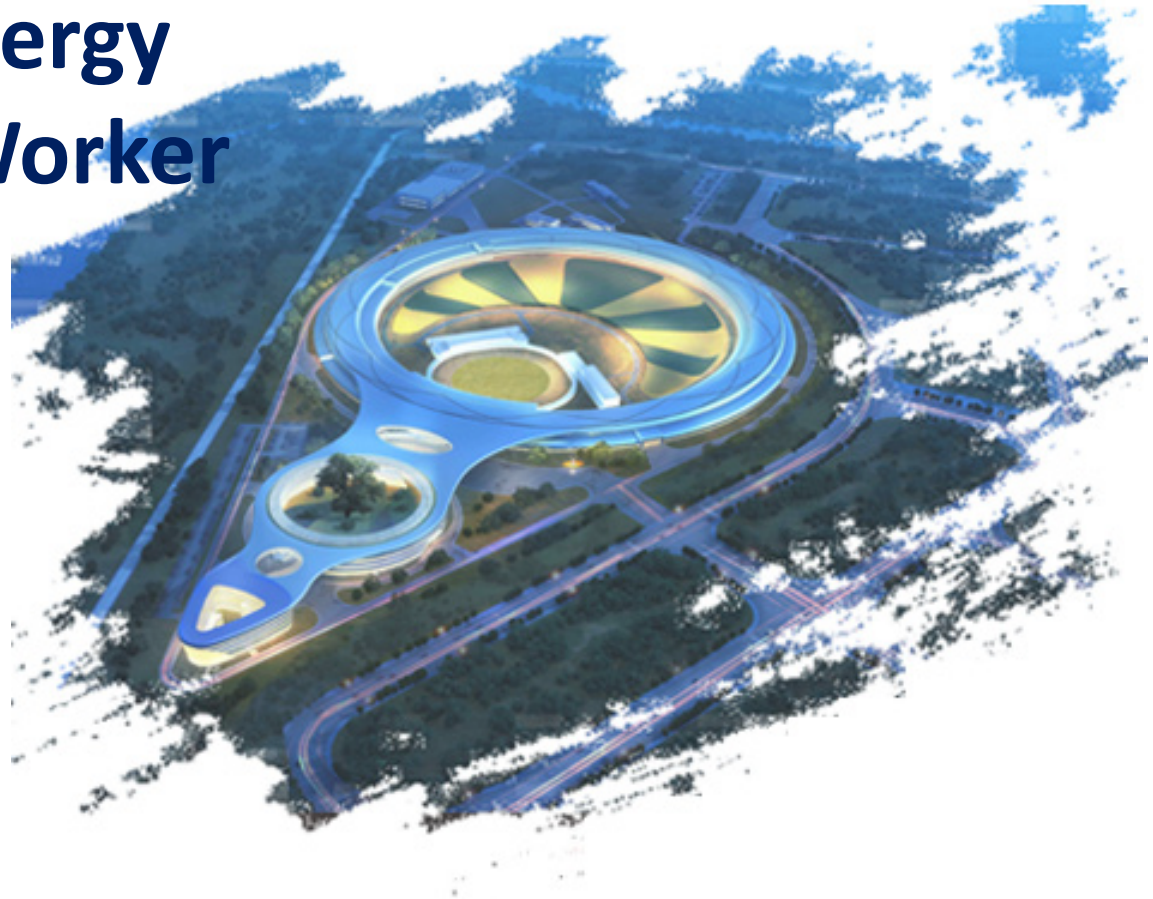
高能同步辐射光源
High Energy Photon Source

High-Throughput Data Orchestration and Streaming System for High Energy Photon Source: Mamba Data Worker



Cheng long Zhang

Multidisciplinary Center

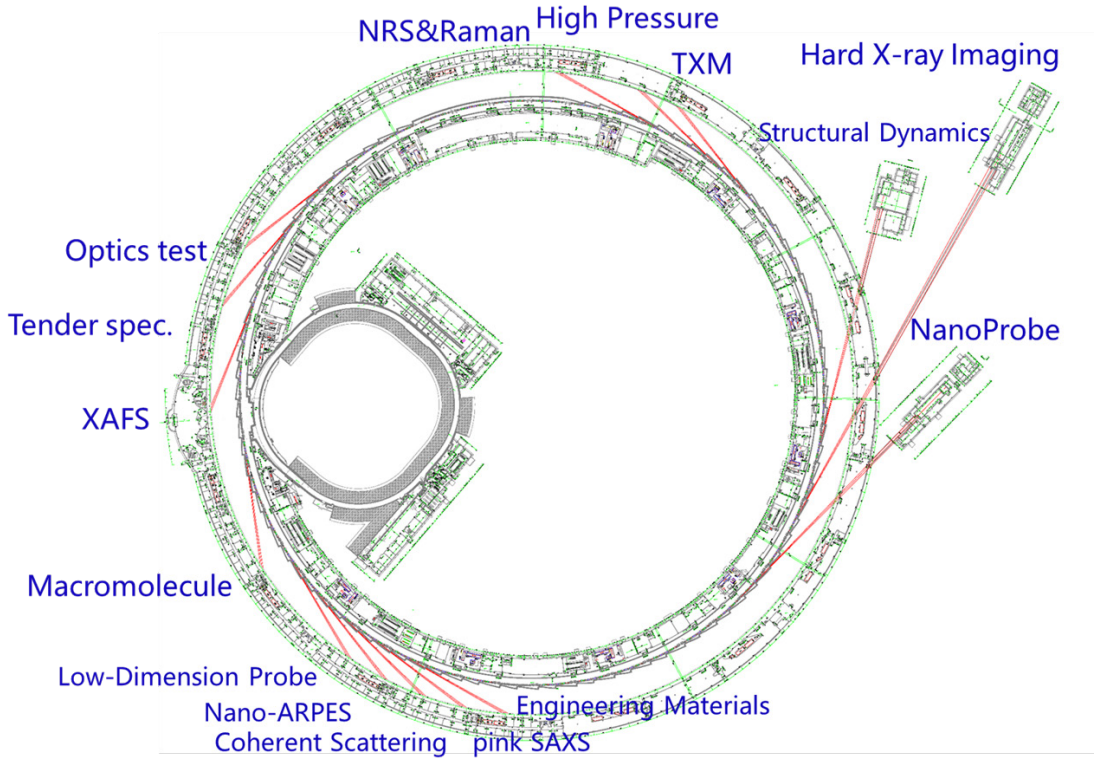
Institute of High Energy Physics

8/11/2023

Outline

- **Project Background**
- **The Design of Mamba Data Worker**
- **Progress status of Mamba Data Worker**

New experimental features in HEPS



Phase I Beamlines

The experimental features at HEPS:

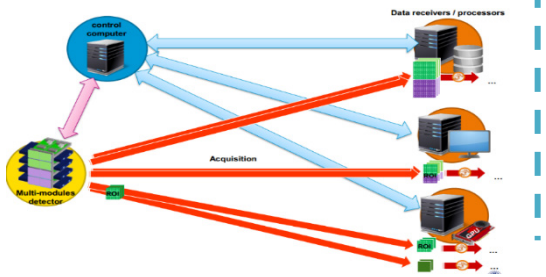
- High data throughput
- Multimodality
- Ultrafast
- In situ and dynamic

Revolutionary requirements in :

