

An Implementation of a Reliable Message Broadcast for the CMS Event Builder System

**I. Suzuki, M. Litmaath, V. O'Dell,
S. Pavlon, K. Sumorok**

Fermi National Accelerator Laboratory, USA

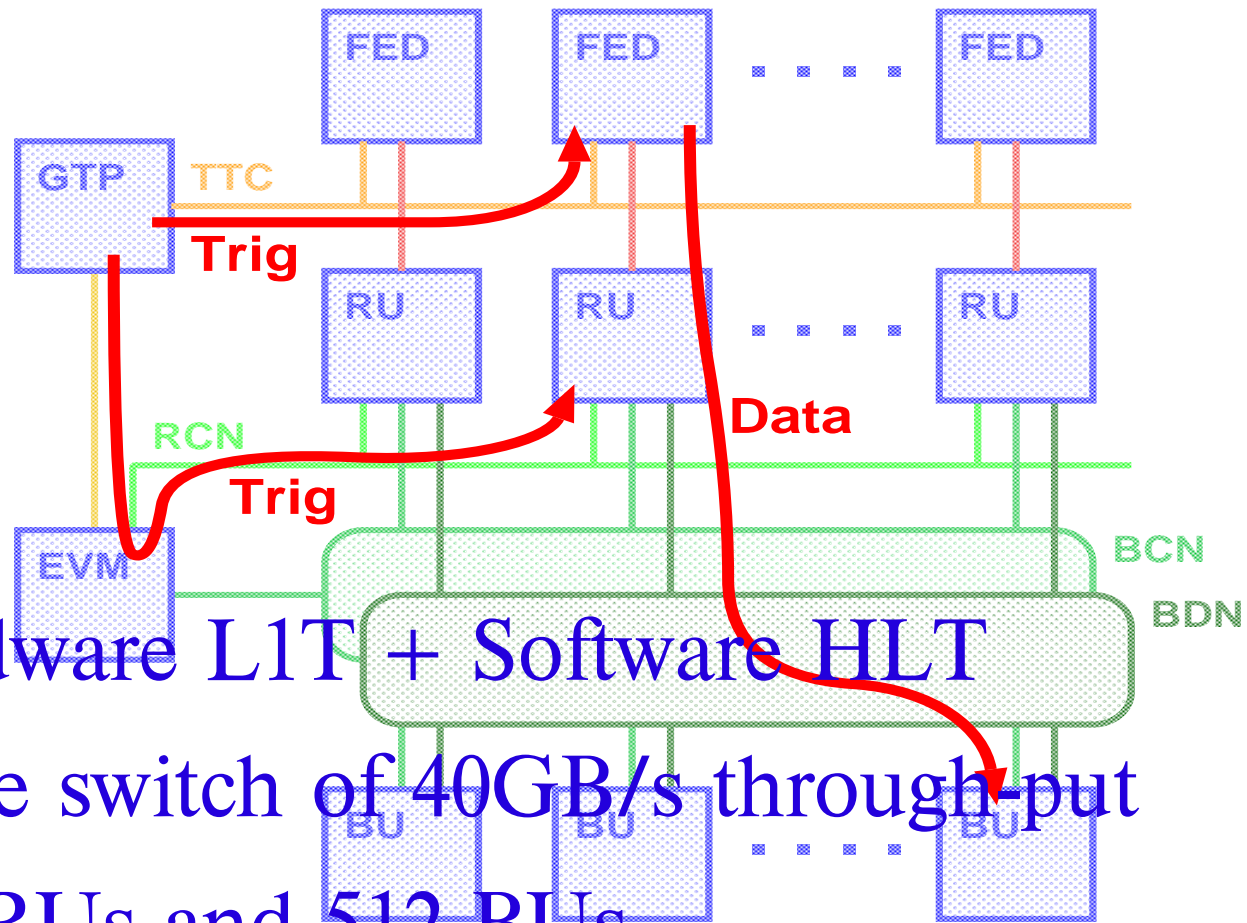
Massachusetts Institute of Technology, USA

