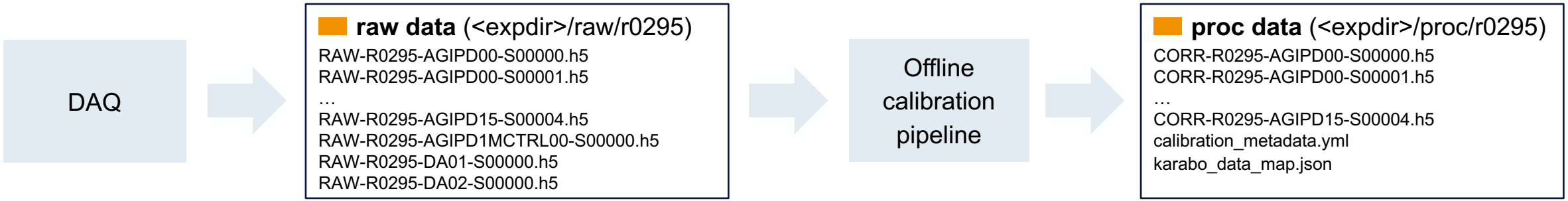
Update on European XFEL data reduction activities

Egor Sobolev
Data Analysis Group, European XFEL

LEAPS-INNOV Annual Meeting 2022
Barcelona, May 3–5, 2022

Data acquisition and processing at European XFEL



- Offline calibration pipeline was dramatically **improved** in last year
  - **Success**: more users work with calibrated data
  - **Price**: the amount of data **almost doubling (we additionally store only calibrated images)**

Data management & reduction

- **Options**:
  - Archive raw data if users can work with calibrated data
  - Remove proc data if they aren't used actively, and reproduce on demand

- **Solution:**
  - **Data Management Plan**
    (Task force was created for 6 month to update Data Policy)
  - **Data Reduction**
    - ► we offer methods, users decide which and when to apply
    - ► Address proc data
    - ► Implement simplest methods first
    - ► **Automated reports**
    - ► **Goal**: 10 time reduction

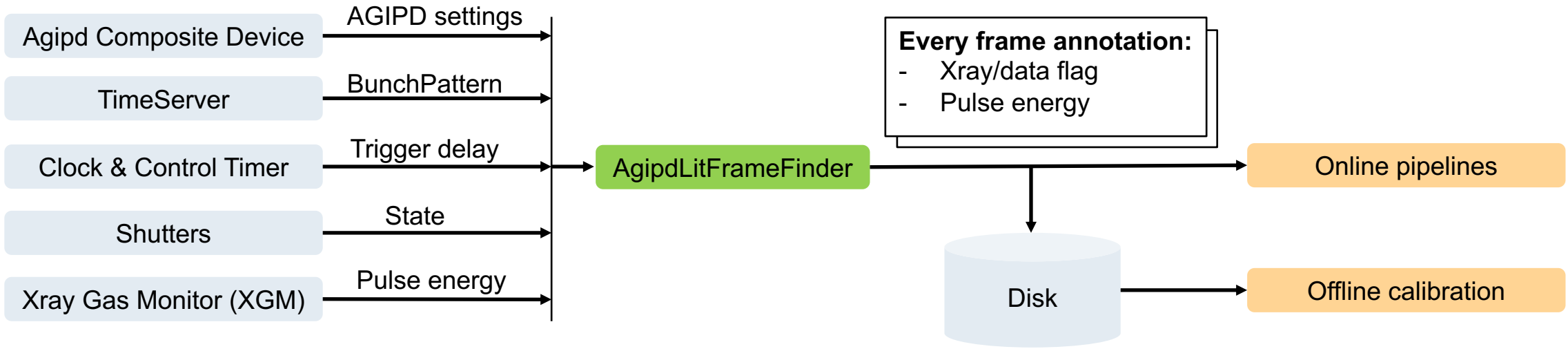
Dark frames annotation and deletion

- **Problem:** detectors expose frames which are not illuminated with Xray
  - Internal VETO mechanism cannot be used due to technical limitation
  - Difficult to reconfigure
- **Solution**: generate **annotations** of frames and use them for filtering

proposal	Instr/Det	Pulse/Frame	Lit, TiB	Total, TiB	Ratio, %
2316	SPB/AGIPD	100/128	213.6 TiB	283.9 TiB	75.25
2542	MID/AGIPD	50/250*	142.1 TiB	651.0 TiB	21.83
2568	MID/AGIPD	10/250*	9.1 TiB	203.2 TiB	4.46
2543	MID/AGIPD	1/250*	2.1 TiB	311.7 TiB	0.66
2595	MID/AGIPD	300/352*	399.0 TiB	506.0TiB	78.92
2671	SPB/AGIPD	348/352	937.0 TiB	948.0 TiB	98.84
2832	MID/AGIPD	2/64	13.0 TiB	256.0TiB	5.08

