

KEK Site Report

**T. Nakamura, G. Iwai, K. Murakami,
T. Sasaki, S. Suzuki, W. Takase**

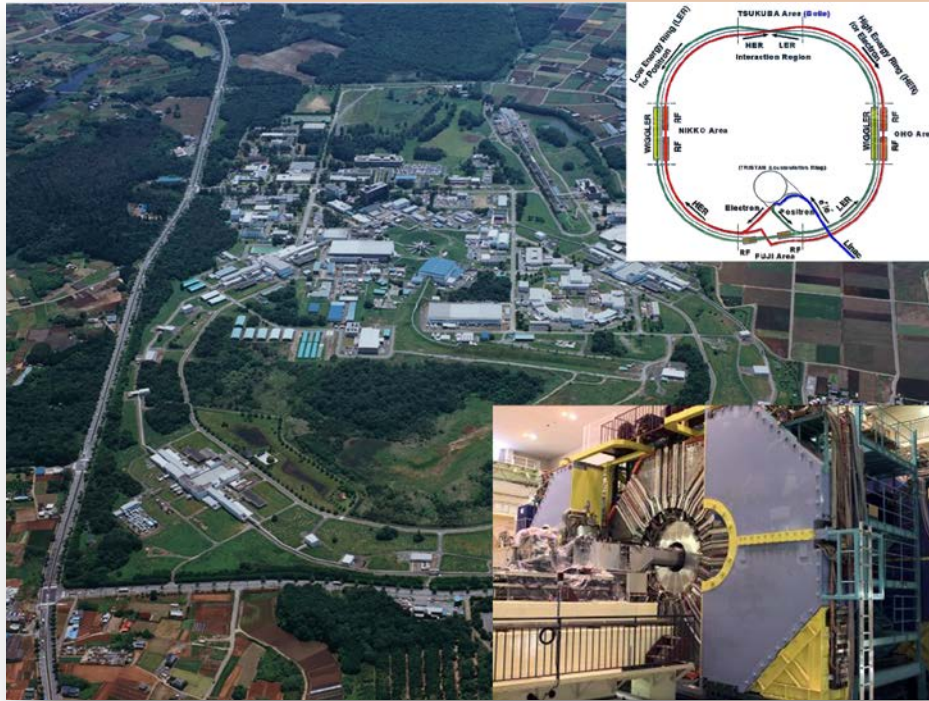
Computing Research Center
Applied Research Laboratory
HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION, KEK



Computing Research Center



SuperKEKB: e^+e^- intensity frontier



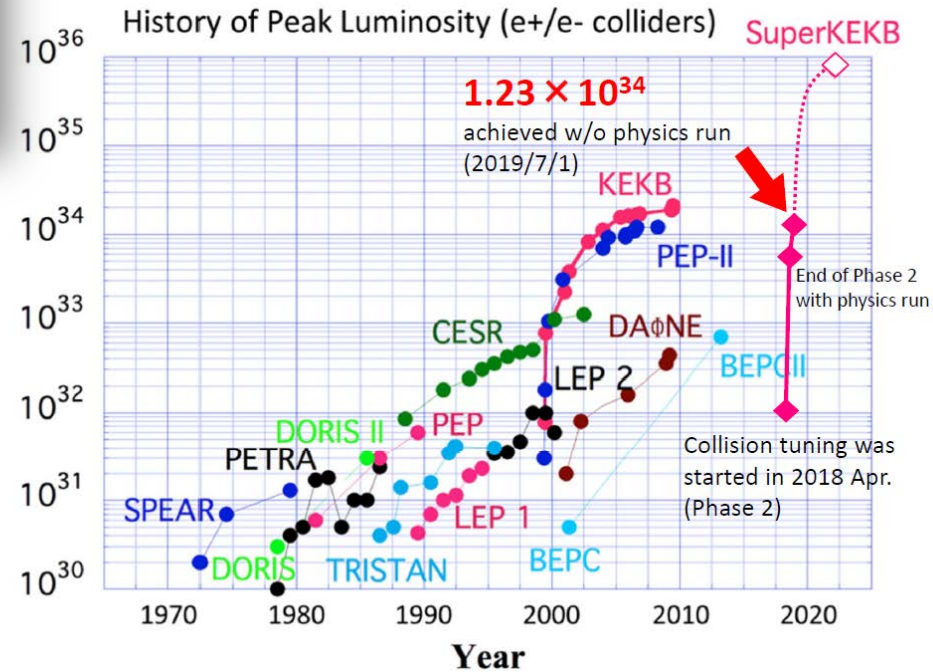
The first collision (Apr. 26th, 2018)



Y. Yusa

Design Luminosity: $8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$
factor 40 higher than the previous KEKB

