Detector Calibration

Karabo Control and Data Analysis at the European XFEL
24.01.2019

Steffen Hauf for DET

Calibration process

Described in Read-the-docs
https://in.xfel.eu/readthedocs/docs/detector-documentation/en/latest/

- Raw detector data is an array of ADUs, which contains various effects does not related to the detected photon flux.
  - Several examples are given in this presentation
- Calibration transfers ADUs to photons and corrects for these effects.
- Correction processing
  - Offline
  - Online
- Calibrated data sets are the main user-accessible data
  - Quality of the calibrated data is done to the best knowledge of the facility
  - Data can be re-calibrated at a later time
  - Speak to DET first if raw data is to be accessed

Detectors at the European XFEL

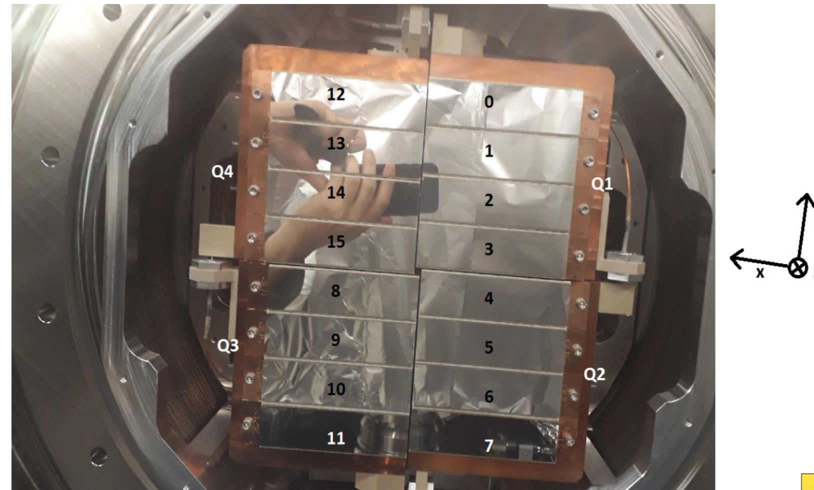
2D detector	Pixel size *(μm^2)*	Energy *(keV)*	Dynamic range	Frame rate	Instrument
AGIPD	200x200	3 - 16	10^4 ph @ 12 keV	4.5 MHz	SPB, MID
DSSC	204x204*	0.5 - 16	10^4 ph @ 1 keV	4.5 MHz	SCS, SQS
Epix 100a	50x50	3 - 20	100 ph @ 8 keV	10 Hz	HED, MID
Epix 10Ka	100x100	3 - 25	10^4 ph @ 8 keV	10 Hz	HED
FastCCD	30x30	0.25 - 6	10^3 ph @ 0.5 keV	10 Hz	SCS
JUNGFRAU	75x75	3 - 25	10^4 ph @ 12 keV	10 Hz	HED
LPD	500x500	5 - 20	10^5 ph @ 12 keV	4.5 MHz	FXE
pnCCD	75x75	0.03 - 25	$6x10^3$ph @ 1 keV	10 Hz	SCS, SQS