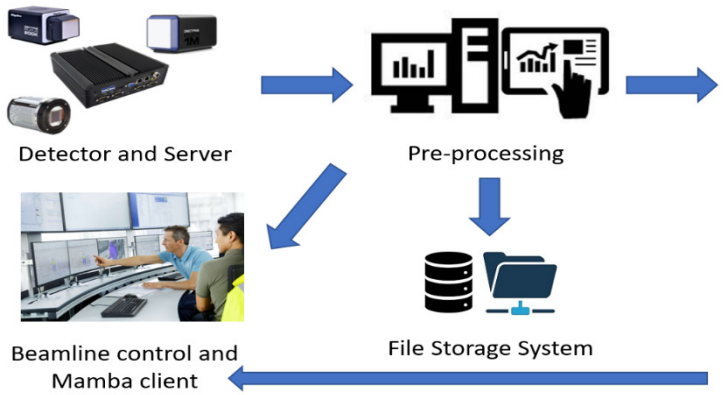
- Scientific software (control, data acquisition, analysis and management)
- Computing resources

Challenges of high-throughput multimodal data acquisition

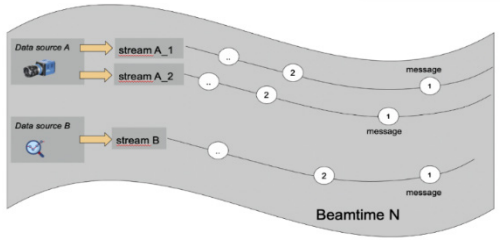
Beamlines	Burst output (TB day^{-1})	Average output (TB day^{-1})
B1	600	200
B2	500	200
B3	8	3
B4	10	3
B5	10	1
B6	2	1
B7	1000	250
B8	80	10
B9	20	5
BA	35	10
BB	400	50
BC	1	0.2
BD	10	1
BE	25	11.2
BF	1000	60
Total		805.4

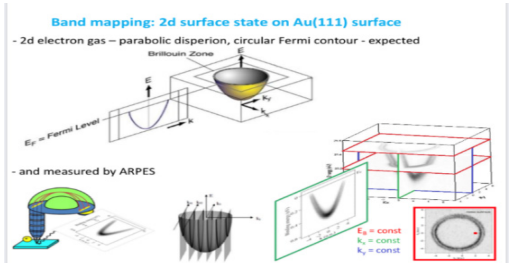


Parallelization of large multi-module detectors DAQ



High throughput multimodal data monitoring, acquisition, transmission, assembly, and disk writing





Data reduction and assembly related to methodology

- With the explosive growth of experimental data volume, **the data throughput of a single beamline will reach 800T/day,** and data acquisition is facing challenges
- **data pipeline has shifted from a single channel from detector to storage in the past to a complex network of multiple data generators and multiple applications**

Outline

- **Project Background**
- **The Design of Mamba Data Worker**
- **Progress status of Mamba Data Worker at BSRF**

Outline

- MDW Framework Architecture
- MDW online data acquisition: Fly scan as an example
- MDW online data assembling and disk writing
  - Data and Metadata storage format specification
  - Customized parameter subsystem
  - Online disk writing architecture

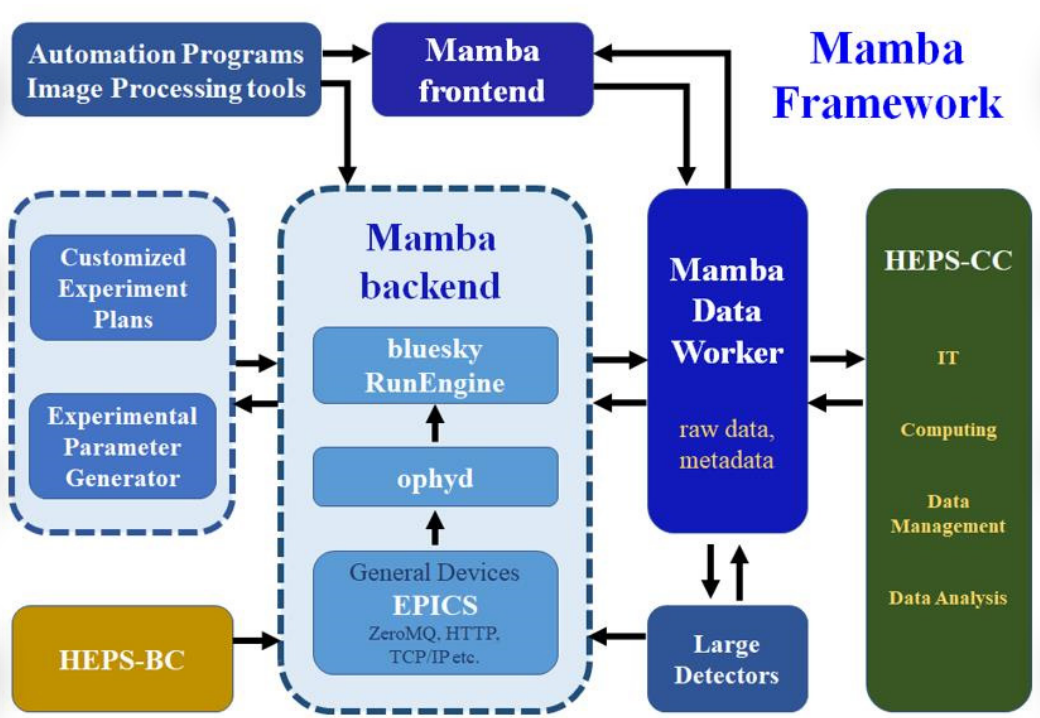
A new generation synchrotron experiment operating software system (*Mamba*)

Beamline automation

Real-time feedback and autonomous control

Custom experimental procedures and commands for all experiment methods

Smart *in-situ* experiment



Friendly GUIs

Easy experiment mode switch

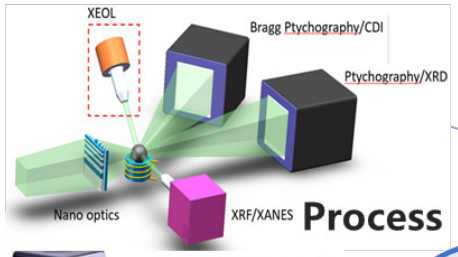
Fast and multi-modal DAQ

Rich metadata, high interoperability

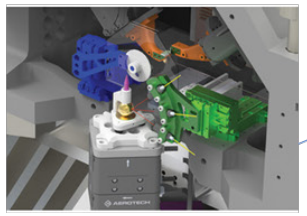
Multi-dimensional, accurate, fast fly scans

Mamba: a systematic software solution for beamline experiments at HEPS. ***Journal of Synchrotron Radiation***, 2022

