NICA/SPD computing and possible inputs for the STCF

Alexey Zhemchugov
JINR Dubna

The 6th International Workshop on Future Tau Charm Facilities
18 November 2024

The NICA project

Timeline

2009 – first proposal

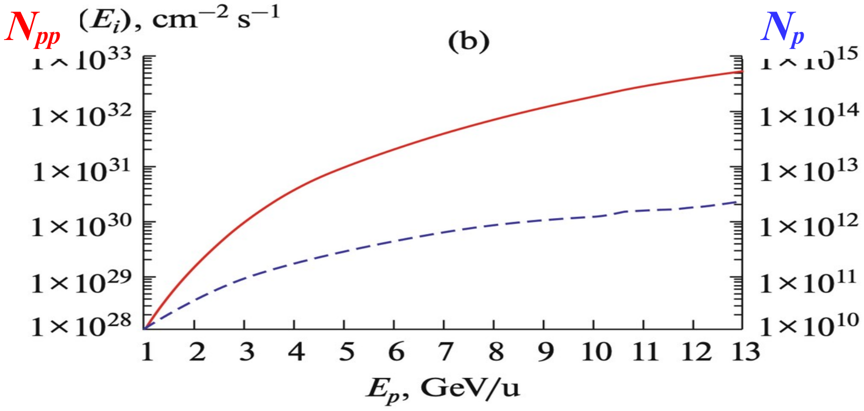
2016 — construction started

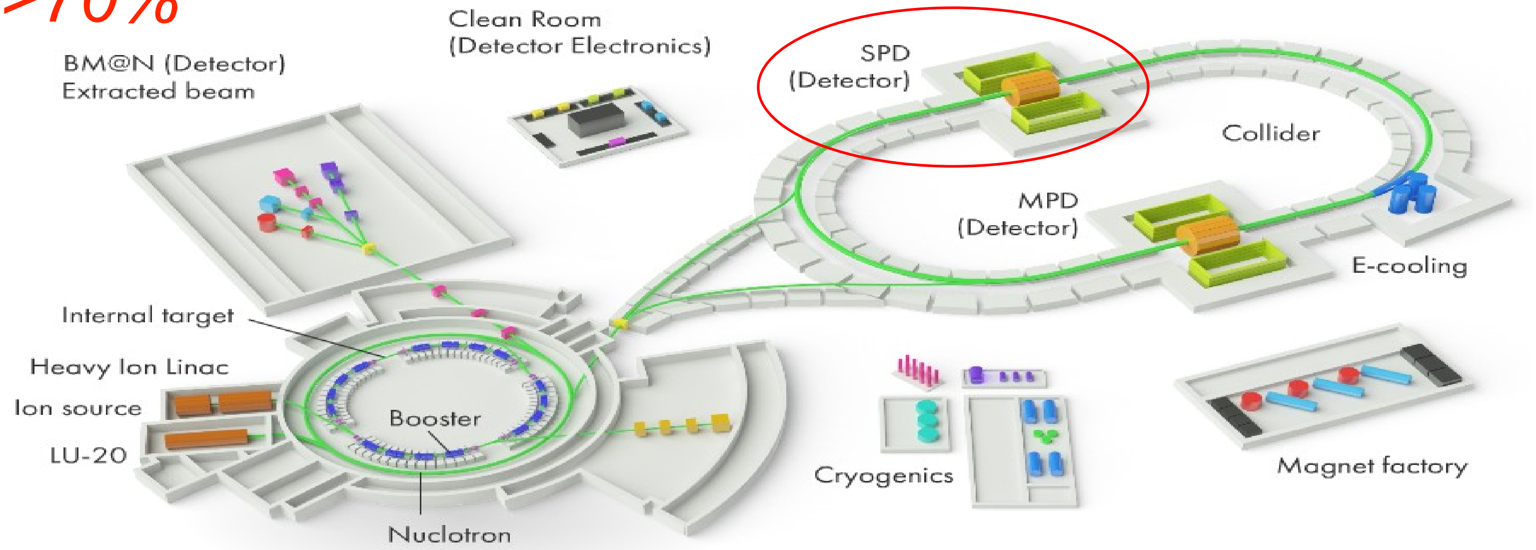
2025 — beam circulation and maybe the first collision

SPD at NICA

p ↑ p ↑ : $\sqrt{s} \leq 27$ GeV
d ↑d ↑ : $\sqrt{s} \leq 13.5$ GeV
d ↑p ↑ : $\sqrt{s} \leq 19$ GeV

U, L, T
|P|>70%





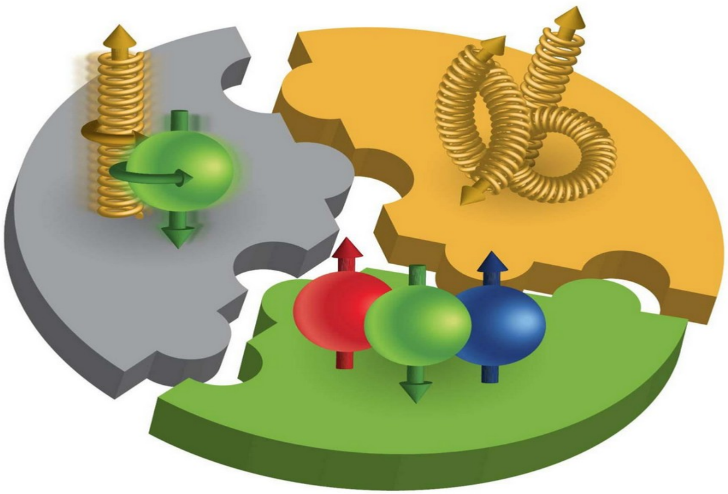
The SPD Collaboration

SPD Collaboration now consists of more than ~400 scientists from many countries.

