P1796
Resilient Backplane Ring (RBR)
"congestion management"
2004Nov04b

David V. James
dvj@alum.mit.edu

Current status

- **Study group authorized by MSC January, 2004**
- **PAR approved June 24, 2004**
- **Scope:**
  - Resilient backplane ring (RBR) is a backplane interconnect based on the dual-ring resilient topology of resilient packet ring (RPR) and the 802 MAC addressing structure. RBR includes features appropriate for the low-latency backplane environment: destination-based flow control, low-power short-haul PHY, backplane-to-backplane links, transport of IEEE-1394 isochronous data, and support of IEEE-1596 memory-update operations.
- **Purpose:**
  - The purpose of this project is to leverage the benefits of network-compatible resilient interconnects within low-latency backplane environment.

Reasons for RBR

- High speed backplanes are oftentimes used within the networking environment, where designs can be simplified by sending network frames and card-to-card communications over the same links.
- Although the resilient packet ring (RPR) has the quality of service (QOS) needed for card-to-card communications, other facilities associated with a low-latency backplane environment are missing.
- When RPR like protocols are supplemented with latency-critical backplane services, the resulting backplane interconnect should be sufficient for many mixed application backplane designs.
- Affected sectors would include enterprise networking and computer server industries; perhaps 100s or hopefully 1000s of companies.

RBR's IEEE heritage

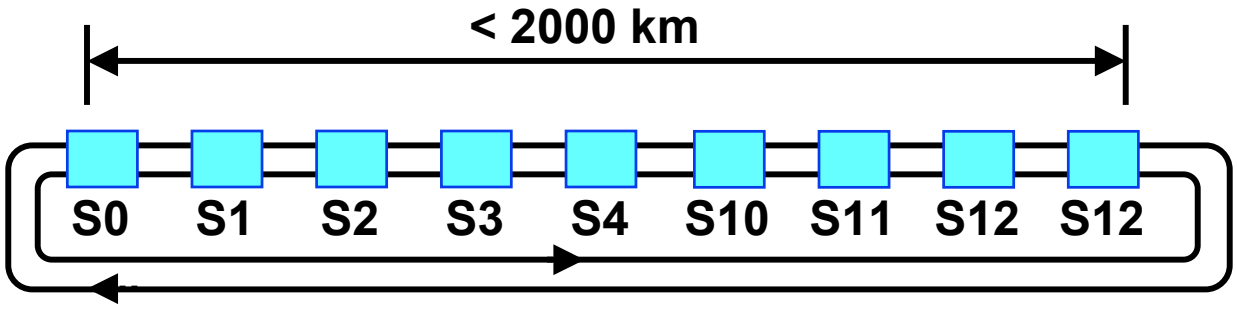
- IEEE Std 1212 – 1991 CSR Architecture
  Indivisible memory-mapped update operations
- IEEE Std 1596 – 1992 Scalable Coherent Interface (SCI)
  Busy-retry destination-based flow control
- IEEE Std 1394 – 1995 Serial Bus
  Isochronous path reservations, time-sync, and per-cycle transmissions
- IEEE Std 802.17 – 2004 Resilient packet ring (RPR)
  Scalable network-on-a-ring, classes of service, resiliency
- IEEE 802.3ap Backplane Ethernet Task Force
  Physical layers for the backplane (PHYs)
- IEEE 802 (CE) Study group
  Isosynchronous path reservations, time-sync, and frame formats

Similar industry technologies

- Infiniband
- HyperTransport
- PCI-express
- Rapid I/O
- Others?
  - Fiber-channel, serial ATA, serial SCSI, FDDI