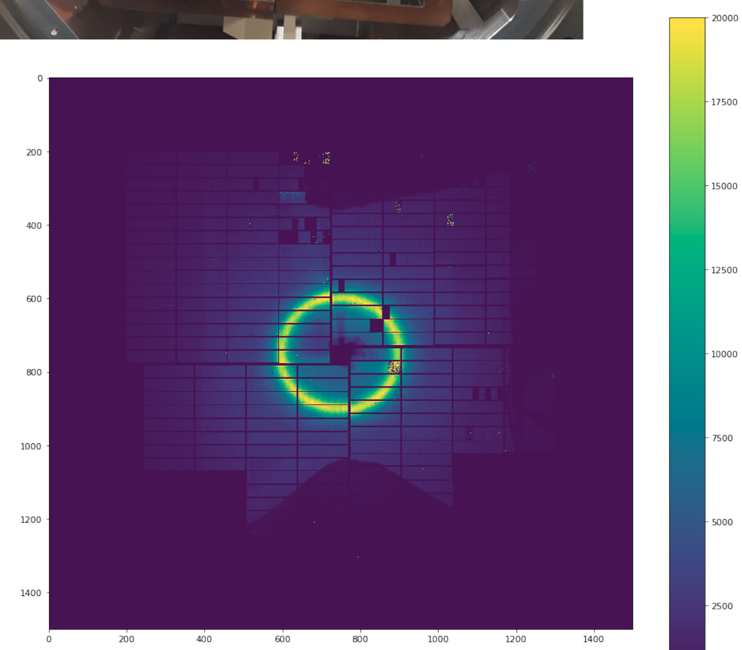
* - hexagonal pixels

Detectors at the European XFEL – Challenges for Calibration and Data Processing

- Different technologies:
  - (hybrid) active pixel
  - Multiple gain stages
  - CCD
- Single module and segmented detectors
  - Modules, Supermodules and tiles
  - e.g. LPD: 256 sensor modules
- 10 Hz and 4.5 MHz
  → Mbytes/s – 10 Gbytes/s

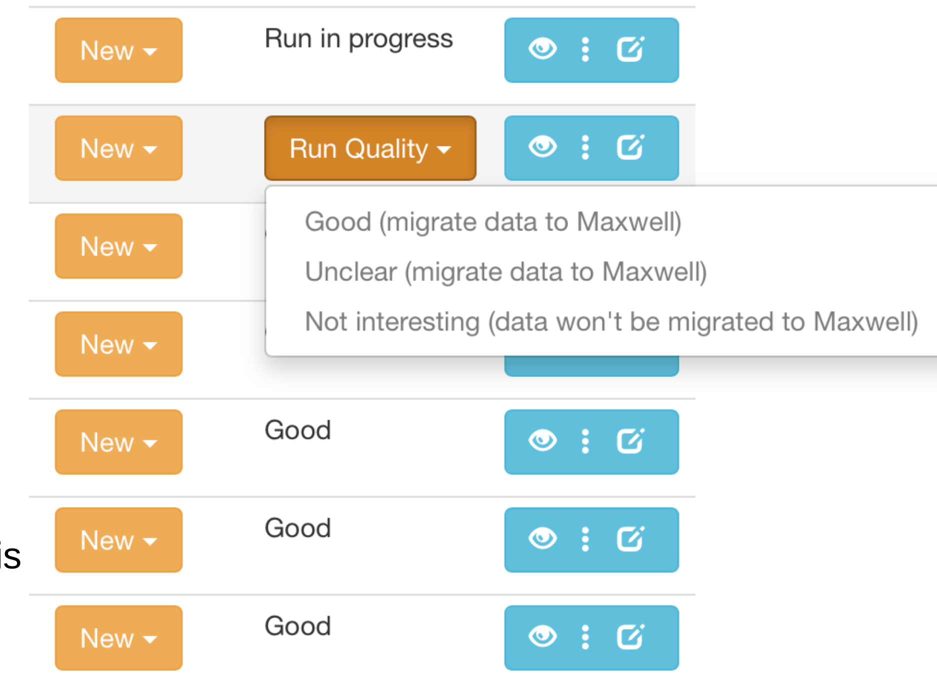
→ **Common tools as much as possible, specialize where necessary**



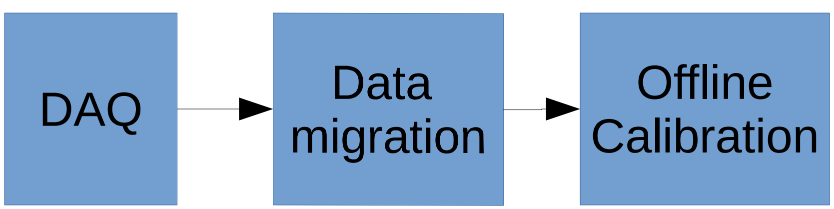


Offline Calibration

- Jupyter notebooks containing correction and characterization code
  - Less overhead
  - Less opaque
- Common front-end script across detectors
- Run both module concurrent on server maxwell cluster nodes
- Features automated processing whenever data is migrated from online to offline processing via Metadata catalog

