The data taking network for COMET Phase-I

IGARASHI Youichi, KEK, Japan
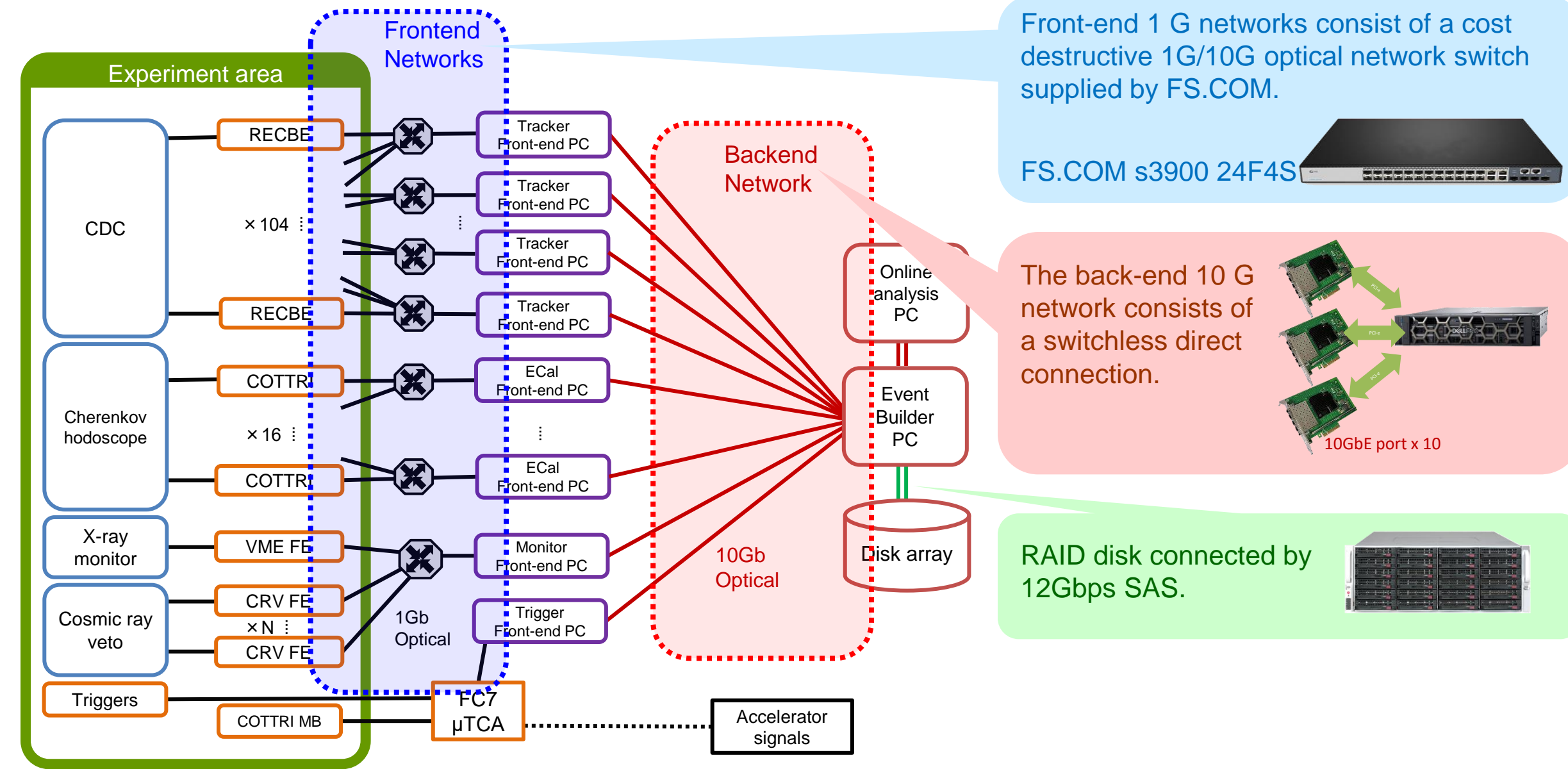
J-PARC COMET Phase-I [1]

COMET Phase-I detector   J-PARC

$$\mu^- + (A, Z) \rightarrow e^- + (A, Z)$$

A trial to discover the lepton flavor violation using $\mu N \rightarrow eN$ will start at J-PARC Hadon facility.
The experiment aims to search with the sensitivity of $O(10^{-15})$.

The data taking network for COMET Phase-I

Front-end boards will be installed on the detectors in the solenoid, which has 1 Tesla magnet field.
→ Front-end boards use 1000BASE-SX optical links with FPGA based TCP/IP engine[2].
→ DAQ front-end networks take care of a large number of 1000BASE-SX optical links.

Front-end 1 G networks consist of a cost destructive 1G/10G optical network switch supplied by FS.COM.

FS.COM s3900 24F4S

The back-end 10 G network consists of a switchless direct connection.

10GbE port x 10

RAID disk connected by 12Gbps SAS.

Network switch evaluation using CDC CRT setup

To evaluate the low price network switch, we put in the switches in a CDC detector's cosmic ray test setup.

RECBE x 105

Optical 1Gb   Optical 1Gb

FS.com S3900 24F4S

Optical 10Gb

From 105 to 1

DAQ PC
Xeon E5-1650 v2 @ 3.50GHz
Mellanox Connect X-3 Pro

CDC cosmic ray test setup

One hundred five readout cards (RECBE) and five low price optical network switches worked on the setup. The data were taken with the full event building and without the BUSY control.

RAW mode (6144 B) — 337MB/s
The speed of the event building limited the performance.

SUPPRESS mode (8B〜400B) — 34.4MB/s
The EB success rate goes down to 0 after a while because of no BUSY control.

A evaluation of a switchless event building

To evaluate the switchless event building, we used the actual DAQ PCs for the experiment. Front-end PCs have two 10Gb network ports, and each port connects to the event building PC. The event building PC has ten 10Gb network ports. MIDAS[3] DAQ software is used for the DAQ.

DAQ PCs	Processor	Memory	10 Gbps NIC	SAS/RAID system	HDD
Event building PC	Xeon Gold 6126 @ 2.6GHz	64 GB	Intel Network Adapter X710-DA4	Broadcom / LSI MegaRAID SAS-3 3108	SEAGATE 10 TB 7200RPM
Front-end PC	Xeon E-2134 @ 3.5GHz	32 GB	Mellanox Connect-X 3 Pro		

MIDAS

12Gbps SAS

10 Gb optical link x 8 for event building

Working processes
- Dummy FE Process working on the frontend PCs.
  - Dummy FE generates 16kiB data and sends them to the backend PC.
- MIDAS Processes working on the event-building PC
  - FE : reading data from network, 8 FE processes were working.
  - EB : building events from event fragments of MIDAS FEs
  - Logger : recording data to files

Logger process limited the recording speeds.
Total throughput : 830 MiB/s

W/O Event building, W/O recording
W/ Event building, W/O recording
W/ Event building and recording
W/ Event building and recording (compressed)
Compression rate : 70% with lz4

Performance test of the RAID HDD

RAID disk system is connected EB PC via 12 Gbps SAS (Serial attached SCSI).
- 10 HDD available
  - 10 TB HDD x 10, 74 TB data space
- RAID system has 44 HDD bays.
- RAID system was configured as RAID5+0.
  - RAID5+0 : Striping RAID5, it has redundancy bit and dual data accessing.

A writing test by "dd" command

File system	Speed
Xfs	1.84 GiB/s adopted
Ext4	1.66 GiB/s

Radius dependency of the writing speed
The writing speeds are faster than 1 GB/s anywhere on the disks.
outside   inside

Long distance data transport

The final destination of the experimental data is KEK Computer research center's tape storage.

Data copy test using "scp"

KEK (Tsukuba city) and J-PARC (Tokai village) are connected by SINET (Science Information network, Japan) with 10 Gbps speed. We confirmed the speed of the data copy by "scp" command.
The speed of the one "scp" is around 280 MiB/s. The total data copy speed achieved 1.03 GiB/s using over five connection.

