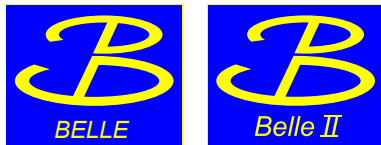




## ***Belle I & II / KEK***

- 1. Belle I data and analysis preservation status*
- 2. discussion on Belle II data and analysis preservation*
- 3. data and analysis preservation of experiments at KEK*

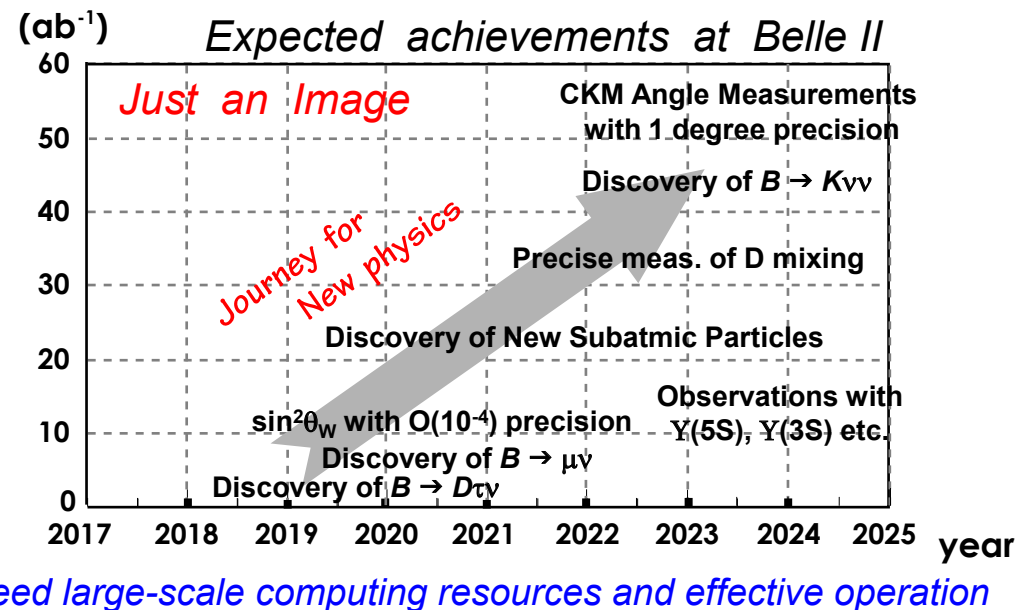
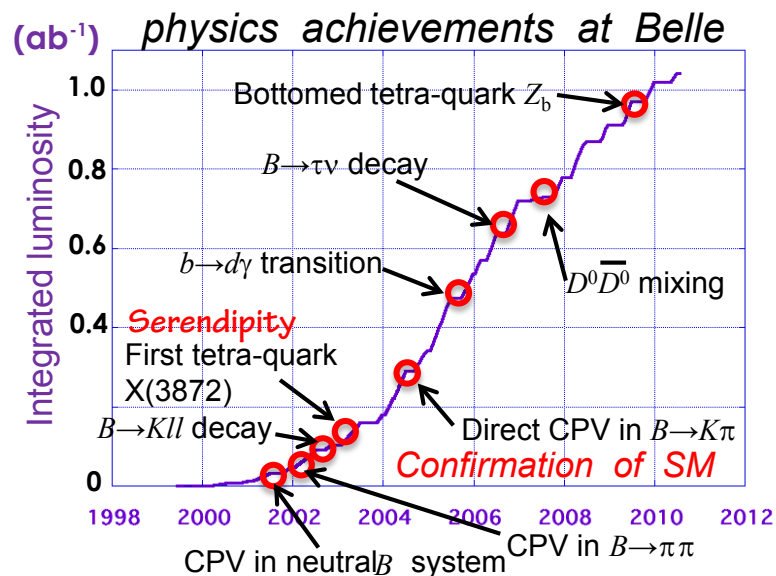
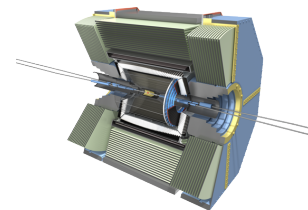