2000/02/09 CHEP2000, Padova

Contents

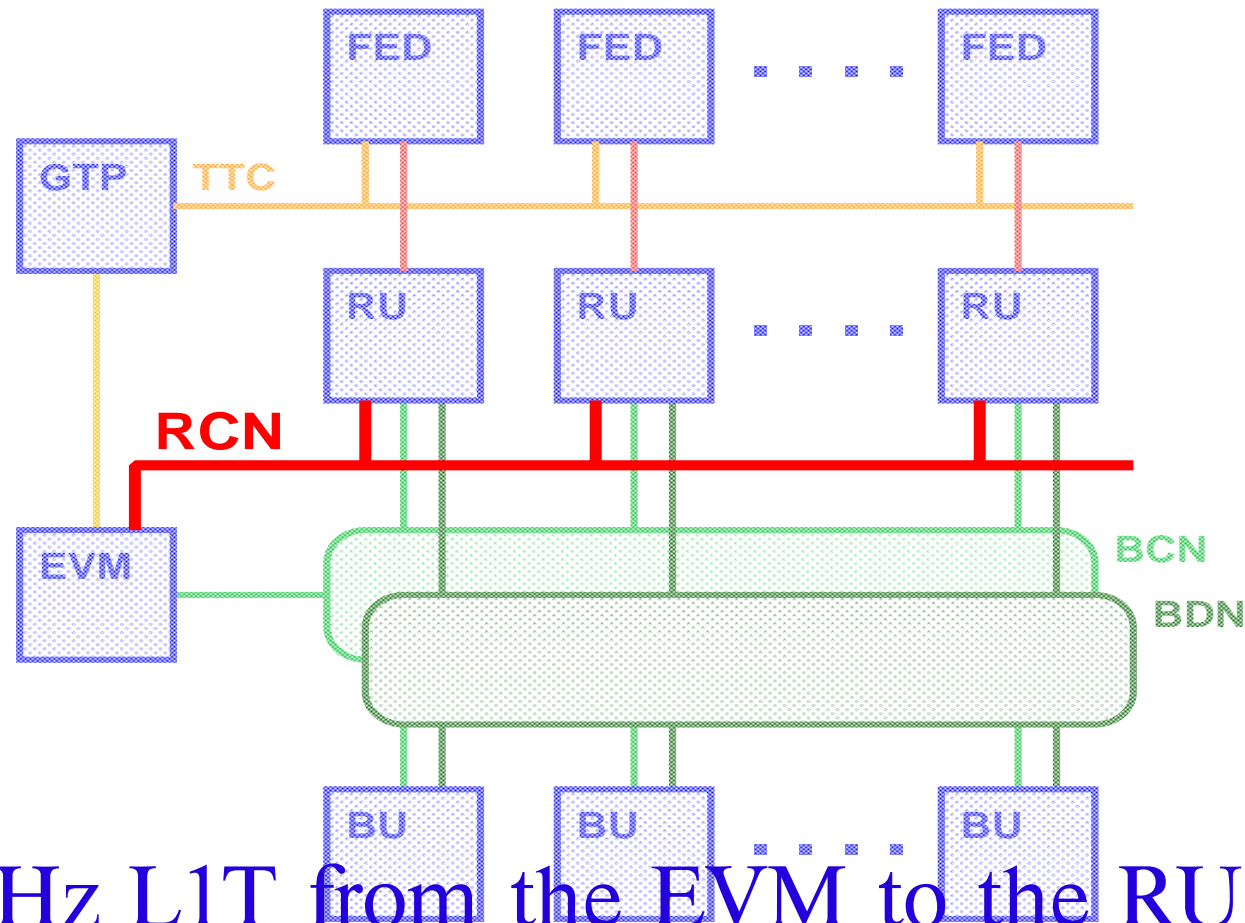
- Introduction
 - CMS data acquisition system
 - Readout Control Network
- Reliable broadcast protocol
 - Protocol details
- Implementation
- Benchmark tests
- Summary

CMS Data Acquisition System



- Hardware L1T + Software HLT
- Large switch of 40GB/s throughput
- 512 RUs and 512 BUs
- EVM controls/monitors the event flow