RBR summary

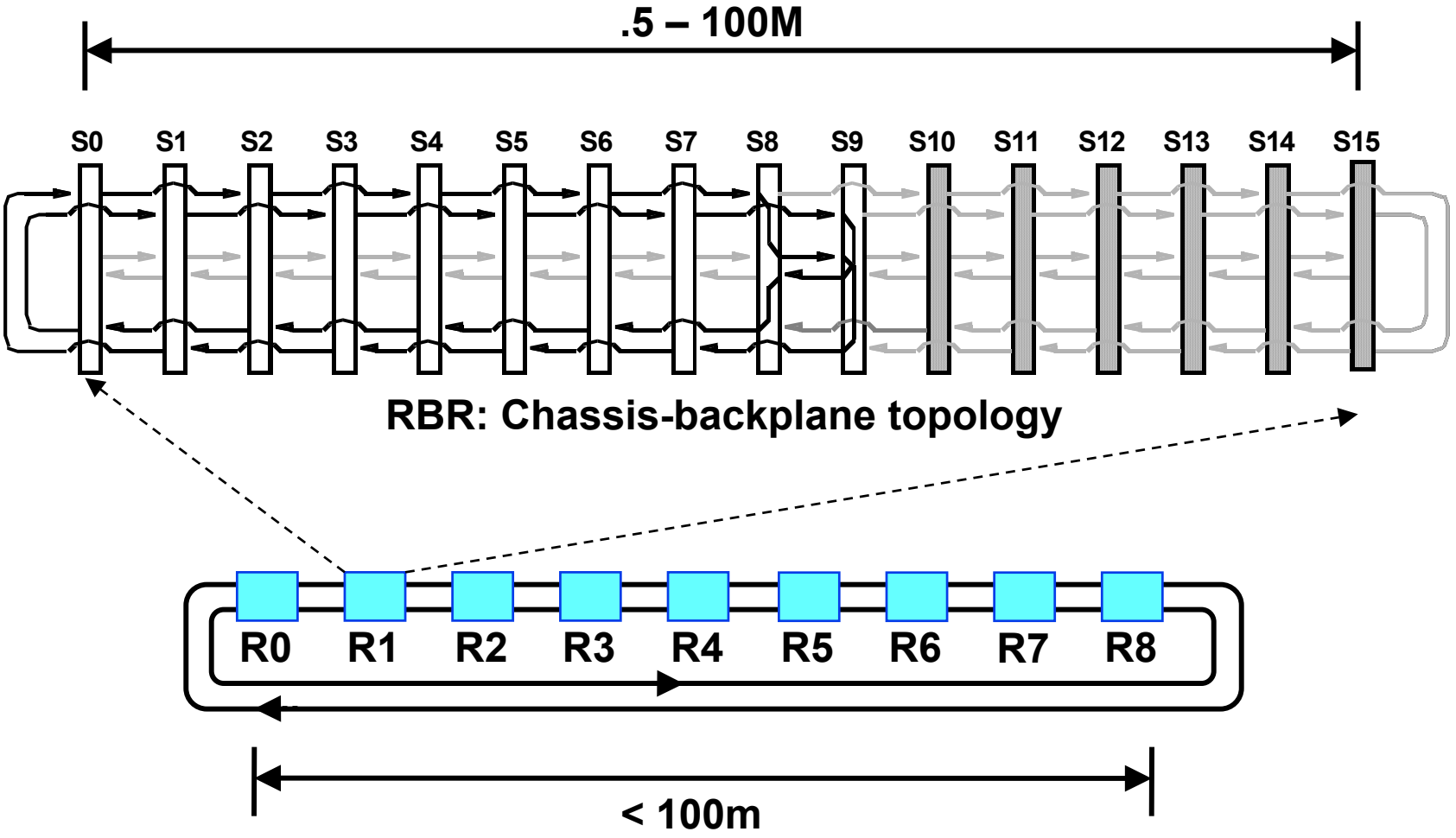
RPR topologies



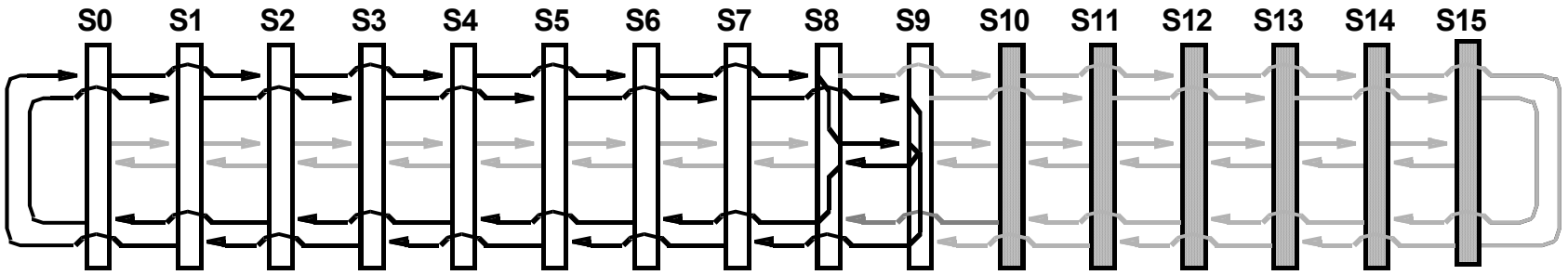
RPR: Metro-area topology

SONET environment applications
Duplex counter-rotating rings with spatial reuse
IEEE 802 frames, with ring-routing supplements
Several product-in-field constraints