*June 23, 2021 @ the 3rd DPHEP*  
*Takanori HARA (KEK IPNS)*  
*takanori.hara@kek.jp*

# Motivation from Physics

Accelerator	KEKB	SuperKEKB
Beam Energy (GeV)	3.5 x 8 (g = 0.425)	4 x 7 (g = 0.28)
CM energy	....., Y(4S), .....	....., Y(4S), .....
Luminosity ( $\text{cm}^{-2} \text{s}^{-1}$ )	$2.1 \times 10^{34}$	$\sim 6 \times 10^{35}$
Total data ( $\text{ab}^{-1}$ )	1	50
raw data : ~1PB		raw data : ~50PB (in total ~200PB including physics datasets and replicas)
Computing	one big center @ KEK (non-grid)	world-wide distributed computing

$\xrightarrow{\text{dozens of times}}$   
 $\xrightarrow{\times 50}$   
**Higher intensity**



# Belle Data preservation policy

## Beam Data

RAW data	960TB
mDST	130TB

## MC

mDST	600TB
------	-------

(10 streams BB + 6 streams udst)

mDST is stored on both Disk and Tape

RAW data is stored on Tape

## Belle collaboration

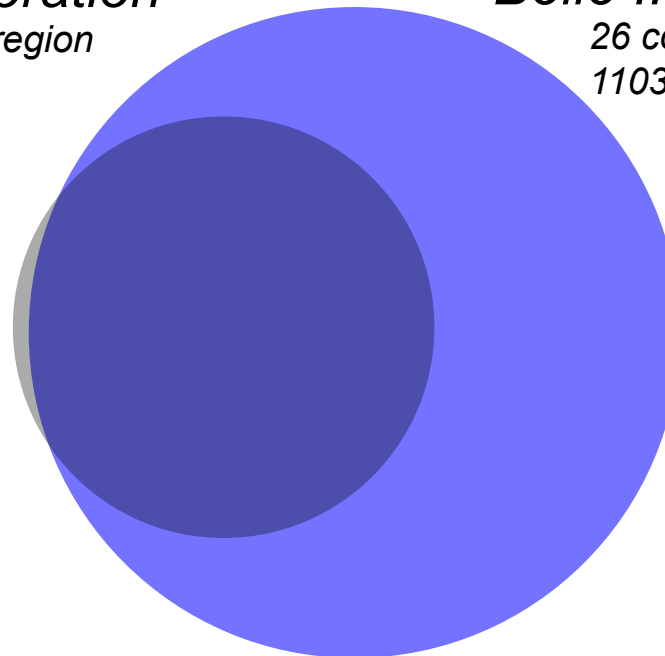
23 countries and region

461 members

## Belle II collaboration

26 countries and region

1103 members



## Belle I Data preservation policy :

- + preserve all RAW and mDST
- + Keep the current software environment as much as possible
- + Data will be used by the Belle community until Belle II data exceeds Belle I data

*The Belle management and the Belle II management agreed to make the Belle data and software accessible to Belle II collaborators*

# Belle Analysis preservation strategy

*Belle I Data preservation policy is*

- + preserve all RAW and mDST
- + Keep the current Belle analysis software framework ("basf") as much as possible
- + Data will be used by the Belle community until Belle II data exceeds Belle I data