Readout Control Network



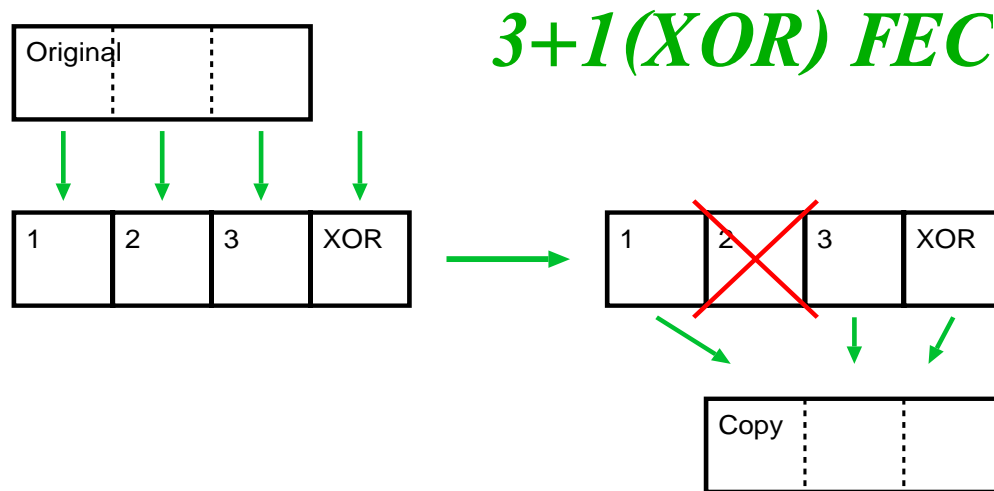
- 100kHz L1T from the EVM to the RUs
- Trigger throttling information from the RUs

Requirements of the RCN

- Network
 - ➔ Broadcast
 - ➔ Bandwidth: 10Mbps
 - ➔ Latency:
 - ★ Front-End buffer depth $\sim 10\text{ms}$
 - ★ $O(\text{ms})/(\text{a packet of } \sim 100 \text{ triggers})$
- Protocol
 - ➔ Total reliability