Physics goals

- SPD - a universal facility for comprehensive study of gluon content in proton and deuteron at large x
  - Prompt photons
  - Charmonia
  - Open charm
- Other spin-related phenomena
- Other physics

More details:
Prog.Part.Nucl.Phys. 119 (2021) 103858
arXiv:2011.15005

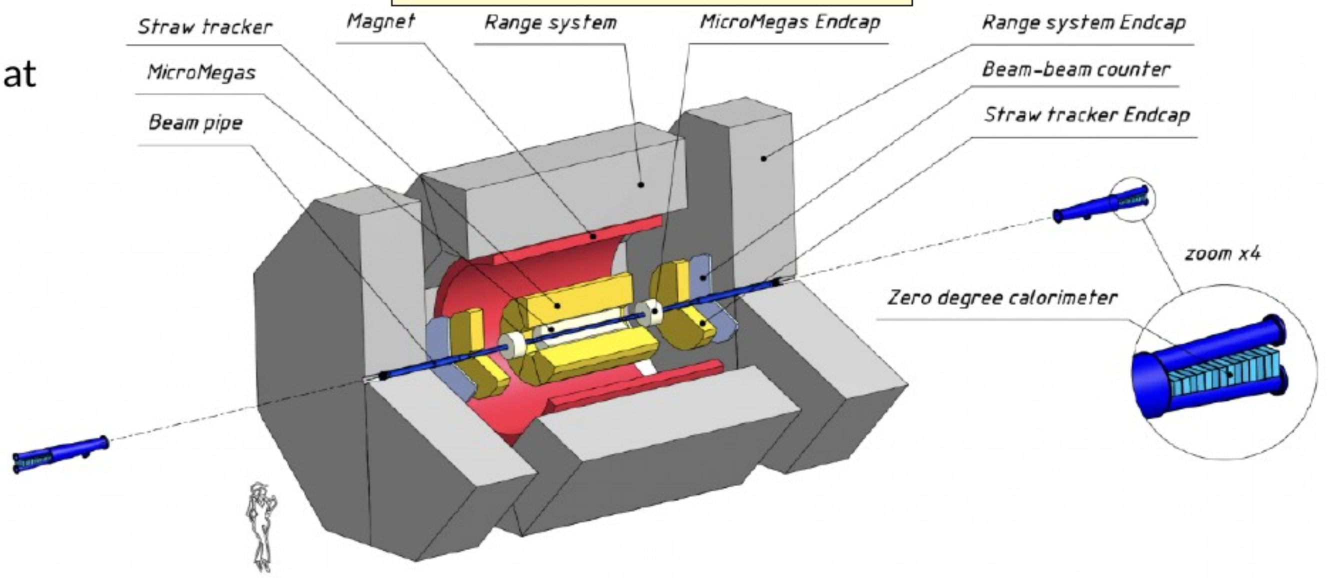


The first stage

The SPD TDR can be found at arXiv:2404.08317

Magnetic field up to **1.2 T**

- Polarized and unpolarized phenomena at **low energies** (3.4 GeV < $\sqrt{s_{pp}}$ < 9.4 GeV) and **reduced luminosity**
- p-p, d-d, and ion collisions (up to Ca)
- Simplified detector set-up
- Up to 2 years of data taking