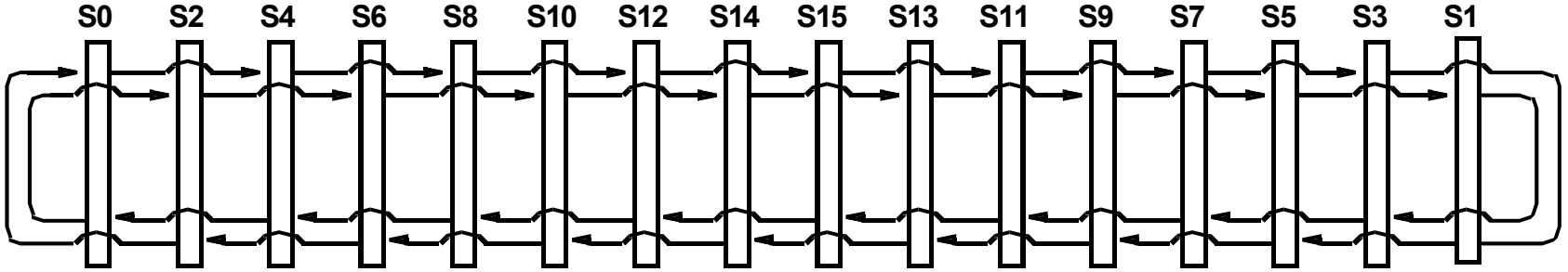
Hierarchical topologies



Topology equivalents



Physical chassis-backplane topology

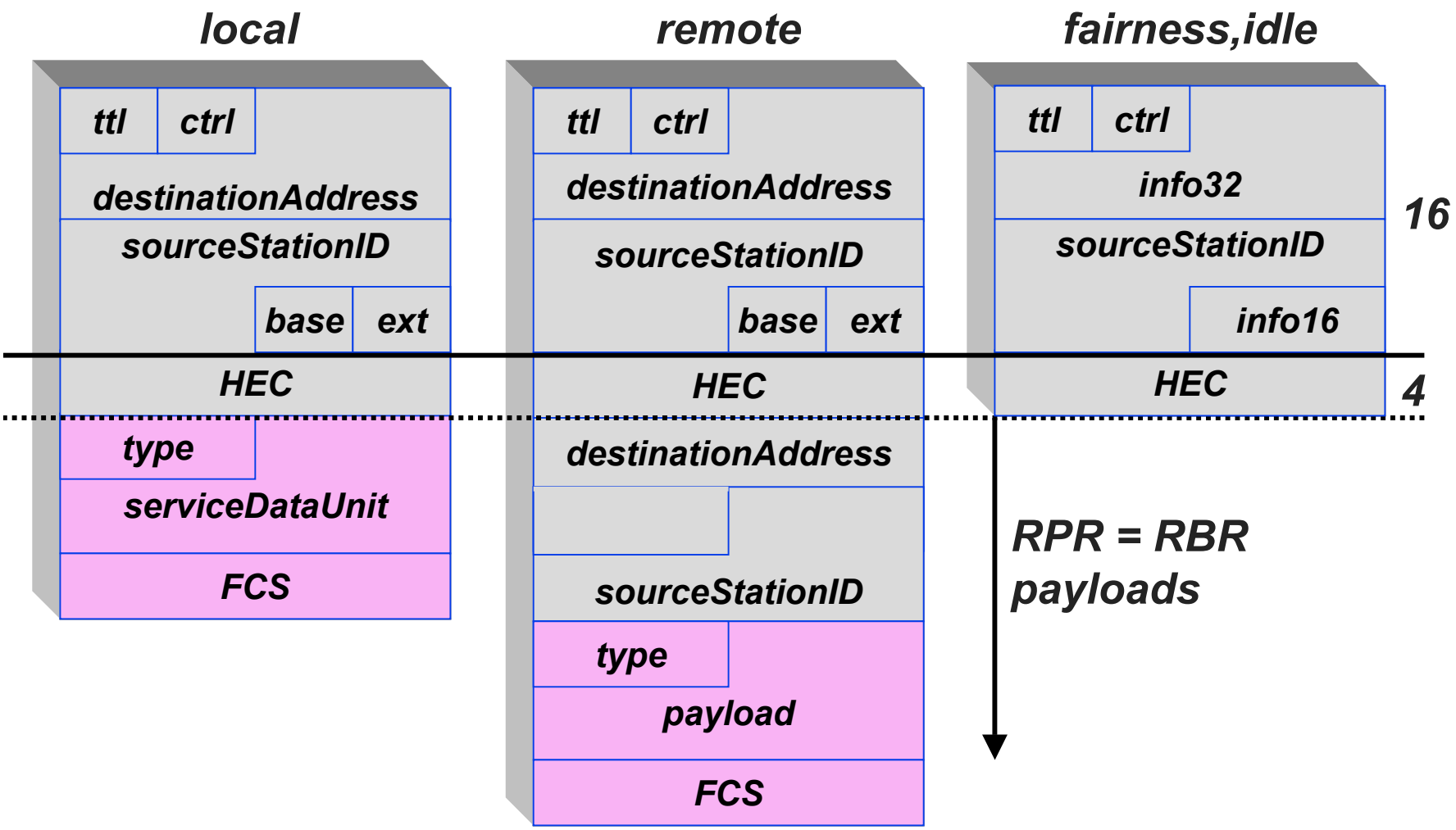


Logical chassis topology

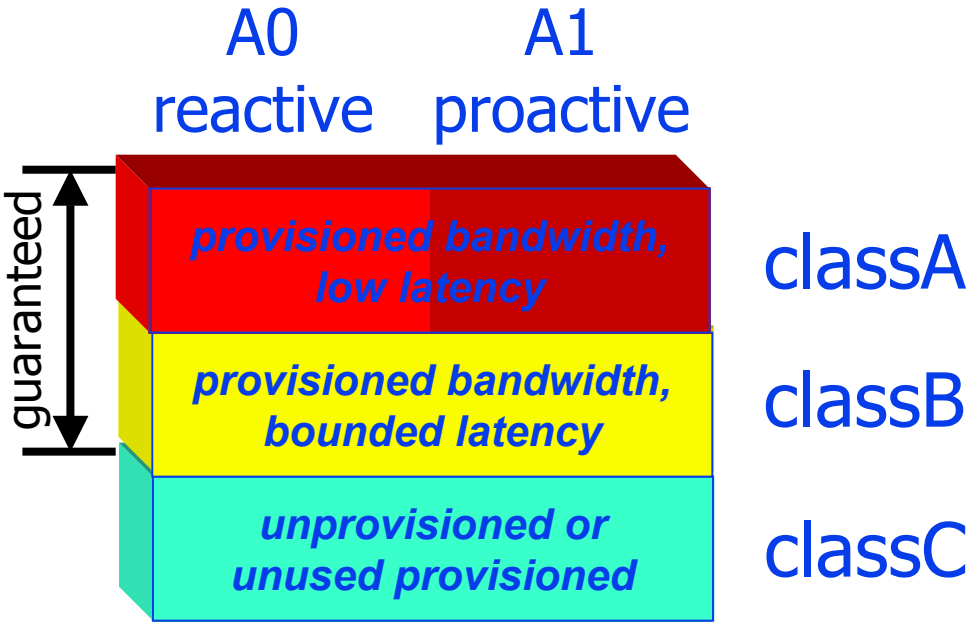
RBR protocol summary

- Leveraged RPR values:
  Ethernet frames with QOS delivery
  Ring efficiency and resiliency
- QOS enhancements
  Accurate time-of-day synchronization
  Revised/verified classA1/classB guarantees.
  Quasi-synchronous isochronous transfers
  Negotiated access controls.
- Lossless transactions
  Destination-asserted flow control
  Hard-coded memory-access commands
  Request/response queuing options
- Backplane PHY definitions