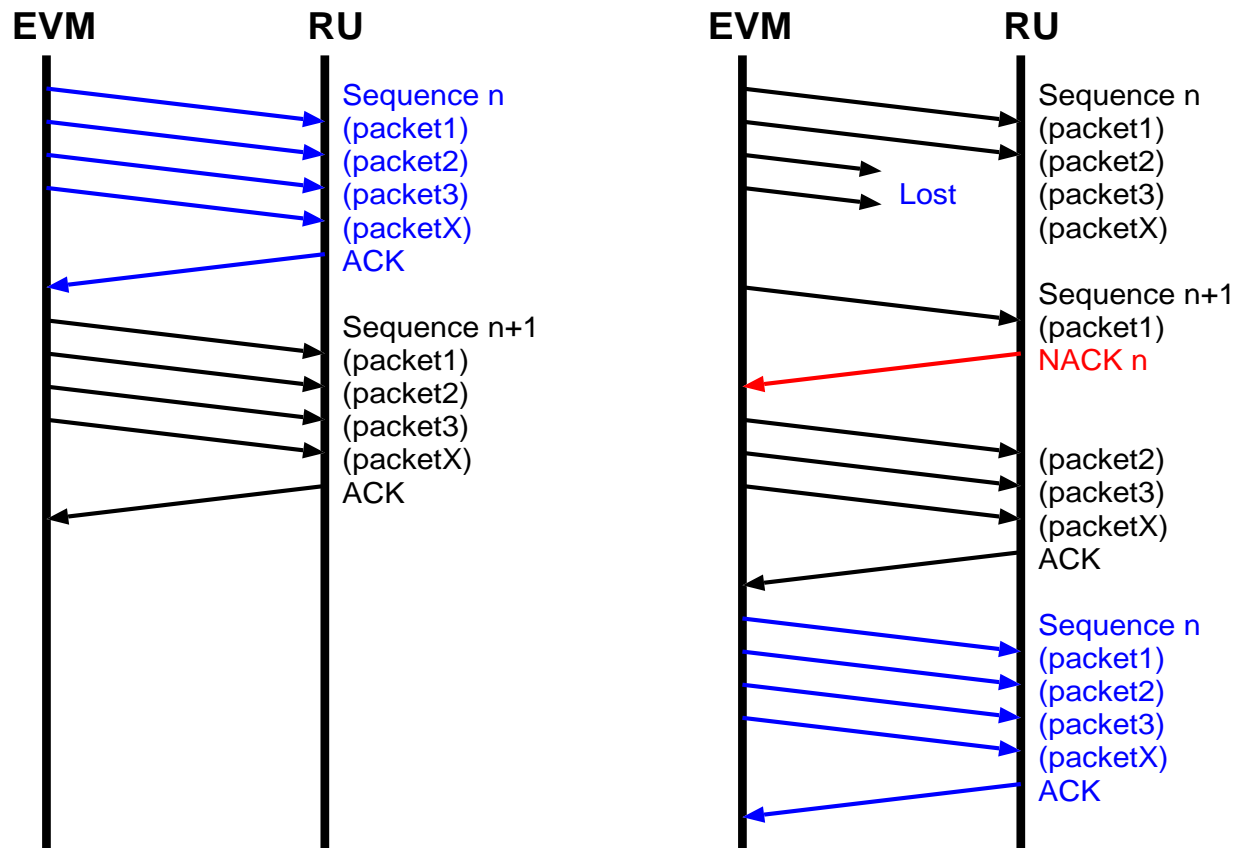
Reliable Broadcast Protocol

- NACK based
- Sequential packet ID + CRC
- Forward Error Correction
- Traffic regulation with ACK packets



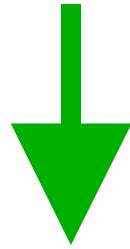
Protocol (cont'd)