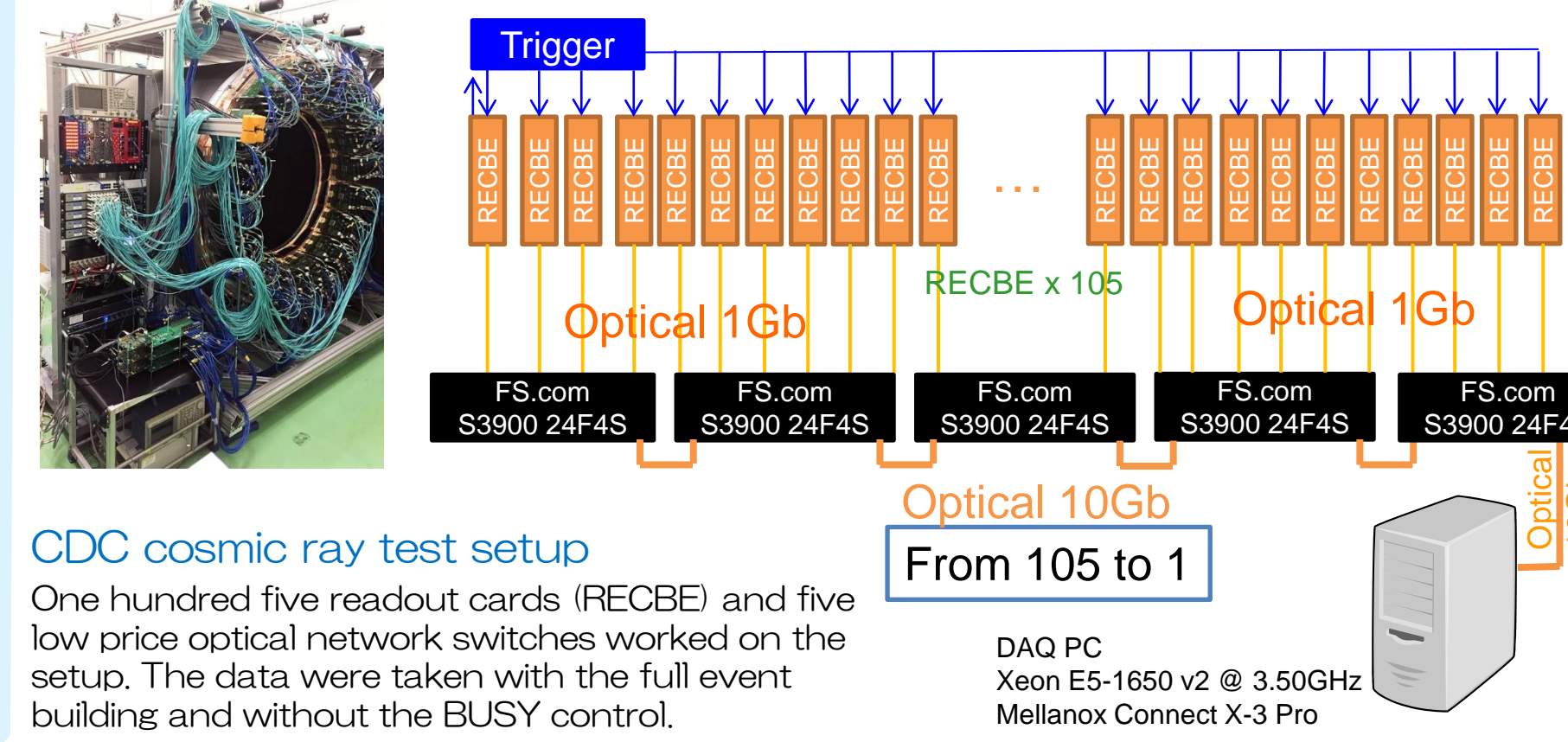~60 km   1.03 GiB/s

Summary

- COMET Phase-I DAQ system is a two-layer network-based system.
- The network of the DAQ assembled with the low price network switches for the front-end and the switchless connection for the back-end.
- The total throughput of the DAQ achieved 800 MiB/s.
  - All data throughputs of the DAQ system were evaluated. The current bottleneck is the local data recording process.