The MDW Software Feature and Design Hierarchy

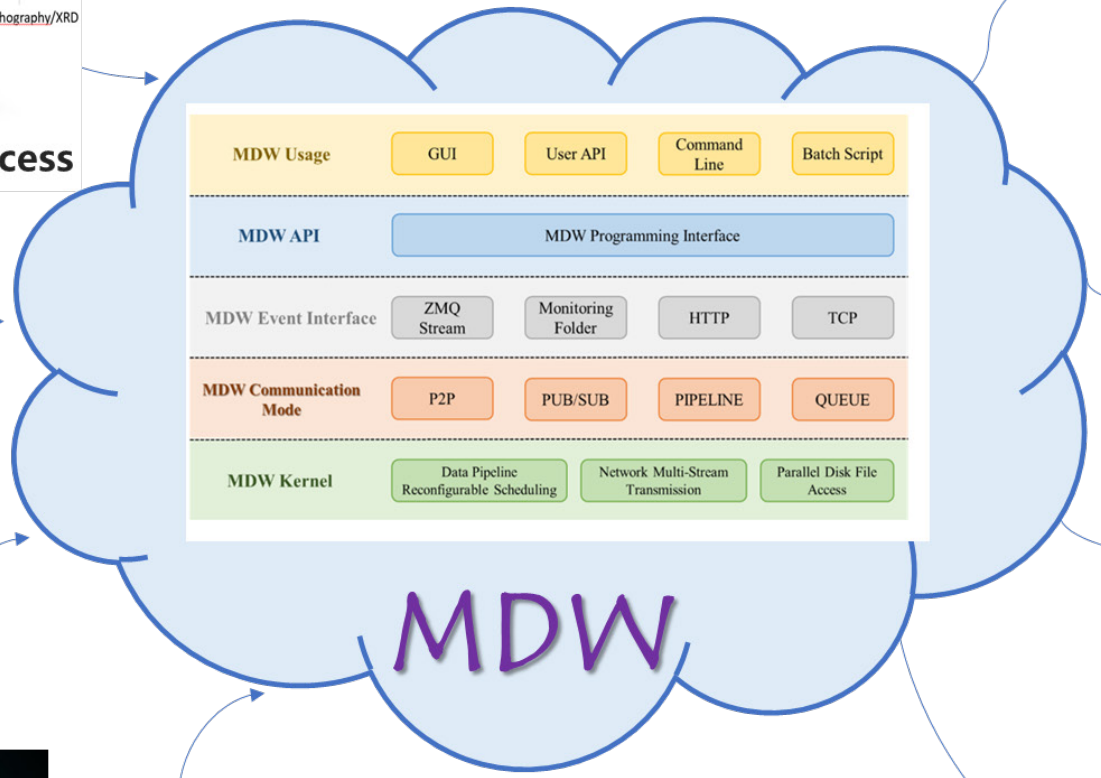


Detectors

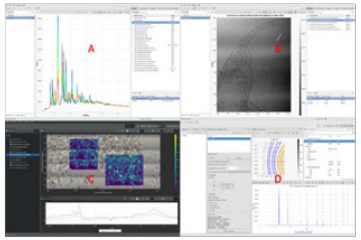


Motion

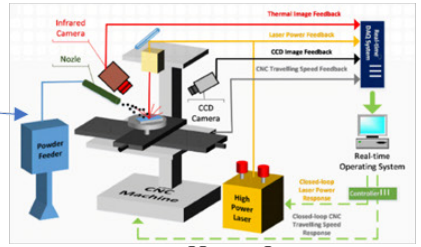
Metadata
(sample, environment etc.)



Scientific database



Data visualization

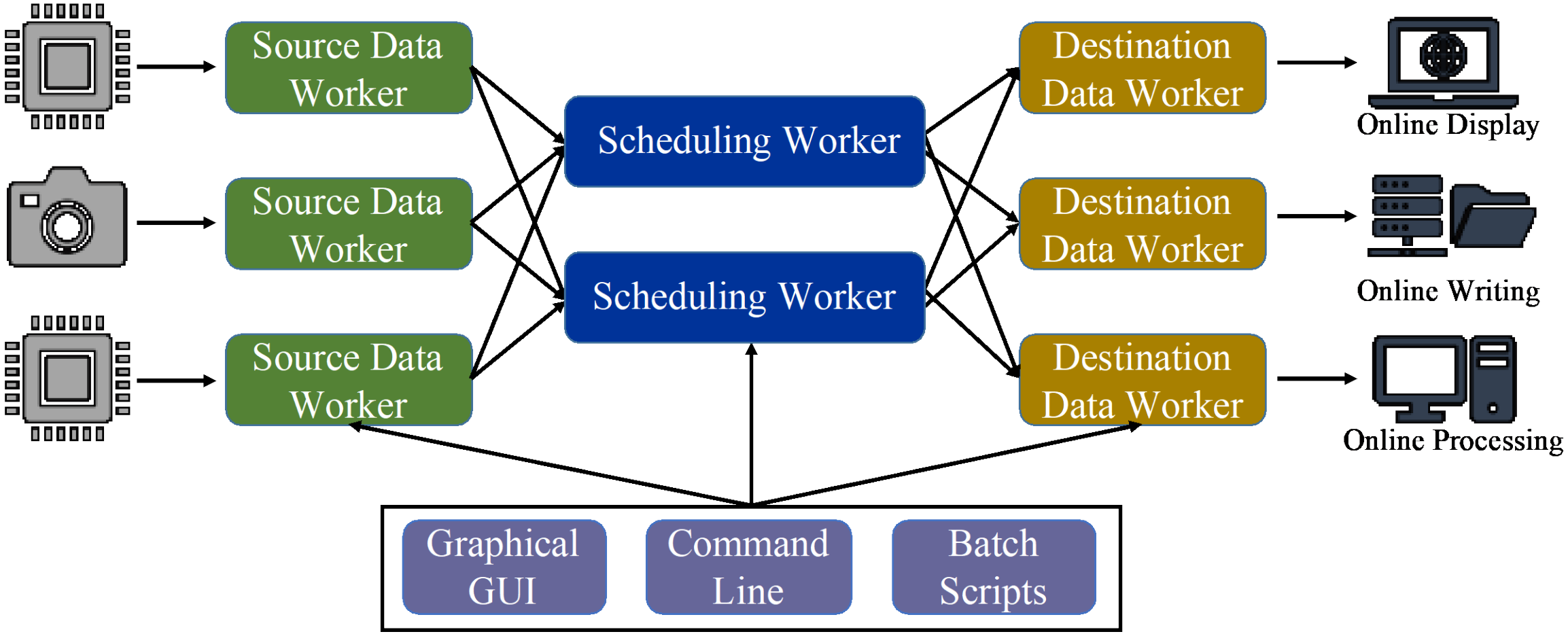


Feedback control

Storage

- DAQ
- Distribution
- Scheduling
- Assembly
- Reduction
- Disk writing
- Visualization

Software deployment, Data and control separation

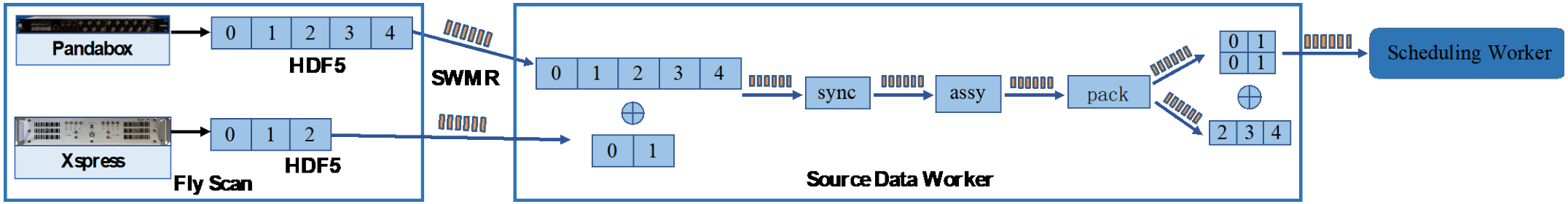


Outline

- MDW Framework Architecture
- MDW online data acquisition: Fly scan as an example
- MDW online data assembling and disk writing
  - Data and Metadata storage format specification
  - Customized parameter subsystem
  - Online disk writing architecture

MDW online data acquisition: Fly scan as an example

- Hardware: Pandabox (motors, etc.), Xpress3 (data)
- Real-time writing to two HDF5 files via swmr
- Source data worker: Swmr, synchronize, assemble, batch pack, send to scheduling worker