Range System
muon identification and coarse hadron calorimetry

Straw tracker:
- **σ ~ 150 μm**
- **σ(dE/dx) = 8.5%**

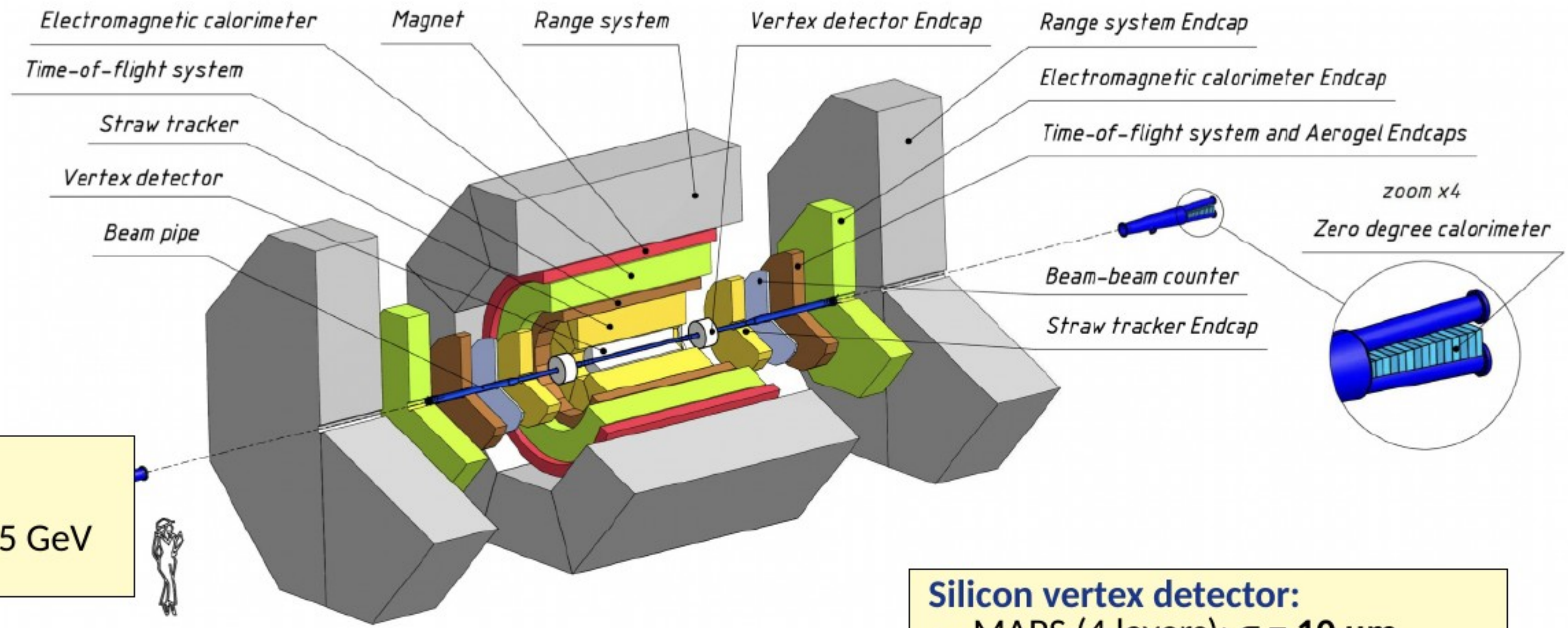
Micromegas central tracker:
σ ~ 150 μm

BBC and **ZDC** for online polarimetry

The second stage

The SPD TDR can be found at arXiv:2404.08317

Electromagnetic calorimeter:
$\sigma E/E = 5\%/\sqrt{E} \oplus 1\%$

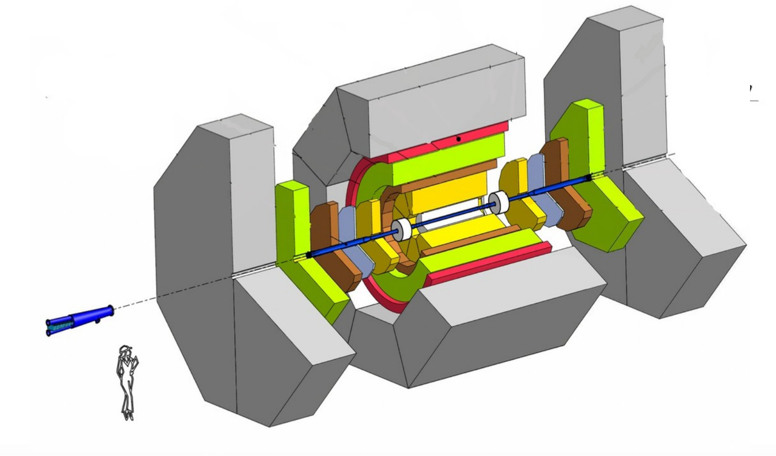


Time of flight system:
σ = 50 ps
3σ π/K separation for p < 1.5 GeV

FARICH in *endcaps* for pion/kaon separation for particle momentum up to **5.5 GeV**

Silicon vertex detector:
- MAPS (4 layers): **σ = 10 μm**
- DSSD (3 layers): **σ_φ = 27.4 μm, σ_z = 81.3 μm**

SPD as a data source



- Bunch crossing every 76.3 ns = crossing rate 13 MHz
- ~ 3 MHz event rate (at 10^{32} $cm^{-2}s^{-1}$ design luminosity)
- 20 GB/s (or 200 PB/year (raw data), $3*10^{13}$ events/year)
- Selection of physics signal requires momentum and vertex reconstruction → no simple trigger is possible

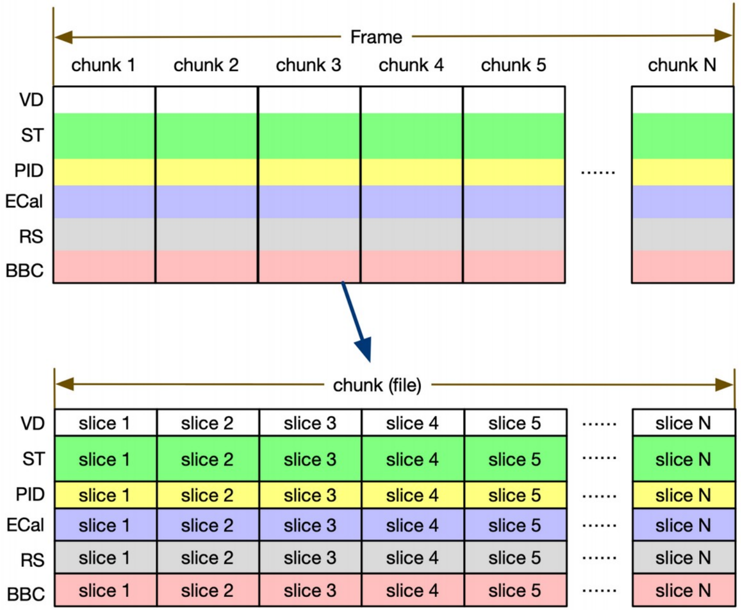
The SPD detector is a medium scale setup in size, but a large scale one in data rate!

Input data

• RAW event size	7 kB
• RECO event size	15 kB
• Time for Reconstruction (1 ev)	100 HepSPEC
• Time for Simulation (1 ev)	500 HepSPEC
• Event rate at maximum luminosity	3000 kHz
• Event rate after online data filter	150 kHz
• Operation time	50000 seconds/day
• Operation time	200 days/year

