SACLA/SPring-8 Data Reduction Pipeline

Haruki Nishino[1,2]

[1]JASRI, [2]RIKEN

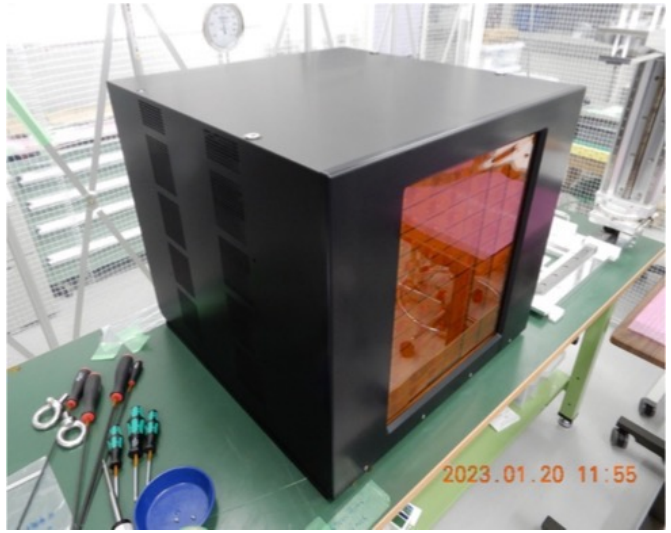
Overview

- Large data volume from high-speed imaging detector
  - CITIUS detector for diffraction experiments
- Data reduction strategies in applications of CITIUS detectors at XFEL and synchrotron radiation facilities

Large data volume from high-speed imaging detector

- **High-speed imaging sensor generates a lot of data.**

CITIUS detector for diffraction



20.2 Mpixels
17.4 kfps

DIFRAS detector for transmission

150 Mpixel

- **This talk will focus on the data reduction strategies and experiences of CITIUS.**

Data reduction for CITIUS at SACLA/SPring-8

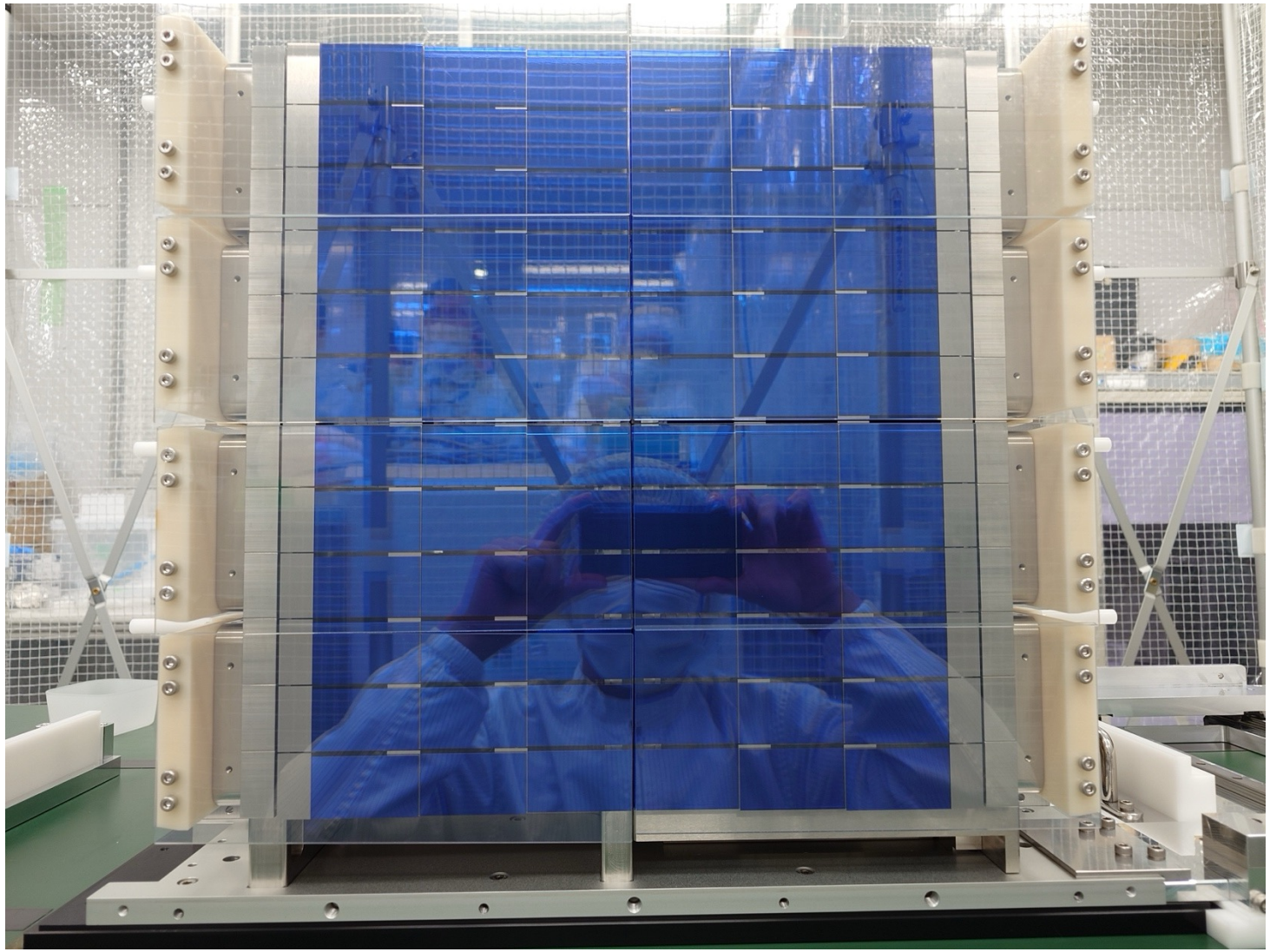
- Applications of CITIUS detectors
  - SACLA (XFEL)
    - 20.2 Mpixel CITIUS detector for Serial Femtosecond Crystallography (SFX)
  - SPring-8 (SR)
    - 840 kpixel CITIUS detector for quasi-elastic scattering experiment at BL35XU
- Hybrid data infrastructure
  - Edge computing + On-site PC cluster and storage + Offsite HPC