- In case of the 3+1(XOR) FEC

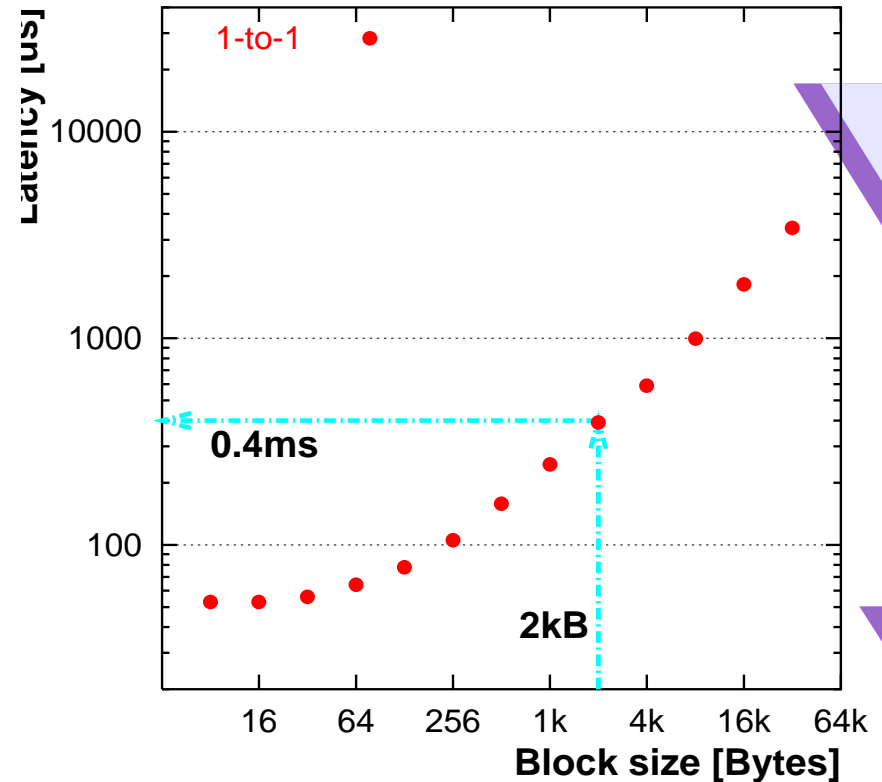


Test System

- To implement/understand the protocol with minimal effort

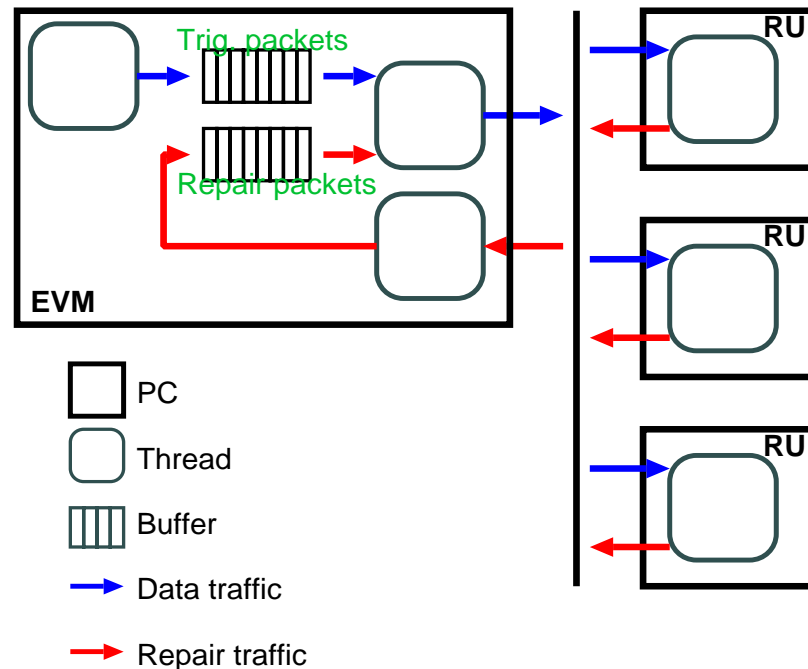


- PCs + FastEthernet SW
- Linux 2.2.12
- UDP/IP



Implementation

- 1 EVM PC + 3 RU PCs
- Internal L1T generation
- Artificial packet drops
- 3 words/trigger
- 64 triggers/packet
- No CRC