* Different number of Xray pulses were used, indicated only most representative

Dark frames annotation and deletion




```
xfel-calibrate AGIPD CORRECT \
        --use-litframe-device SPB_IRU_AGIPD1M1/REDU/LITFRM \
        --energy-threshold 300 \
        [ … ]
```

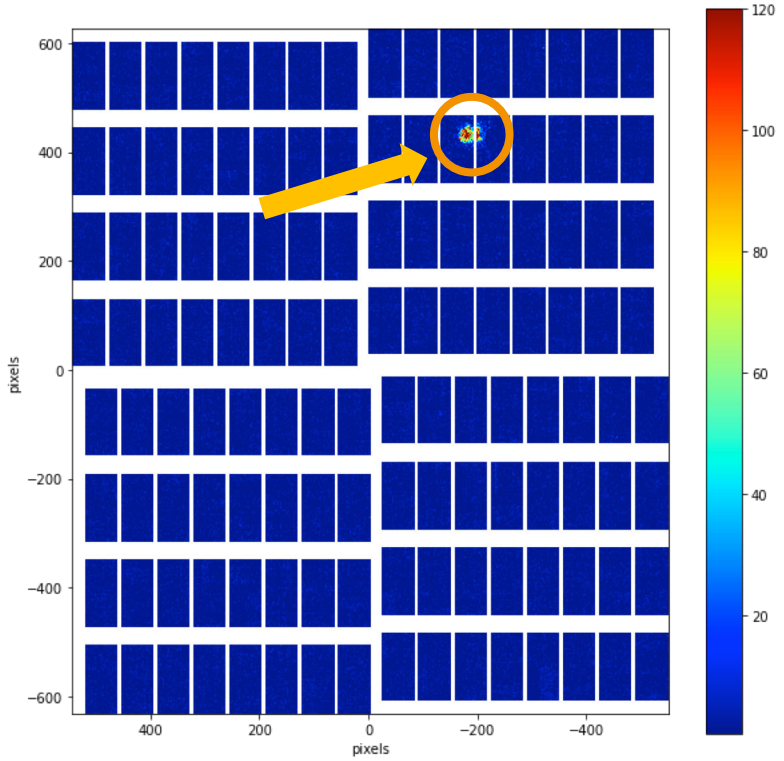
+

- Data filtering at later data management stages

Delete dark modules of multimodule detectors

- **Problem**: only part of multimodule detector is used
  - Signal location can changing during experiment
  - DAQ cannot be accordingly adjusted on-fly
- **Solution**:
  - Record **manual selection** as annotations
  - Automated detection (at least as computer-aided tool)

diffuse scattering around the crystal Bragg peaks

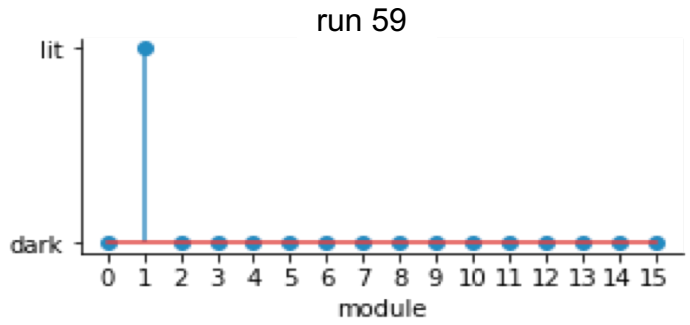


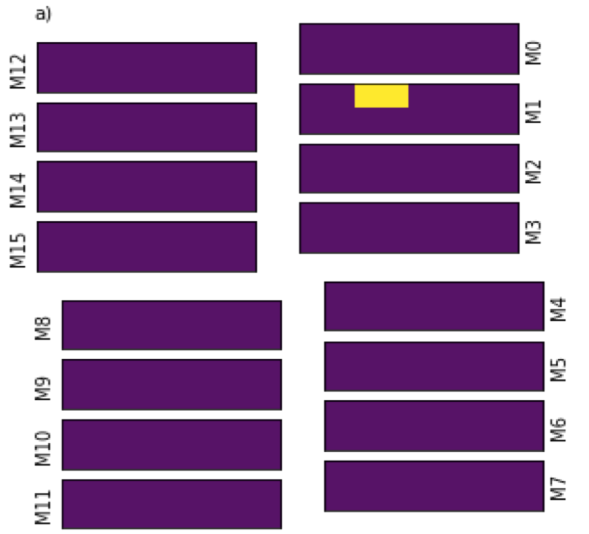
Pulse/Frame pattern used:
1/64, 63/64, 300/352

Total size: **506 TiB**
Only lit frames: **399 TiB (1.27:1)**
One module: **25 TiB (20:1)**