20.2 Mpixel CITIUS detector for SFX at SACLA

- SFX at SACLA
  - Recent publications in high-profile journals in 2023
    - M. Maestre-Reyna et al., *Visualizing the DNA repair process by a photolyase at atomic resolution*, **Science** 382, 1014 (2023).
    - H. Li et al., *Oxygen-evolving photosystem II structures during S1–S2–S3 transitions*, **Nature** 626, 670 (2024).
    - T. Gruhl et al., *Ultrafast structural changes direct the first molecular events of vision*, **Nature** 615, 939 (2023).
    - C. D. M. Hutchison et al., *Optical control of ultrafast structural dynamics in a fluorescent protein*, **Nat. Chem.** 15, 1607 (2023).
    - A. M. Wolff et al., *Mapping protein dynamics at high spatial resolution with temperature-jump X-ray crystallography*, **Nat. Chem.** 15, 1549 (2023).
    - K. Takaba et al., *Structural resolution of a small organic molecule by serial X-ray free-electron laser and electron crystallography*, **Nat. Chem.** 15, 491 (2023).
  - Number of publications since 2013: ~70
  - **One of the workforce instruments**

- We have developed a 20.2 Mpixel CITIUS detector detector for Serial Femtosecond Crystallography (SFX) at SACLA.

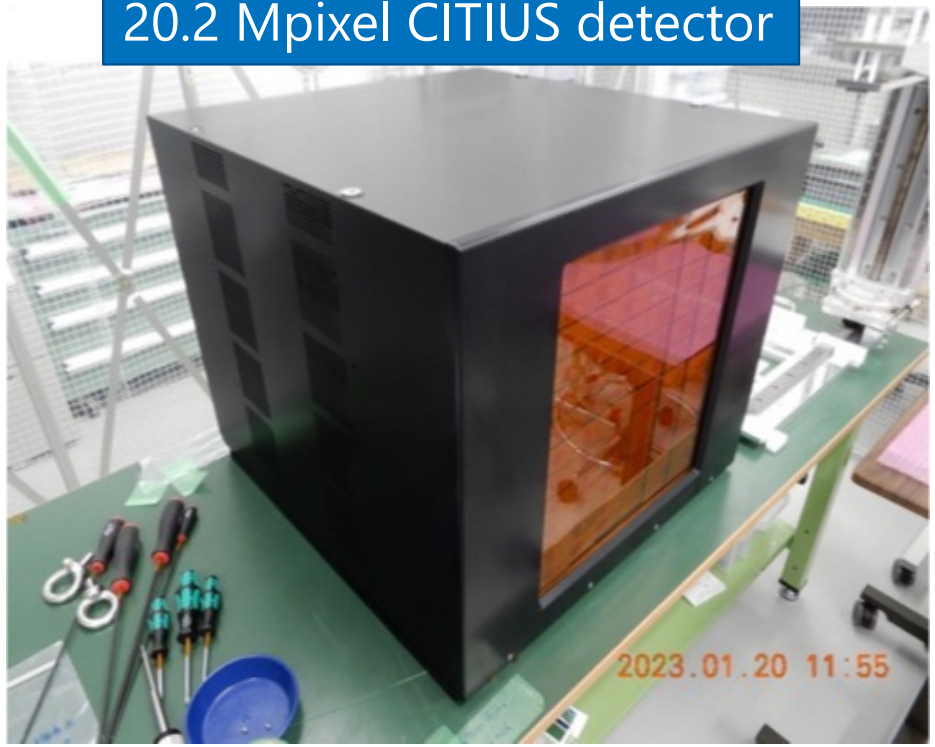
20.2 Mpixel CITIUS detector



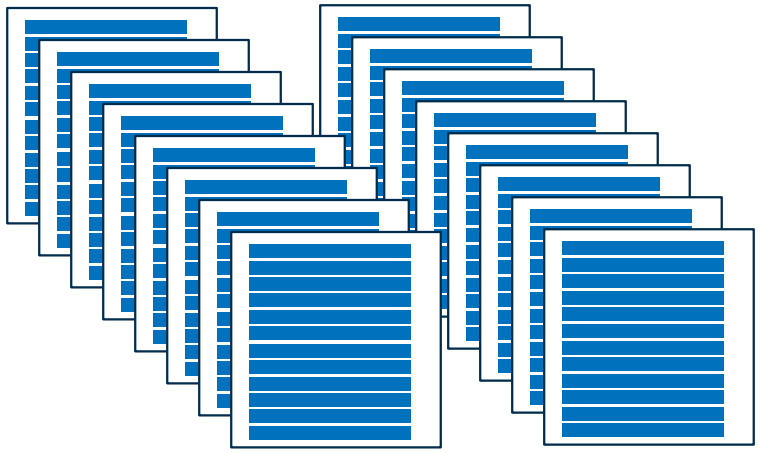
We will install the detector in SACLA BL2EH3 in April 2024.