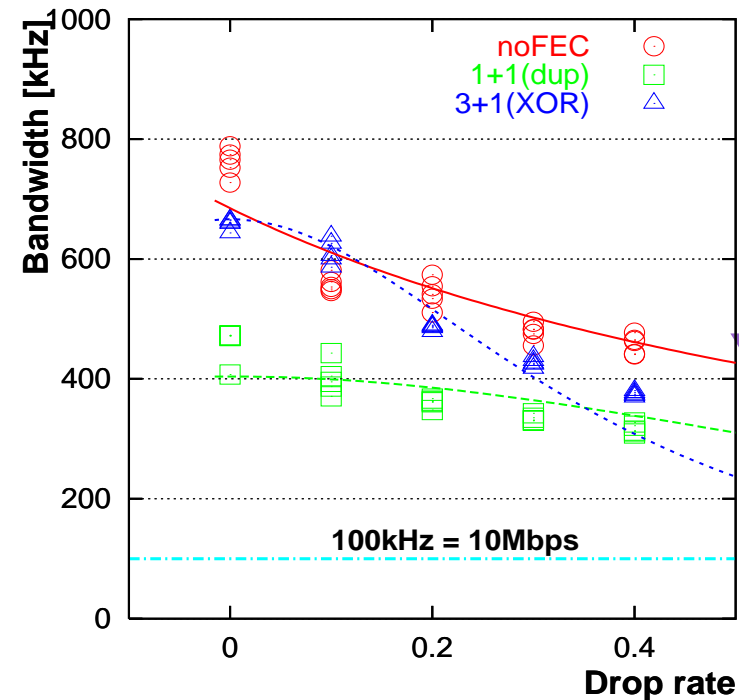
Bandwidth

- Global fit with a simple model

$$f_{noFEC}(\mu) = \frac{a}{(1+b\mu)} \frac{1}{(1+c/1)}$$

$$f_{1+1}(\mu) = \frac{1}{2} \frac{a}{(1+b\mu^2)} \frac{1}{(1+c/2)}$$

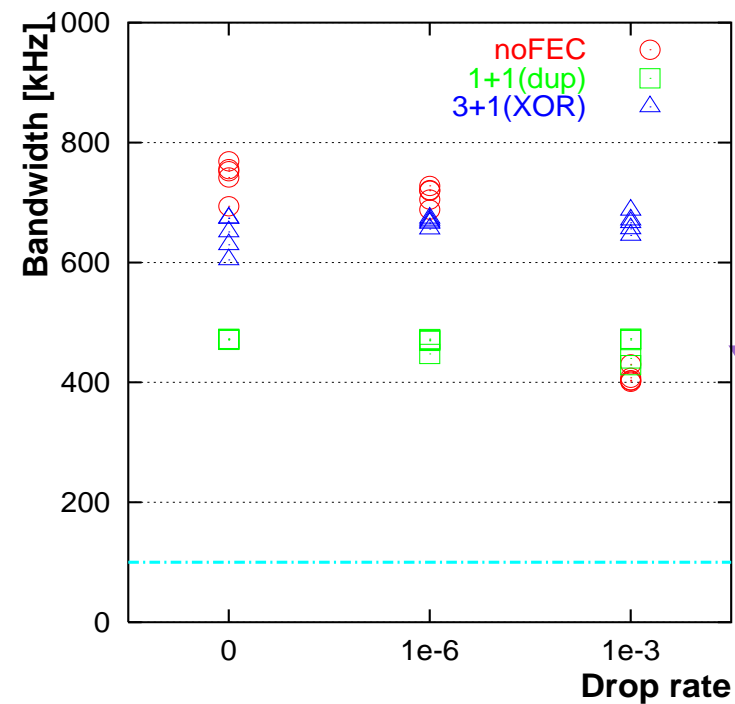
$$f_{3+1}(\mu) = \frac{3}{4} \frac{a}{(1+6b\mu^2)} \frac{1}{(1+c/4)}$$



- Bandwidth is better than 10Mbps
- Effect of packet drops is understood with a simple model

Bandwidth

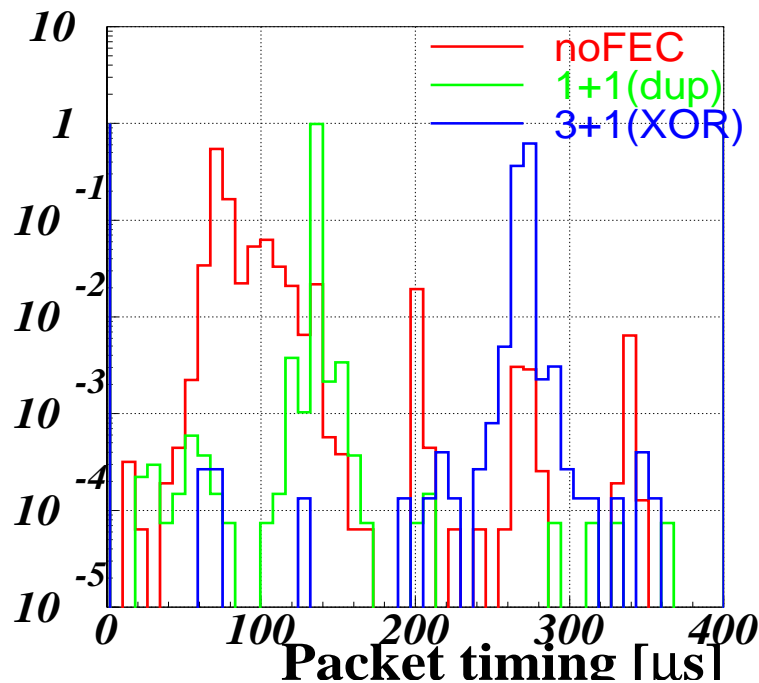
- Effect of FECs
 - ➔ No FEC
 - ➔ 1+1(dup)
 - ➔ 3+1(XOR)
- Measured under BG traffic