Automated signal location

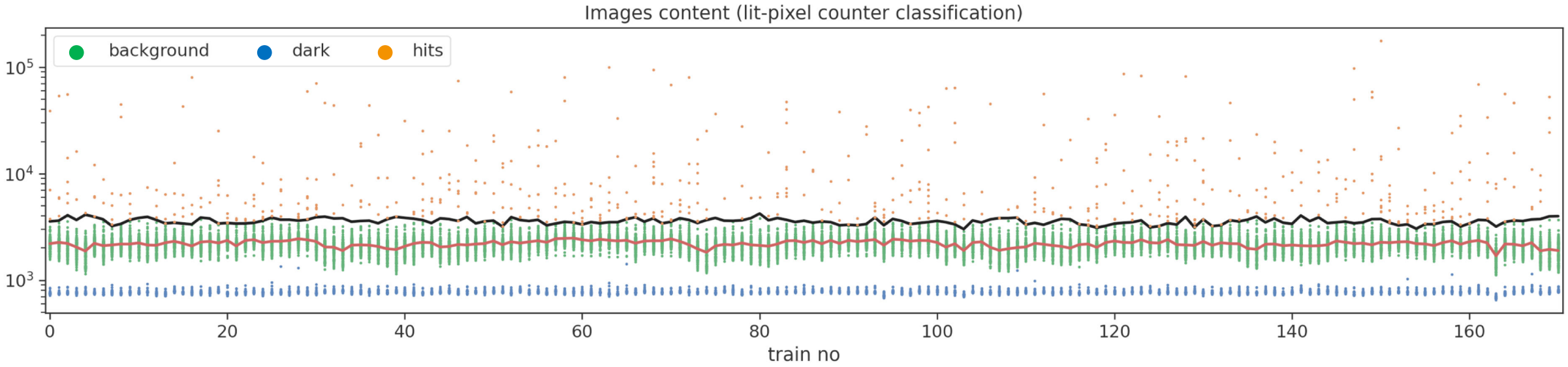
- **Lit-pixel counter statistics**:
  - compute the number pixels with photons in given region
  - used in SPI for hit-finding
- **Advantages**:
  - robust
  - low complexity (~pedestal correction)





Dark/lit/hit image classification

- SPB, Proposal 900230, run 275, **single particle diffraction**, Au bipyramid
- For **other experimental techniques** may be used as **diagnostic tool**



Red line – median of lit-pixels number (in train)
Black line – median + 4 IQR (in train)

Lossy compression

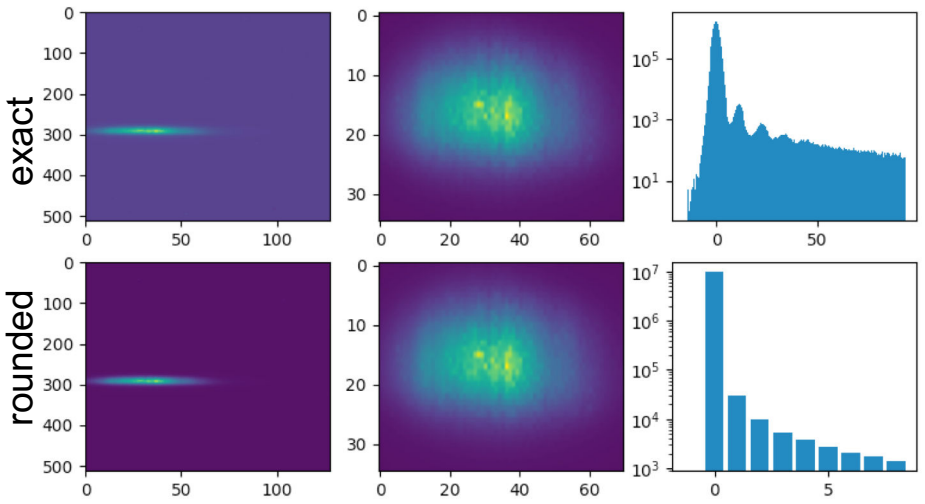
- Error bound rounding + Lossless compression
  - Photon conversion (may work as zero suppression and reach CR=50)
  - Rounding up to few highest significant digits (e.g. SFX, SAXS, etc)

SFX data

dtype	f32	f16	f32	f32	f16	f32	f16	f32	f16
compression	uncomp	uncomp	gzip 1	gzip 1	gzip 1	gzip 1	gzip 1	gzip 1	gzip 1
rel. prec				5 bit	5 bit			5 bit	5 bit
abs. prec						int	int	int	int
Size (MB)	22440	11220	18861	5211	4982	4319	3754	4306	3739
CR	1	2	1,19	4,31	4,5	5,2	5,98	5,21	6
read time, s	9	5	148	111	88	109	71	105	69
read ratio	1	0,55	16,17	12,14	9,54	11,87	7,73	11,4	7,55

Inspired by M. Galchenkova & O. Yefanov, CFEL

25:1 ratio in uint16, 10:1 ratio in float32



Summary

- **Introduction of Data Management Plan**
- Dark frame deletion
- Introduce on-fly selection of frames and modules
- Run report on signal location and dark/lit frames
- SPI hit-finding
- Lossy compression

Thank you for you attention. Questions?

Egor Sobolev, egor.sobolev@xfel.eu
Data Analysis Group, da@xfel.eu, www.xfel.eu/data_analysis
Support: da-support@xfel.eu