Data rate from 20.2 Mpixel CITIUS detector at SACLA

20.2 Mpixel CITIUS detector



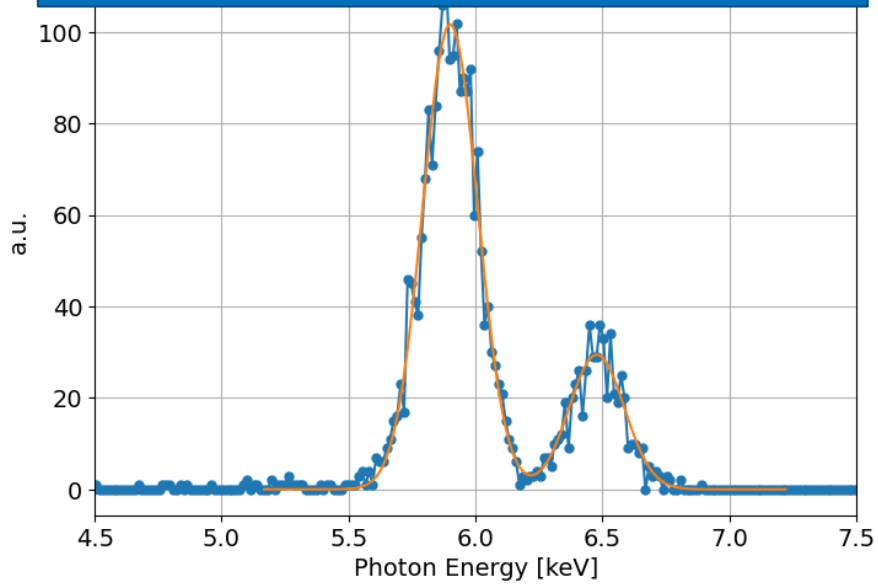
x 16 AD samplings for each pulse



➔ **107 GB/s**

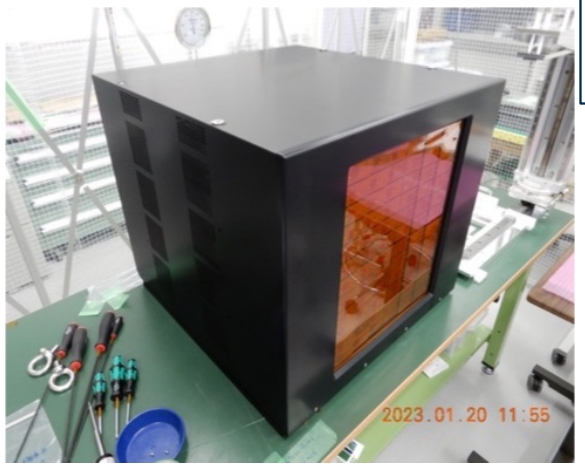
- XFEL mode of CITIUS sensor
  - Maximum frame rate: 5 kHz
- CITIUS sensors run at a frame rate of 960 Hz at SACLA
  - SACLA's beam repetition rate: 60 Hz
  - 16 frames for each pulse
    - resulting in a lower noise floor of 25 e-rms

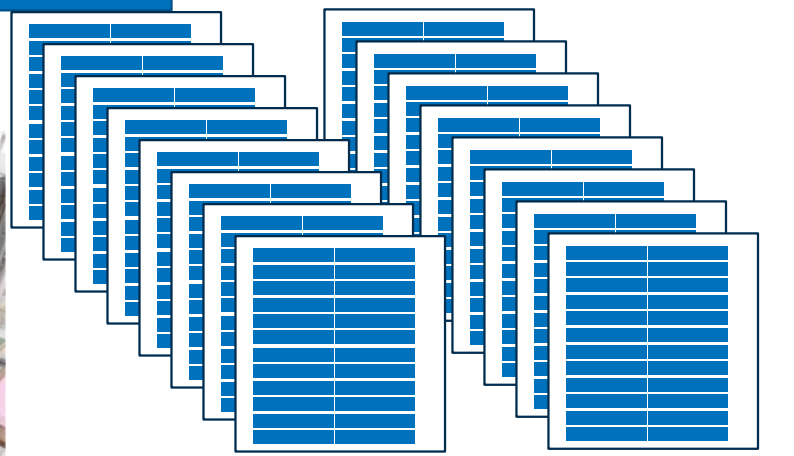
Energy spectrum of Mn fluorescence



Edge data reduction

20.2 Mpixel CITIUS detector



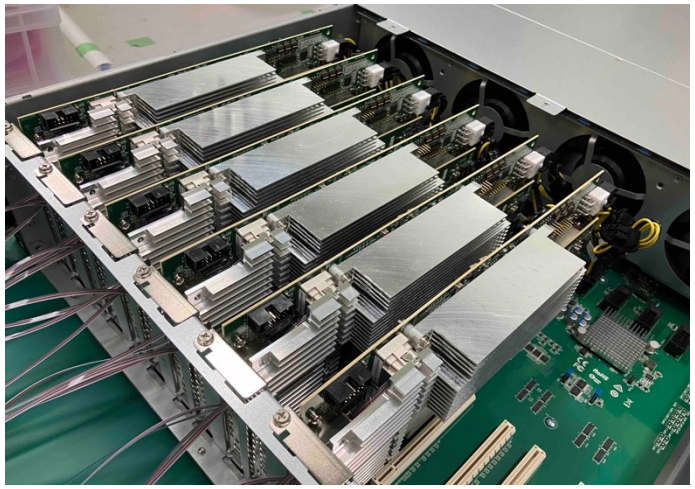


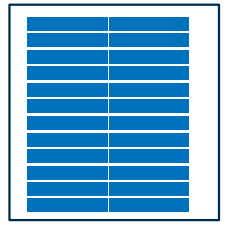
(14 bits/frame/pixel)