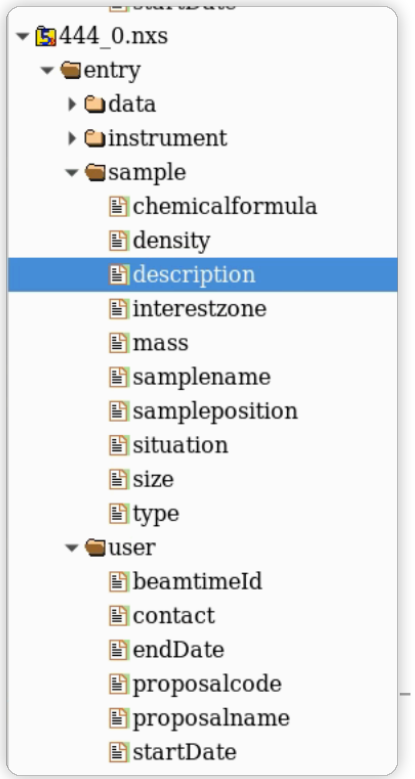


Outline

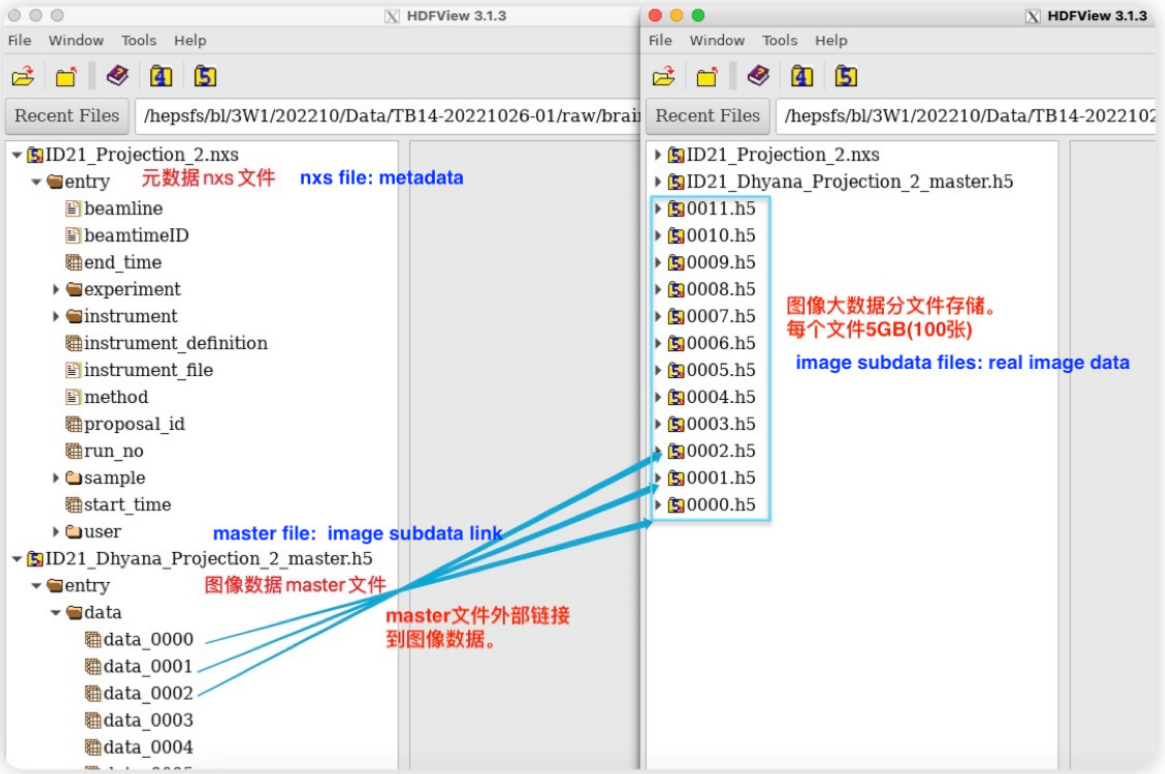
- MDW Framework Architecture
- MDW online data acquisition: Fly scan as an example
- MDW online data assembling and disk writing
  - Data and Metadata storage format specification
  - Customized parameter subsystem
  - Online disk writing architecture

data and metadata storage format specification

- nexus-like format, in order to provide data compatibility
- Allow beamline to define their own data storage format



Single nxs file



nxs+master+ data0+data1+…

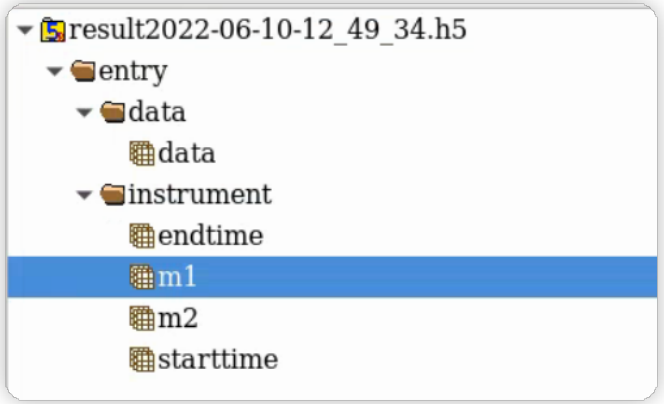
Outline

- MDW Framework Architecture
- MDW online data acquisition: Fly scan as an example
- MDW online data assembling and disk writing
  - Data and Metadata storage format specification
  - Customized parameter subsystem
  - Online disk writing architecture

Customized parameter mechanism

- Problem: different beamline stations has a different storage format.
- Purpose: Flexible configuration of which parameters of the metadata and data are written into specific location in HDF5 according to a configuration file defined by different beamline.

```yaml
storage:
  enable: true
  count: *aa
  files:
    file1:
      filename: result.h5
      directorypath: /home/mamba/mdw_result
      time_suffix: True
      stepscan:
        start:
          time: {datasetname: /entry/instrument/starttime}
        key2dataset:
          M_m2: {datasetname: /entry/data/m2, shape: [5,1], dtype: float64}
          M_m1: {datasetname: /entry/data/m1, shape: [5,1], dtype: float64}
          D_xsp3_image: {datasetname: /entry/data/data, shape: [5,4,4096], dtype: uint32}
        stop:
          time: {datasetname: /entry/instrument/endtime}
```



Pros:

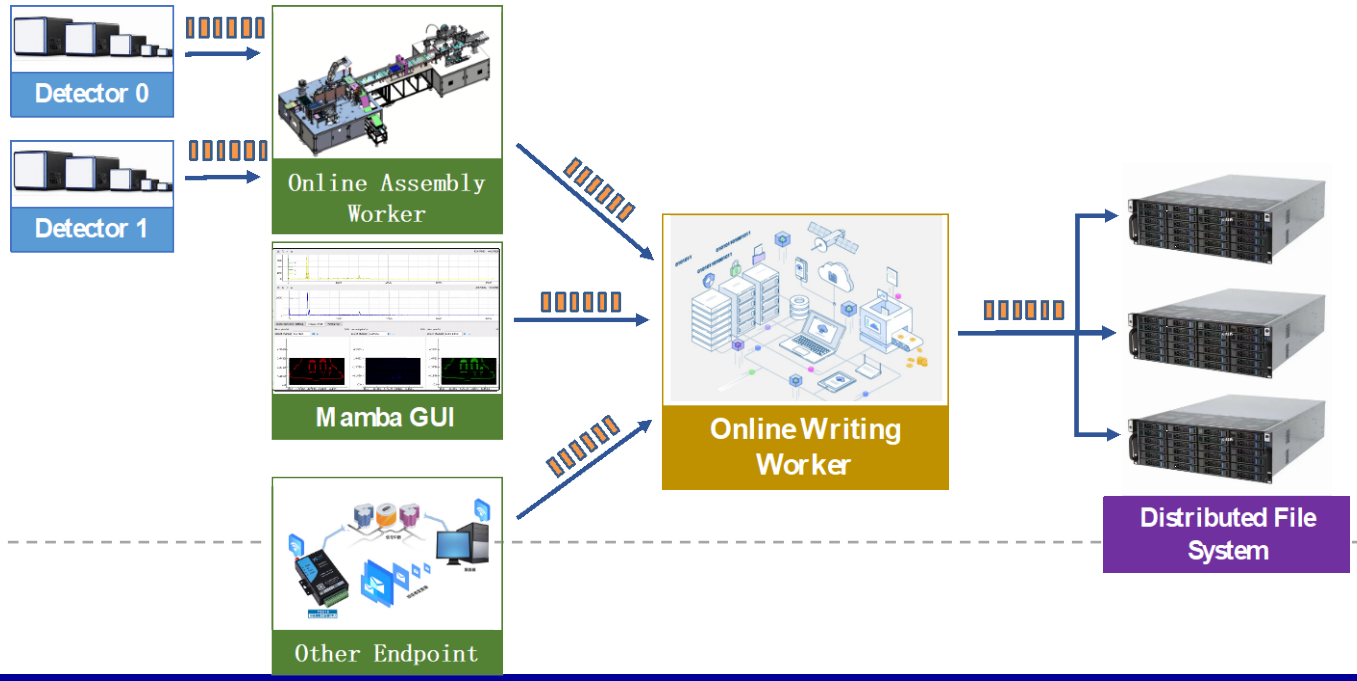
- ✓ Simple and easy to understand syntax
- ✓ Reduces duplicate development and improves development efficiency