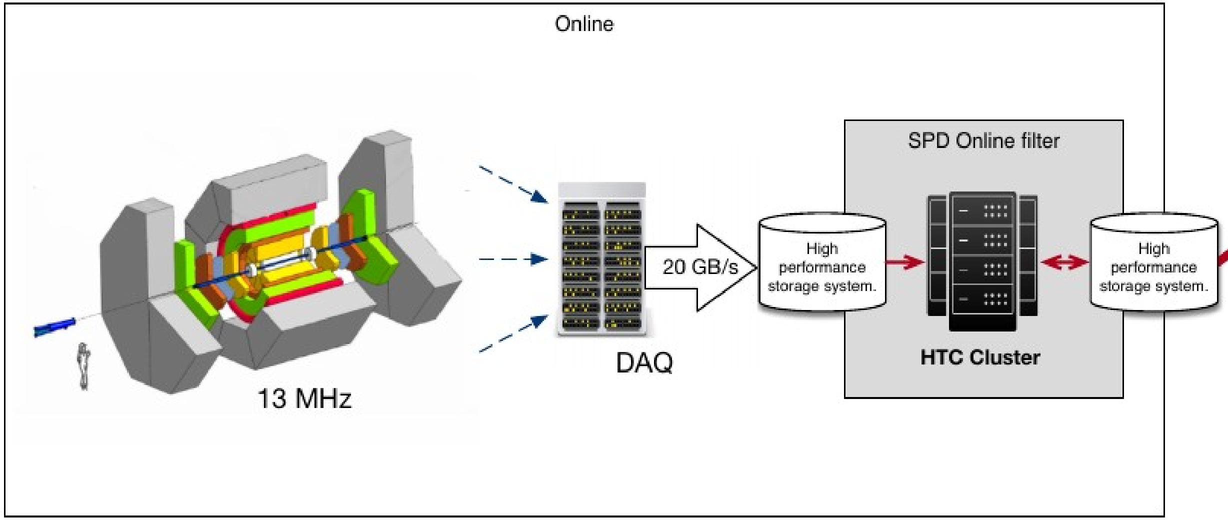
Free running DAQ

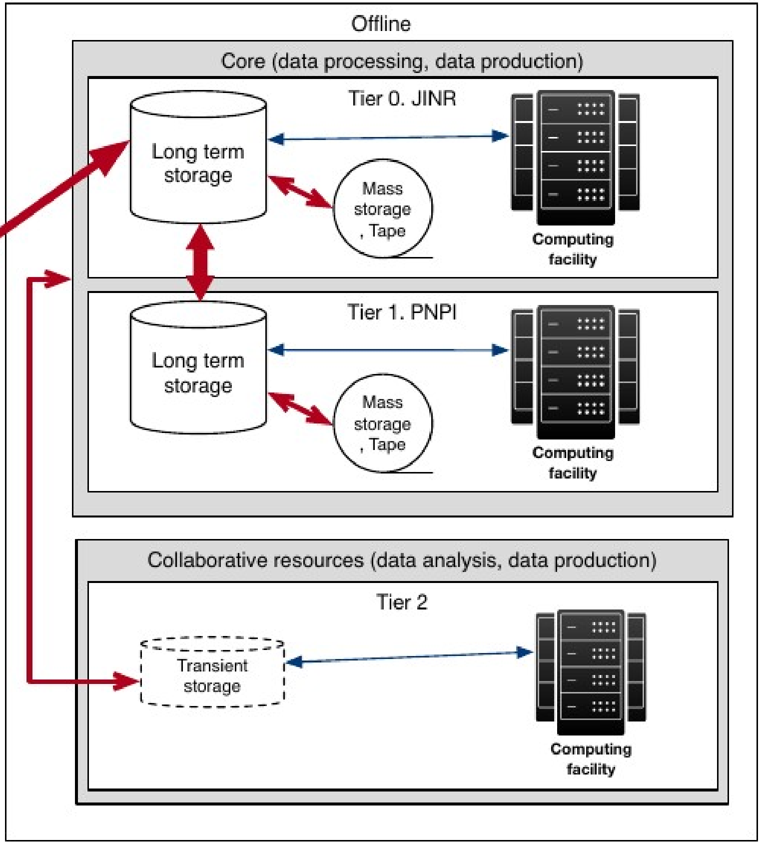
No trigger = No classical events anymore



- Primary data unit: **time slice** (1 us — 8.3 ms)
- Time slices combined in **time frames** (up to 549 s, 16 GB max, < 160 MB to fullfil 20 GB/s limit)
- Intermediate units — **time chunks** of 0.1-0.2 s (2-4 GB or ~10^5-10^6 events) are being discussed now
- Every time slices will contain signals from a few to many collisions (events)
- Event building have to unscramble events from a series of time slices.

Data workflow



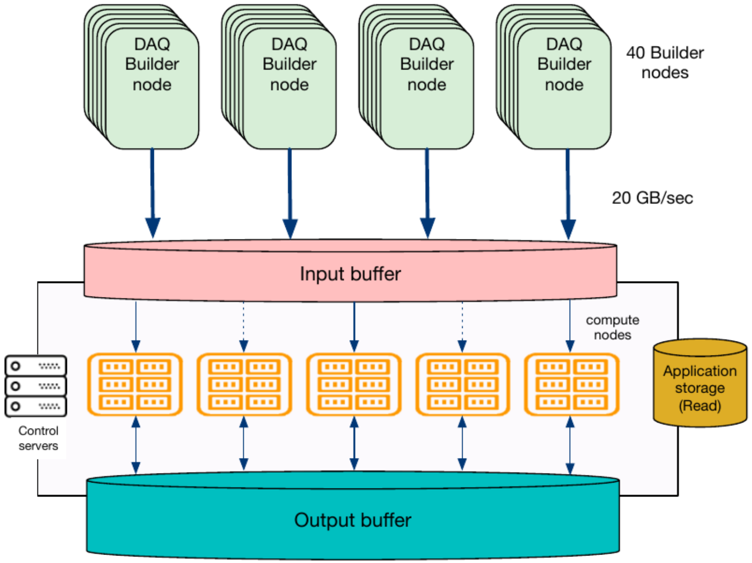


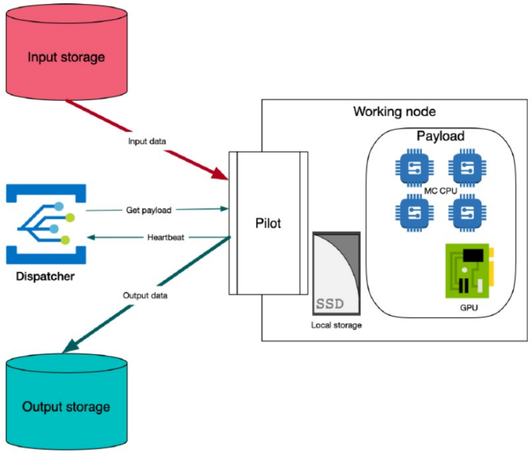
Continuous data reduction:

- DAQ: noise suppression
- Online filter: event building, partial reconstruction, software high-level trigger
- Offline computing: data analysis and long term storage