*No DST / MC mass-production is planned anymore*

*mDST level datasets are enough for physics analysis*

*Offline software has been frozen*

*no update of the detector simulation (based on Geant3)*

but, *analysis users like to use new event generators and new analysis tools*

*The number of people who knows Belle software is decreasing...*

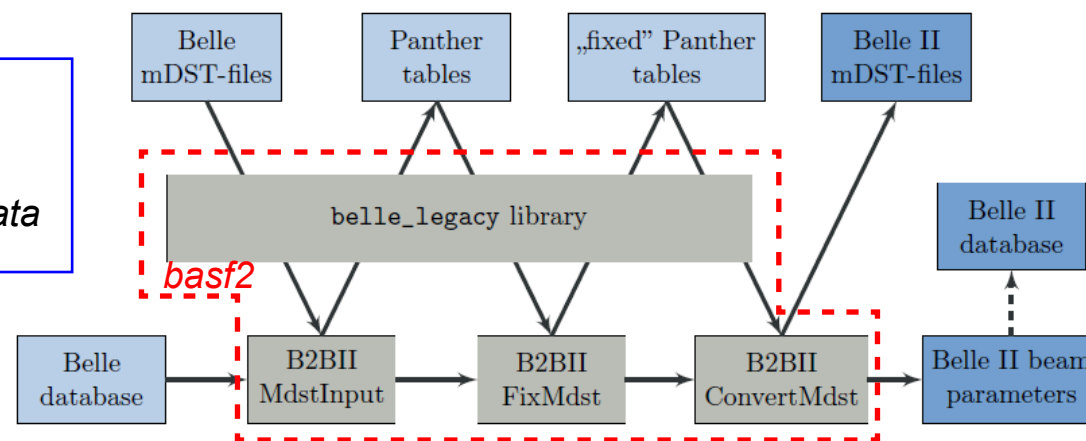
*But still "basf" is required  
for signal MC production...*

*Belle II event generator output can be fed to  
Belle detector simulation in basf*

*Belle data is converted to Belle II format*

*New analysis technologies and ideas can be applied  
Belle II analysis software ("basf2") can be checked with Belle data*

M. Gelb, et al. "B2BII - Data conversion from Belle to Belle II"  
<https://arxiv.org/abs/1810.00019>



# History of Belle computers

*Belle computing model is established on the KEK local computing system  
(data archive, software development, analysis environment, etc.)*

	1999- (4 years)	2001- (5 years)	2006- (3 years)	2009- (3 years)	2012/4- (~4.5 years)	2016/9- (4 years)	2020/9- (4 years, planned)
<b>CPU</b> [SI2K ---> HS06]	~100 SI2K	~1200 SI2K	~42500 SI2K	~115200 SI2K	~40 kHS06 (~3500 cores)	~240kHS06 (~10,000 cores)	~274kHS06 (~15,000 cores)
<b>Disk</b> [TB]	4	9	1000	1500	7000	10000 + 3000 HSM cache	17000 +8500 HSM cache
<b>Tape</b> [TB]	160	620	3500	3500	16000	70000	100000

(Belle dedicated) (Belle dedicated) (Belle dedicated) (Belle dedicated) (Belle + other KEK exp.) (Belle + other KEK exp.) (Belle + other KEK exp.)

## *KEK computing system*

*is a rental system with operation support service provided by vendor  
has to be replaced every ~4 years*

*Data and analysis software environment*

*Cyber infrastructure and services (such as Database)*

*have to be migrated to the new system*

# Data and Analysis preservation status

*new KEKCC*

*+ started operation since 2020 September (and will run until 2024 summer)*

*Belle I Data*

*+ all necessary data were migrated*

*+ database (condition, data searcher, etc.) was migrated*

*Belle Software*

*+ libraries were migrated from SL6 to CentOS7 (as KEKCC OS was changed to CentOS7)*

*at least, Belle physics analysis environment will be preserved until 2024 summer*