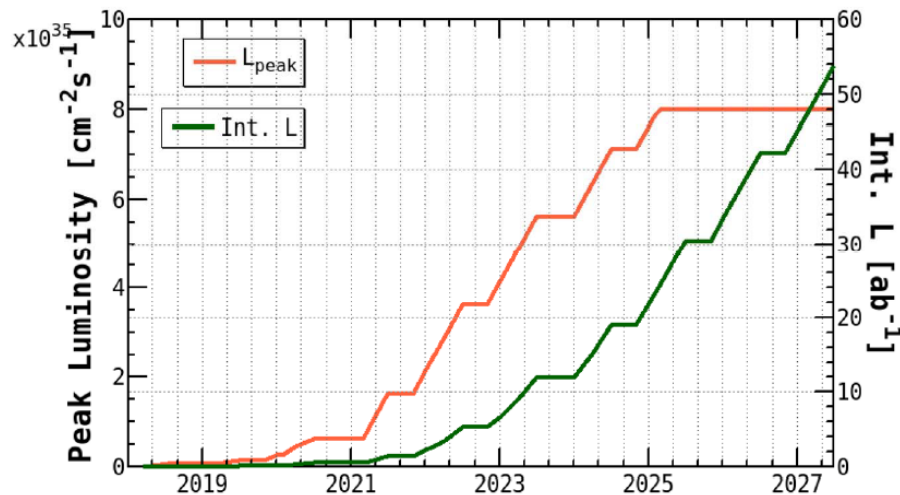
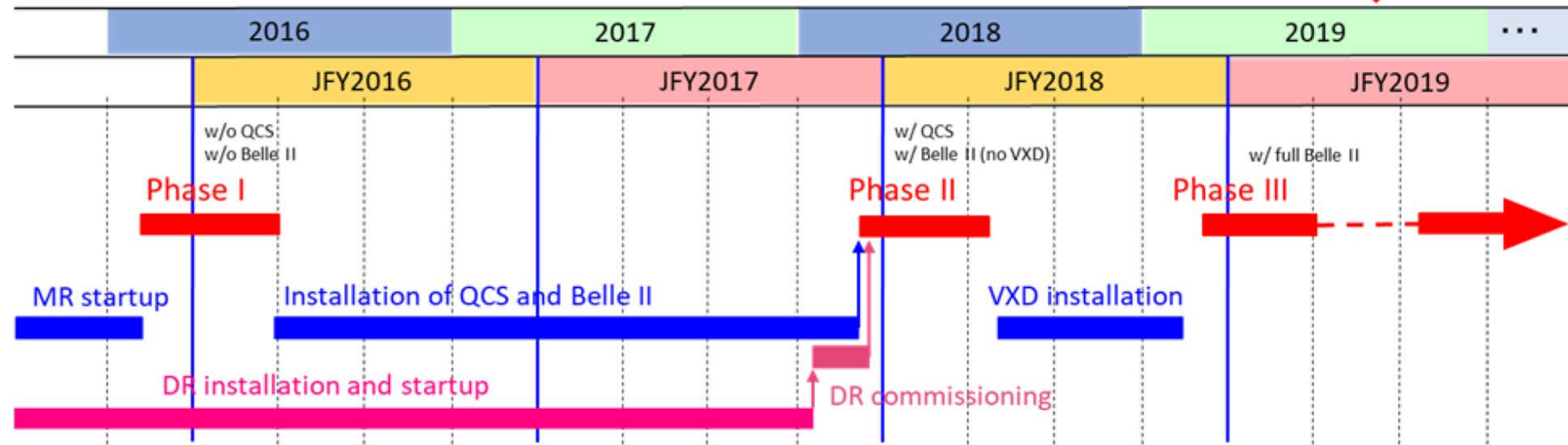
Luminosity [$\text{cm}^{-2} \text{ s}^{-1}$]



H. Ikeda



Schedule of SuperKEKB/Belle II



H. Ikeda

The Phase II operation was over on Jul. 18th, 2018. The first collisions were observed during the Phase II.