RBR format summary

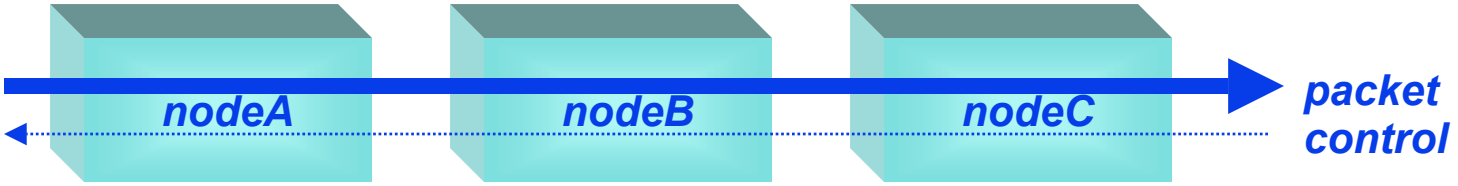


Arbitration classes



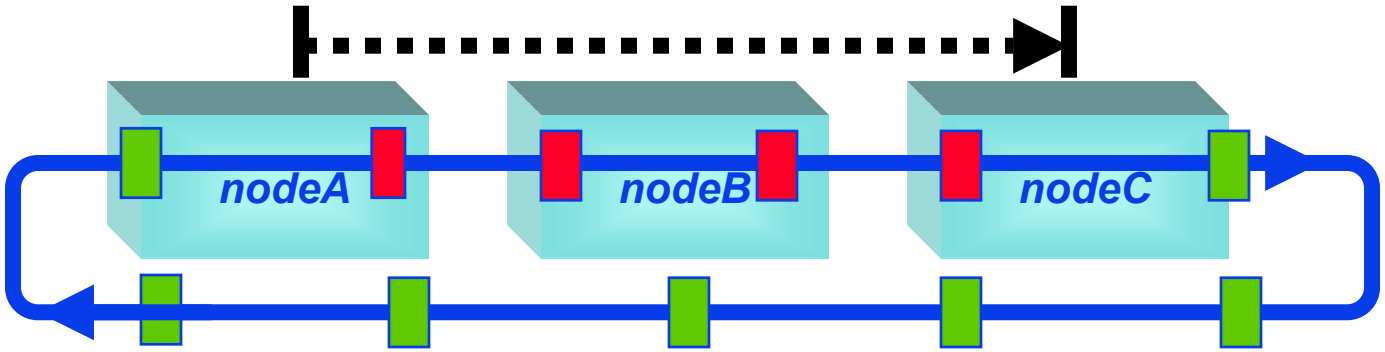
Flow control

Opposing arbitration



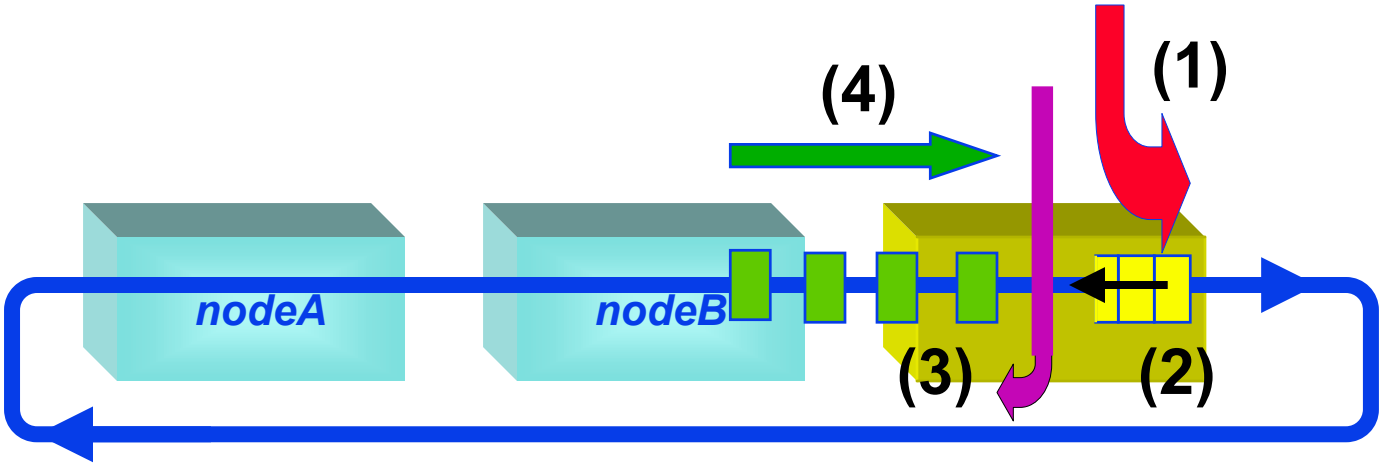
- Data packets flow in one direction
- Arbitration control flows in the other*

Proactive class-A0 partitions



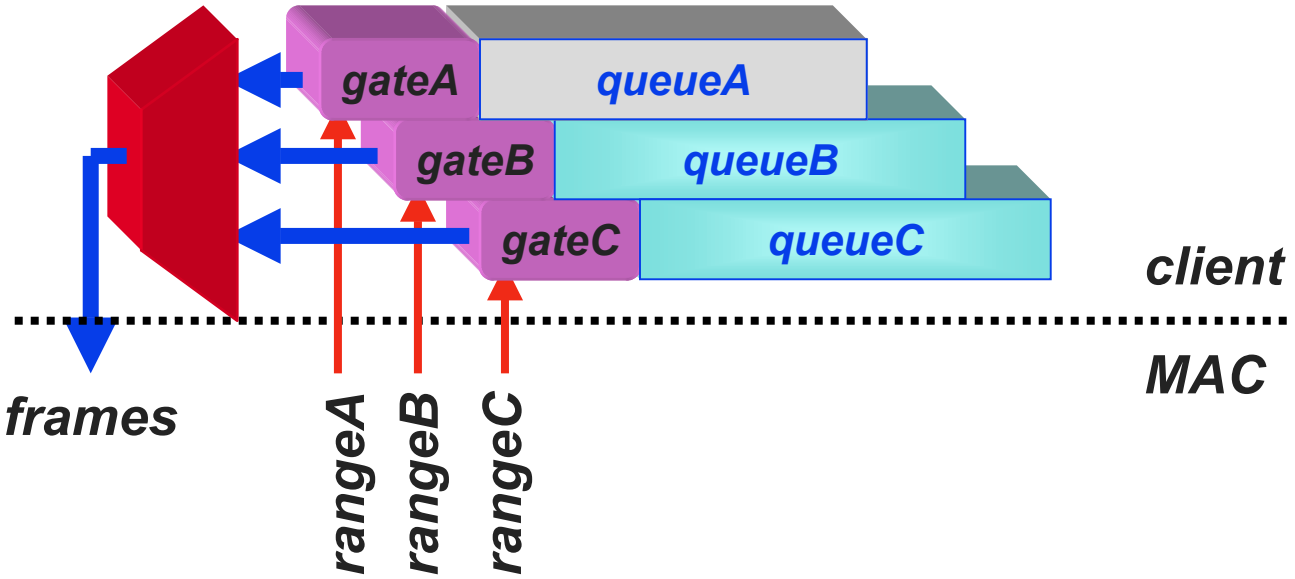
- Data packets go source-to-destination
- Residue returns destination-to-source to provide subsistence for transmissions