References

[1] COMET collaboration, "Experimental Proposal for Phase-I of the COMET Experiment at J-PARC", 2012, [Online]. Available: http://j-parc.jp/researcher/Hadron/en/pac_1207/pdf/E21_2012-10.pdf
[2] T. Uchida, "Hardware-Based TCP Processor for Gigabit Ethernet", IEEE Trans. Nucl. Sci., vol. 55, no. 3, pp. 1631 – 1637, 2008. DOI: 10.1109/TNS.2008.920264
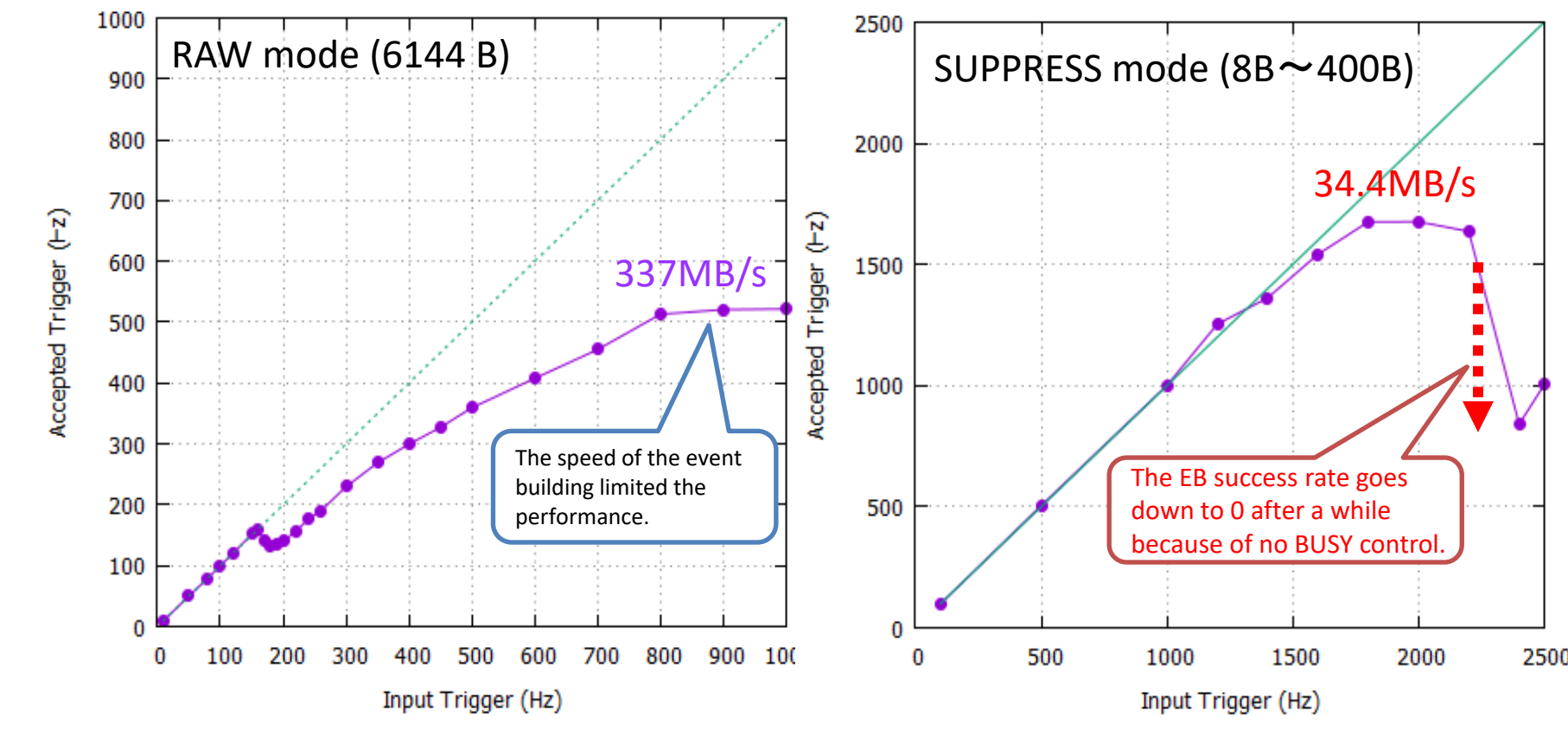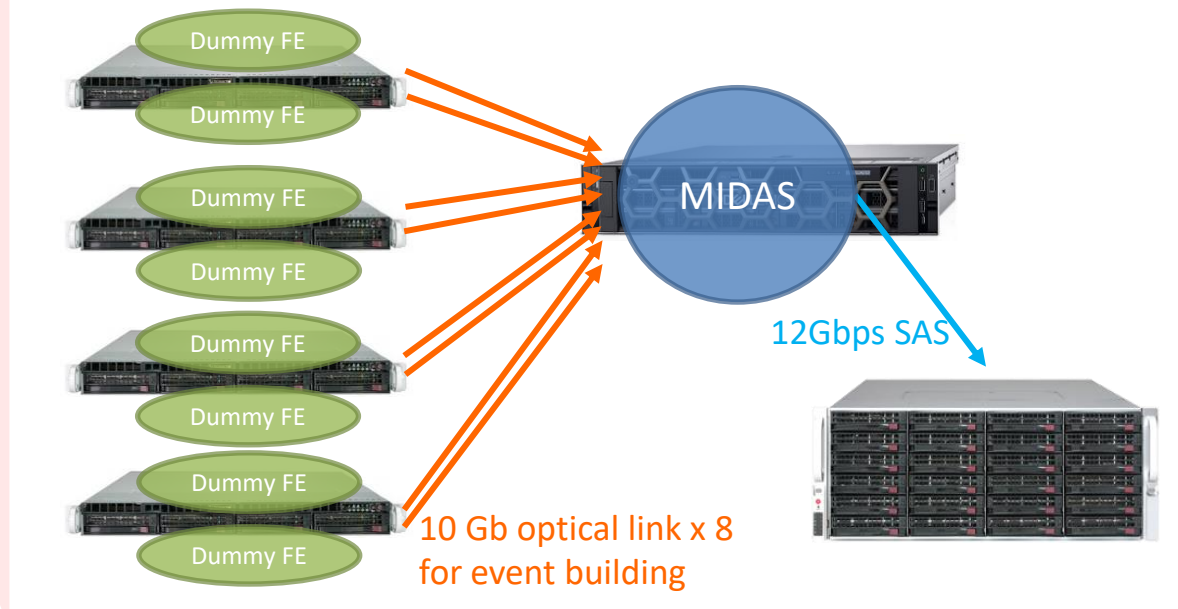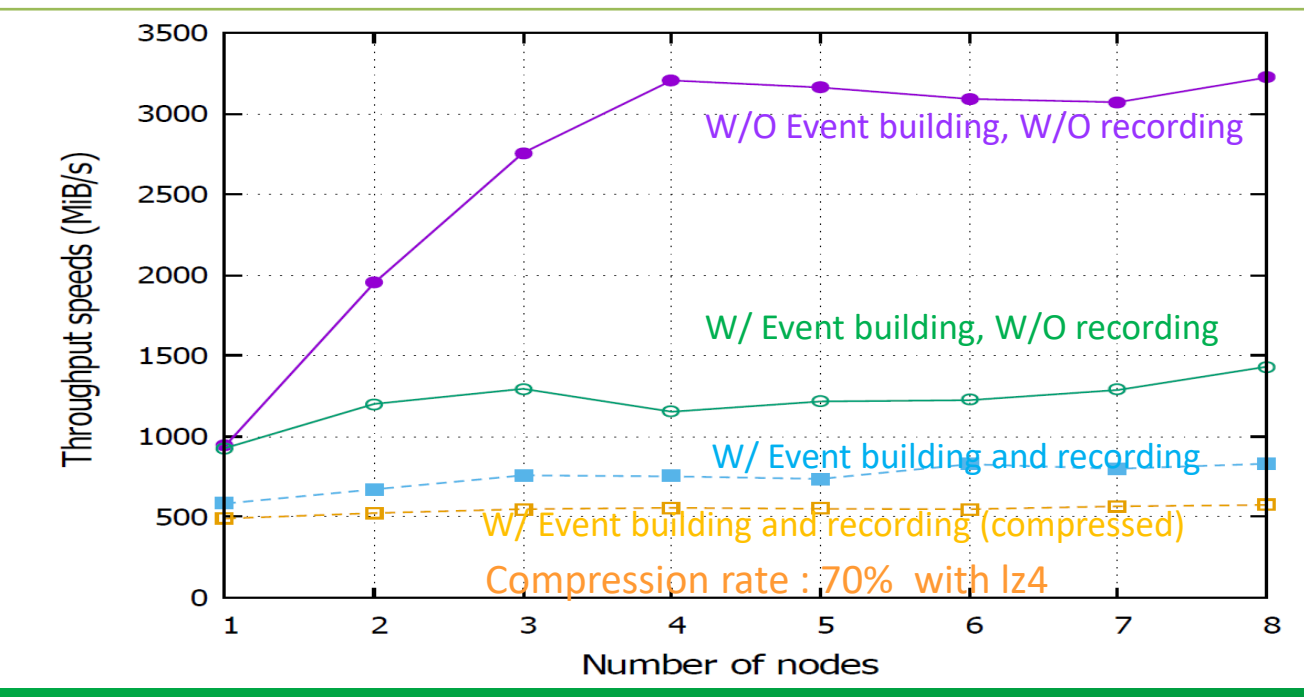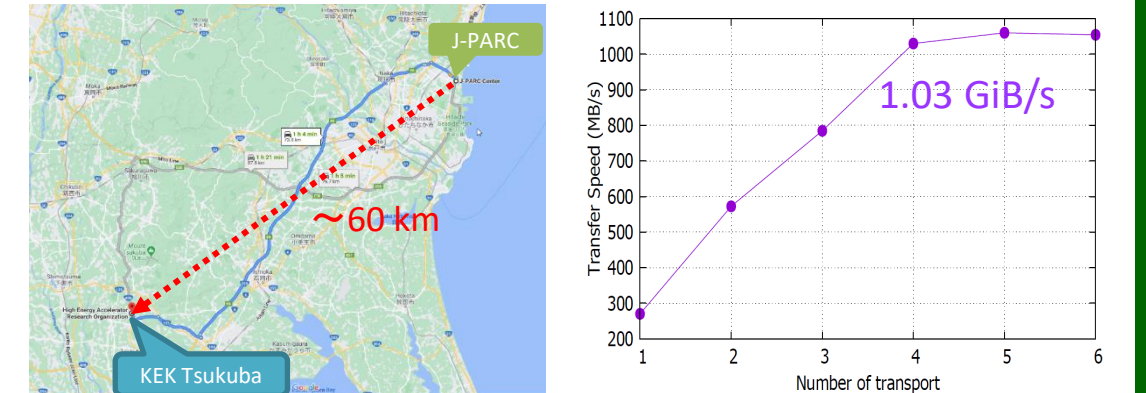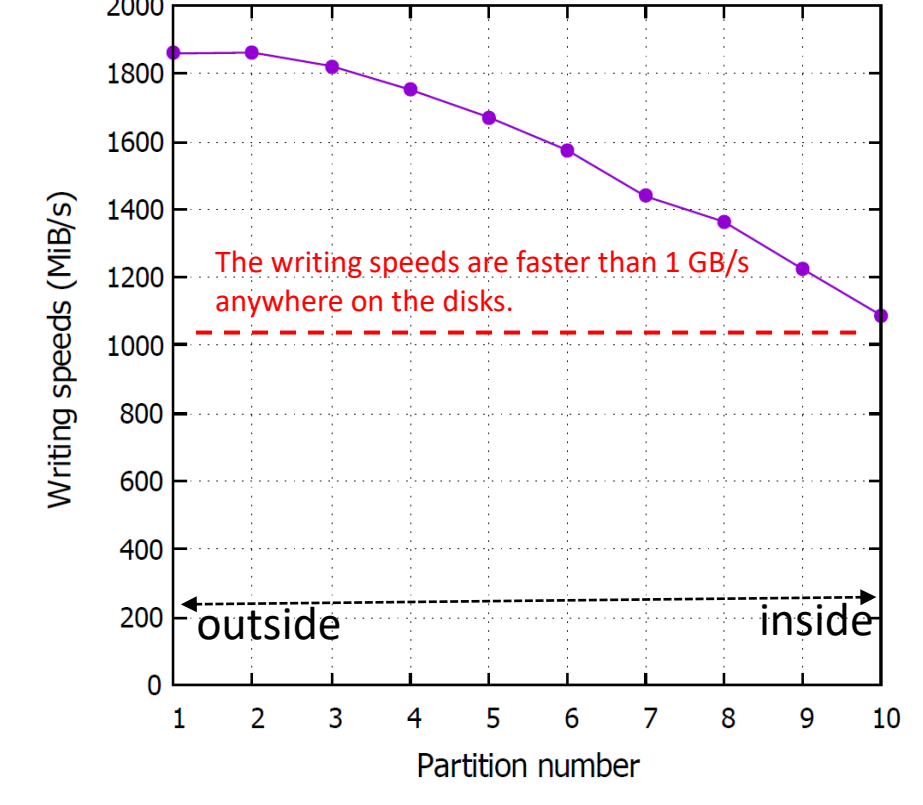[3] PSI and TRIUMF, "MIDAS wiki", [Online]. Available: https://www.triumf.info/wiki/DAQwiki/index.php/MIDAS
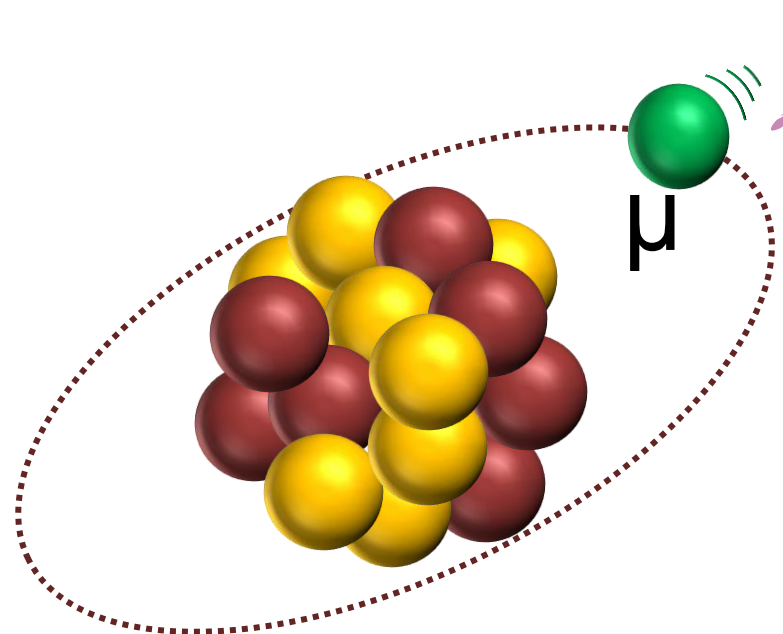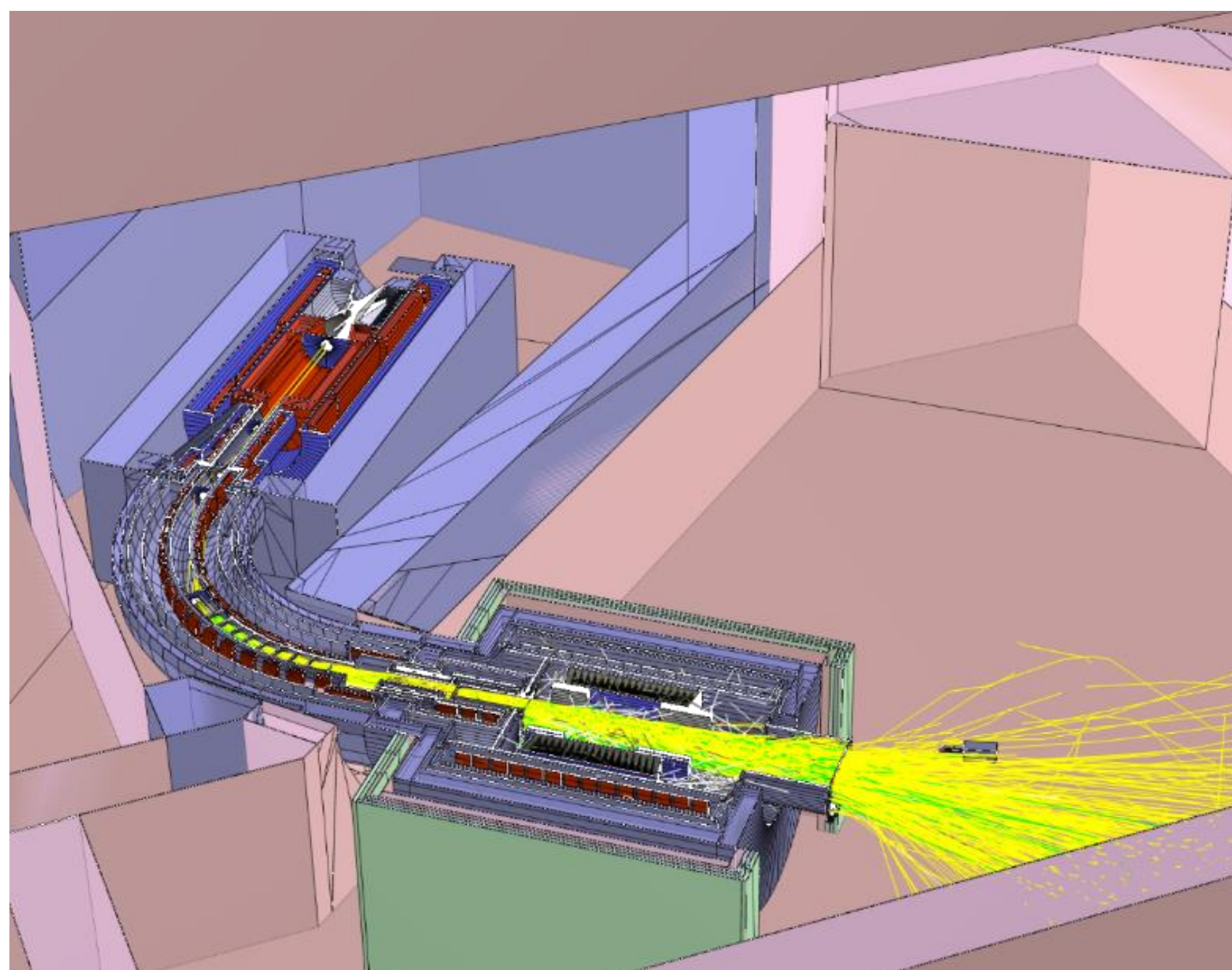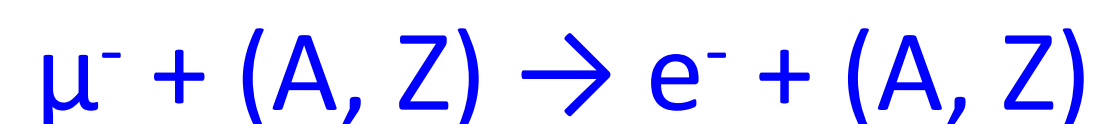
The data taking network for COMET Phase-I
IGARASHI Youichi, KEK, Japan