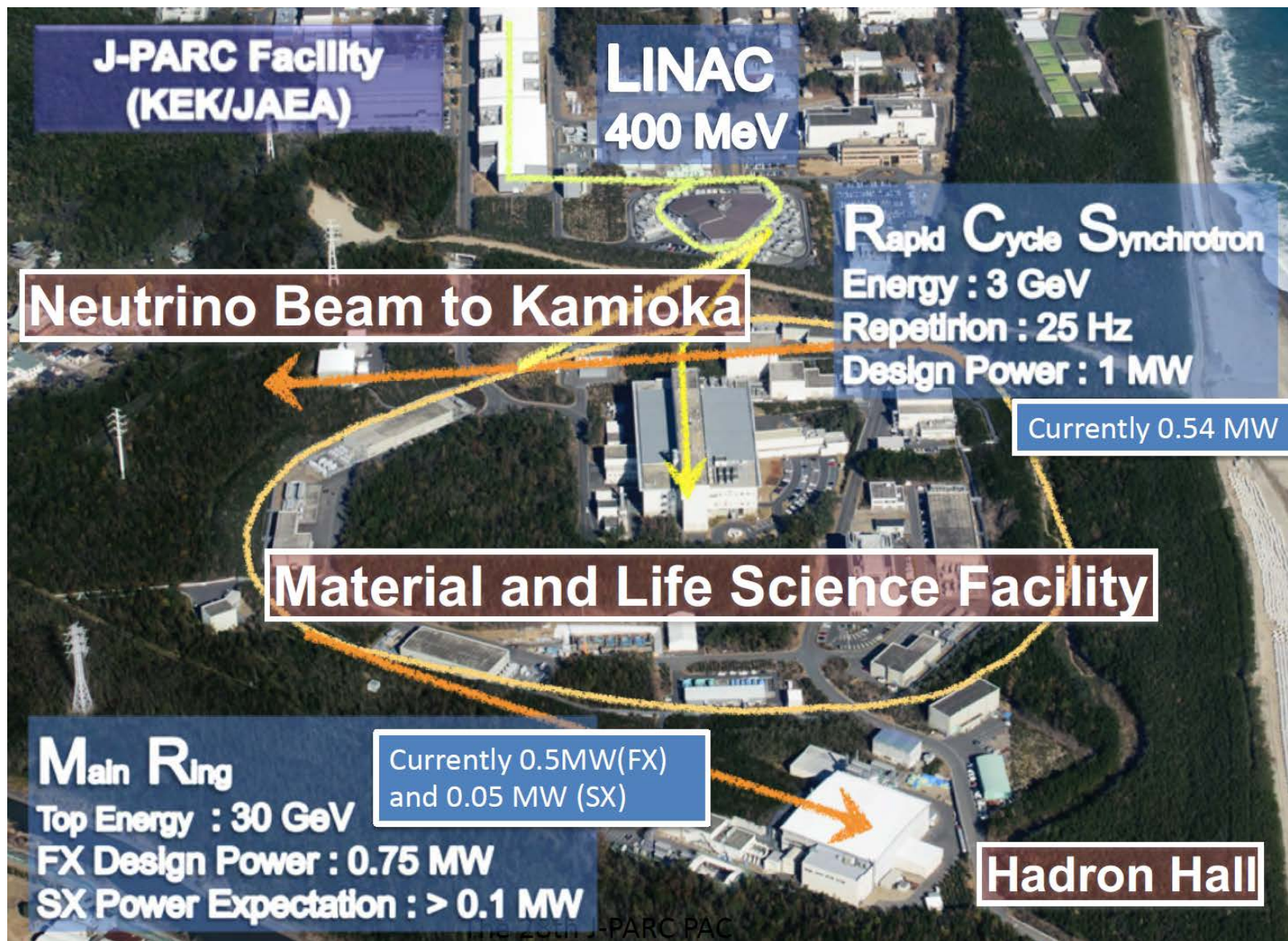
The beam operation has been re-started with full Belle II detector (silicon VTX) as Phase III run on Mar. 11, 2019.

After the summer shutdown period, SuperKEKB will re-start the operation from Oct. 15th, 2019.

Final goal of the accumulating statistics is 50 ab^{-1} which corresponds to roughly 100PB of raw data. (50 times larger data than the Belle experiment)