Reactive class-A1 control

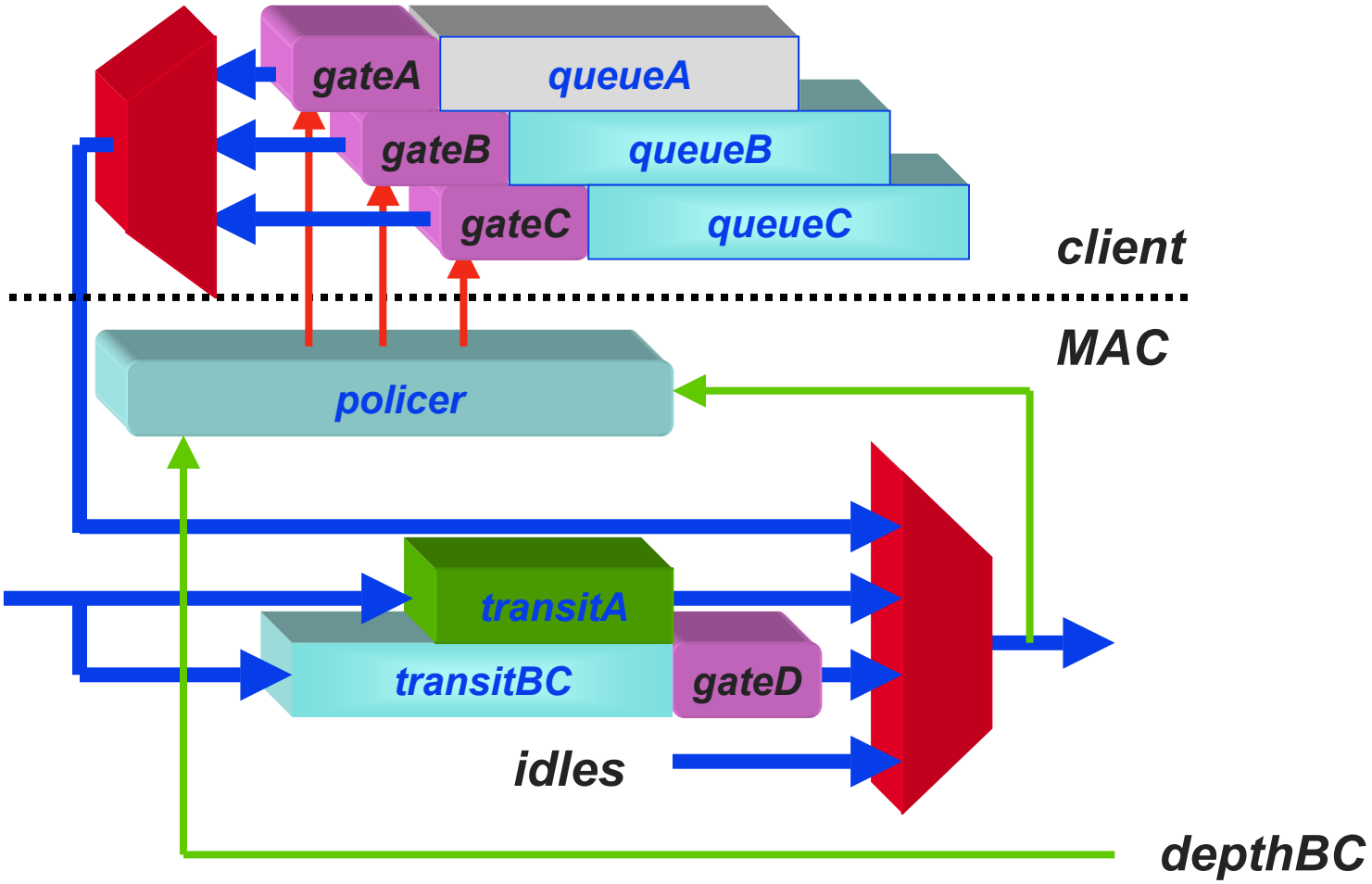


- Transmission of packets causes
- Backup of passBC FIFO that
- Returns flow-control information that
- Provides consumable idle packets

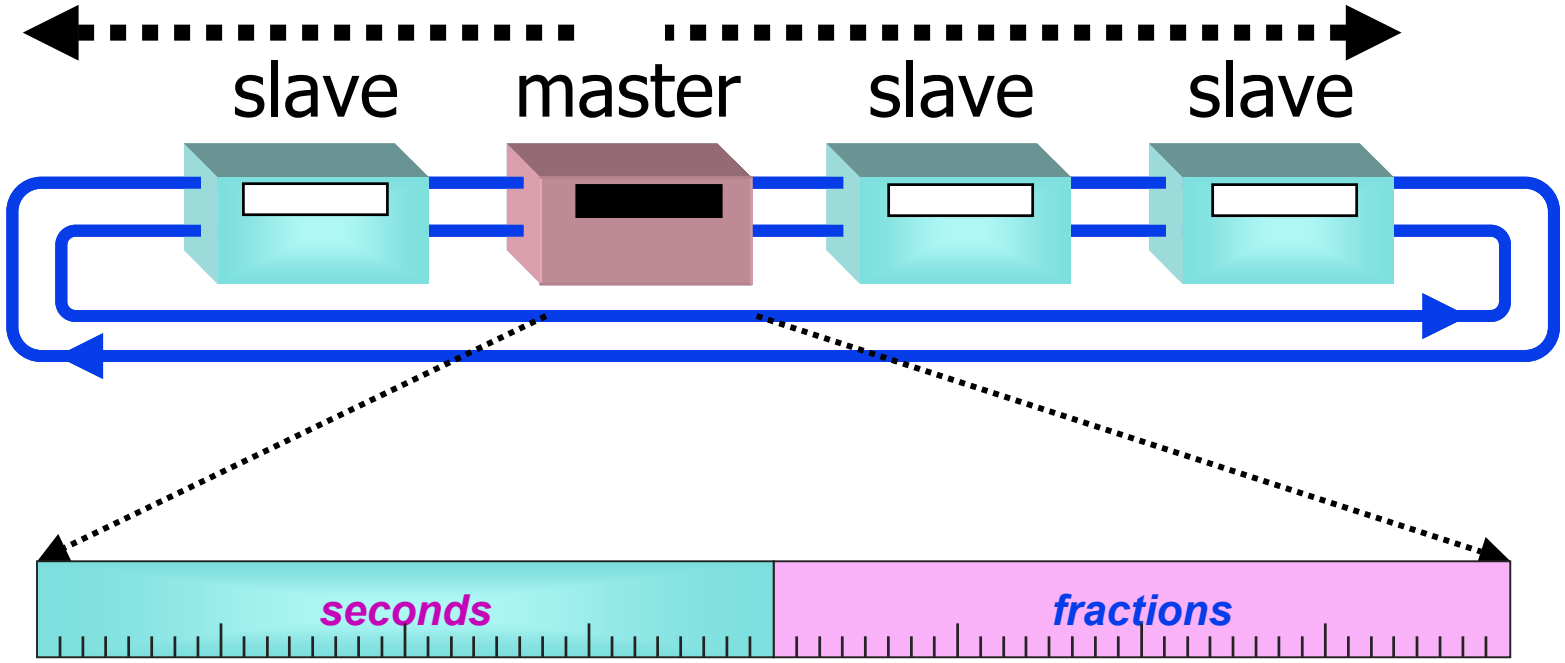
MAC-Client interface signals



Arbitration components



Time-of-day synchronization (not bit-clock synchronization)