J-PARC COMET Phase-I [1]

A trial to discover the lepton flavor violation using $\mu N \rightarrow eN$ will start at J-PARC Hadon facility. The experiment aims to search with the sensitivity of $O(10^{-15})$.
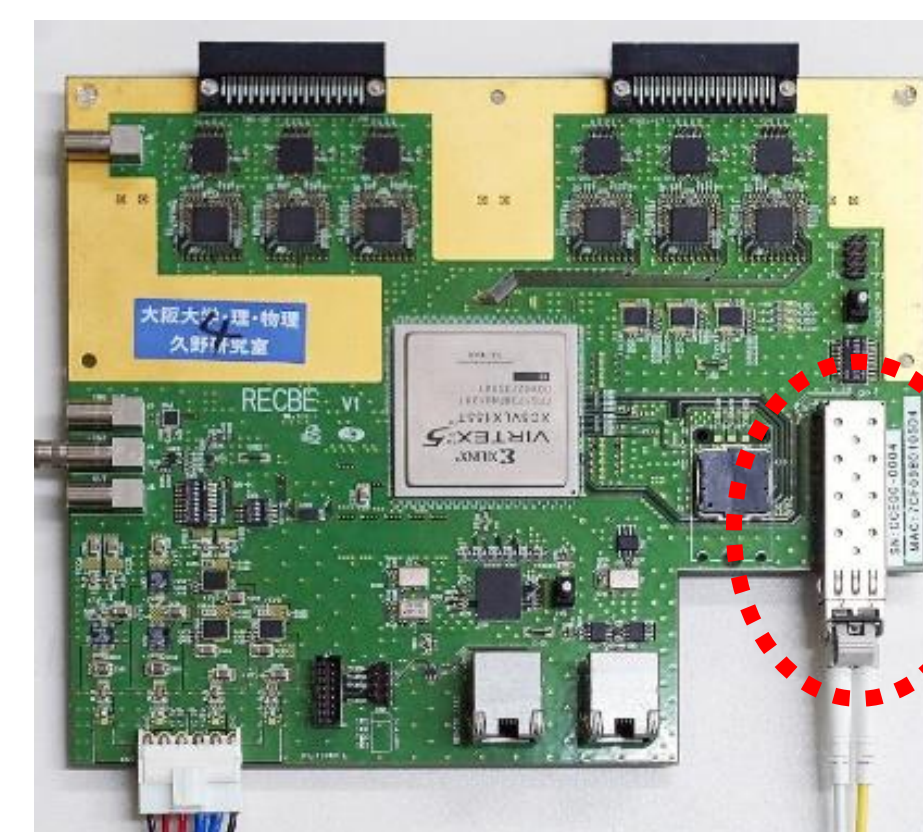
$$\mu^- + (A, Z) \rightarrow e^- + (A, Z)$$

COMET Phase-I detector

J-PARC

On detector readout concept

Front-end boards will be installed on the detectors in the solenoid, which has 1 Tesla magnet field.

→Front-end boards use 1000BASE-SX optical links with FPGA based TCP/IP engine[2].

→DAQ front-end networks take care of a large number of 1000BASE-SX optical links.
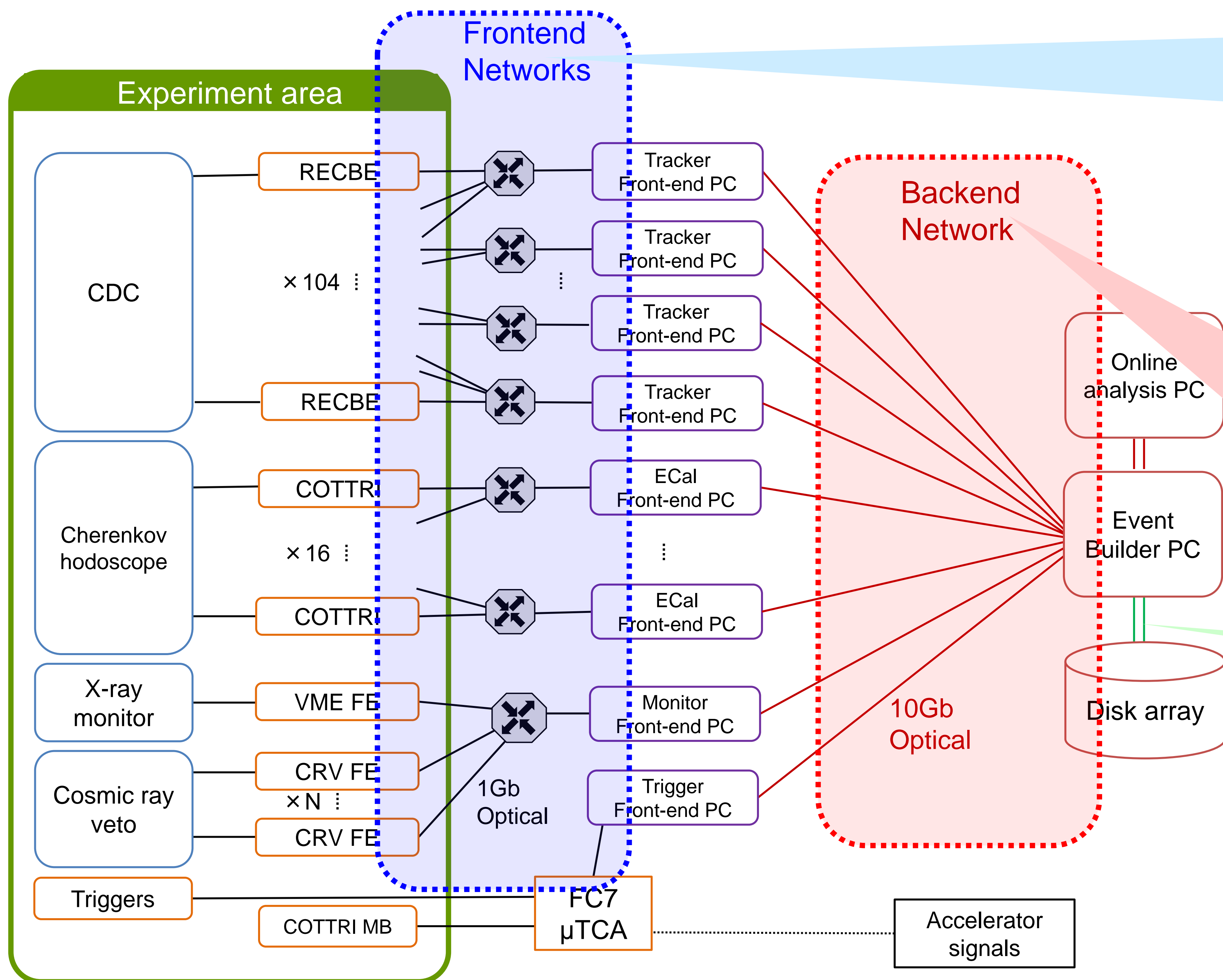
Optical links