Outline

- MDW Framework Architecture
- MDW online data acquisition: Fly scan as an example
- MDW online data assembling and disk writing
  - Data and Metadata storage format specification
  - Customized parameter subsystem
  - Online disk writing architecture

MDW Online disk writing architecture

- Separate online assembly from disk writing. Online assembly assembles source data into beamline station custom data format.
- Remote data storage. Workers can run on any machine.
- Batch writing data to distributed file system
- Supports multimodal data disk writing

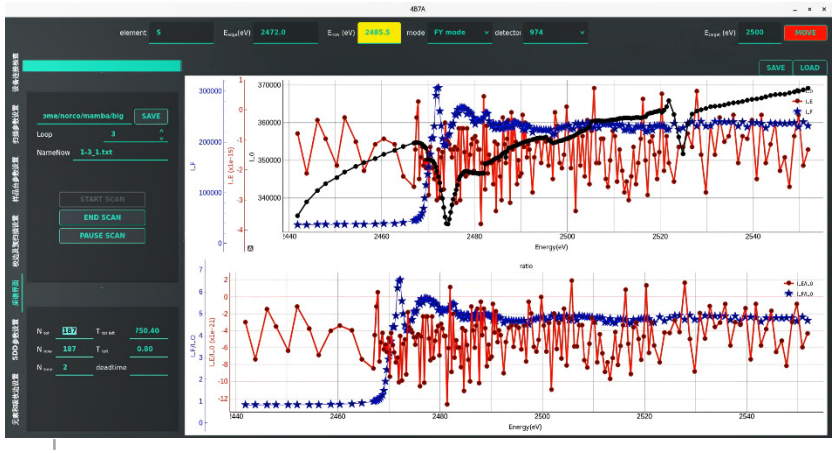


Outline

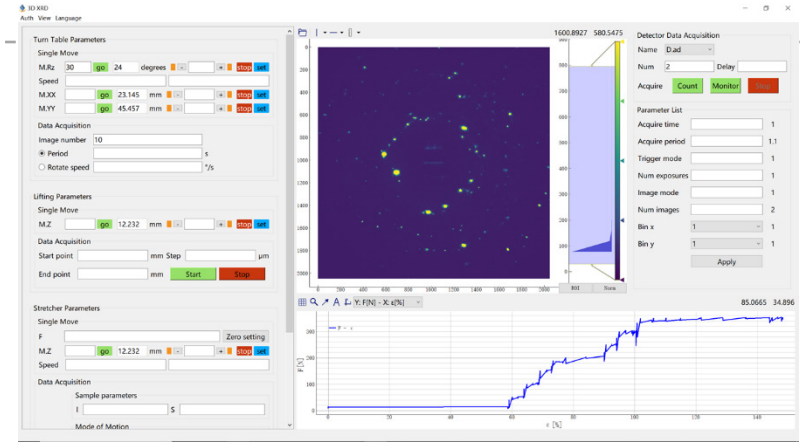
- **Project Background**
- **The Design of Mamba Data Worker**
- **Progress status of Mamba Data Worker at BSRF**