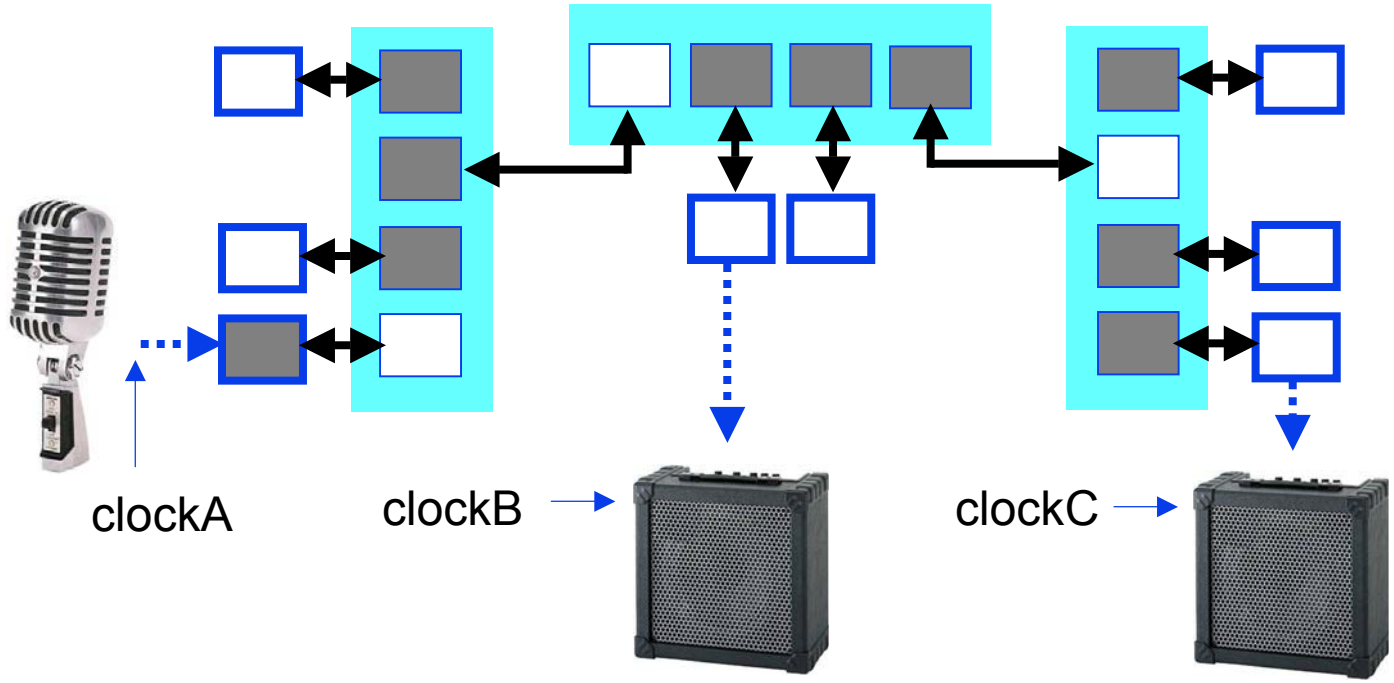
Difficult remaining problems

- Classes of service
  - Tight classA latency guarantees
  - Unconstrained classB levels
- Destination-based flow control
  - Busy retry has the right properties
    per-source feedback is simple
    output-port feedback is possible
  - Overhead must not exceed 1 retry

Remaining difficult problems

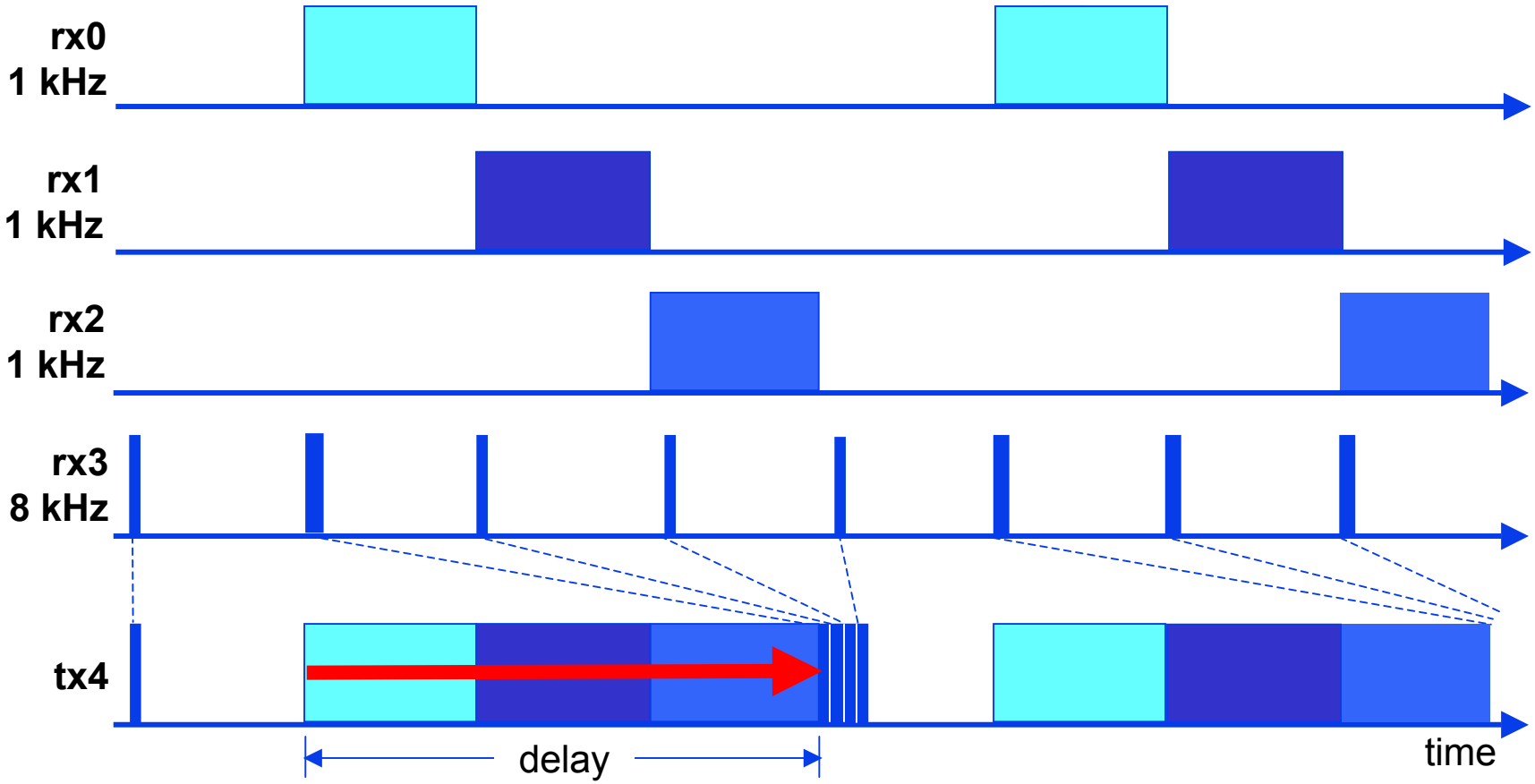
Guaranteed classA service (latency & bandwidth)