- 3+1 FEC shows good BW for all drop rates

Latency

- Jitter: Variance of the packet receive timing

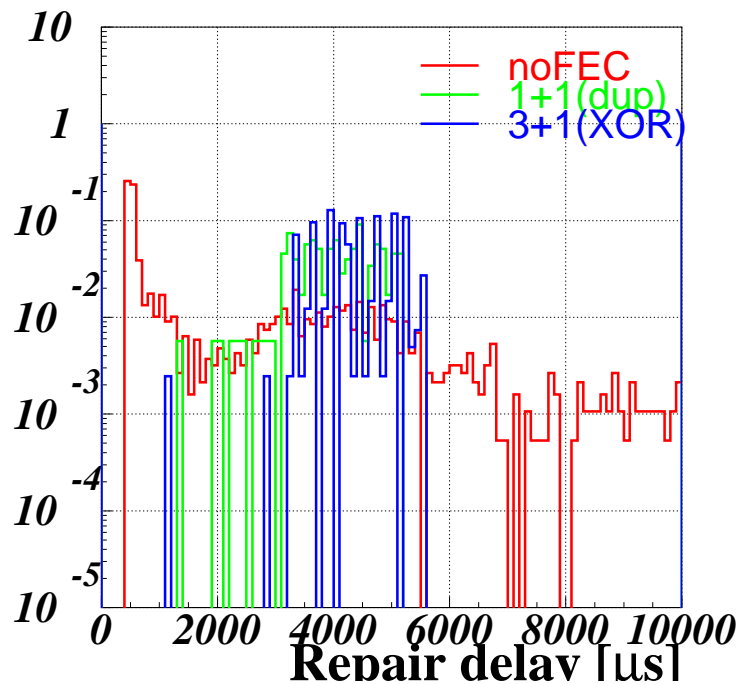


- ➔ noFEC: 17.1 [μs]
- ➔ 1+1(dup): 3.6
- ➔ 3+1(XOR): 3.4

- Jitter is smaller than the requirement.
- Long tails of Jitter is not understood.

Latency (cont'd)

- Delay: Time spent to get a repair packet

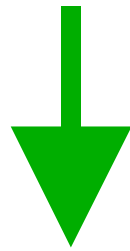


- ➔ noFEC: 2 [ms]
- ➔ 1+1(dup): 4
- ➔ 3+1(XOR): 4

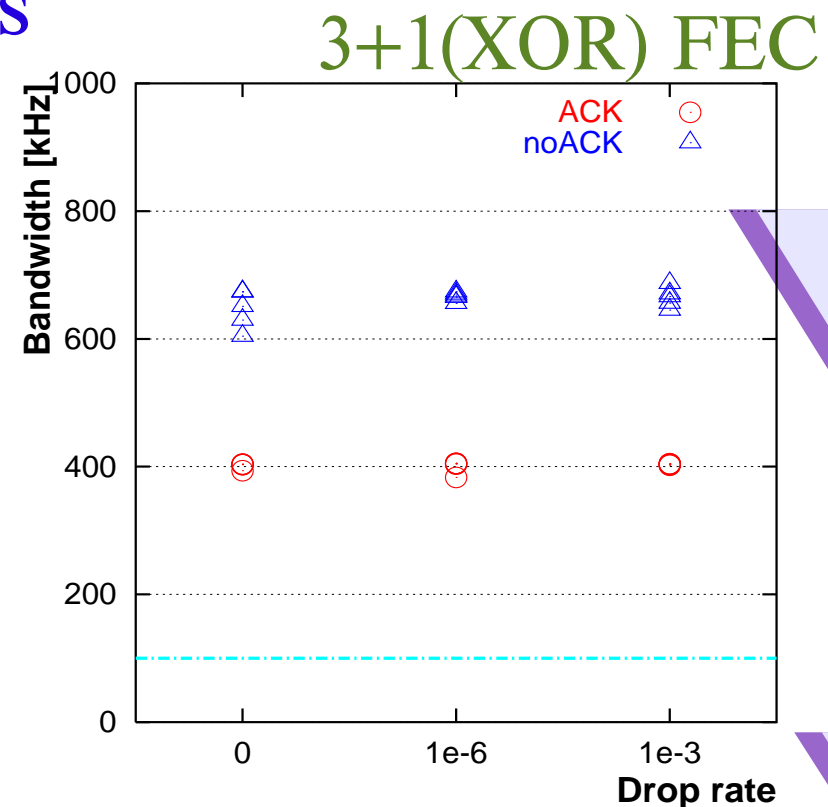
- The delay is smaller than 10ms (FED buffer).
- Distribution of the delay is not understood.

Traffic regulation

- How to avoid network congestions



- The EVM requires ACK packets from one RU for each trigger packet.



- 40% BW drop

Summary

- A reliable broadcast protocol was proposed and implemented on a Linux test system
- Characteristics of FEC and traffic regulation mechanism were shown with benchmarks
- Further studies on long latency tail are necessary
- Re-design and implementation for the real system are planned. (more robust protocol, performance tuning, IEEE1394)