ROESTI : A readout board for a straw chamber
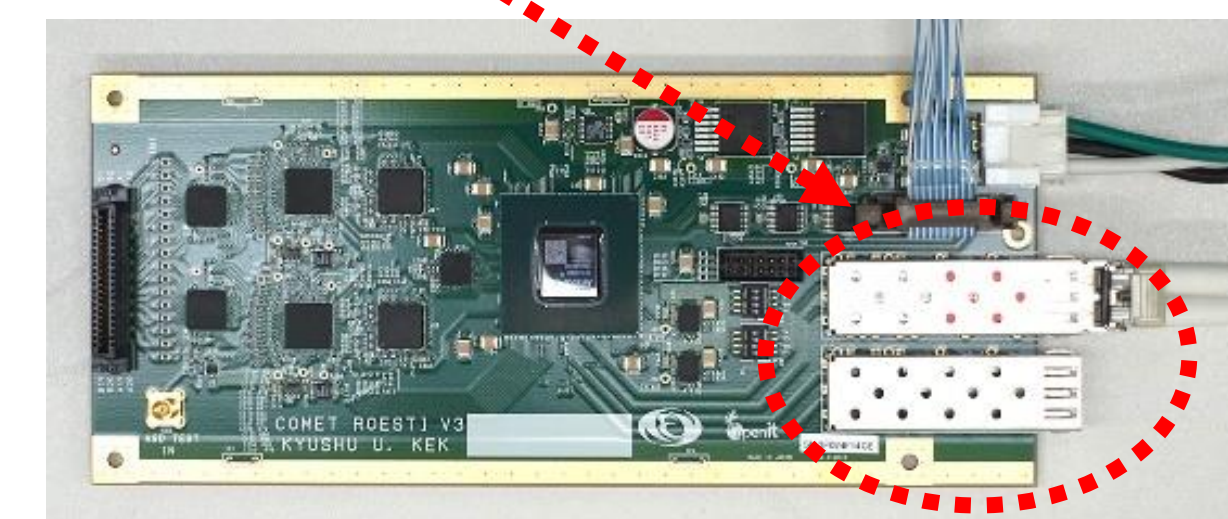
RECBE : A readout board for CDC

High-performance optical network switches push up the DAQ assembly cost!

What's our option?

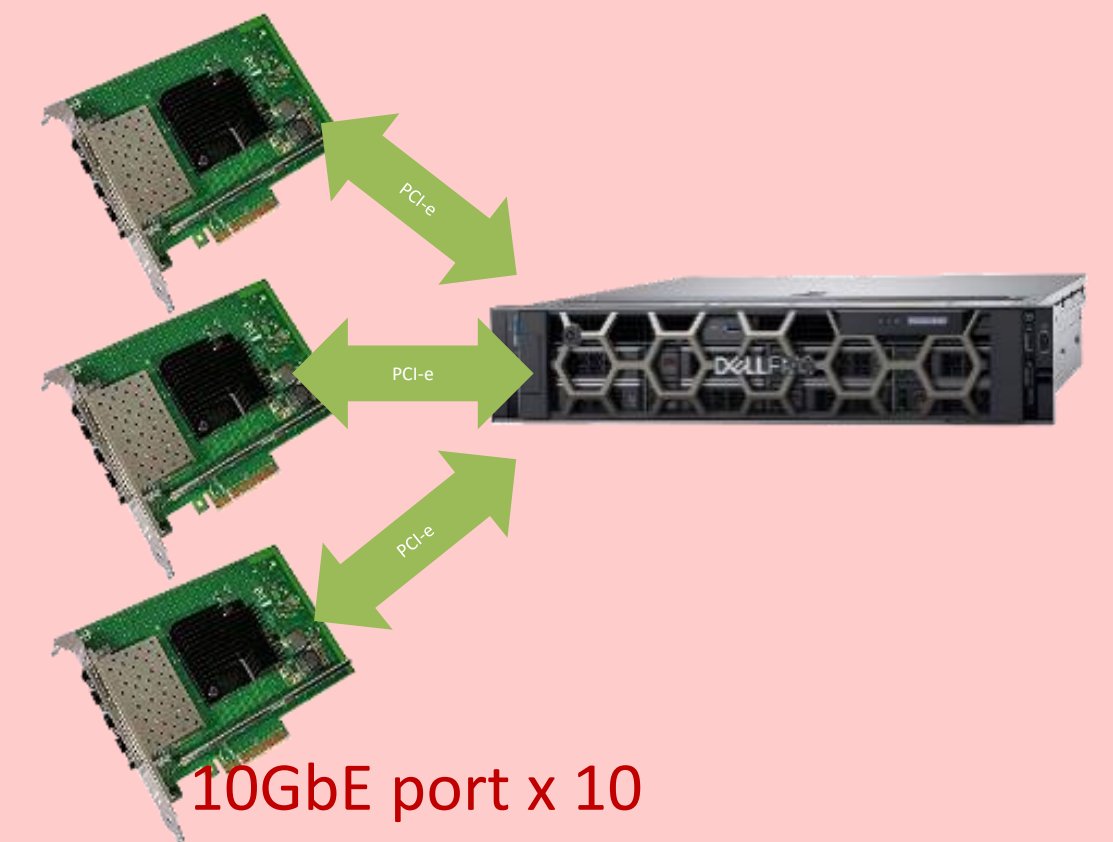# The data taking network for COMET Phase-I

Front-end 1 G networks consist of a cost destructive 1G/10G optical network switch supplied by FS.COM.

FS.COM s3900 24F4S

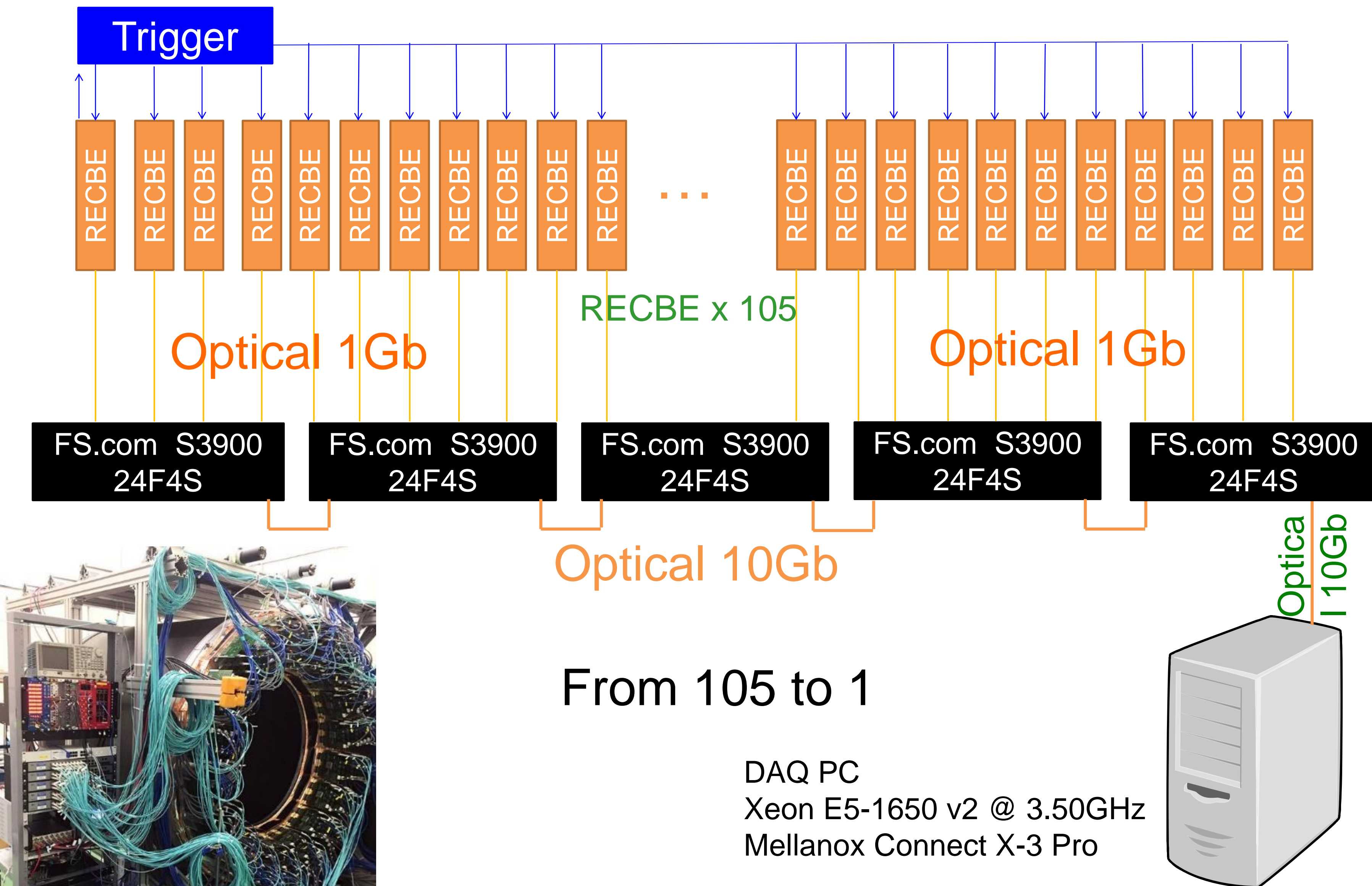The back-end 10 G network consists of a switchless direct connection.