SPD Online Filter

High-performance heterogeneous computing cluster





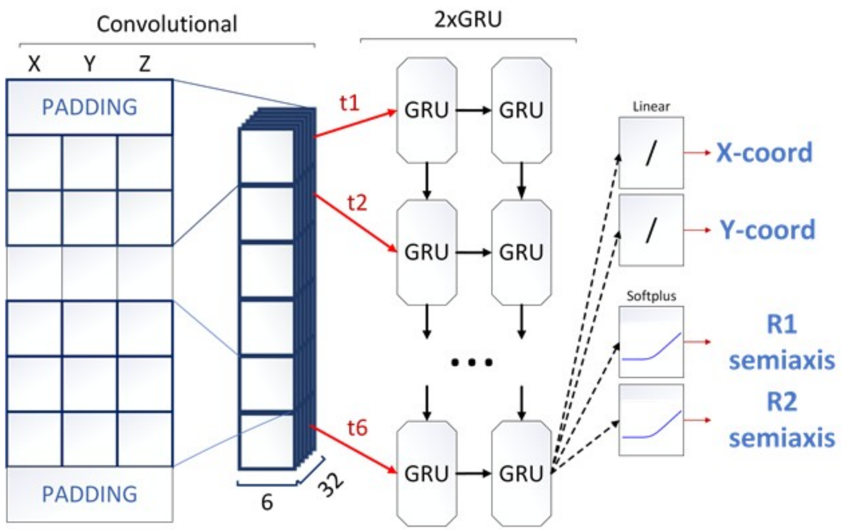
Onlne reconstruction and event filtration

- Partial reconstruction
  - Fast tracking and vertex reconstruction
  - Fast ECAL clustering
  - RS reconstruction
  - FARICH reconstruction
  - Particle ID (π^0, muons, kaons)??
- Event unscrambling
- Software trigger
  - several data streams
- Monitoring and Data quality assessment
- Local polarimetry

Machine learning is
a promising technology

Example: TrackNETv3 for track recognition

JINST 17 (2022) 12, P12023



- Network predicts an area at the next detector layer where to search for the track continuation
- If continuation is found the hit is added to the track candidate and the procedure repeats again
- Essentially reproduces the idea of the Kalman filter: track parameters are predicted by synaptic weights determined by network training

	Time slices of 40 events
Track efficiency (recall) (%)	96,54
Track purity (precision) (%)	94.75
Time slices / sec	63.74 (*40 = 2549.6)