*In addition, analysis users can utilize the latest event generators*

*and analysis technologies with Belle II software framework (thanks to B2BII)*

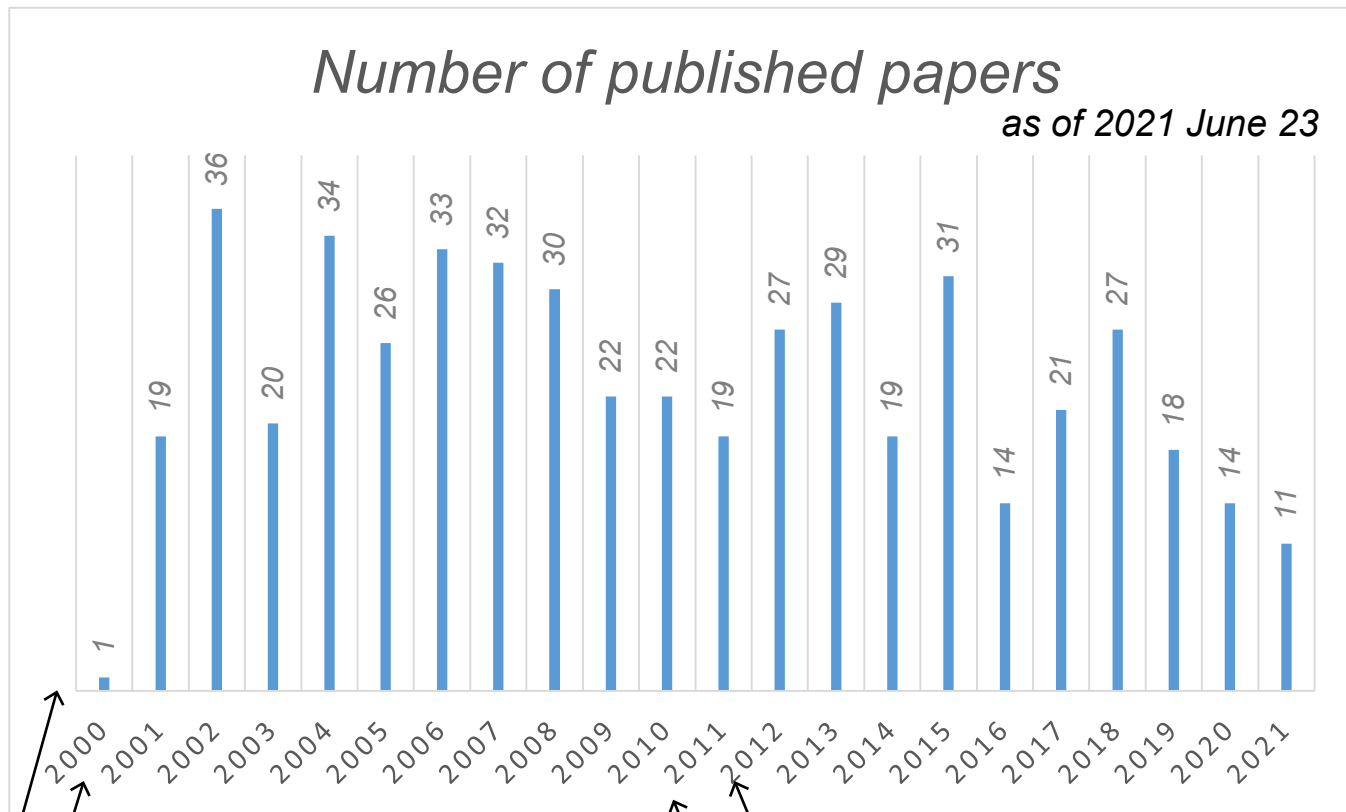
## **But**

*We need to consider the data / analysis preservation beyond 2024 summer*

*Belle II data exceeds the Belle I ?, OS for signal MC production ?,*

*necessary computing resources (CPU, Storage) must not be a big problem*

# Keeping the momentum for physics analysis



*Not in the left plot*

*Accepted : 5*

*Submitted : 6*

*To be submitted : 16*

*(and many on-going analyses)*

*Data taking stopped in 2010*

*Big earthquake in 2011*

*First physics paper (CP violation in B meson system)*

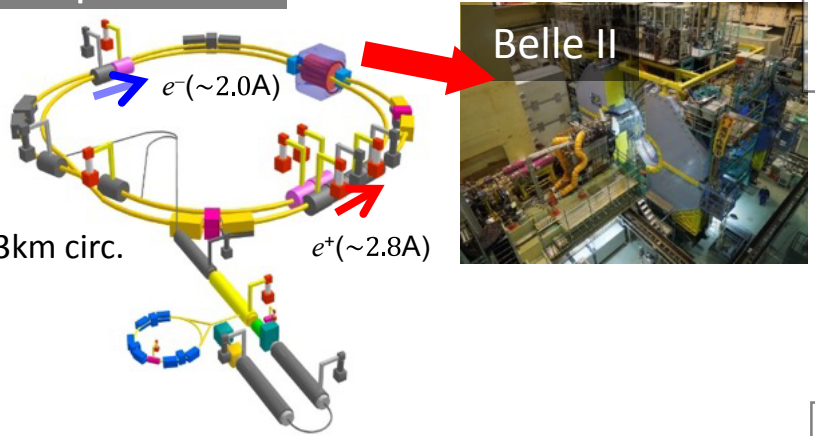
*Experiment started in 1999*

*Belle physics analysis activities  
are quite vital even 10 years  
after data-taking was finished !*