Application Progress of MDW in Mamba project

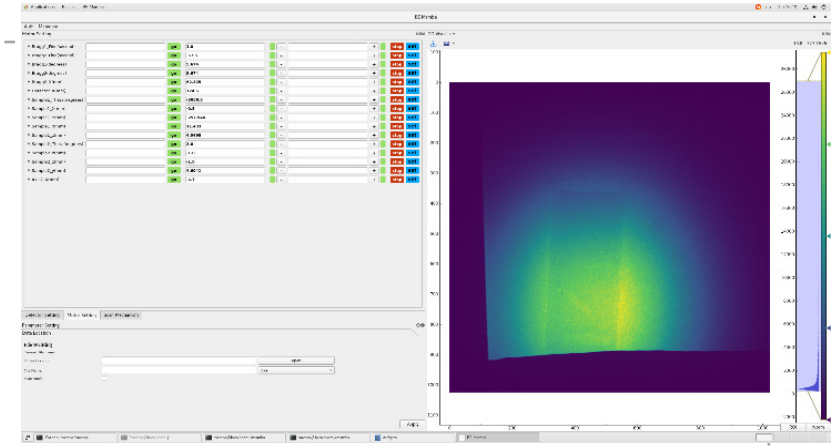
BSRF – 4B7A



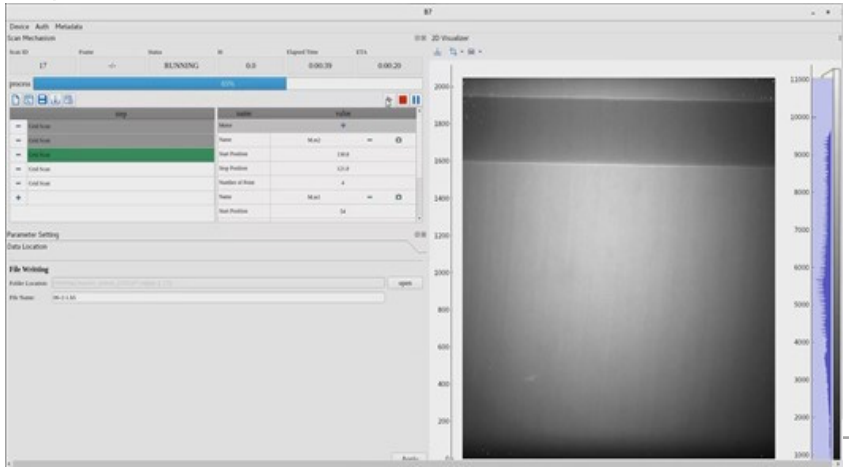
BSRF – 1W2A



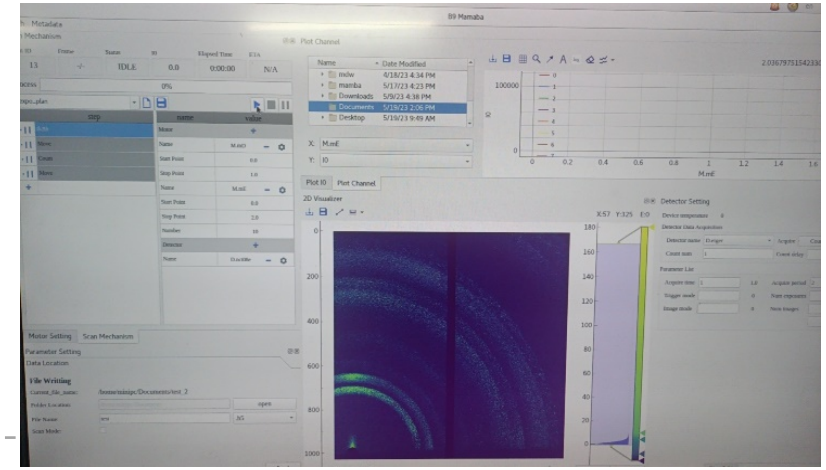
BSRF – 4W1A



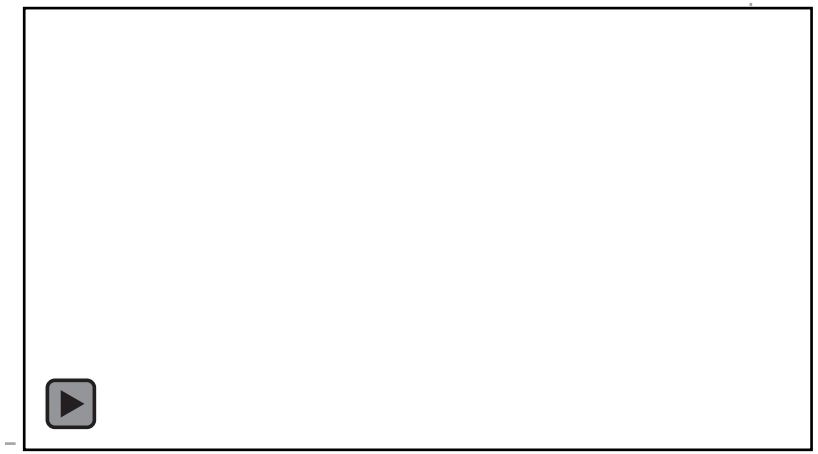
BSRF – 3W1



BSRF – 1W1A

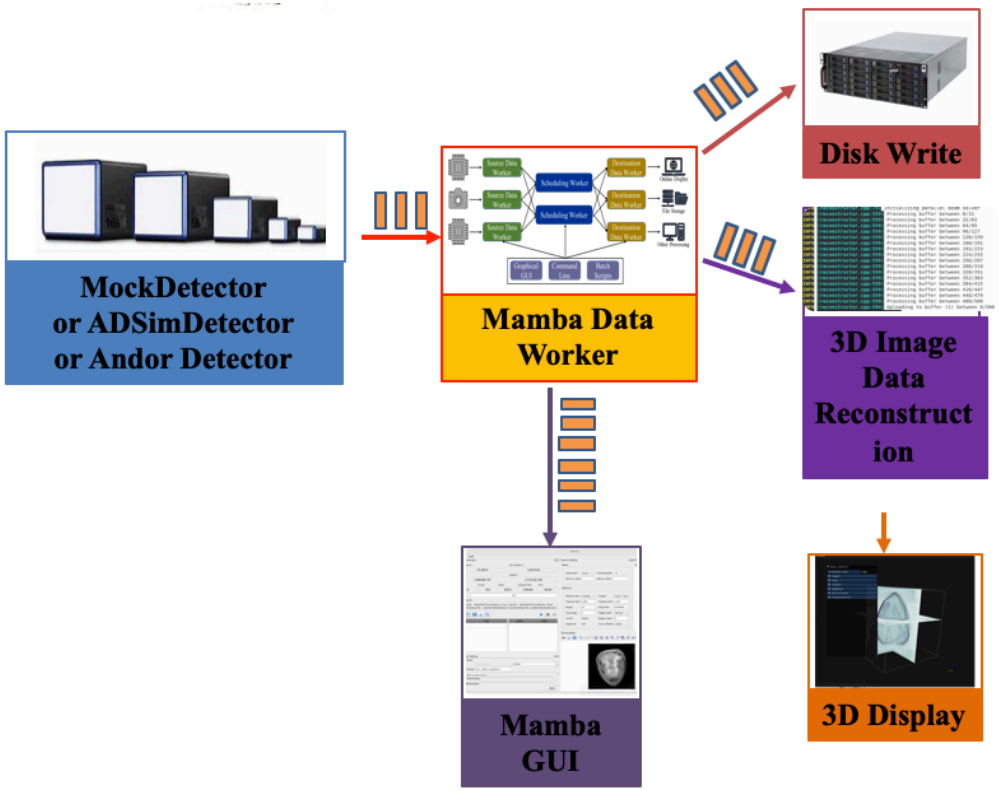


BSRF – 4B9B



- **Intensive development and test of *MDW* at *BSRF* beamline**

Case1: Real-time tomography reconstruction



Development stage:

1. DAQ Simulation testbench stage
   - Mock detector as source data worker
   - Data stream Application
2. Beamline deploy stage
   - Andor Detector

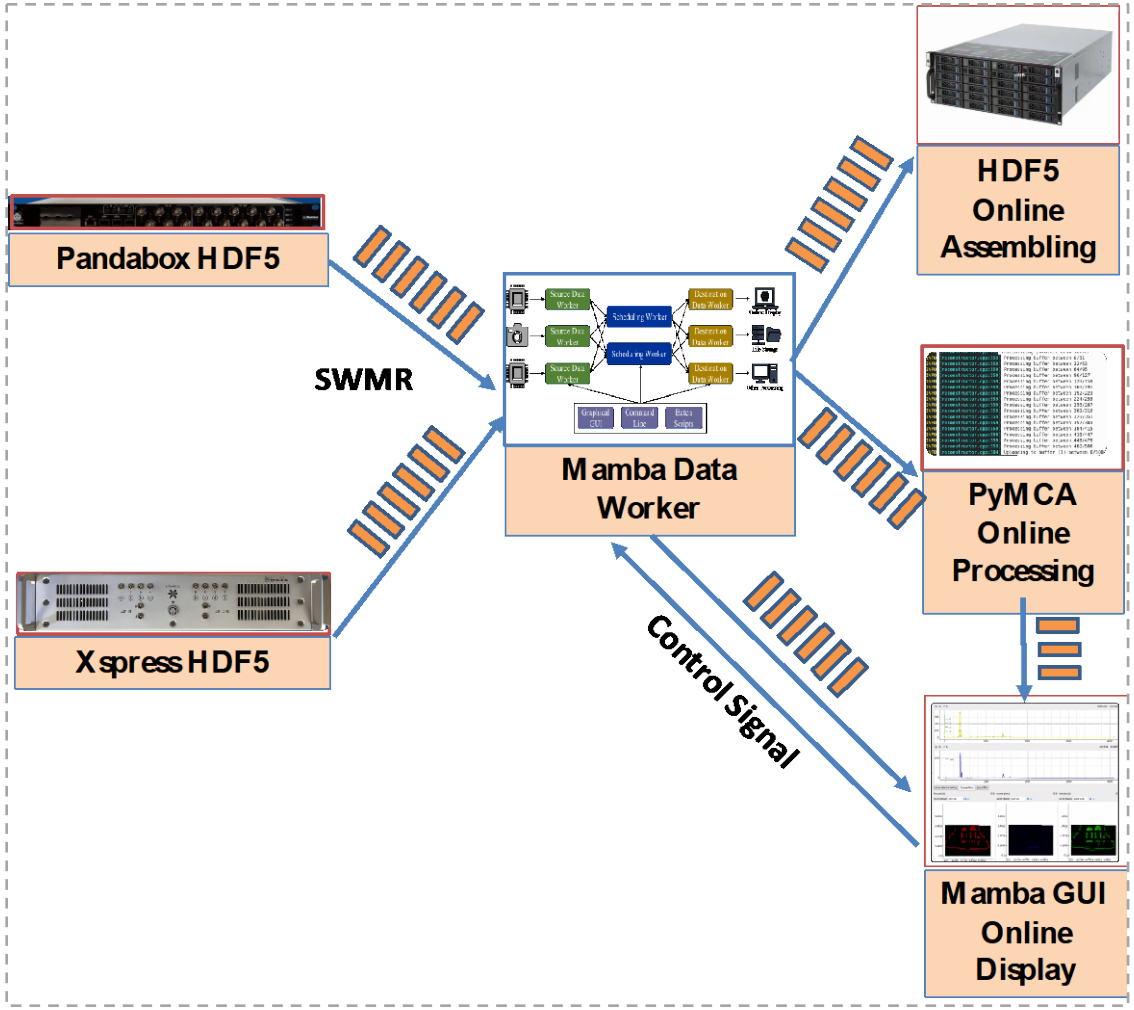
Three Data Stream Application:

