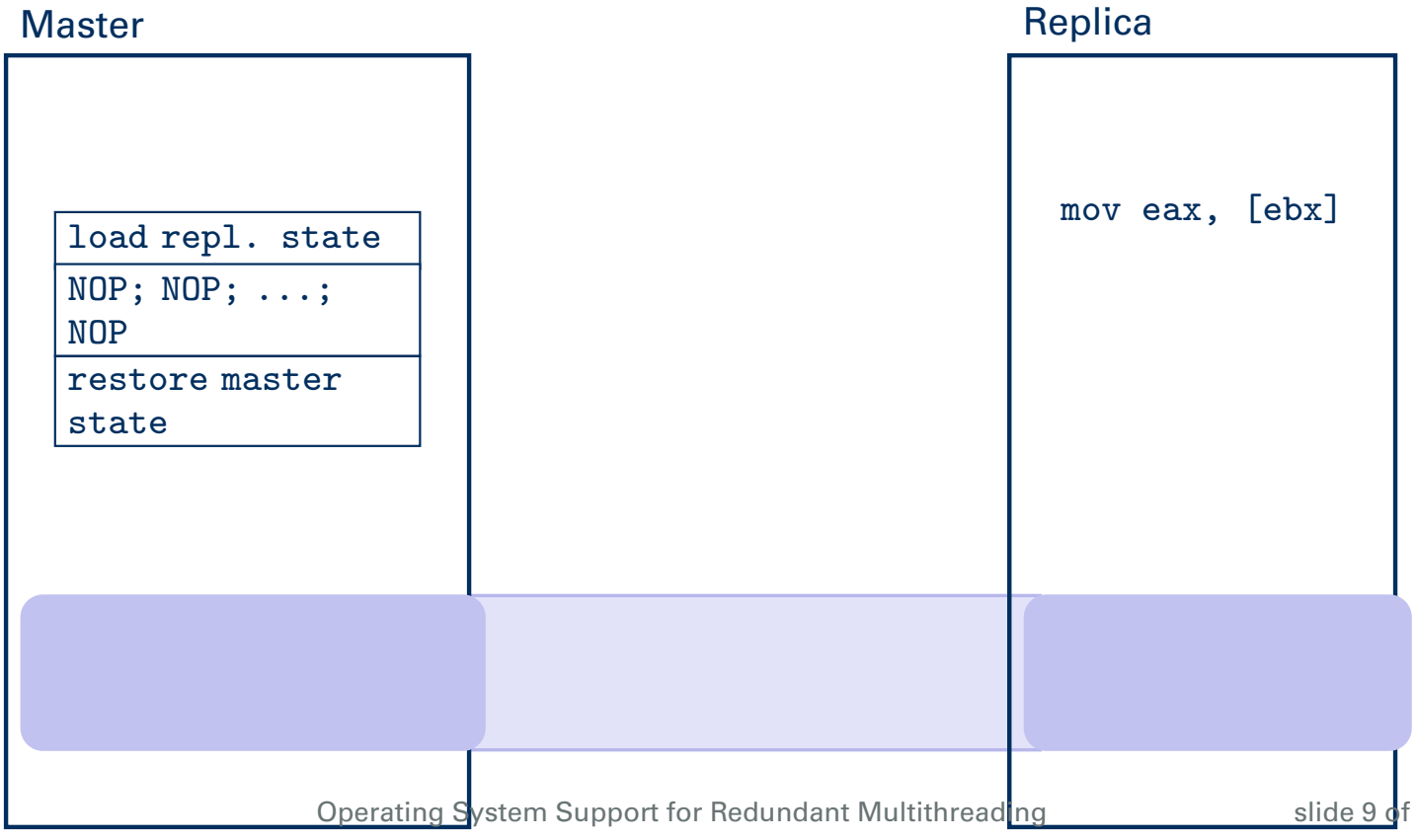
Copy&Execute



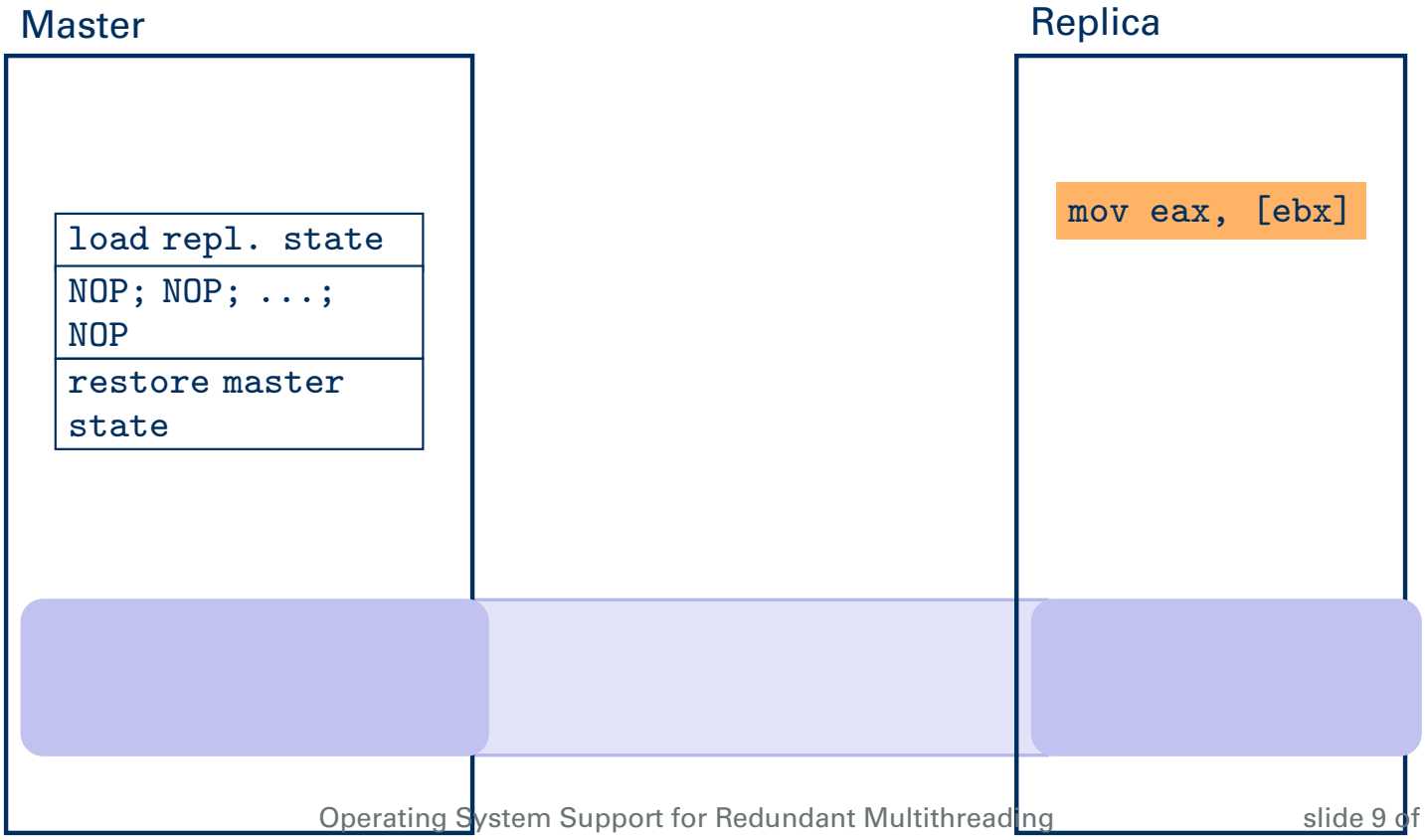
Copy&Execute