PRELIMINARY

Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz + GPU Nvidia V100 32Gb

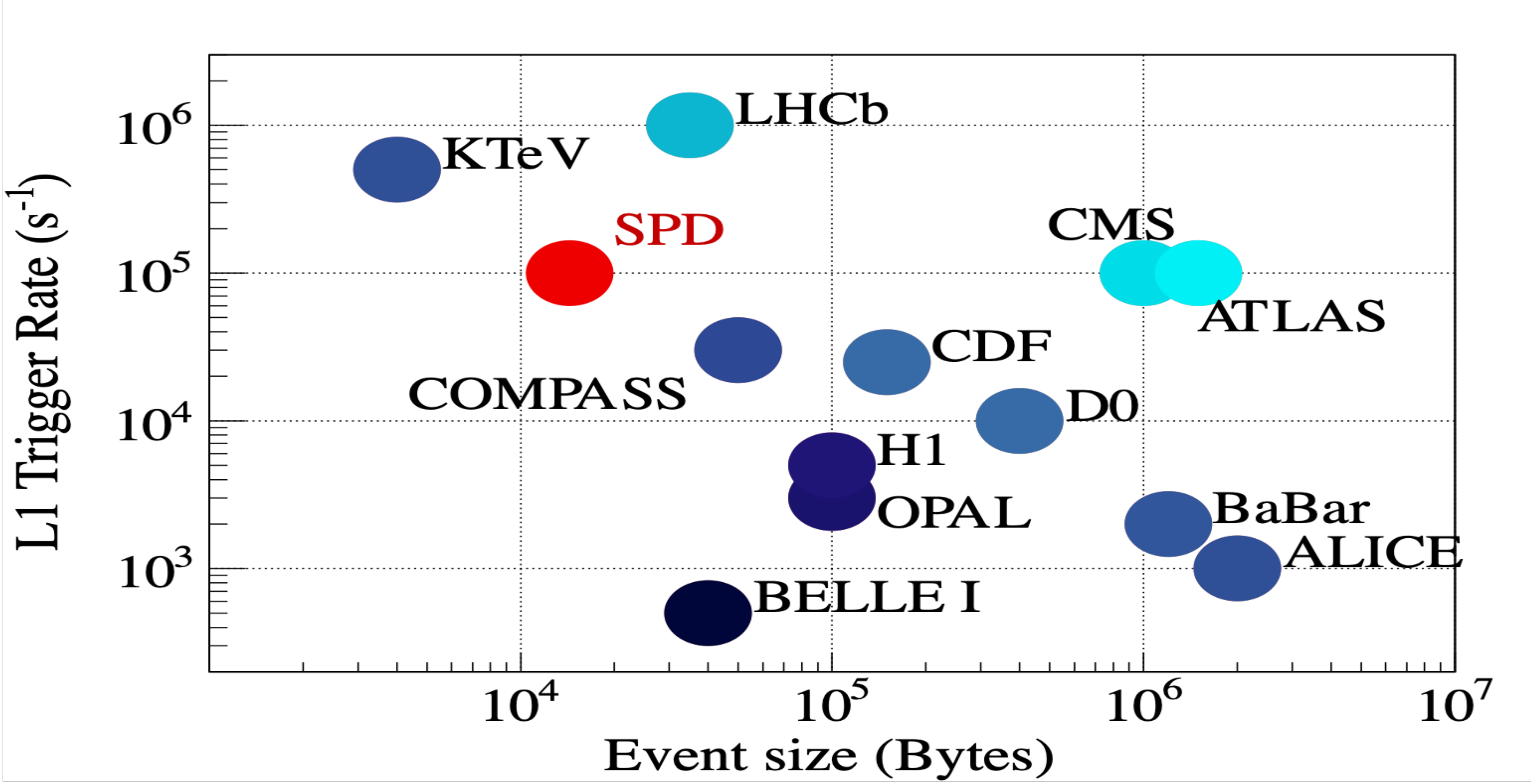
Software Framework

- SAMPO - a Gaudi-based software framework is being developed:
  - Geometry description: GeoModel
  - Generators: Pythia8, FTF, UrQMD + capability to add more generators
  - Simulation: Geant4
  - Reconstruction: ACTS or GenFit for tracking, Kfparticle for vertex reconstruction, own algorithms for other subsystems
- The same framework is planned for the online filter and for the offline data processing
  - Integration with ML tools for the online filter (inference only)
- Current simulation and performance studies are done by another framework SpdRoot, based on FairRoot software (frozen)

Databases

- Several databases are needed:
  - Data taking conditions and calibration data
  - Physics metadata (including MC input configurations)
  - EventIndex: catalog of physics events, both collected from the detector and simulated
  - Hardware database and mapping
  - Monitoring and logging
  - Collaboration management data.
- Designed as a complex information system that includes data collection and transfer tools, APIs for access from the production and analysis software, client software, supervisors, and monitoring.
- A PostgreSQL RDBMS is considered as a database platform

After the online filter



Expected data volumes

- **Preparation for the experiment.**
  - Monte Carlo simulation from 2024 to 2028 will provide 2 PB per year.
  - Total per stage: 10 PB.
- **Stage I: running at low luminosity of the NICA collider.**
  - Monte Carlo simulation and real data taking from 2028 to 2030 will provide 4 PB per year. Reprocessing: 2 PB per year.
  - Total per stage: 18 PB.
- **Upgrade of the setup for operation at high luminosity.**
  - Monte Carlo simulation from 2031 to 2032 will provide 2 PB per year. Reprocessing: 2 PB per year.
  - Total per stage: 8 PB.
- **Stage II: running at maximum design luminosity of the NICA collider.**
  - Monte Carlo simulation and real data taking from 2033 to 2036 will provide 20 PB per year. Reprocessing: 10 PB per year.
  - Total per stage: 120 PB.

Total for all stages: 156 PB

Difficult (and unsafe!) to manage in a single computing center → need for the distributed computing system!

Computing resources

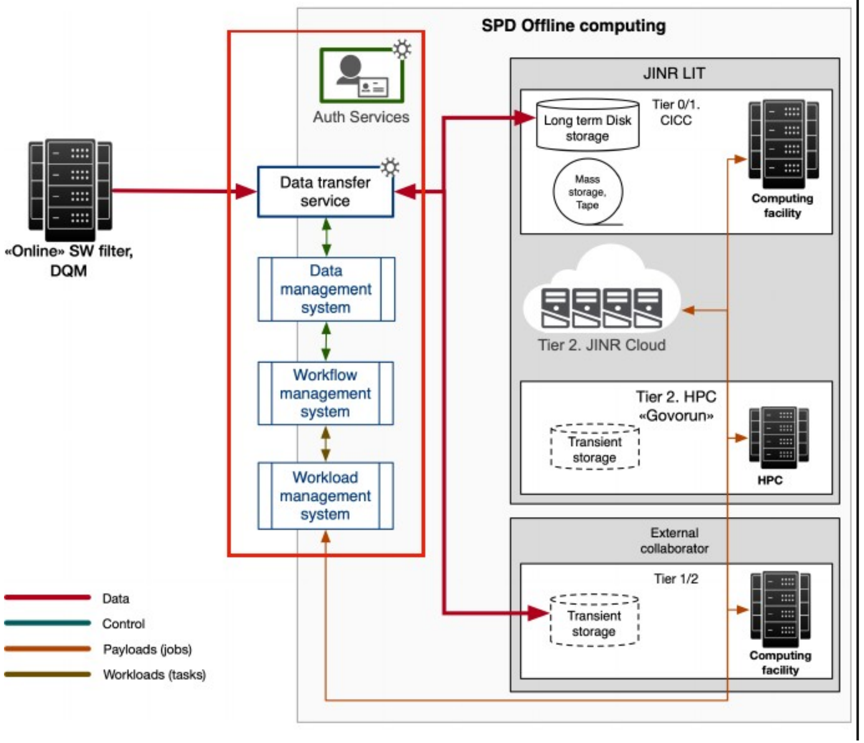
	CPU (cores)	Disk storage (PB)	Tape storage (PB)
SPD Online filter (stage 1)	3000	2	
Offline computing (stage 1)	20000	5	6 per year
SPD Online filter (stage 2)	6000	4	
Offline computing (stage 2)	60000	15	30 per year

- Tier-0 at JINR will provide about 25-30% of all computing resources
- Tier-1 at PNPI is going to contribute about 25%
- The rest should be distributed between the participating institutes

Distributed computing system

All basic components are already available from LHC experiments:

- *Authorization service: IAM*
- *Workload management: PANDA*
- *Data management and transfer: RUCIO and FTS*
- *Information service: CRIC*
- *Software distribution: CVMFS*