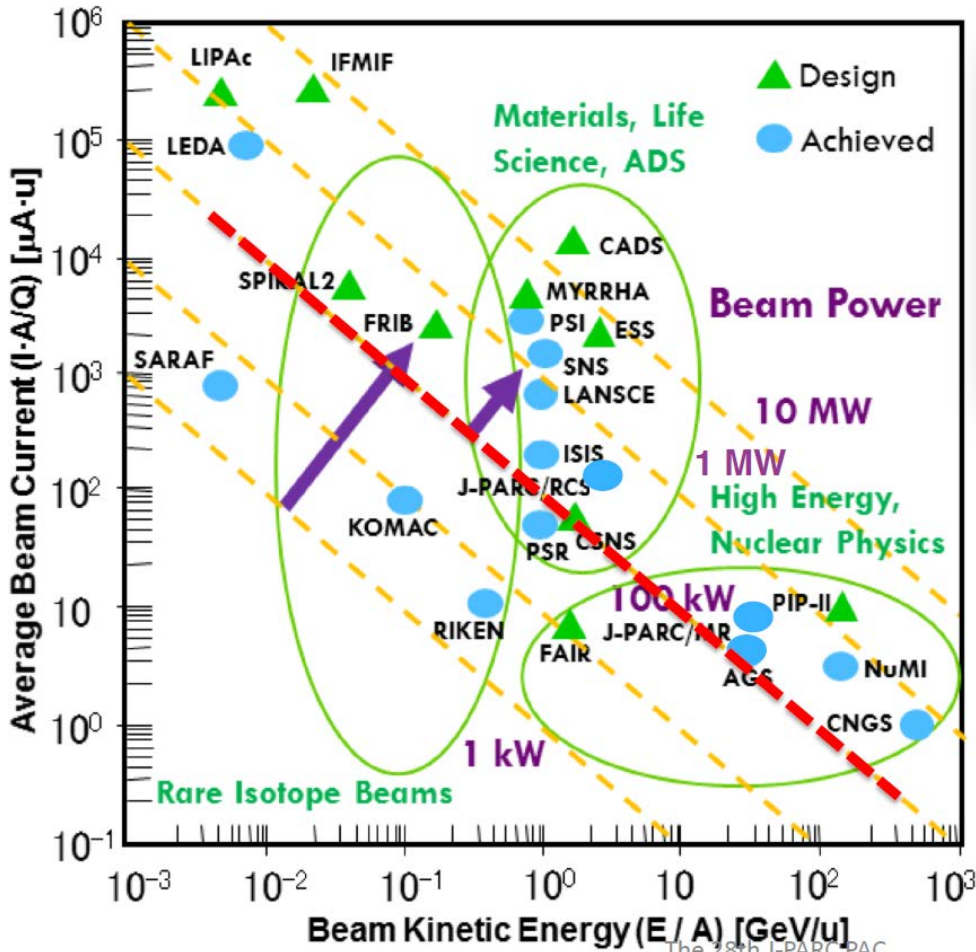


J-PARC: Japan Proton Accelerator Research Complex





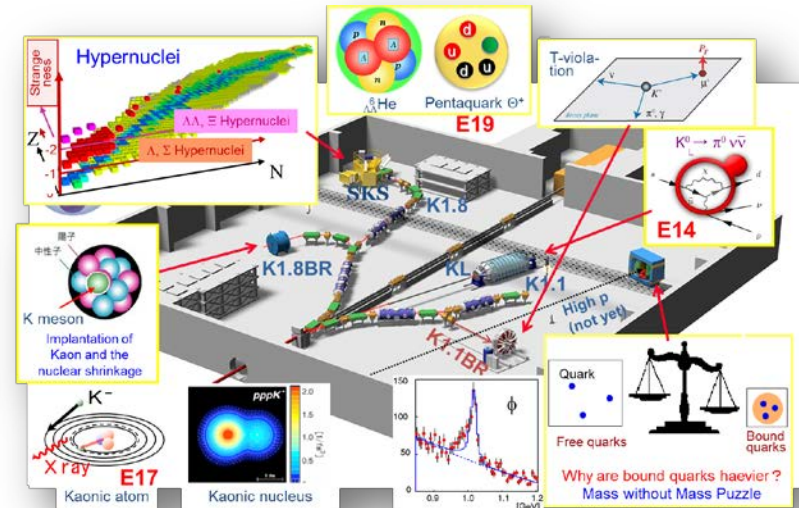
J-PARC: proton intensity frontier



Fast extraction



Slow extraction



Jie Wei / Y. Yamazaki

The 28th J-PARC PAC



J-PARC upgrade



Mid-term plan of MR

FX: The higher repetition rate scheme : Period 2.48 s → 1.3 s for 750 kW.
(= shorter repetition period) → 1.16 s for 1.3 MW