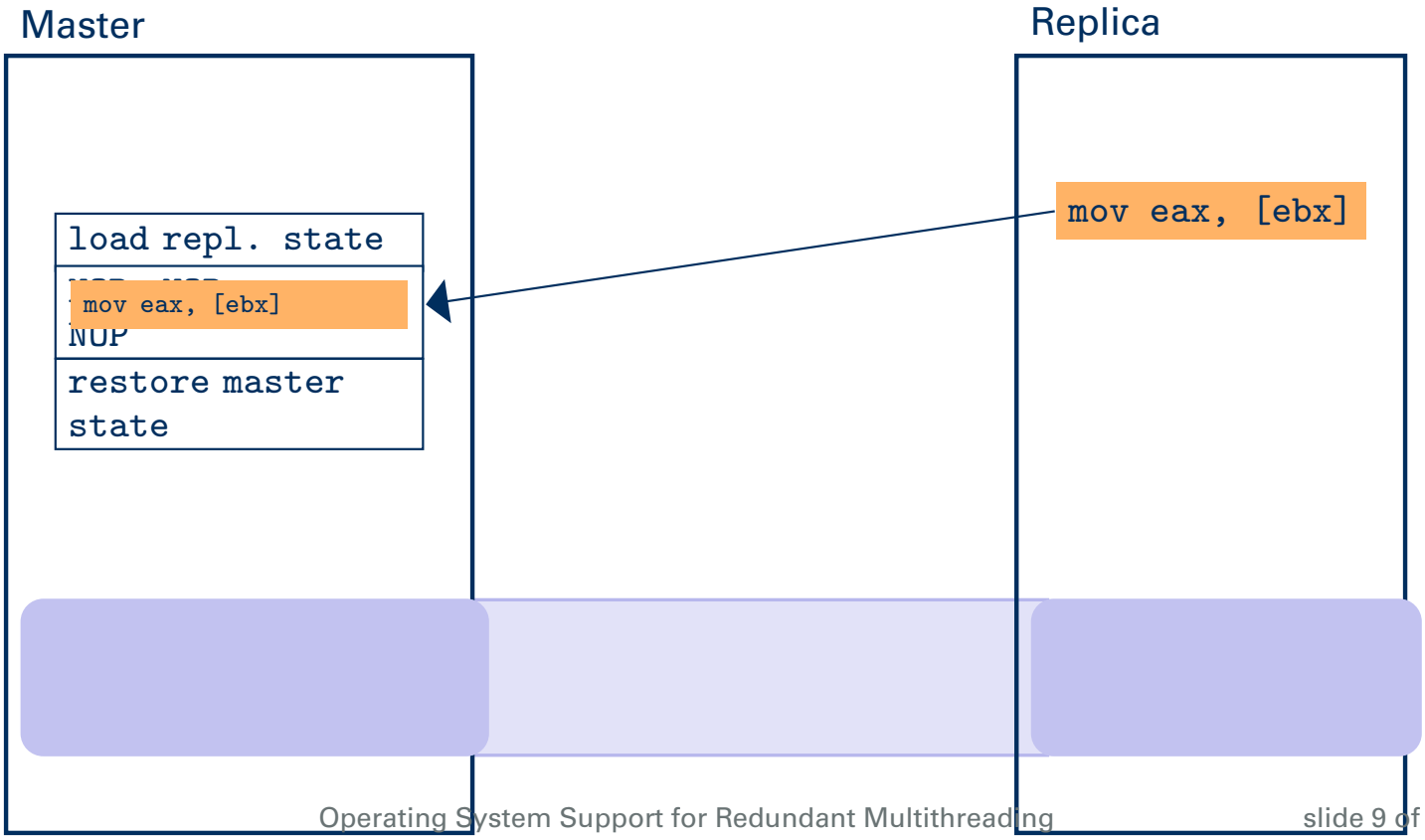
Copy&Execute



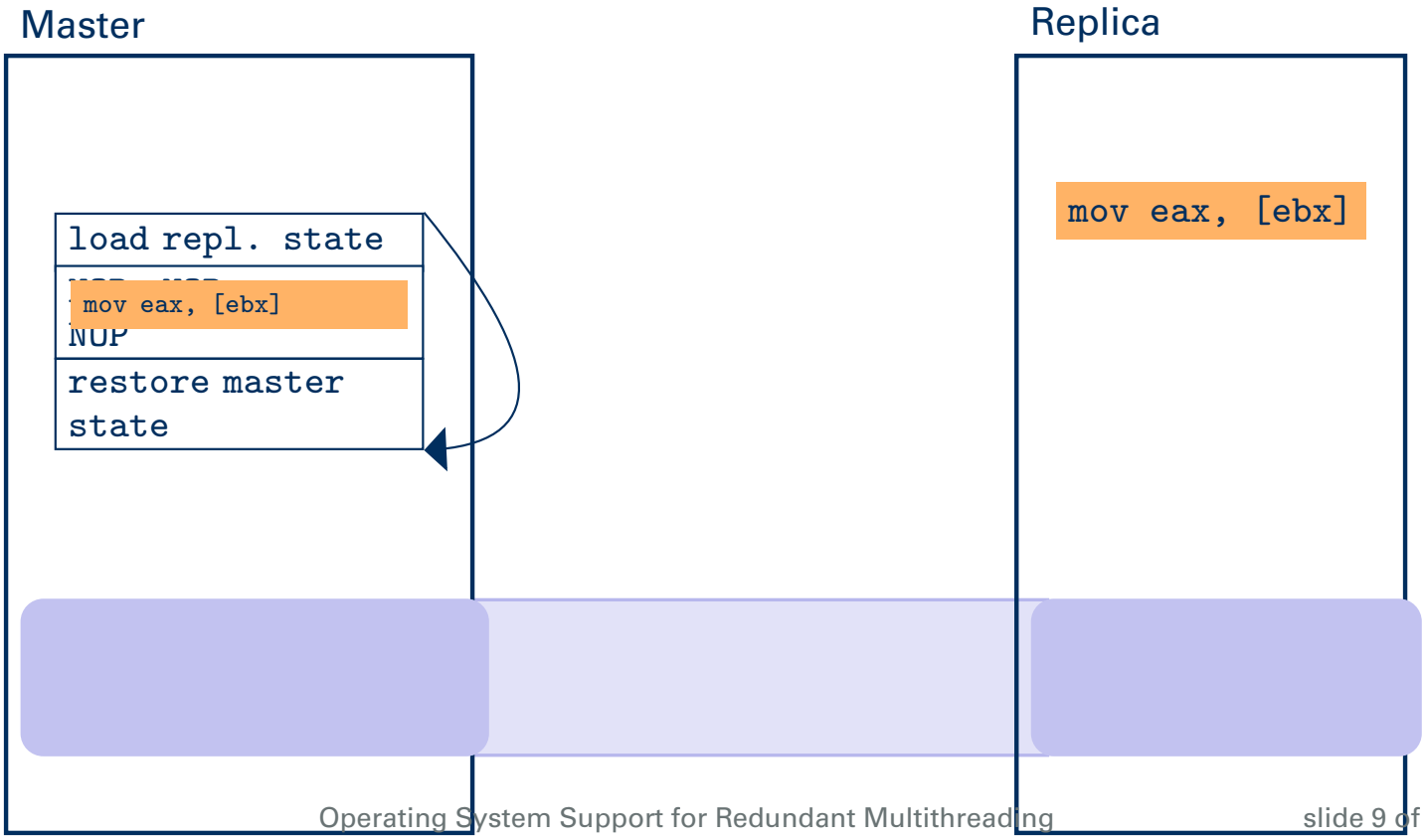
Copy&Execute



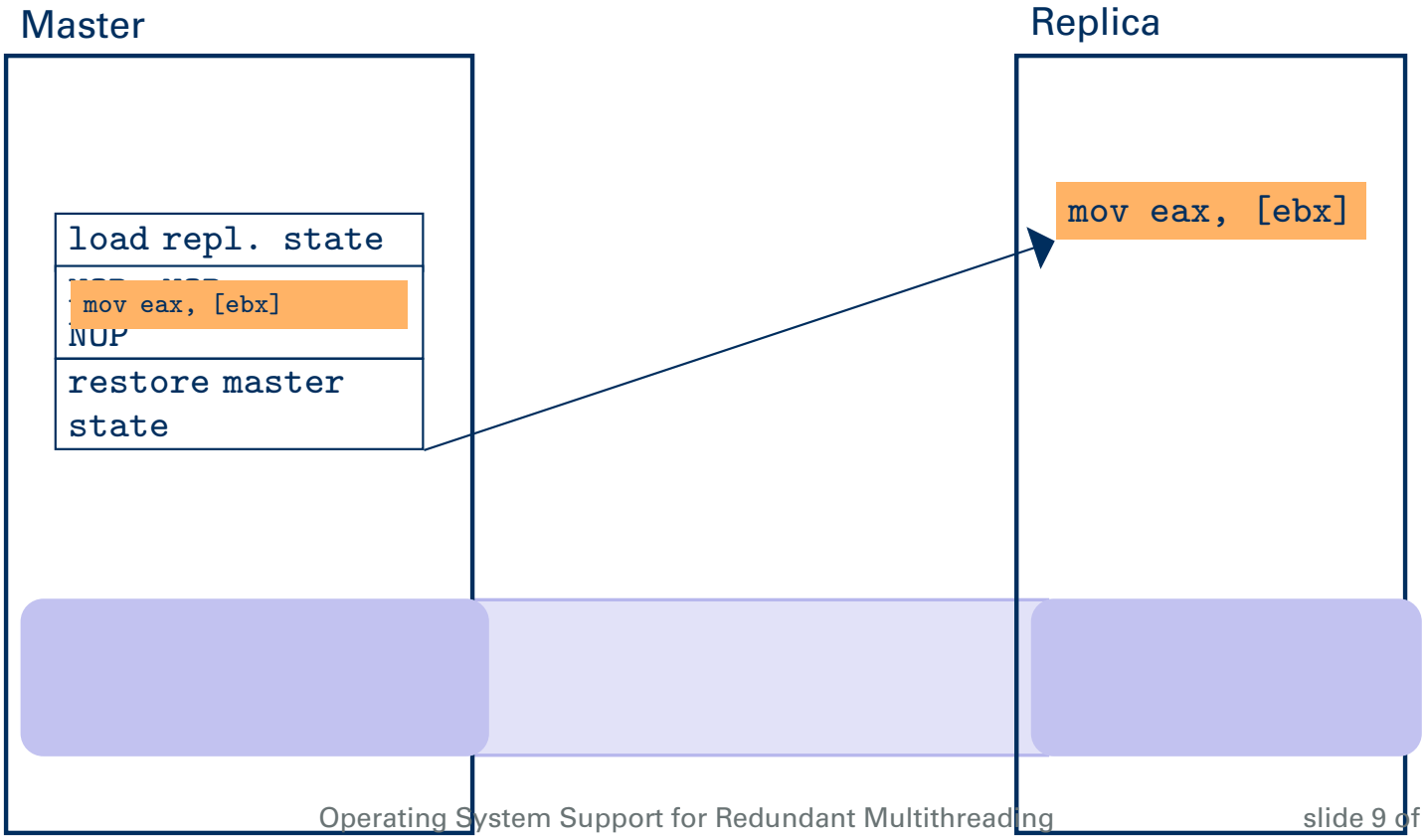