10GbE port x 10

RAID disk connected by 12Gbps SAS.
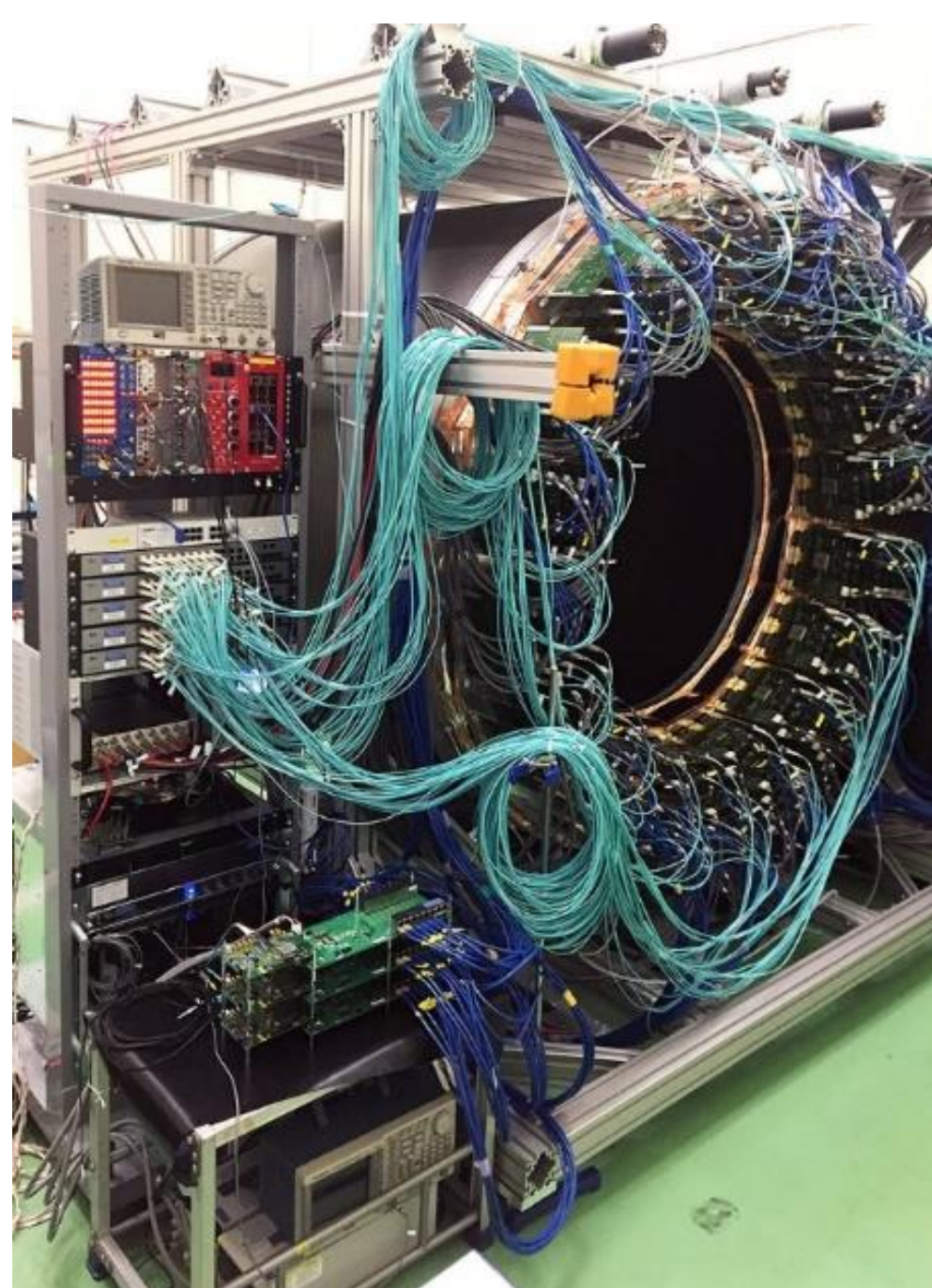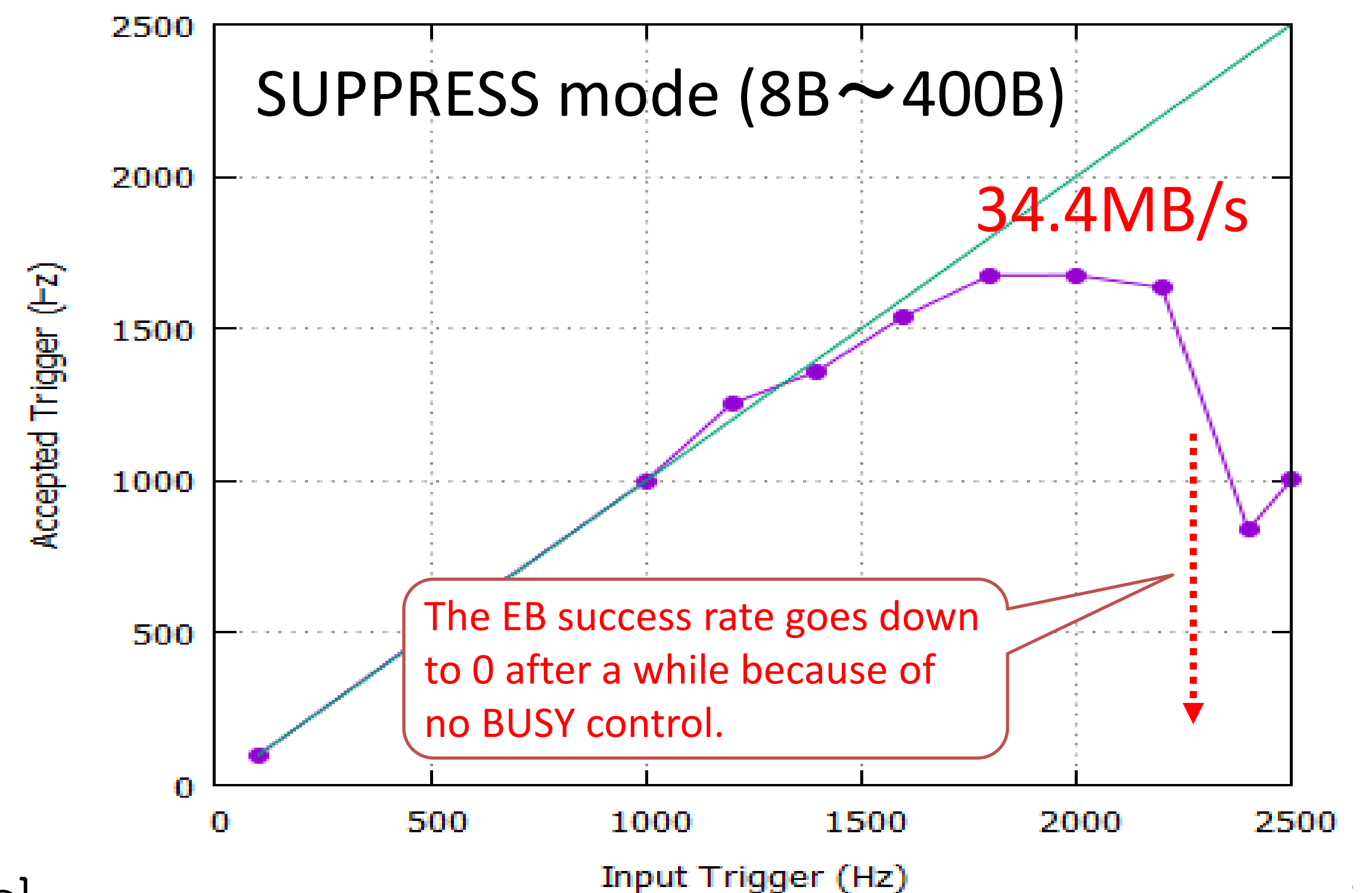
Network switch evaluation using CDC CRT setup

To evaluate the low price network switch, we put in the switches in a CDC detector's cosmic ray test setup.

Trigger

RECBE (×105)

Optical 1Gb ... Optical 1Gb

RECBE x 105

FS.com S3900 24F4S | FS.com S3900 24F4S | FS.com S3900 24F4S | FS.com S3900 24F4S | FS.com S3900 24F4S

Optical 10Gb

Optical 10Gb

From 105 to 1

DAQ PC
Xeon E5-1650 v2 @ 3.50GHz
Mellanox Connect X-3 Pro

CDC cosmic ray test setup

One hundred five readout cards (RECBE) and five low price optical network switches worked on the setup. The data were taken with the full event building and without the BUSY control.

RAW mode (6144 B)

337MB/s

The speed of the event building limited the performance.

Accepted Trigger (Hz) vs *Input Trigger (Hz)*

SUPPRESS mode (8B〜400B)

34.4MB/s

The EB success rate goes down to 0 after a while because of no BUSY control.