SX: Mitigation of the residual activity for 100kW

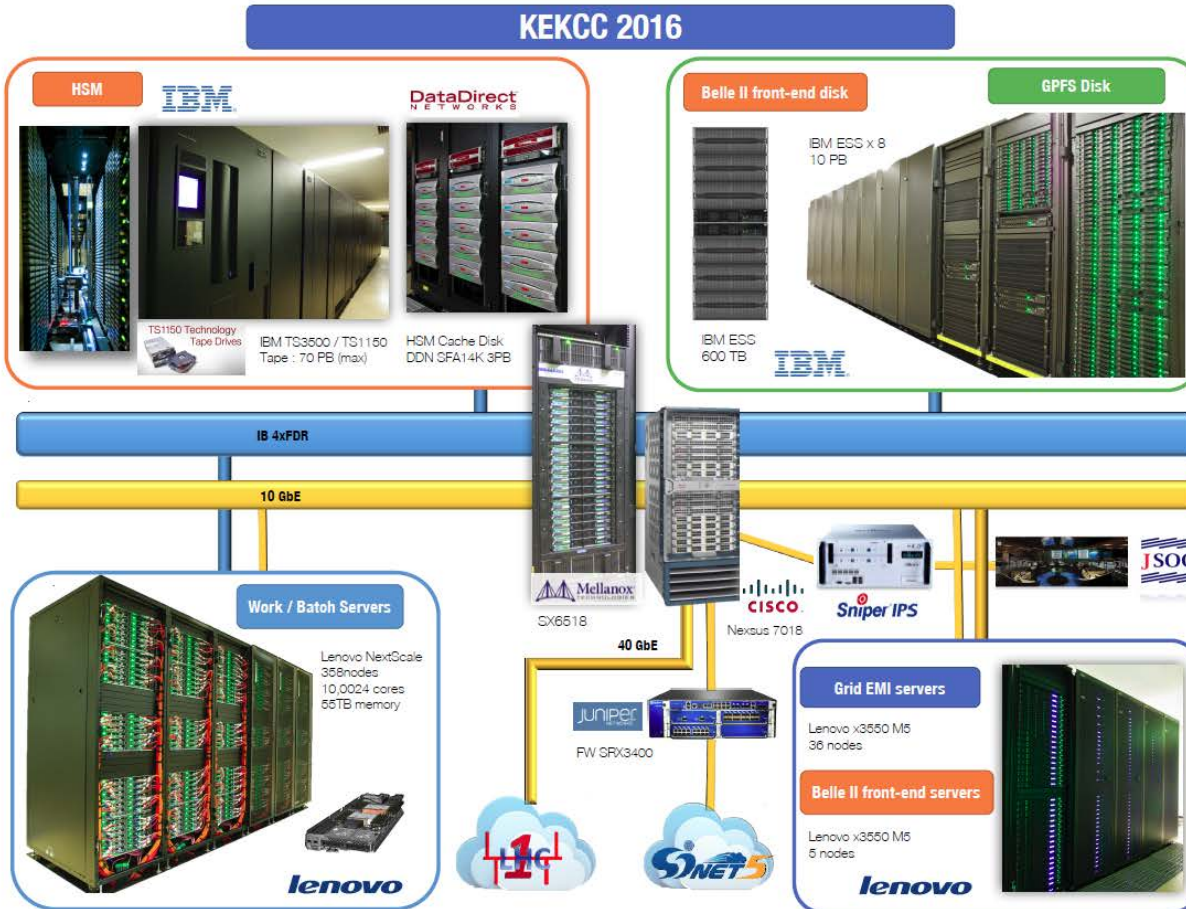
JFY	2017	2018	2019	2020	2021	2022	2023	2024
Event	New buildings		HD target		Long shutdown			
FX power [kW]	475	>480	>480	>480		>700	800	900
SX power [kW]	50	50	50	70		>80	>80	>80
Cycle time of main magnet PS	2.48 s	2.48 s	2.48s	2.48s		1.32s	<1.32 s	<1.32 s
New magnet PS	Mass production installation/test							
High gradient rf system	Manufacture, installation/test							
2nd harmonic rf system	Manufacture, installation/test							
Ring collimators	Add.collimators (2 kW)				Add.coll. (3.5kW)			
Injection system	Kicker PS improvement, Septa manufacture /test							
FX system	Kicker PS improvement, FX septa manufacture /test							
SX collimator / Local shields	Local shields							
Ti ducts and SX devices with Ti chamber	Ti-ESS-1	(Ti-ESS-2)						

F. Naito



KEKCC: KEK Central Computer System

Launched at Sep. 2016, All system component has been fully in production. No major upgrade in terms of the HWs since then. Quite a stable phase.



SYSTEM RESOURCES

- CPU :** 10,024 cores
- ❑ Intel Xeon E5-2697v3 (2.6GHz, 14cores) x 2 358 nodes
 - ❑ 4GB/core (8,000 cores) / 8GB/core (2,000 cores) (for app. use)
 - ❑ 236 kHS06 / site

Disk : 10PB (GPFS) + 3PB (HSM cache)

Interconnect : IB 4xFDR

Tape : 70 PB (max cap.)
HSM data : 8.5 PB data, 170 M files, 5,000 tapes

Total throughput : 100 GB/s (Disk, GPFS), 50 GB/s (HSM, GHI)

JOB scheduler : Platfrom LSF v9

K. Murakami



Breakdown of CPU consumption



Compute node

CPU: Intel Xeon E5-2697v3 (2.6GHz, 14cores) x 2
 358 nodes, 10,024 cores, 236kHS06/site

Memory: 4GB/core (8,000 cores)
 8GB/core (2,000 cores)

Storage

Disk: 10PB (GPFS, IBM ESS x8 racks)
 3PB (HSM cache)

Interconnect: InfiniBand 4xFDR (56 Gbps)