# Belle II Experiment Status

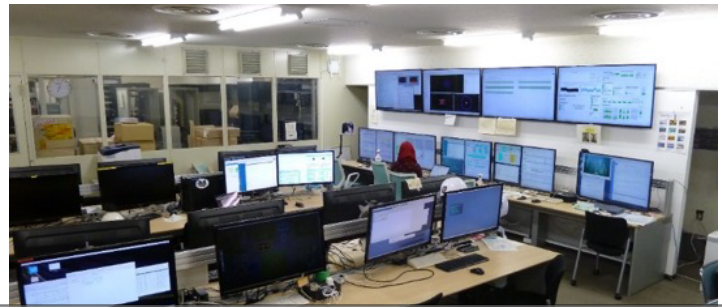
based on Y.Ushiroda and N.Saito's slides

## SuperKEKB

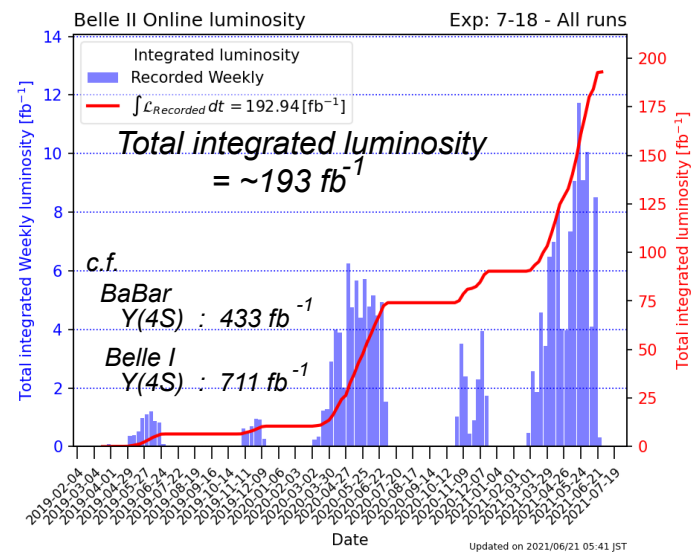


target  $L \sim 6 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$

→ Physics Run started from March 2019  
 Peak luminosity achieved so far  
 $\sim 2.96 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$  (World Record)  
 Integrated Luminosity by 2020:  $\sim 90 \text{fb}^{-1}$   
 More to accumulate before Long Shutdown

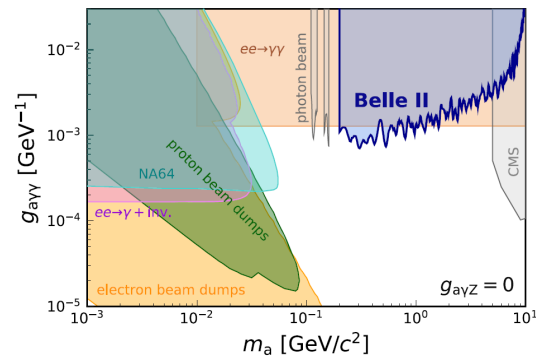


↑ Belle II Control Room under COVID-19  
 More automatized, more remote system, and more dedication of local collaborators enabled **highly efficient data taking even during pandemic.**



Rucio handles 100 k files/hour.  
 Similar level with ATLAS

← Efforts in software and computing in addition to detector operation enable **physics analyses with high sensitivities.**  
 → Two papers on Dark Sector published. More to come.