1. Calculation: online reconstruction of 3D images
2. Display: 3D interactive display
3. Storage: HDF5

One-click data collection:

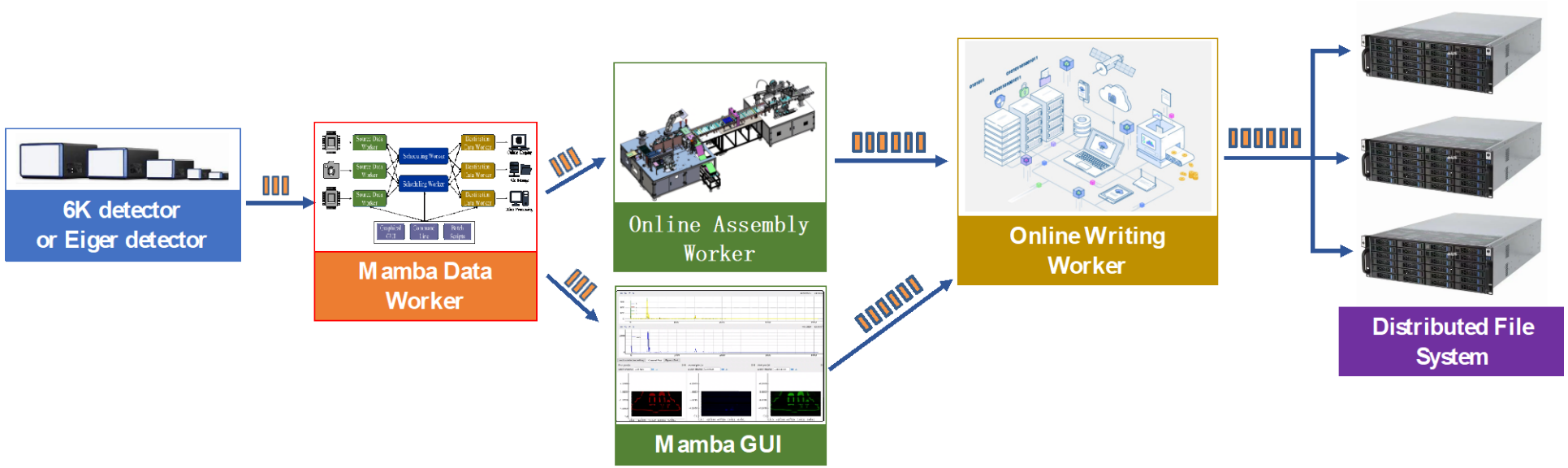
Online file storage、Online reconstruction、Online interactive display