Accepted Trigger (Hz) vs *Input Trigger (Hz)*

COMET DAQ PCs

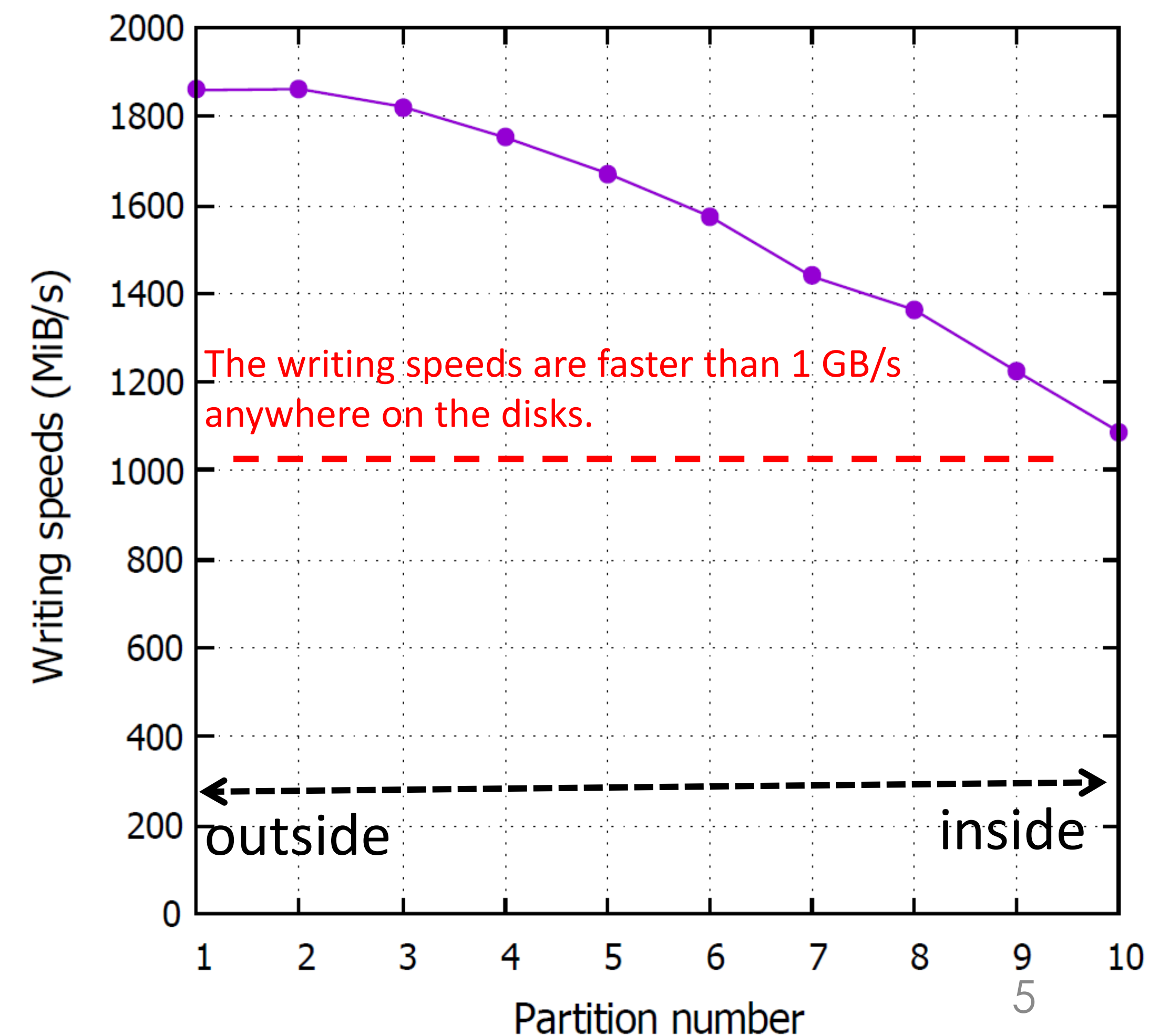|	Processor	Memory	10 Gbps NIC	SAS/RAID system	HDD
Event building PC	Xeon Gold 6126 @ 2.6GHz	64 GB	Intel Network Adapter X710-DA4	Broadcom / LSI MegaRAID SAS-3 3108	SEAGATE 10 TB 7200RPM
Front-end PC	Xeon E-2134 @ 3.5GHz	32 GB	Mellanox Connect-X 3 Pro		

Performance test of the RAID HDD

- RAID disk system is connected EB PC via 12 Gbps SAS (Serial attached SCSI).
- 10 HDD available
  - 10 TB HDD x 10, 74 TB data space
- RAID system has 44 HDD bays.
- RAID system was configured as RAID5+0.
  - RAID5+0 : Striping RAID5, it has redundancy bit and dual data accessing.
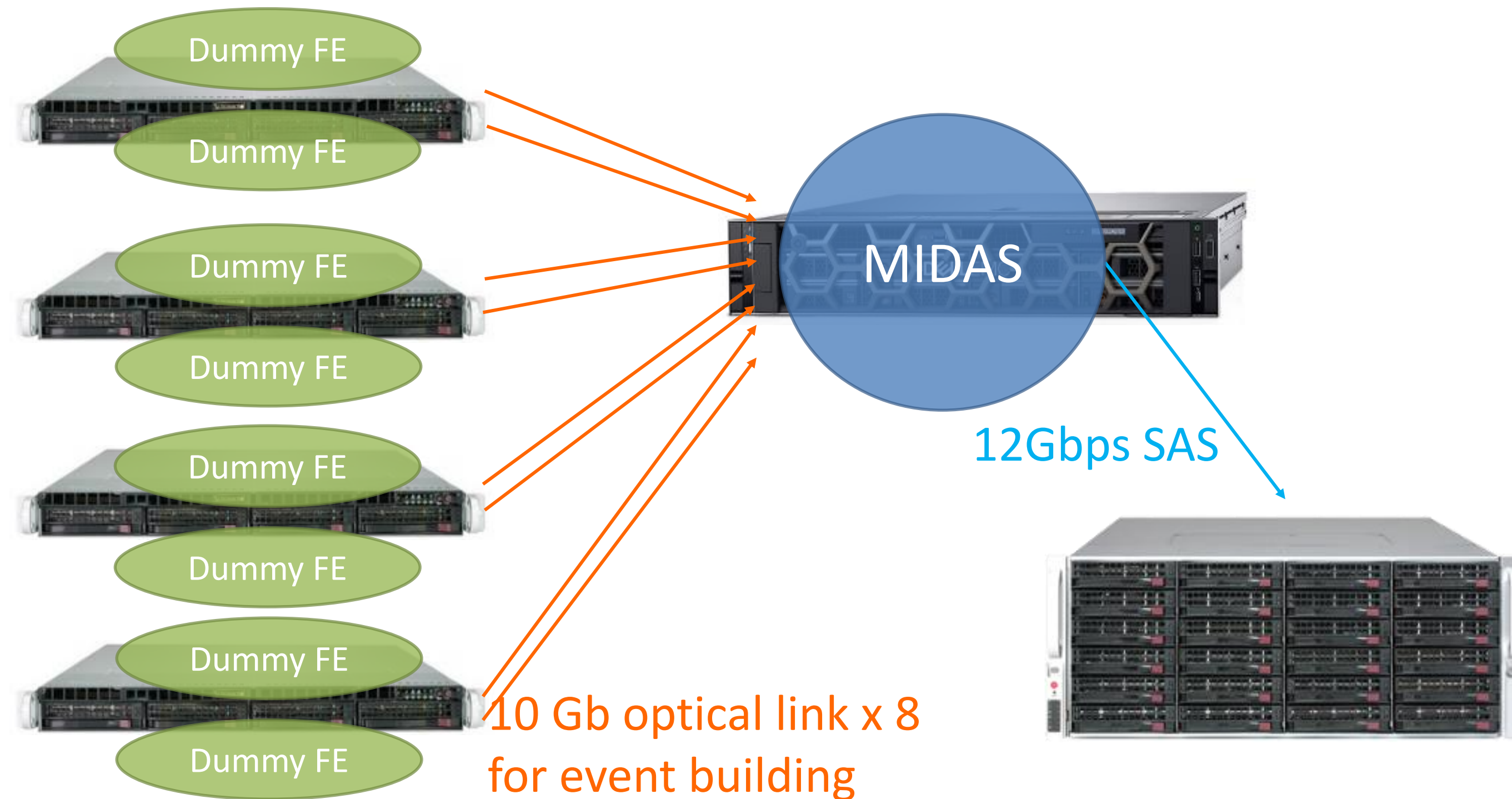
A writing test by "dd" command

File system	Speed	
Xfs	1.84 GiB/s	adopted
Ext4	1.66 GiB/s	

DAQ PCs are installed on the 3rd floor of COMET experimental building.
- 2 EB PC and 4 FE PC (for the main detector readout).
- RAID disk system

Radius dependency of the writing speed

The writing speeds are faster than 1 GB/s anywhere on the disks.

outside   inside

Writing speeds (MiB/s) vs Partition number

A evaluation of a switchless event building

To evaluate the switchless event building, we used the actual DAQ PCs for the experiment. Front-end PCs have two 10Gb network ports, and each port connects to the event building PC. The event building PC has ten 10Gb network ports. MIDAS[3] DAQ software is used for the DAQ.