[ALP] Newly excluded parameter space around a few hundred MeV  
 Phys. Rev. Lett. 125, 161806 (2020)

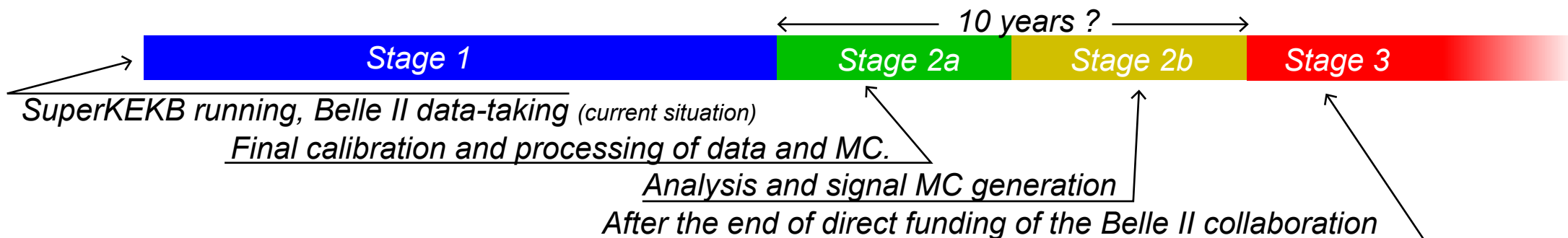


# Belle II Data and Analysis preservation

Last year, Belle II started a discussion on the Data and Analysis preservation and launched a task force

## Charge

1. The expected impact of the data preservation plan on Belle II Physics publications,
2. The computing model required to enable the preservation plan, including raw data reprocessing and MC production, both in the post-SuperKEKB-running period and the post-Belle II lifetime,
3. The data that should be preserved,
4. The period of time for accessibility of the preserved data,
5. The analysis infrastructure that should be preerved,
6. The estimated cost and effort of Belle II data and analysis preservation, and
7. The outreach potential of open Belle II data



# Experiments at KEK

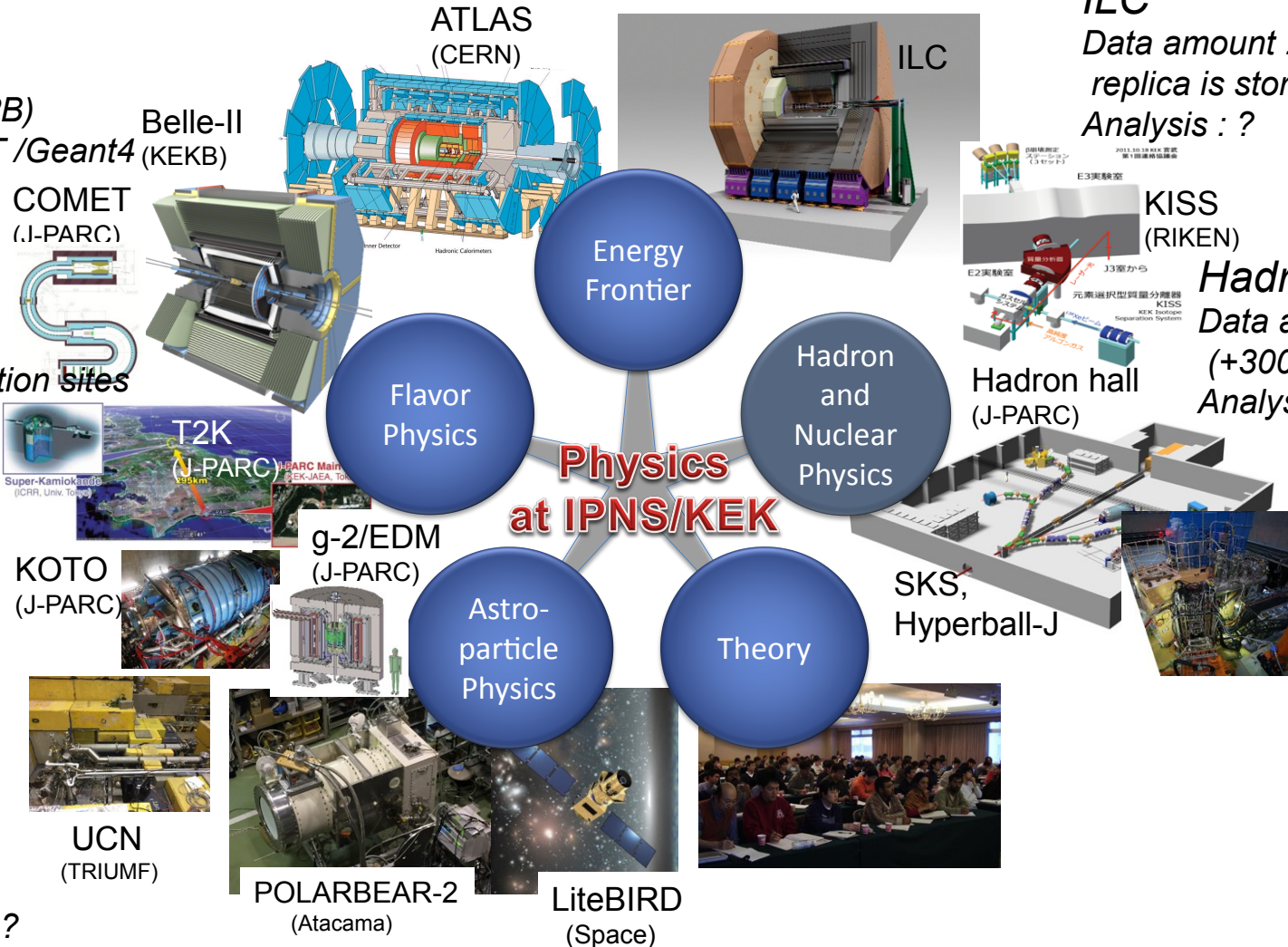
based on M.Yamauchi and N.Saito's slides