Copy&Execute



Copy&Execute



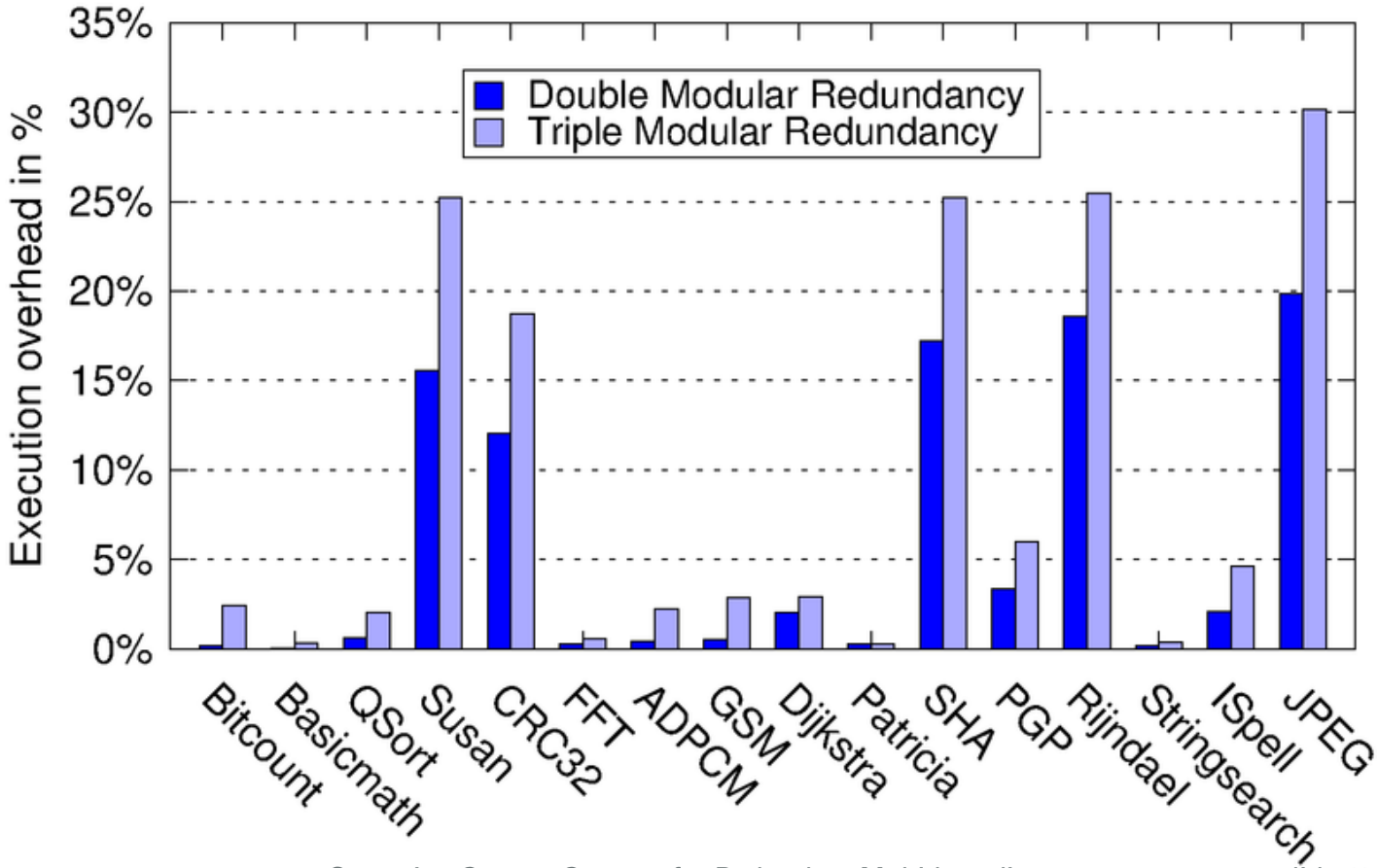
Copy&Execute



Benchmarks

- MiBench suite
- Fault injection to confirm fault distribution ratios
- Overhead for DMR and TMR
- Microbenchmarks for shared memory

Overhead vs. Unreplicated Execution



Romain Lines of Code

Base code (main, logging, locking)	325
Application loader	375
Replica manager	628
Redundancy	153
Memory manager	445
System call proxy	311
Shared memory	281
Total	2,518
Fault injector	668
GDB server stub	1,304