Synchronized presentation

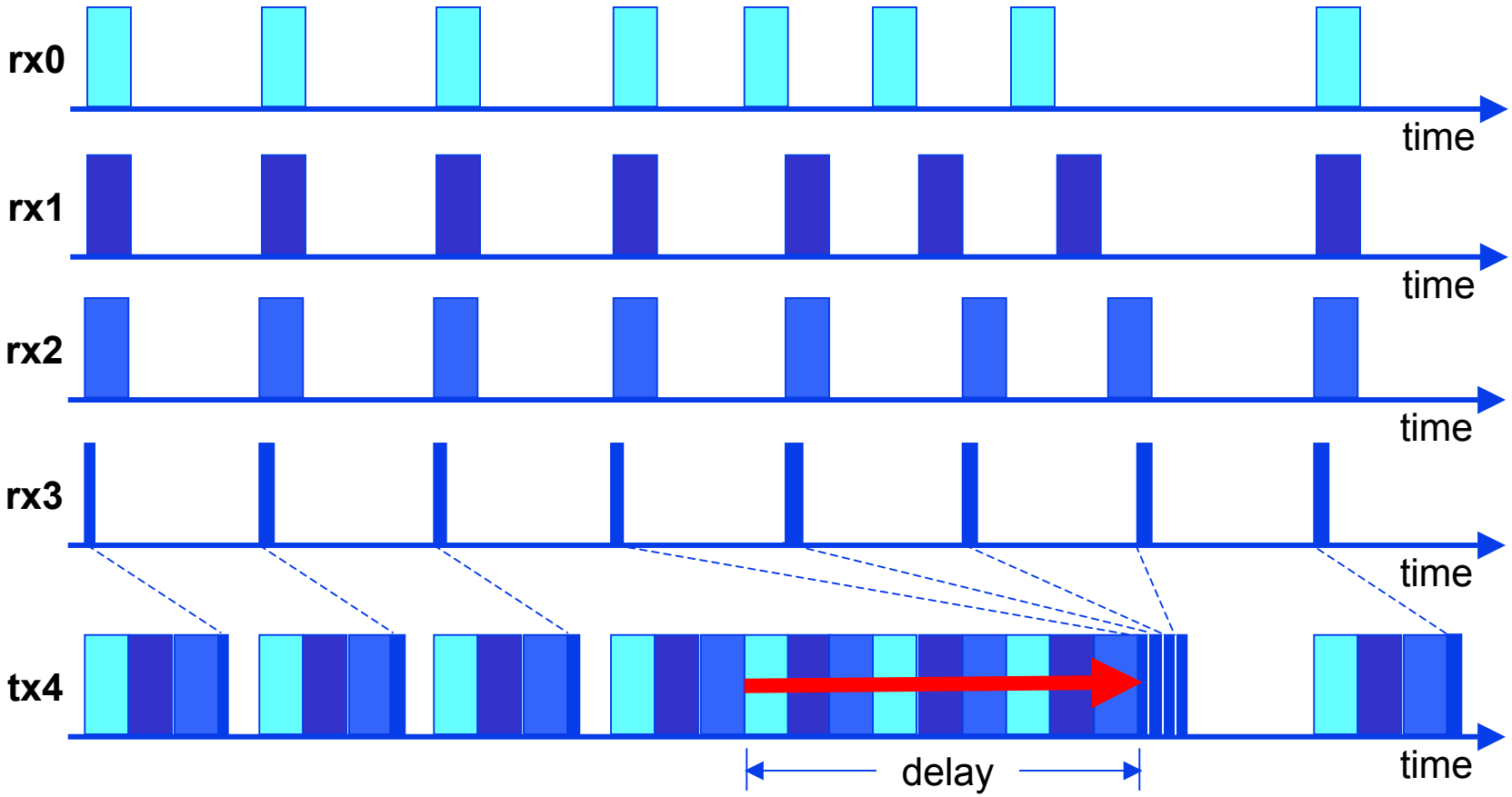


No long-term drift: clockA, clockB, clockC
Clock jitter: sub nanosecond (after PLL)