**COMET**  
 Data amount : 200TB  
 (2023 + several years : 2.5PB)  
 Analysis : depends on ROOT /Geant4

**T2K**  
 Data amount : 620TB  
 (in 2024 : +360TB)  
 replica is stored in collaboration sites  
 Analysis : ?

**g-2/EDM** (future experiment)  
 Data amount : ~1.5PB  
 (preserved until 2038 JFY)  
 Analysis : ?

**KOTO**  
 Data amount : ~10PB  
 (for next 4 years : +~10PB)  
 Analysis : self-developed is preserved ?



**ILC**  
 Data amount : ~200TB  
 replica is stored at DESY  
 Analysis : ?

**KISS**  
 (RIKEN)  
 Hadron hall (J-PARC)

**Hadron hall** (3 experiments)  
 Data amount : ~820TB  
 (+300-400TB / year)  
 Analysis : ?

SKS,  
Hyperball-J

KOTO  
(J-PARC)

Super-Kamiokande  
(ICRR, Univ. Tokyo)

COMET  
(J-PARC)

Belle-II  
(KEKB)

ATLAS  
(CERN)

ILC

Energy Frontier

Flavor Physics

Hadron and Nuclear Physics

Physics at IPNS/KEK

Astro-particle Physics

Theory

Hadron hall (J-PARC)


UCN  
(TRIUMF)

POLARBEAR-2  
(Atacama)

LiteBIRD  
(Space)

# Data and Analysis preservation at KEK

Most of experiments on-going at KEK are  
 using local “KEK computing system (KEKCC)” as the main computing system  
 So, the data and analysis preservation plan is similar to the Belle experiment



*new KEKCC*  
 + started operation since 2020 September (and will run until 2024 summer)

*Belle I Data*  
 + all necessary data were migrated  
 + database (condition, data searcher, etc.) was migrated

*Belle Software*  
 + libraries were migrated from SL6 to CentOS7 (as KEKCC OS was changed to CentOS7)

*at least, Belle physics analysis environment will be preserved until 2024 summer*  
 In addition, analysis users can utilize the latest event generators  
 and analysis technologies with Belle II software framework (thanks to B2BII)

**But**

*We need to consider the data / analysis preservation beyond 2024 summer*  
*Belle II data exceeds the Belle I ?, OS for signal MC production ?,*  
*necessary computing resources (CPU, Storage) must not be a big problem* → KOTO ??