Optical fiber cables

➔ **107 GB/s**

Data-Framing Boards (DFBs) x 36 in Edge servers



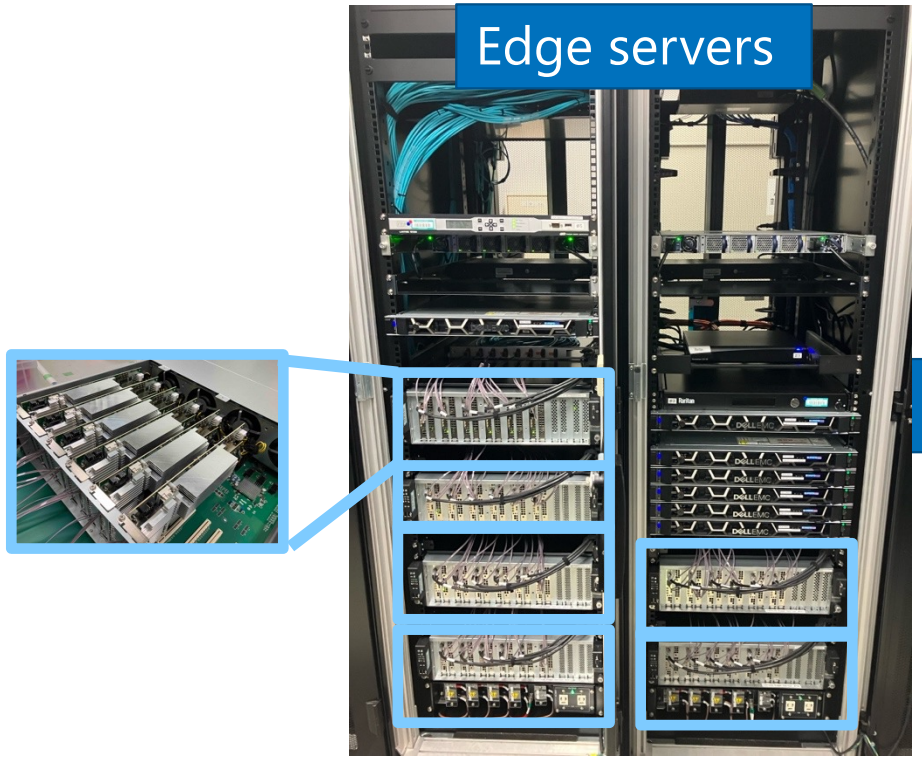


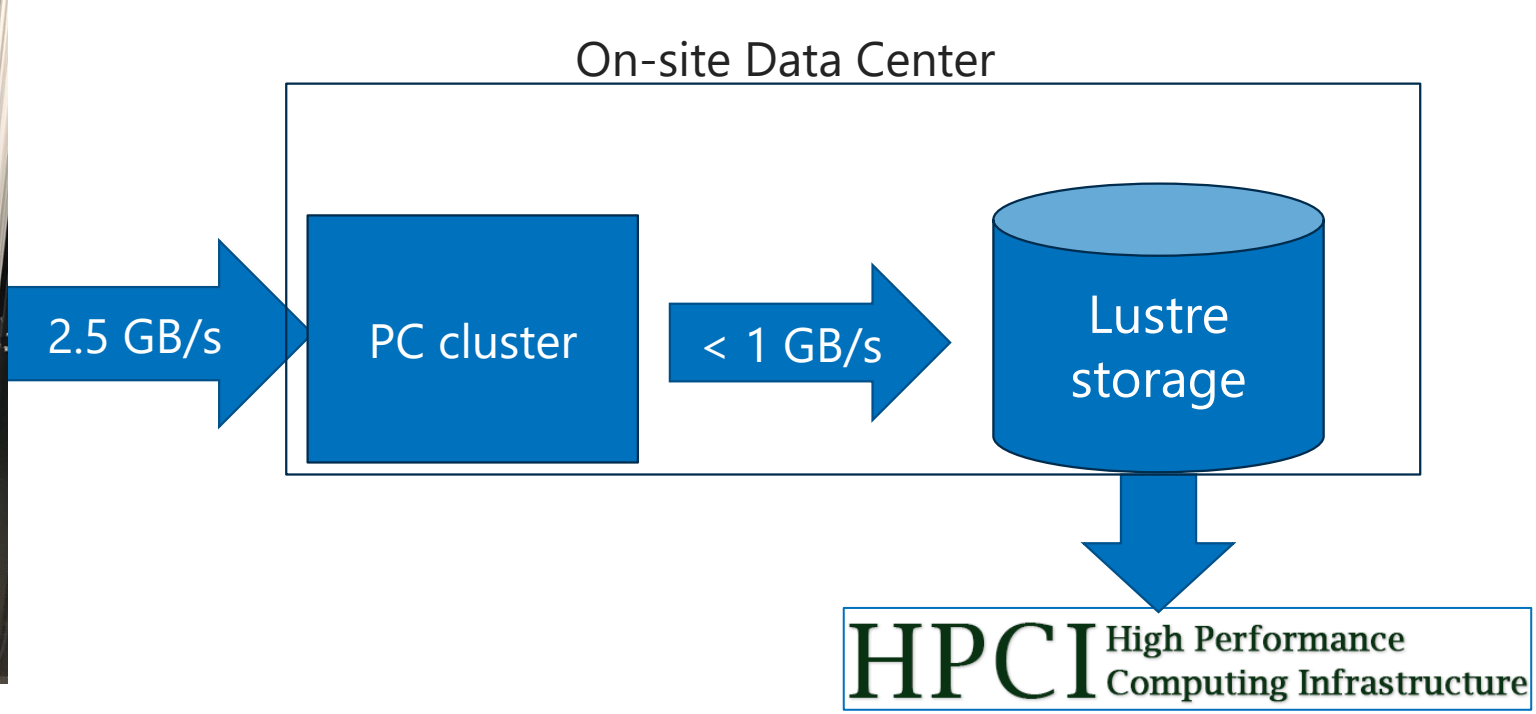
(32 bits/frame/pixel)

➔ **15 GB/s**

- Data reduction on FPGA
  - Data-framing Boards (DFBs) sum 16 frames into one summed frame
    - Output data rate of 15 GB/s

Data handling with hybrid data infrastructure



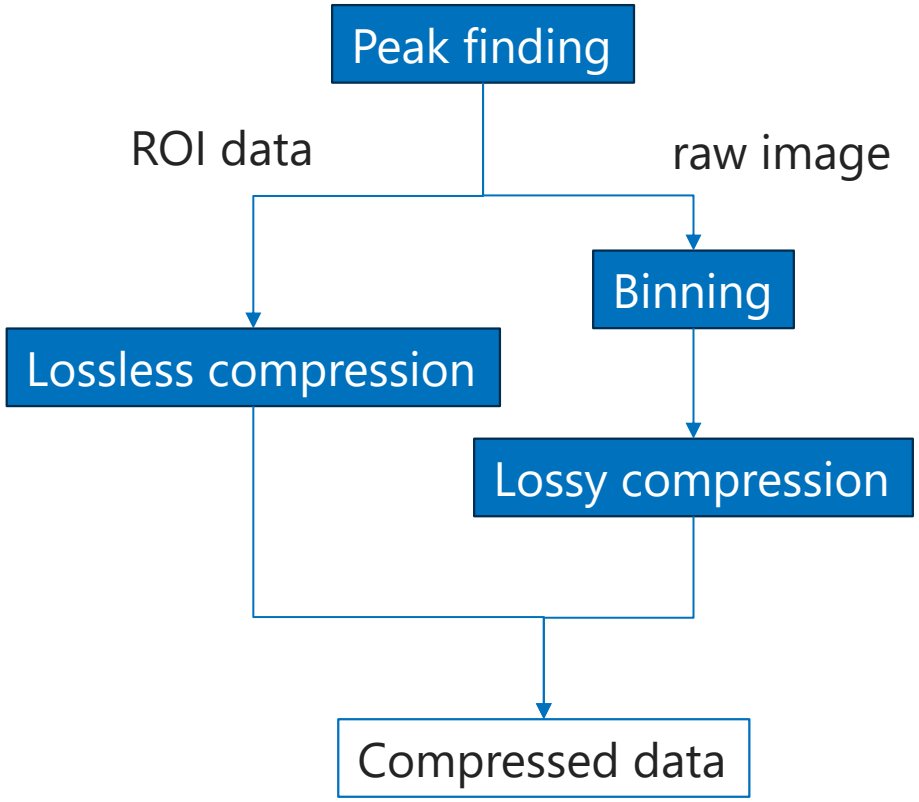


- On-the-fly Lossless data compression on the Edge servers
  - reduced data rate to on-site PC cluster: 2.5 GB/s
- Calibration and statistically-lossless compression (SLC) [1] on the on-site PC cluster
  - Compression ratio of >3
  - Estimated data rate: 7 PB / year
- Connected to offsite-HPC HPCI for CPU-intensive analyses