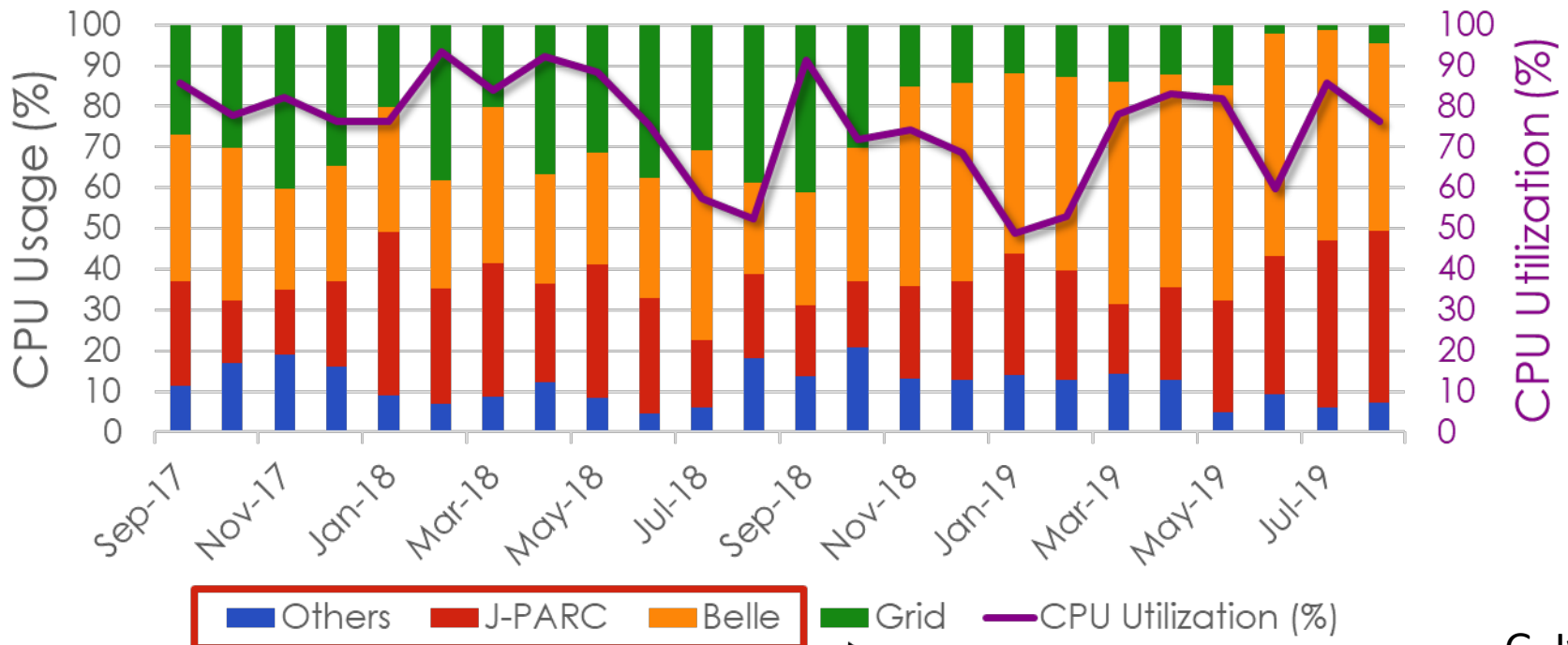
Tape: 70 PB (max cap.)

CPU usage: breakdown by groups,
 normalized by the total CPU usage per month

CPU usage has been reached more than **90 %** of total resource

Throughput

100 GB/s (Disk, GPFS), 50 GB/s (HSM, GHI)



Local batch jobs

Belle2 jobs are dominant.