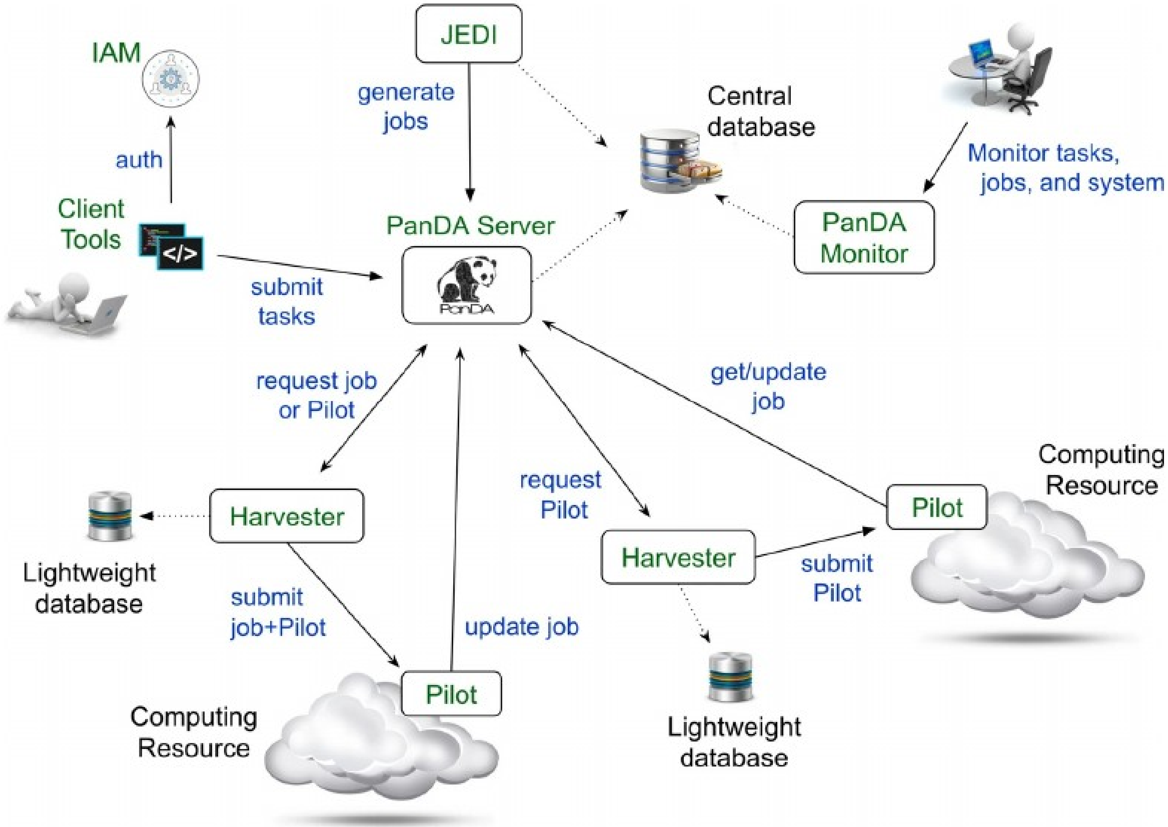


The system is being phased in starting in summer 2024.

PanDA

- The Production and Distributed Analysis (PanDA) is a data-driven workload management system capable of operating at massive data processing scale, designed to have the flexibility to adapt to emerging computing technologies in processing, storage, networking and distributed computing middleware

- The PanDA system has been developed to meet ATLAS production and analysis requirements for a data-driven workload management system capable of operating at the LHC data processing scale



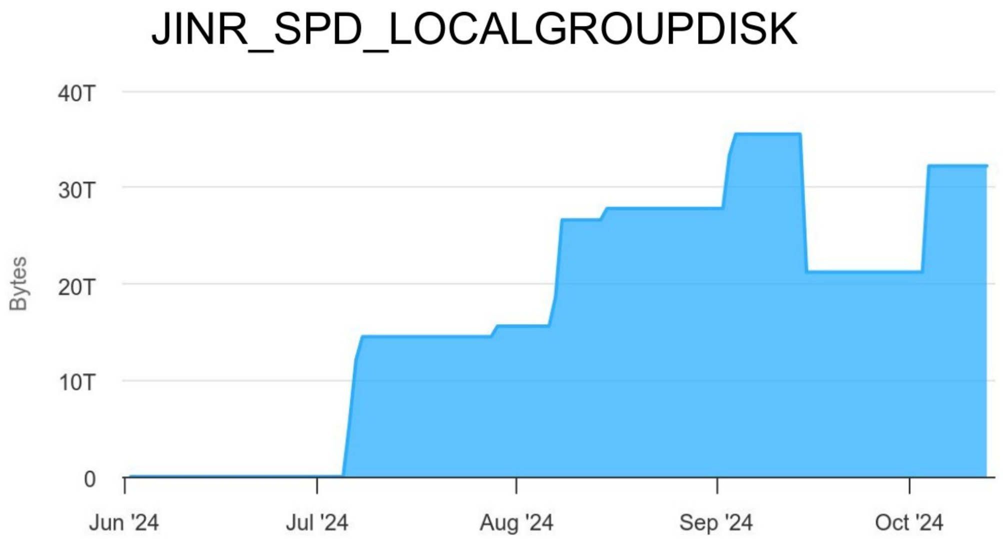
Our team has 10 years experience with PanDA and more than 15 years of the development of distributed computing and distributed data management services for the ATLAS experiment