Case 2: XRF mapping experiment

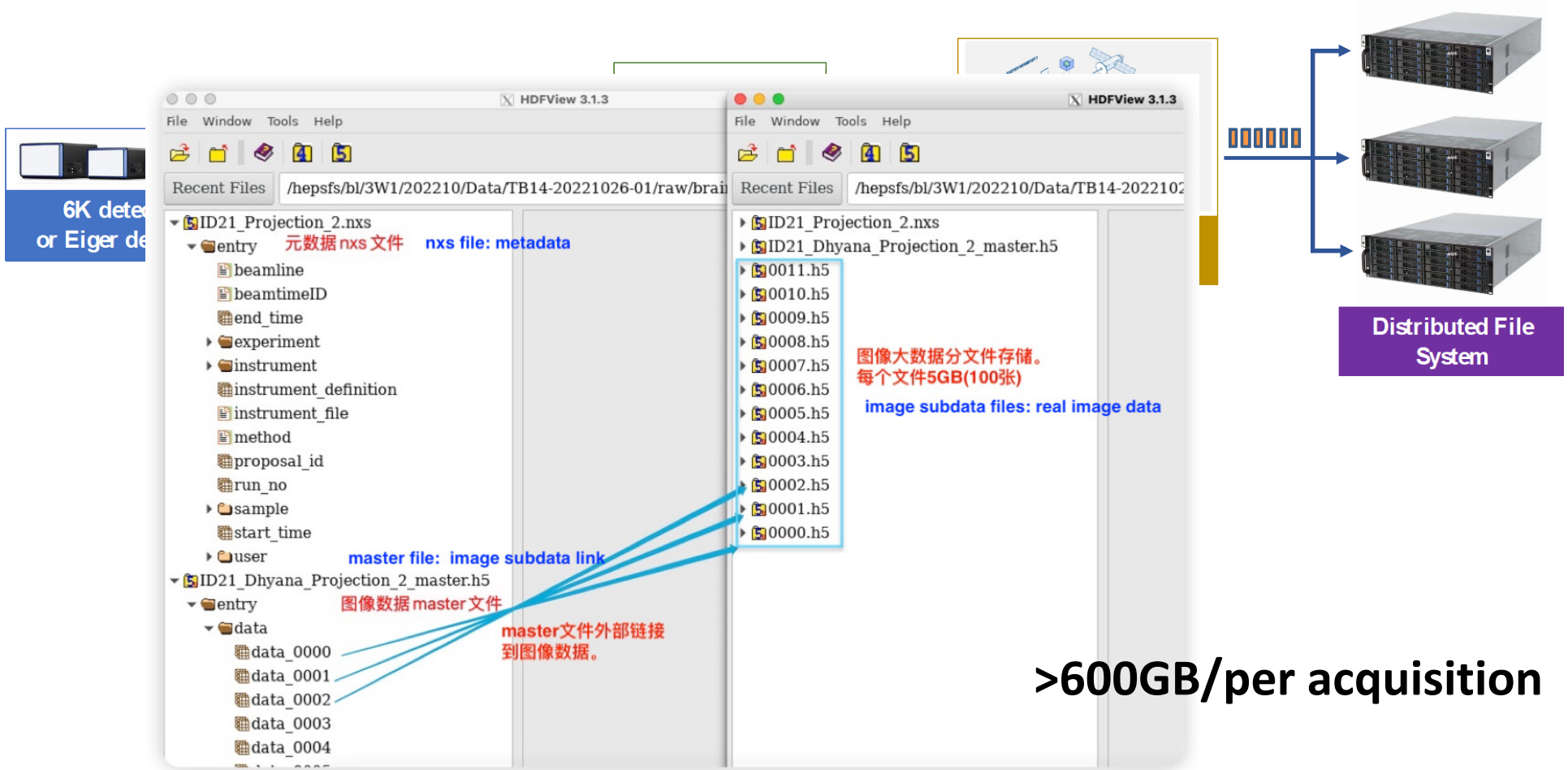


- Fly scan
- Pandabox、Xpress3 HDF5 SWMR
- PyMCA online processing

Case 3 Complex online data assembly &Writing



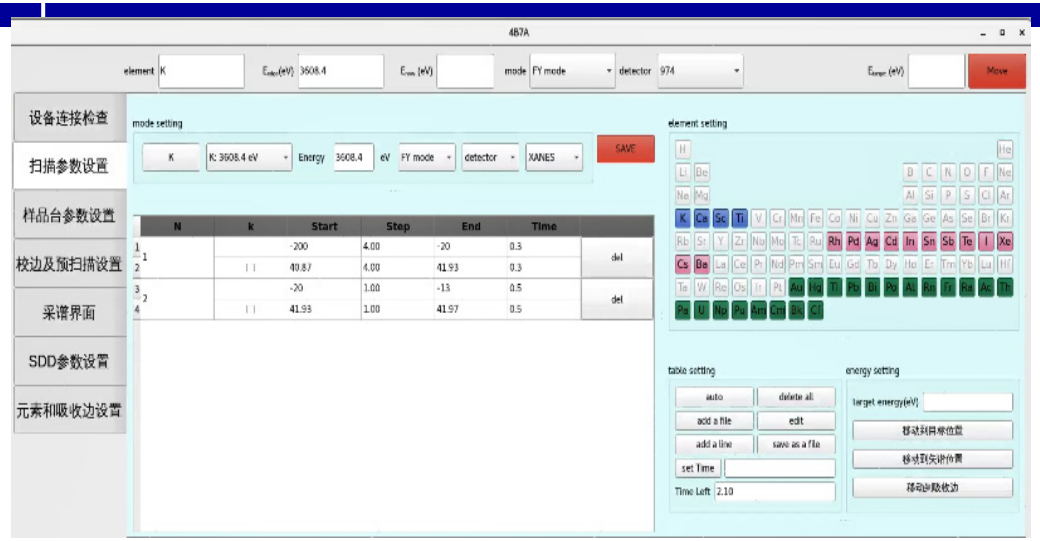
Case 3 Complex online data assembly &Writing



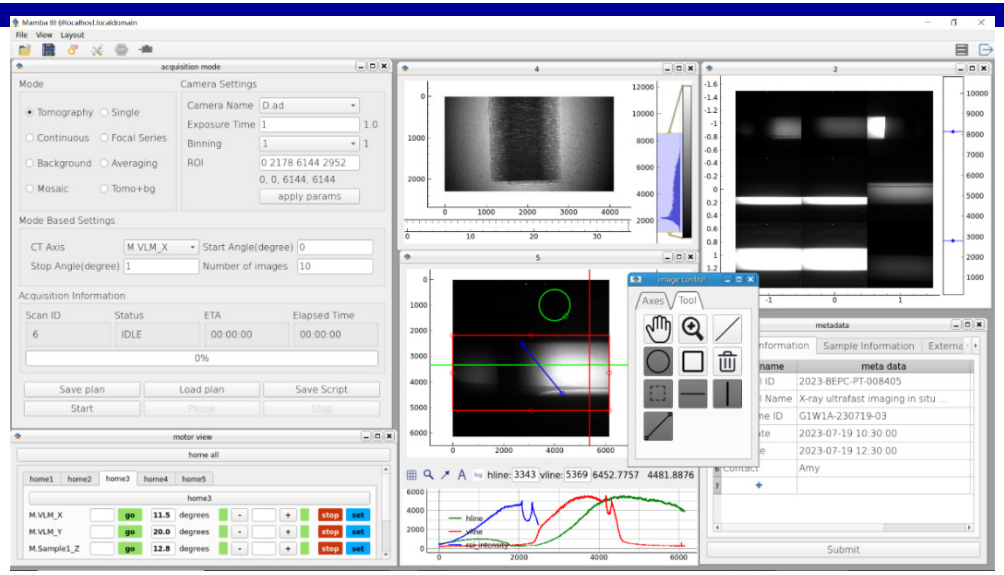
One click acquisition(automatic acquisition Flat, Dark, Projection)

Progress of MDW in Mamba project for HEPS

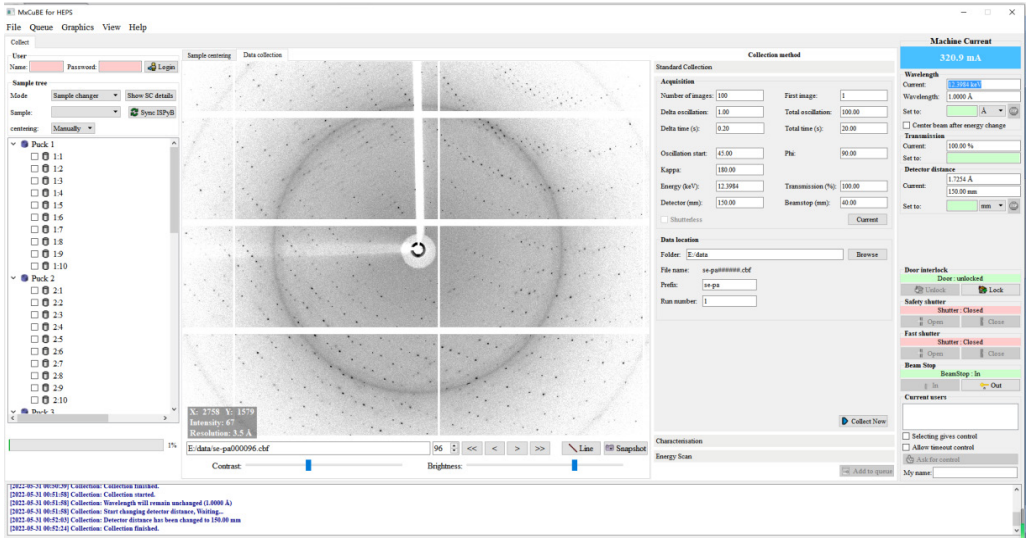
◆ B8



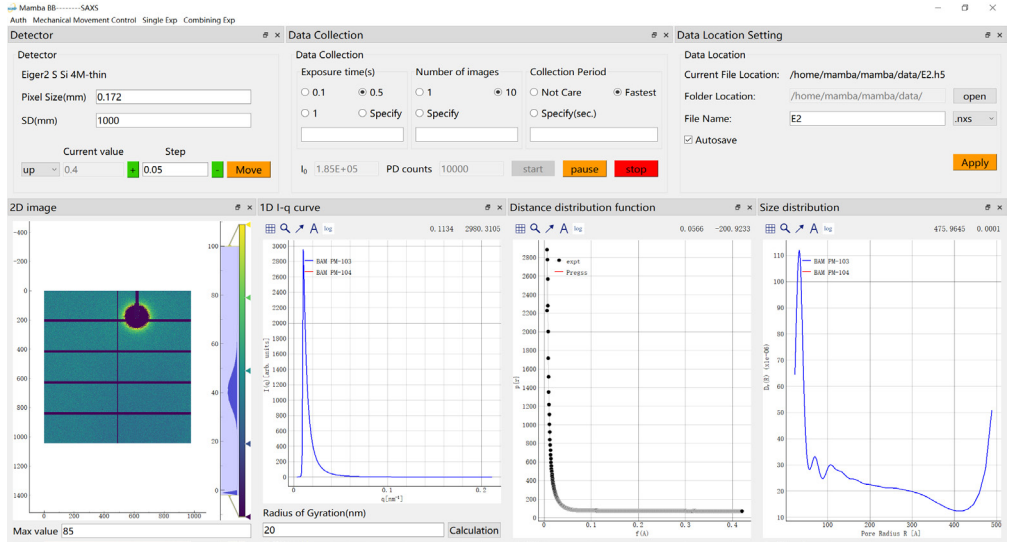
◆ BE



◆ BA



◆ BB



- **The first batch of *MDW* software for *HEPS* beamline is underway**

Thanks for your attention!