10 Gb optical link x 8
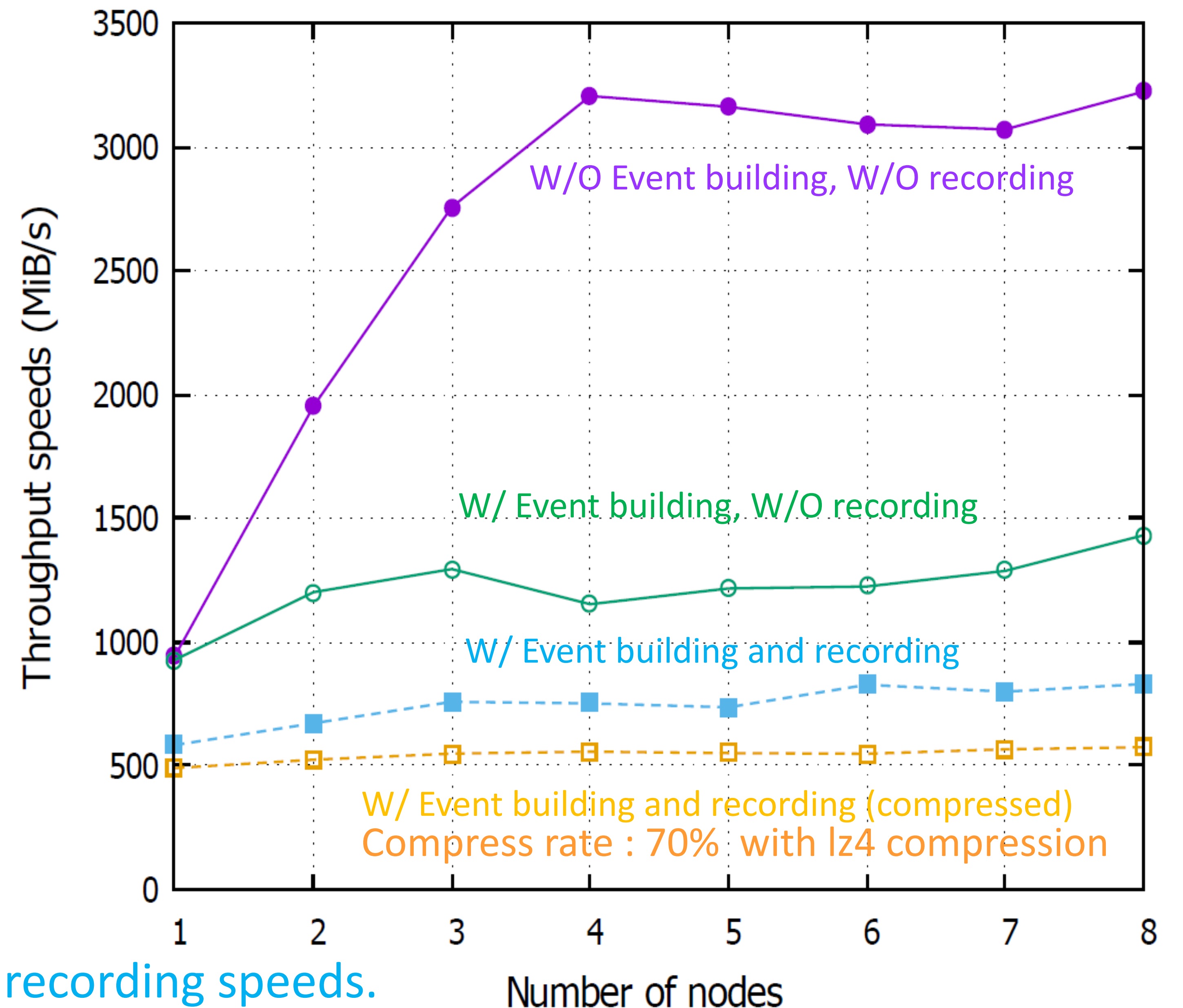for event building

12Gbps SAS

Working processes

- Dummy FE Process working on the frontend PCs.
  - Dummy FE generates 16kiB data and sends them to the backend PC.
- MIDAS Processes working on the event-building PC
  - FE : reading data from network, 8 FE processes were working.
  - EB : building events from event fragments of MIDAS FEs
  - Logger : recording data to files

Logger process limited the recording speeds.
Total throughput : 830 MiB/s

W/O Event building, W/O recording

W/ Event building, W/O recording

W/ Event building and recording

W/ Event building and recording (compressed)
Compress rate : 70% with lz4 compression

Long distance data transport

The final destination of the experimental data is KEK Computer research center's tape storage.

Data copy test using "scp"

KEK（Tsukuba city）and J-PARC（Tokai villadge) are connected by SINET（Science Information network, Japan）with 10 Gbps speed. We confirmed the speed of the data copy by "scp" command.
The speed of the one "scp" is around 280 MiB/s. The total data copy speed achieved 1.03 GiB/s using over five connection.
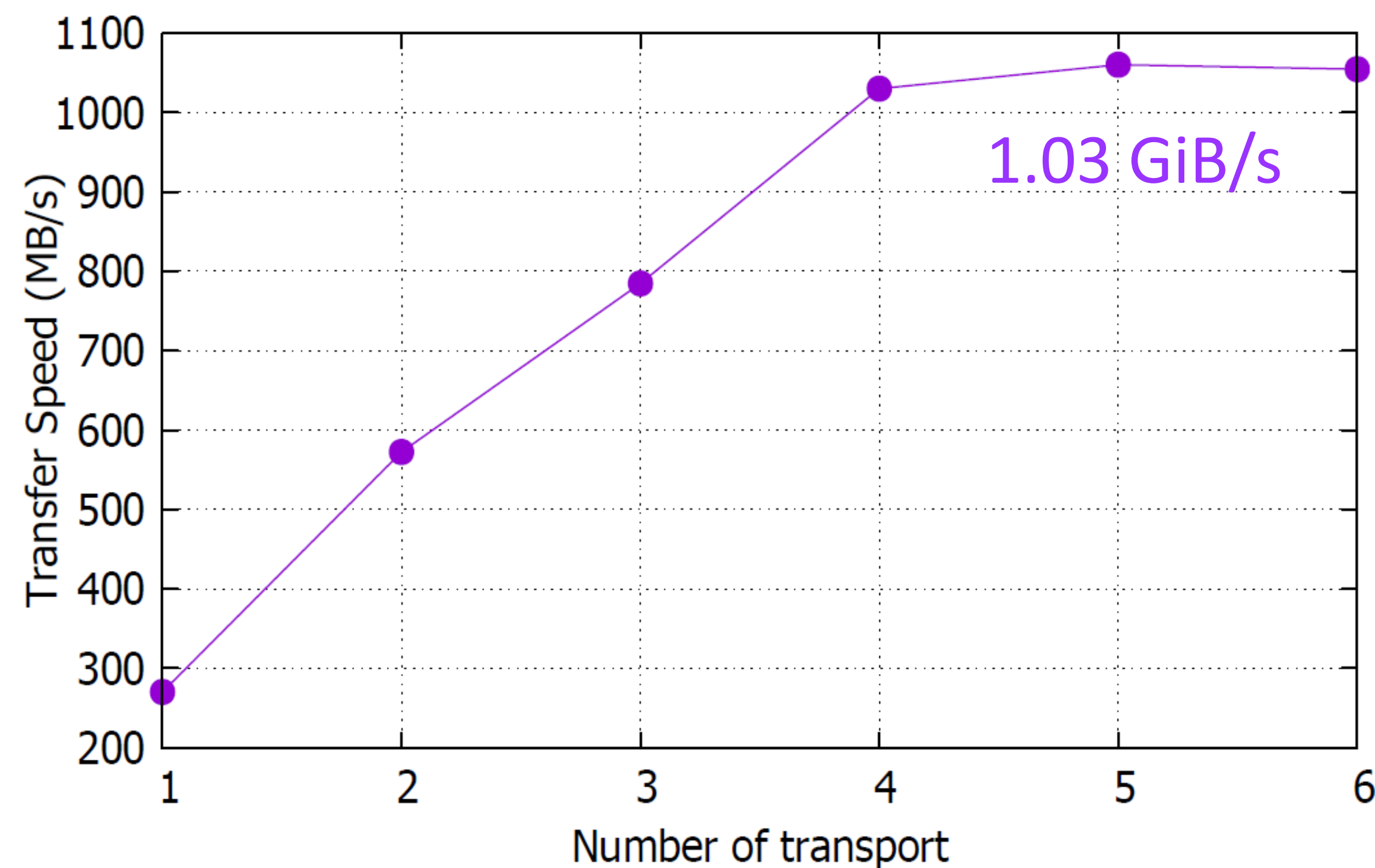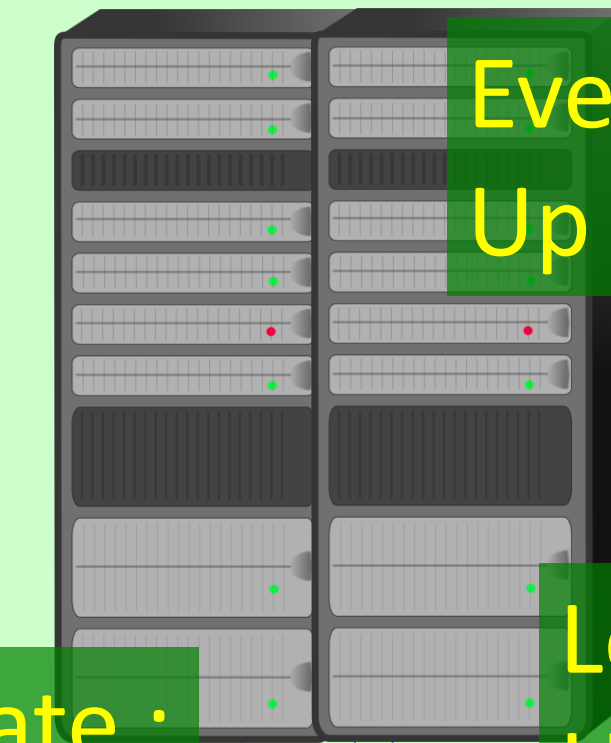
The Internet (SINET)

Data copy :
Up to 1 GiB/s

3F: DAQ Server

Event Building :
Up to 1 GiB/s

Local recoding :
Up to 800 MiB/s

Total data rate :
Up to 3 GiB/s

KEKCC Data storage

1.03 GiB/s

Transfer Speed (MB/s) vs Number of transport

J-PARC

~60 km

KEK Tsukuba

Phase-I

Muon Transport Solenoid

Proton Beam

μ-e conv. Detector

Pion Capture Solenoid

UG: Experimental room

Summary

- COMET Phase-I DAQ system is a two-layer network-based system.
- The network of the DAQ assembled with the low price network switches for the front-end and the switchless connection for the back-end.
- The total throughput of the DAQ achieved 800 MiB/s.
  - All data throughputs of the DAQ system were evaluated. The current bottleneck is the local data recording process.

References

[1] COMET collaboration, "Experimental Proposal for Phase-I of the COMET Experiment at J-PARC'', 2012,
    [Online]. Available: http://j-parc.jp/researcher/Hadron/en/pac_1207/pdf/E21_2012-10.pdf
[2] T. Uchida, "Hardware-Based TCP Processor for Gigabit Ethernet", IEEE Trans. Nucl. Sci., vol. 55, no. 3, pp. 1631 – 1637, 2008. DOI: 10.1109/TNS.2008.920264
[3] PSI and TRIUMF, "MIDAS wiki", [Online]. Available: https://www.triumf.info/wiki/DAQwiki/index.php/MIDAS