[1] T. N. Hiraki et al, in preparation

Next plan: ROIBIN-SZ

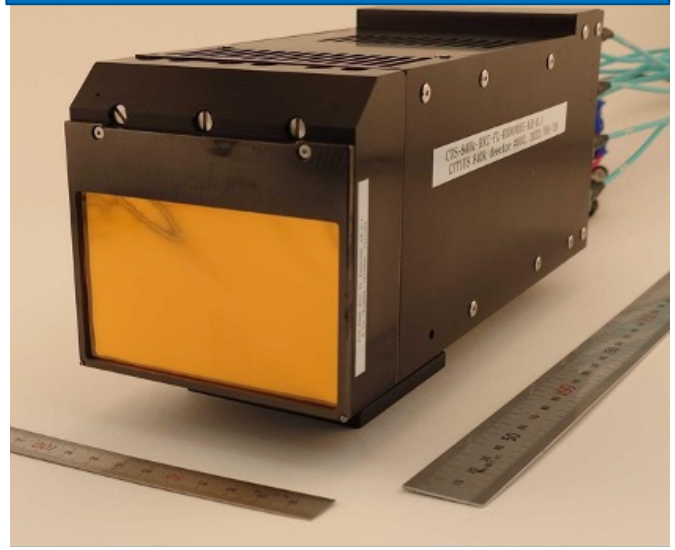
- Need for a higher compression ratio
- Plan to implement ROIBIN-SZ [1]
  - Science-preserving lossy compression
  - Developed for Serial Crystallography
  - Demonstrated Compression ratio
    - 196 on lysozyme data
- We will test the ROIBIN-SZ compression using 20.2 Mpixel CITIUS detector data from the coming feasibility study in July.



[1] R. Underwood et al., arXiv:2206.11297

CITIUS detectors at SPring-8

840 kpixel CITIUS detector



- CITIUS sensors run at a higher frame rate of 17.4 kHz for SR experiments
  - 70 TB/hour/sensor
- I will show the experience of experiments using an 840 kpixel CITIUS detector.

CITIUS for Quasi-Elastic Scattering at SPring-8

- **New method to directly observe microscopic molecular dynamics (Prof. Saito (Tohoku U.))**
  - Quasi-elastic scattering using a nuclear Bragg monochromator technique [1]
    - **X-ray diffraction + Mössbauer spectroscopy**
  - CITIUS is suitable because it is a 2D imaging, high-speed, and high quantum-efficiency detector.

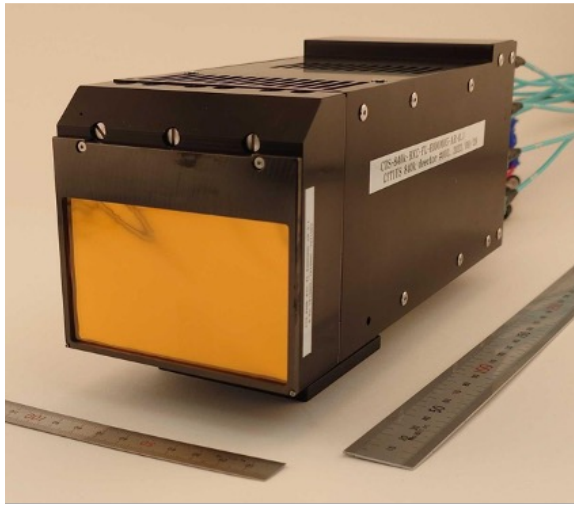
- **Five science experiments since 2021**
  - BL29XU and BL35XU of SPring-8

- **Requirement for data acquisition**
  - High-frame rate
  - Continuous and long data-taking

[1] R. Masuda et al., Jpn. J. Appl. Phys. **47**, 8087 (2008)