Conclusion

- Redundant Multithreading as an OS service
- Support for binary-only applications
- Overheads $<$30%, often $<$5%
- Shared memory handling is slow
- Work in progress:
  - Multithreading
  - Device drivers

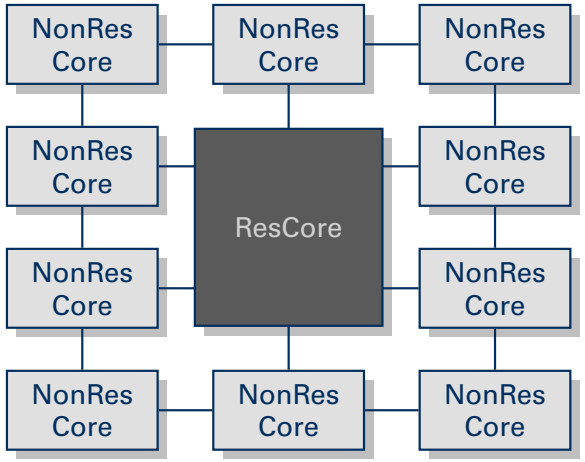
Nothing to see here

This slide intentionally left blank.

Except for above text.

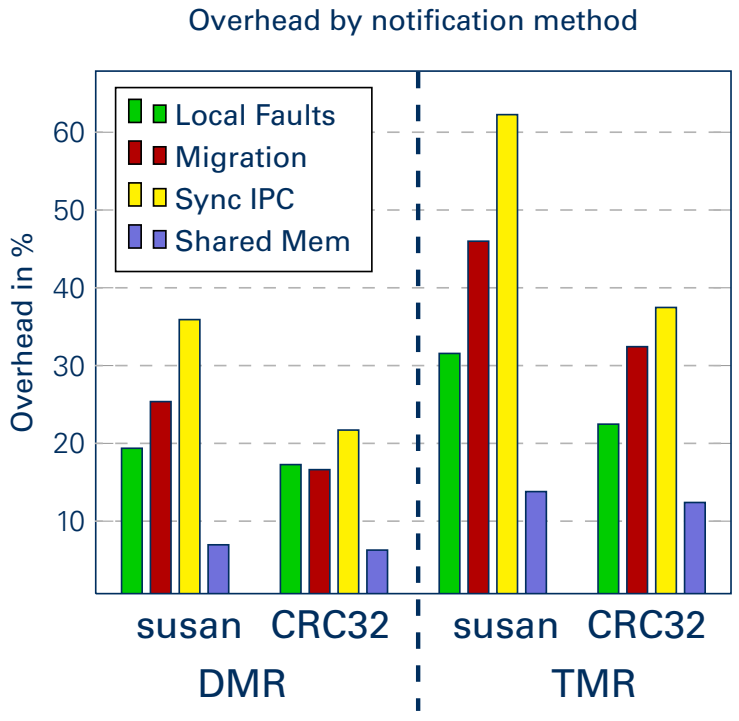
Hardening the RCB

- **We need:** Dedicated mechanisms to protect the RCB (HW or SW)
- **We have:** Full control over software
- Use FT-encoding compiler?
  - Has not been done for kernel code yet
  - Only protects SW components
- RAD-hardened hardware?
  - Too expensive
- **Our proposal:** Split HW into ResCores and NonRes-Cores



Signaling Performance

- Overhead compared to single, unreplicated run
- Benchmarks with highest overhead in EMSOFT paper
- Test machine:
  - 12x Intel Core2 2.6 GHz
  - Replicas pinned to dedicated physical cores
  - Hyperthreading off



What about signalling failures?

Missed CPU exceptions	→	detected by watchdog
Spurious CPU exceptions	→	detected by watchdog / state comparison
Transmission of corrupt state	→	detected during state comparison

Overwriting remote state during transmission

- NonResCore memory
- Accessible by ResCores, but not by other NonResCores
- Prevents overwriting other states
- Already available in HW: IBM/Cell

Romain

http://www.dynamo-dresden.de