G. Iwai

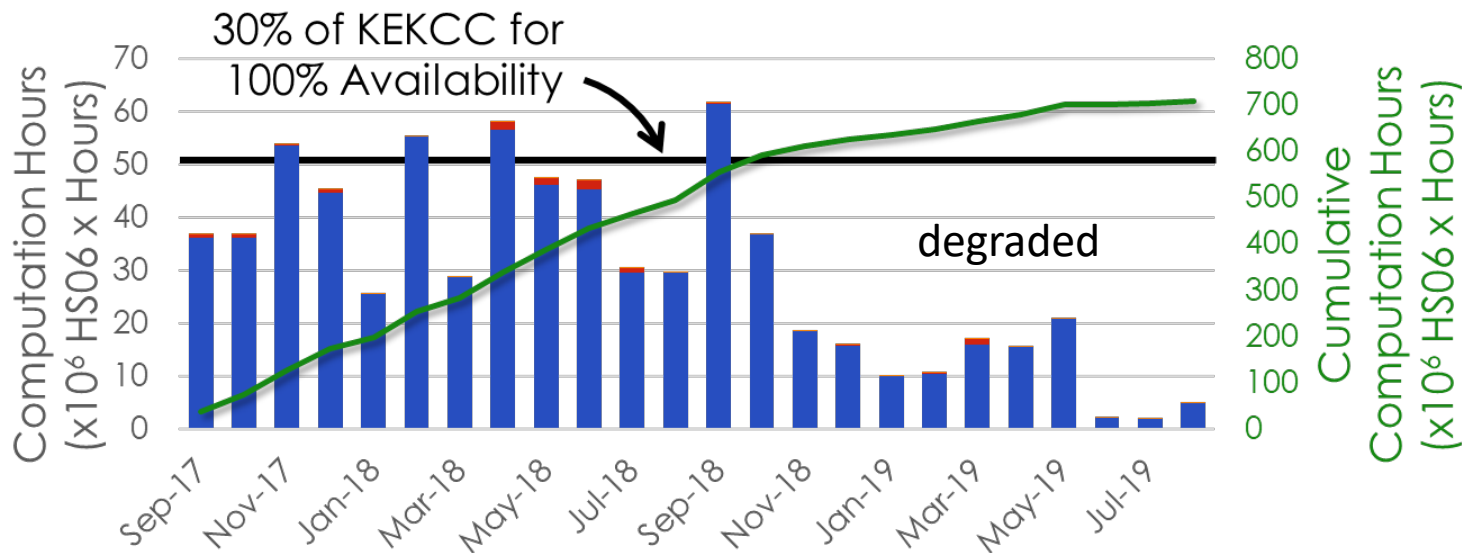


Grid Jobs and Data

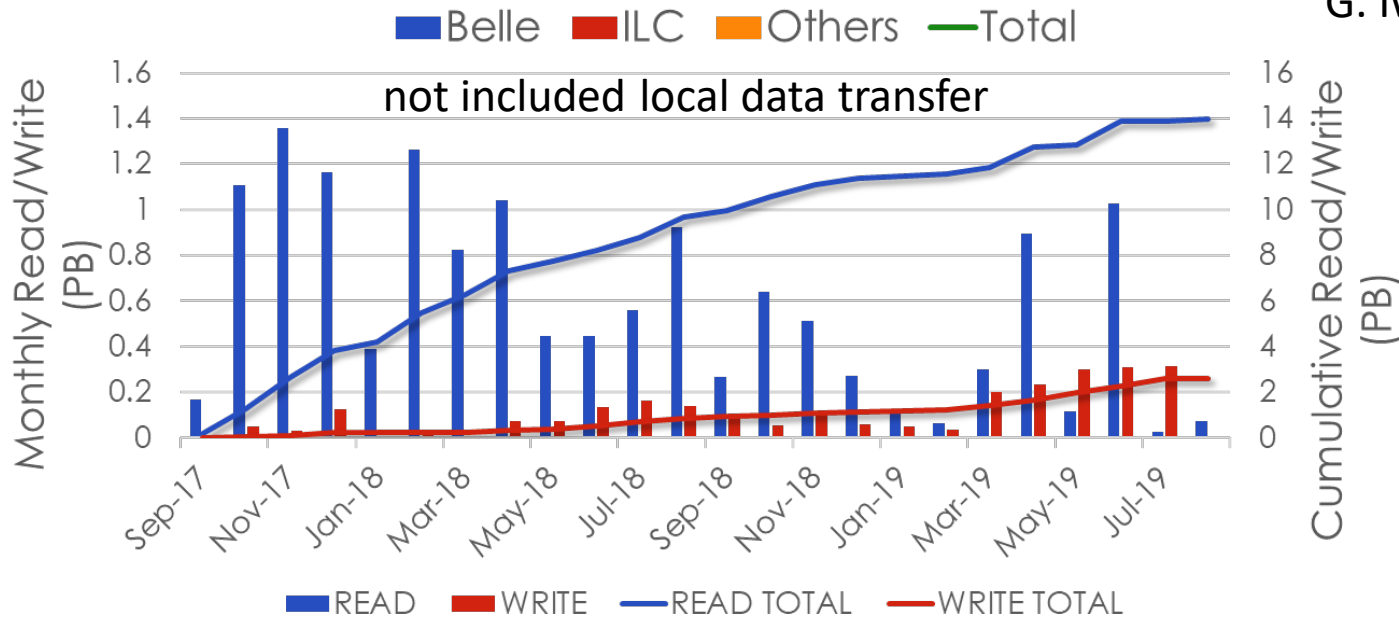


Grid Jobs

168M HS06 hour/month
(23.5 HS06/core)



Grid storage read/write external data transfer



G. Iwai



The CPU consumption of the grid jobs was degraded due to the massive data staging requests from the tape archive system.

- **HSM cache made by the GHI was filled up and files were purged so quickly from the cache.**
- **A lot of jobs occupied job slots with no CPU time, and then timed out.**
- **This problem infected also some grid service, for example, the pile-up of data transfer requests in GridFTP servers and FTS, and then service down.**

This problem is already cleared by several countermeasures:

- **Files repeatedly used were copied to the disk only storage a priori.**
- **Files to be used by the planned data production were pined at the HSM cache.**
- **A scheme (so-called hstage) for the organized data staging has been newly introduced for the user request.**

Now, system is in the normal operations.



System upgrade in 2020

The current KEKCC is three years old system. System replacement and upgrade are planned at the next summer in 2020 so that it can accommodate user's and many group's requirement for the computing resources.

KEKCC is the total system for the KEK projects including:

Analysis and Data processing and MC production, Grid (UMD, iRODS), Grid-CA, e-mail, Mailing list, Web, Indico, Wiki, Online storage etc.

Therefore, the formulation of the specification takes a long time.

The specification have been almost finalized recently.