Generated data from 840 kpixel detector at SPring-8

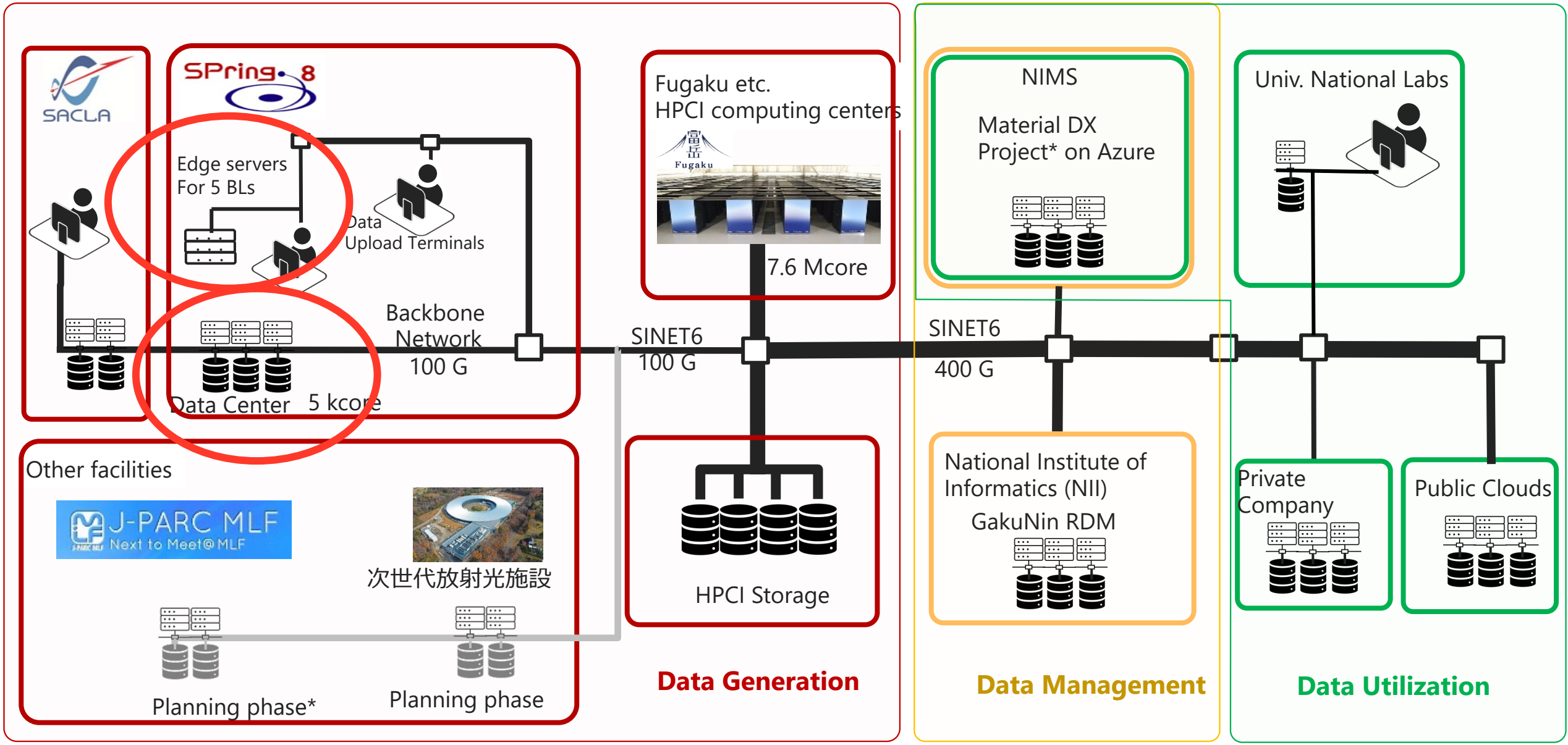
840 kpixel CITIUS detector



840 kpixels at 17.4 kfps
= 60 GB/s
= 0.22 PB/hour
= 5.1 PB/day

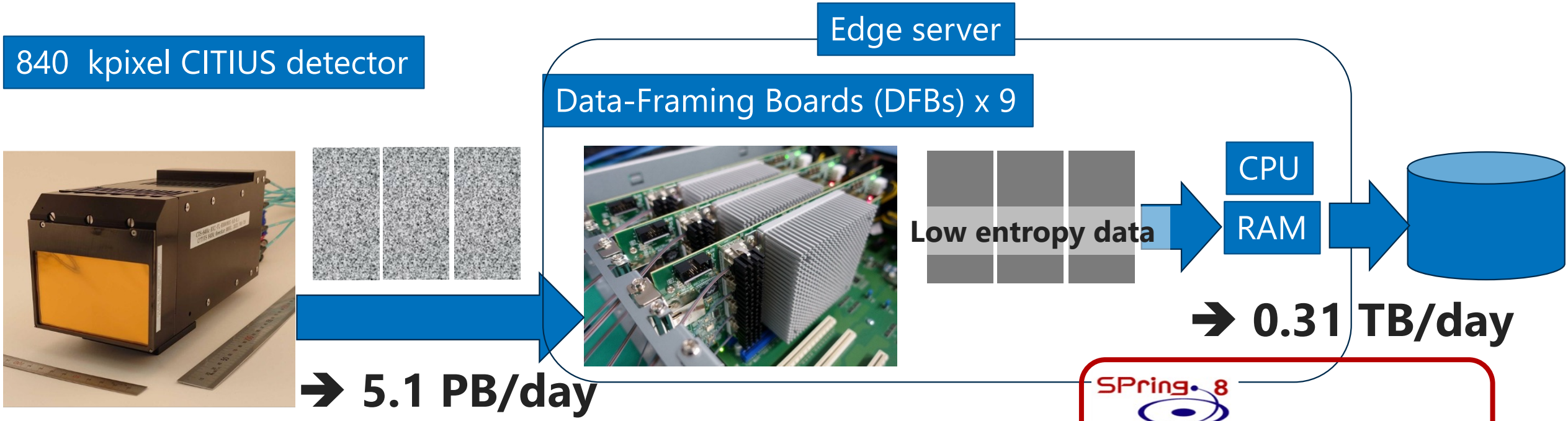
- A quasi-elastic scattering experiment requires continuous data-taking for a week of beamtime.
  - 35 PB / beamtime
- Data reduction/analysis under the SPring-8 Data Center Initiative in operation