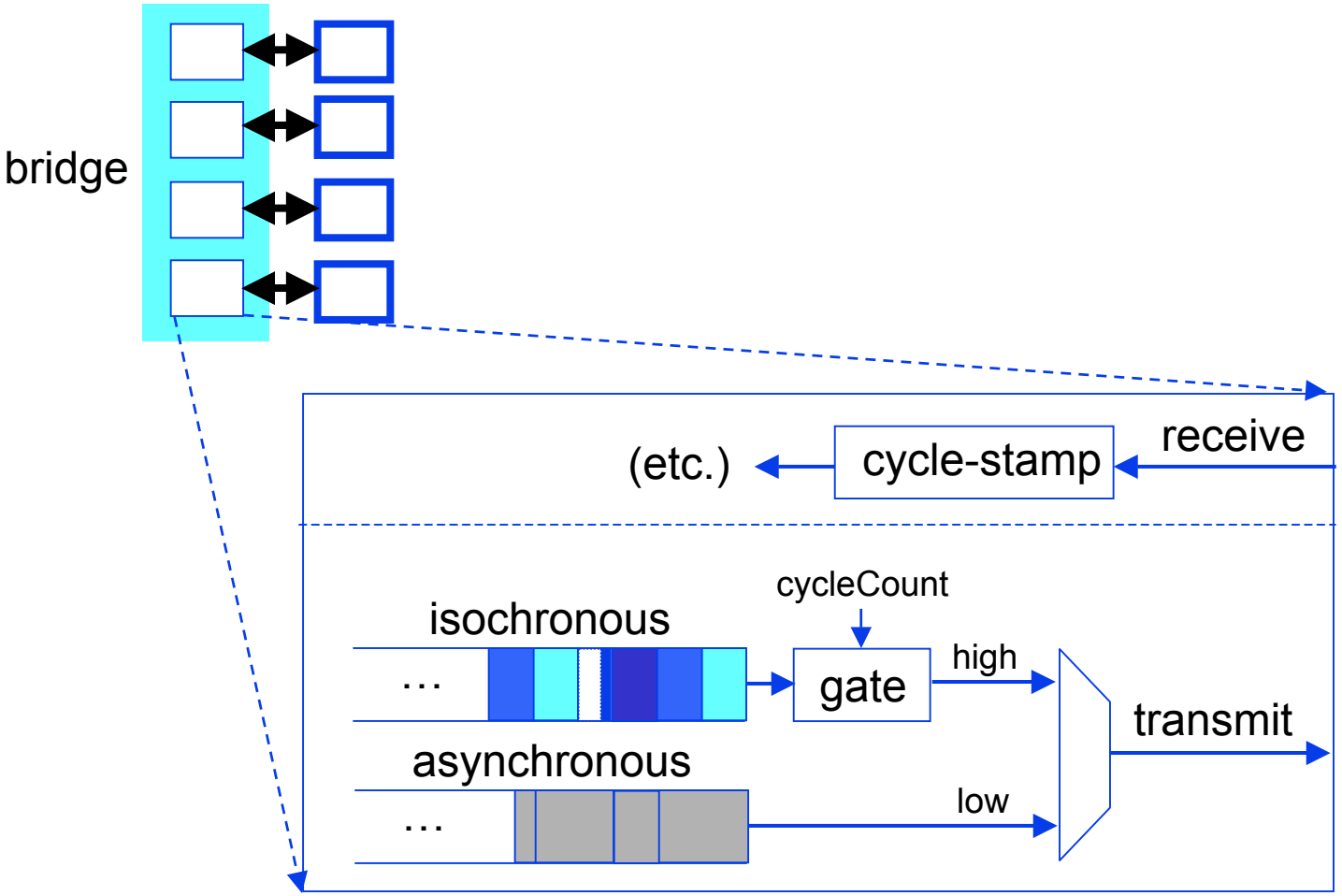
Bursting causes jitter



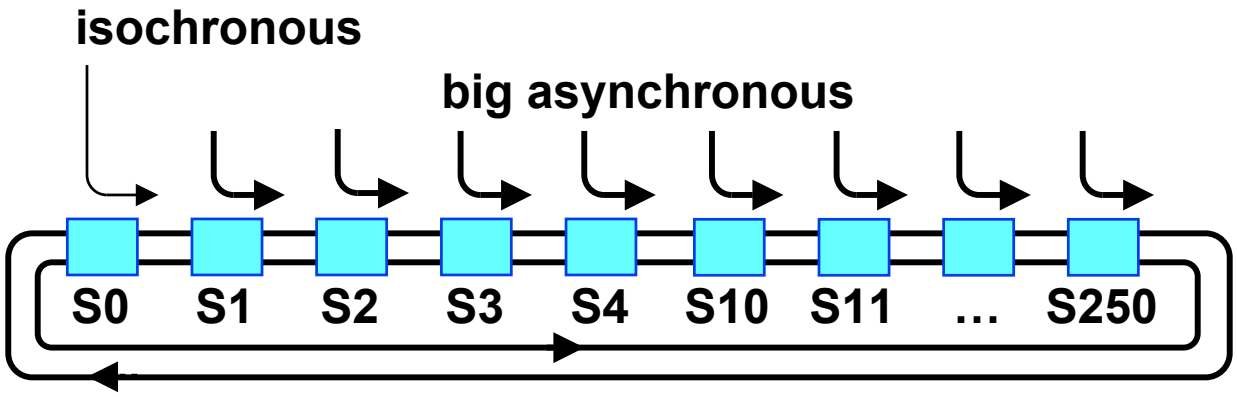
Bunching causes jitter



Bridge re-clocking limits jitter



Worst-case isoch delays



collisionDelay = stations * MAX_SIZE;

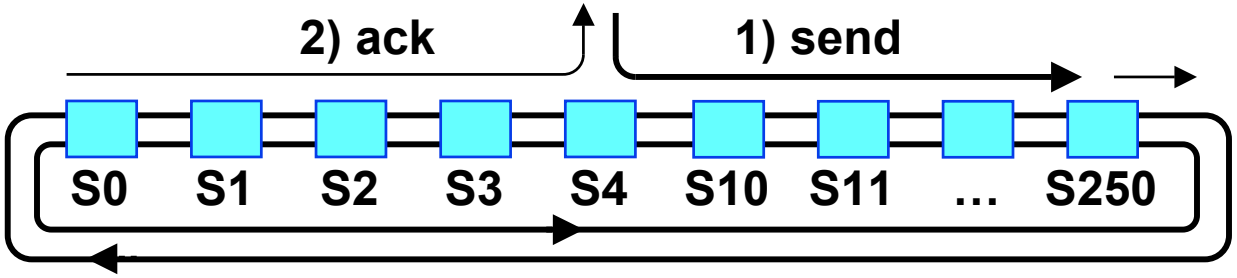
1.5kB @ 1Gb w/250 stations➔ 12us * 250 ➔ 3 ms

8.5kB @ 1Gb w/250 stations➔ 68us * 250 ➔ 17 ms

Remaining difficult problems

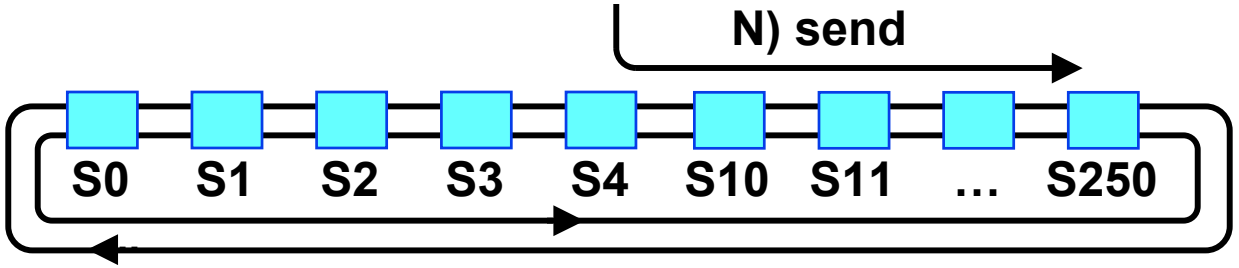
Guaranteed delivery
(classB & classC)

Destination-based flow



(A) Initial try

[(B) TBD signaling]



(C) Final retry

When compared to PAUSE

- Link granularity
  - Based on source/destination
  - Can be class dependent
- Output queue dependent
  - Possible, since destination knows
  - But, knowledge may be inconvenient.

Summary

- Classes of service
  - Not useful unless guaranteed
  - A small number is sufficient
- Guarantees are either:
  - Latency and bandwidths
  - Lossless delivery