RUCIO

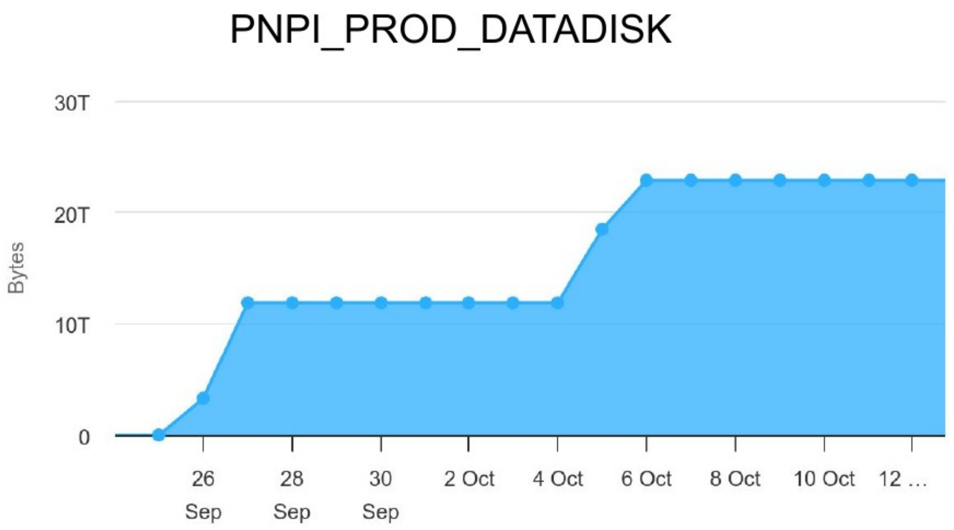
- Rucio is an open-source software framework that provides functionality for da management and access in a distributed storage environment. Rucio also provides protection against data loss and speed up access to data through a controlled number of replicas.
- Currently, the Rucio system can be used to:
  - organize data in a hierarchical structure for easy navigation and management;
  - unified interaction of a heterogeneous network and storage infrastructure;
  - distribute data for storage;
  - adaptive data replication and recovery;
  - automated data transfer between storages;
  - storage of all types of experimental data;
  - data lifecycle management;
  - storage and management of metadata;
  - provides metrics for monitoring data usage, system performance and the status of various components.
- Official Rucio documentation: https://rucio.cern.ch/documentation/

RUCIO deployment at SPD

Data production at JINR



Automatic replication to PNPI



The case of STCF

According to the STCF CDR ***(arXiv:2303.15790v3, 5 Oct 2023)***

- 400 kHz trigger rate
- Event size ~70 kB

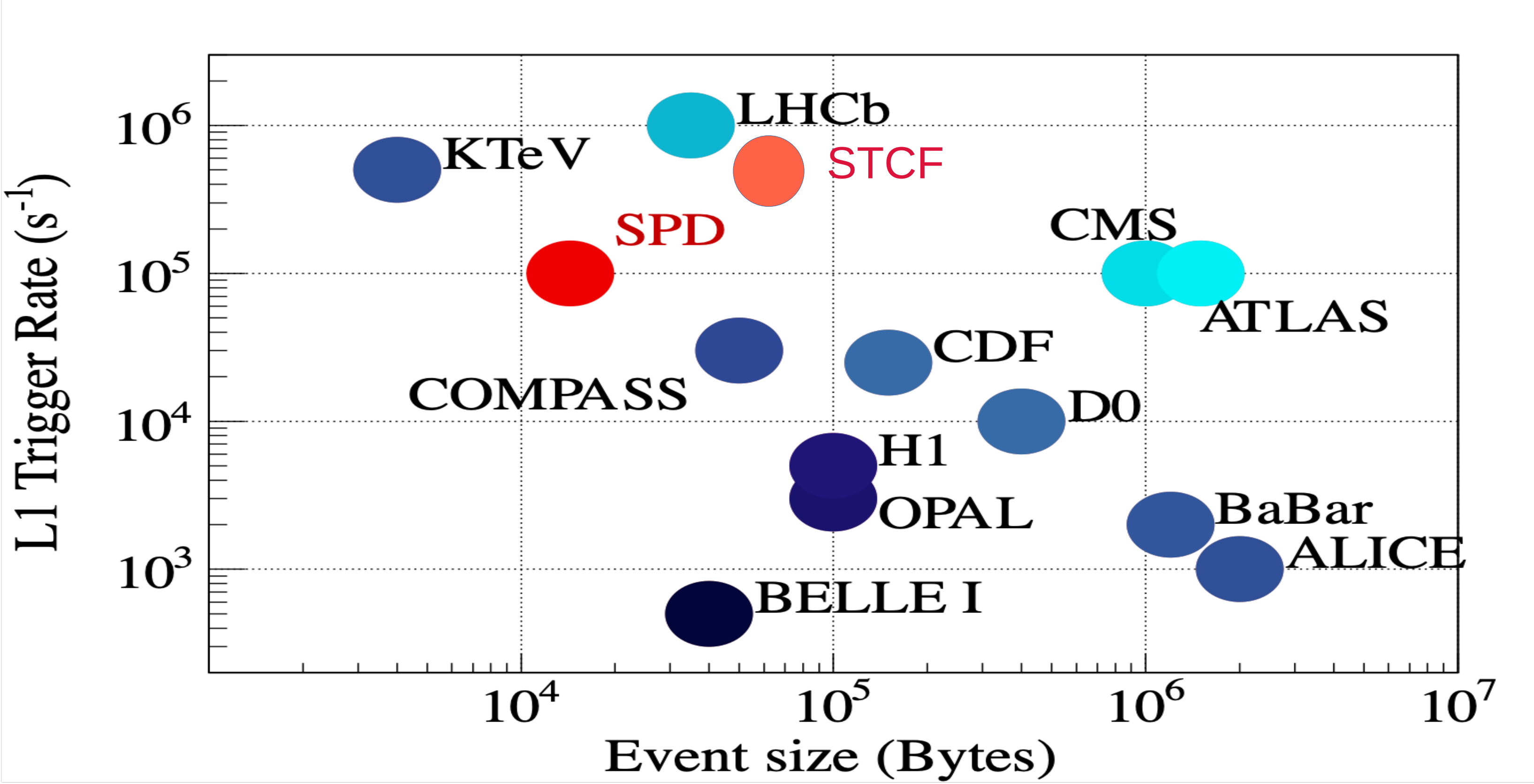
→ about 30 GB/s

STCF is expected to produce a significant amount of data (several tens PB monthly)

Even if the event size is a little bit overestimated, the amount of data will reach much more than 10-20 PB/year anyway (=millions of files, millions of jobs)

A dedicated distributed offline computing system is necessary for STCF, including authorization, workload, data and information management.

SPD vs STCF vs others



Summary

- Large scale offline computing system is necessary to cope with the huge amount of STCF data
- SPD at NICA has similar requirements and already gained experience in the deployment of the distributed computing system
- This experience can be useful for the STCF project as well