SPring-8 Data Center Initiative

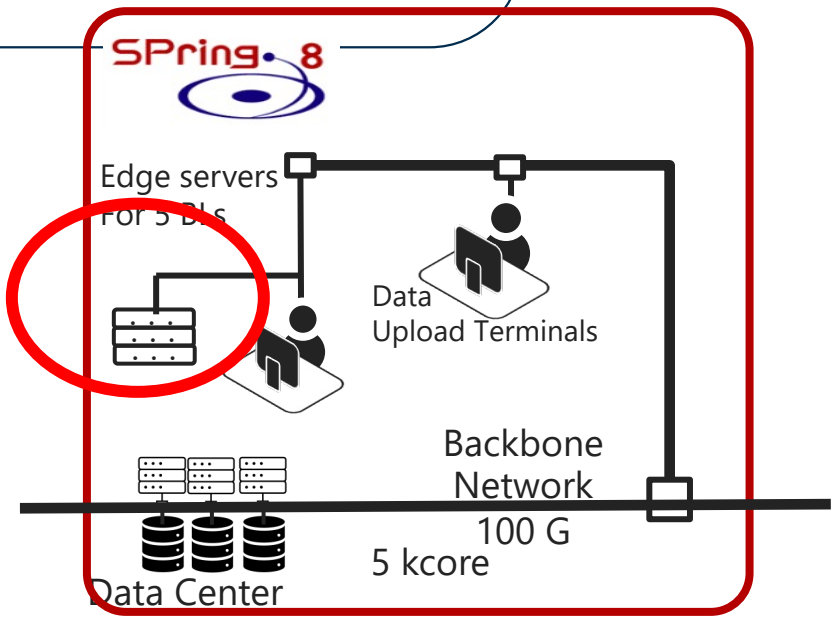


* To be implemented in the cloud

Data reduction for 840 kpixel CITIUS at SPring-8



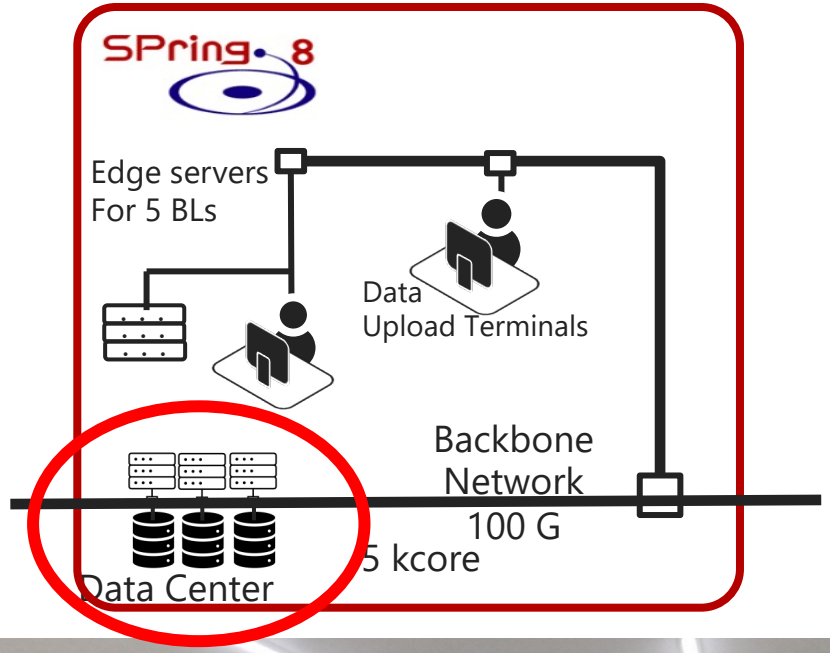
- **On-the-fly statistically-lossless compression (SLC) [1] with the combination of FPGA and CPU achieves a compression factor of > 1000.**
  - DFBs in the DAQ servers output low-entropy data
    - Background subtraction
    - Calibration (gain, non-linearity)
    - Setting pixel values below a certain threshold to zeros
    - Frame accumulation
  - Compression by CPU using a lossless algorithm

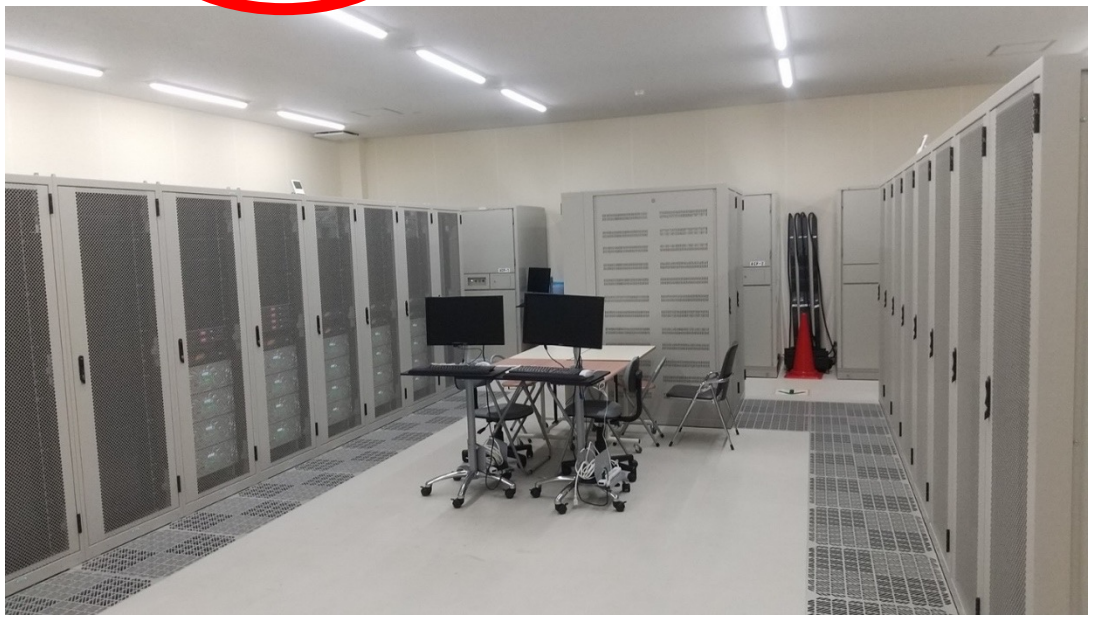


[1] T. N. Hiraki et al, in preparation

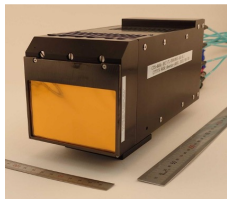
SPring-8 Data Center Initiative: Onsite

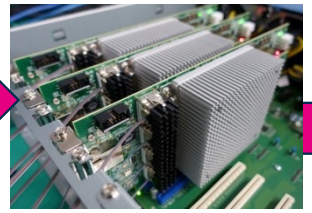
- Compute node
  - CPU nodes (64 nodes) 4k CPU cores
  - GPU nodes (16 nodes) 1k CPU cores, 111k CUDA cores
- Storage
  - 20 PB disk (logical)
- History
  - Installation completed in Mar 2023
  - General use of PC cluster system began in Nov 2023