- **The number of CPU cores will be increased.**
- **Storage capacity will be enlarged for both HSM cache and disk only distributed storage.**
- **Base operating system.**
- **Grid computing element (replacement from CREAM-CE).**
- **Reinforcement of wide area data transfer capability and functionality.**
- **Improvement of network bandwidth for the Grid storage element.**
- **IPv6 and Jumbo frame support for the Grid data transfer.**

The detail specification will be reported at the next HEPiX workshop at Taipei.

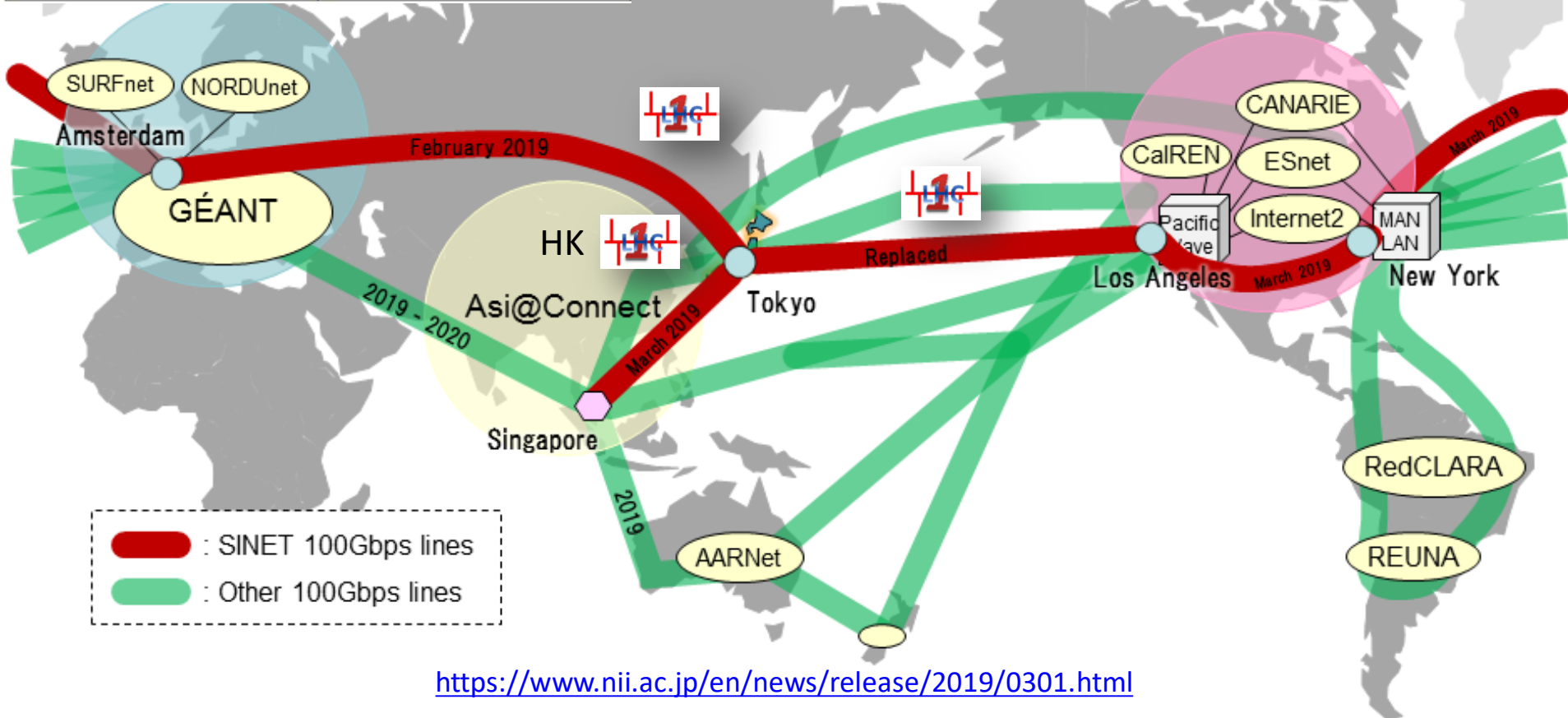


Connectivity of International network



Feb. 2019: 20 Gbps to 100 Gbps to Amsterdam
 Mar. 2019: 100 Gbps to New York via Los Angeles
 Mar. 2019: 100 Gbps New York to Amsterdam
 Sep. 2017: LHCONE for Asian sites at HK by JGN

Planned to connect with AARNet by LHCONE at Singapore



<https://www.nii.ac.jp/en/news/release/2019/0301.html>



Summary

The KEK central computer system (KEKCC) has already become the fourth years operation. No major upgrade was made in terms of hardware. The system was in the quite stable phase.

The problem due to the massive requests of the data staging affected to the degradations of the effective CPU consumption of the Grid jobs. The problem was already cleared by several countermeasures.

The formulation of the next KEKCC specification is going to be finalized soon. We would like to report on the details of the specification and the construction work at the next HEPiX workshop.