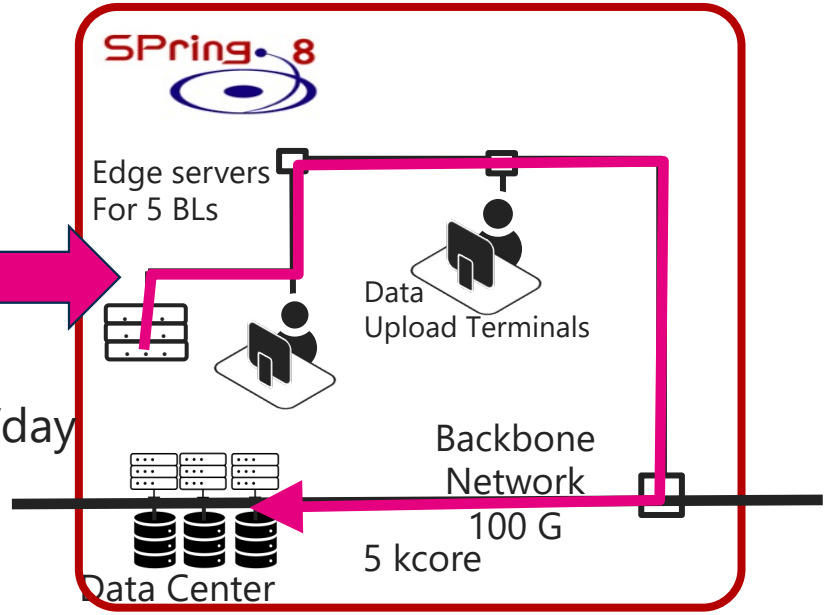
Data analysis integrated with SP8DCI

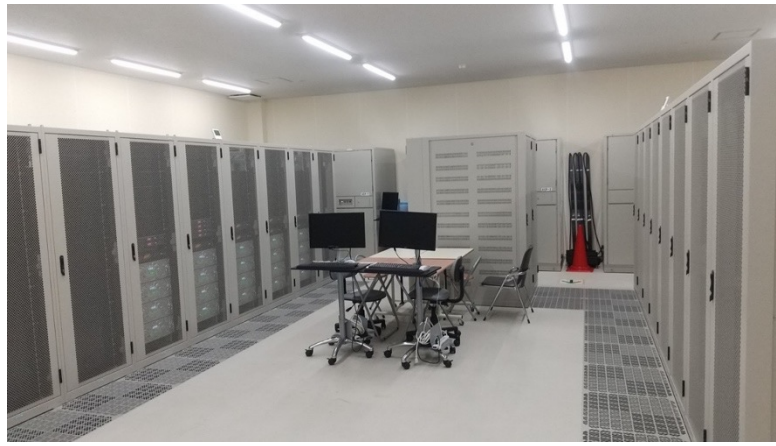




5.1 PB/day

0.31 TB/day





- Compressed data are sent to the on-site data PC cluster through a 100G network while taking data.
- Users obtained analysis results using the on-site pc clusters during beamtime.
  - Fast enough to analyze data at the uncompressed data rate of 5.1 PB/day

Results coming from CITIUS@BL35XU SPring-8

S-crosslinked polybutadiene measurement

2D scattering image

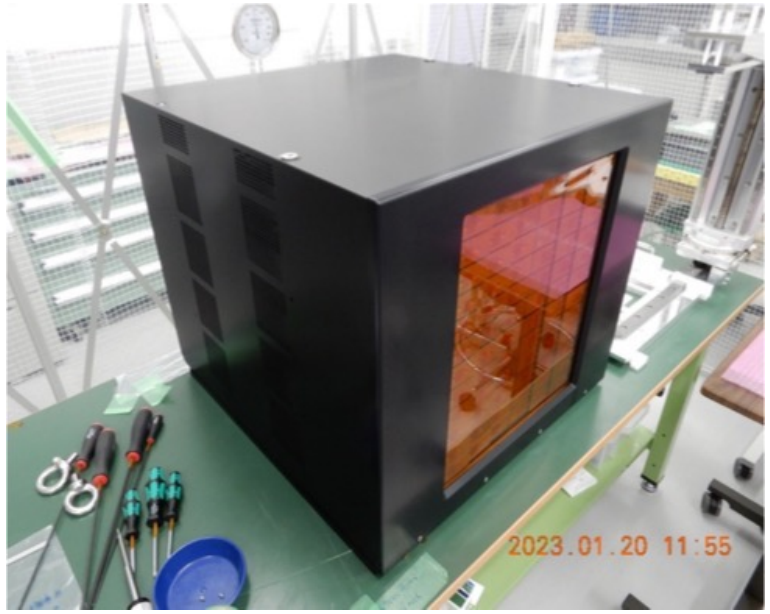
Energy spectrum of scattered photons

Intermediate scattering function

The new system successfully observed the microscopic dynamics in the wide timescales ranging from 300 ps to 30 ns.

We have already conducted five science experiments using this data-handling scheme.
The users obtained scientific results using compressed data.

Future developments



- Target
  - Running a 20.2 Mpixel system with 17.4 kfps
    - generating 122 PB/day of data
- Future developments under investigation
  - Distributed low-latency storage infrastructure by all-photonics network (IOWN)
  - Upgrade of the DFBs with Agilex9 FPGA
    - 1st-generation DFBs are equipped with three Arria10 FPGAs

Summary

- **Data reduction for XFEL and SR experiments**
  - CITIUS detectors for diffraction generate large amounts of data.
  - We have prepared a hybrid data infrastructure with edge computing, on-site PC cluster, and off-site HPC.

- **SACLA**
  - 20.2 Mpixel detector generates 106 GB/s of data
    - FPGA reduces the data rate to 15 GB/s
  - Plan to implement ROIBIN-SZ compression

- **SPring-8**
  - 840 kpixel detector generates 60 GB of data
  - On-the-fly SLC data compression on edge achieved a data compression ratio of over 1000.
  - Data analysis integrated with SPring-8 Data Center Initiative.
  - Science results have been obtained from the compressed data.

Acknowledgements

- RIKEN and JASRI Team
  - K. Ozaki, Y. Honjo, T. N. Hiraki, K. Kobayashi, T. Kudo, T. Sugimoto, M. Yamaga, T. Kameshima, K. Nakajima, N. Yasuda, Y. Inagaki, K. Fujiwara, T. Nakagawa, Y. Oyaki, M. Kimoto, M. Nakamachi, K. Motomura, Y. Joti, T. Hatsui, M. Yabashi, T. Ishikawa
  - I. Inoue, J. Kang, Y. Inubushi, S. Owada, T. Yabuuchi, K. Tono
  - Y. Yoda, N. Nagasawa
- RIKEN Center for Computational Science (R-CCS)
  - S. Matsuoka, K. Sato, K. Sano, F. Shoji, and their division members
- Tohoku Univ.
  - M. Saito, M. Kobayashi
- Private Companies
  - Sony Semiconductor Solutions Corporation
  - GLORY System Create Ltd.
  - GLORY Mechatronics Ltd.
  - NIPPON GIJUTSU CENTER Co. Ltd.
  - Meisei Electric Co. Ltd.
  - JEPICO Corporation
  - Tokyo Electron Device Limited
  - Ryobi Systems Co. Ltd.