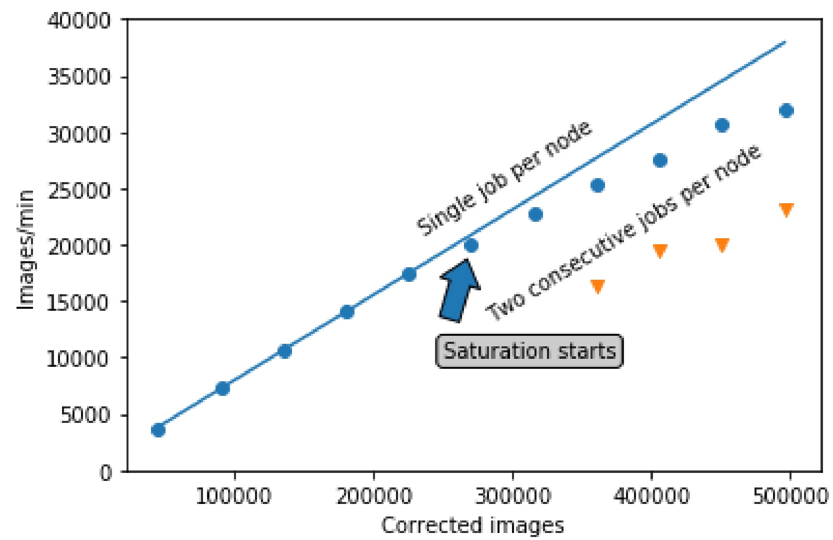


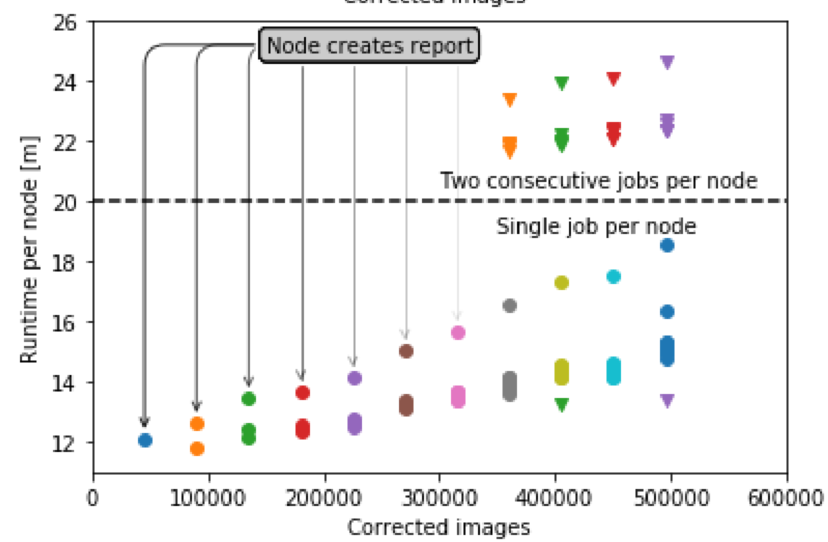
	AGIPD	LPD
Gain evaluation	X	X
Offset correction	X	X
Relative gain correction	X	X
Bad pixels	G/O/N	G/O/N



Offline Calibration - benchmarking

- Benchmarking is performed on Maxwell cluster
  - 5 min DAQ (120 pulses) → 15 min Offline processing
  - About 4000 MPix images per min per node (16 CPUs)
    - Memory limited
  - For multiple subsequent runs in quick succession we might run out of nodes if we simply scale:
    - Feedback latency would increase significantly for later runs
      –> Solution: limit each run to N node maximum (currently 8), have multiple jobs per node
- Good horizontal scaling has been shown
  - No saturation up to 5 nodes
  - Limiting factors:
    - Cal DB access
    - I/O (to be investigated)





Offline Calibration – Validation of calibration algorithms

- Test suite tests against artifacts from previous tests to verify data produced from a defined run has not changed
  - Test structure
  - Test Karabo Data compatibility
  - Test checksums
  - Statistical tests on histograms of data at various level
- If changes are intended, new artifacts can be generate
- Documentation keeps track of test results

Table Of Contents

ad47d9a5 - 2018-11-16 08:52:25.314479

Test execution for TestAGIPDCorrection on 2018-11-16 14:57:12

Git tag: ad47d9a5159e3bc632d64532eef82a6c47130547
Tests: 6
Errors: 1
Failures: 0
Skipped: 4
Duration: 0.116s

Detailed Results

Result	Test	Error	Message	Duration (s)
passed	test_generate_checksums			0.104
failed	test_checksums	NameError	name 'ckfname' is not defined	0.012
skipped	test_generate_histograms	skip	Artifact generation is not requested	0
skipped	test_generate_karabo_data	skip	Artifact generation is not requested	0
skipped	test_histograms	skip	User requested to skip histogram test	0
skipped	test_karabo_data	skip	User requested to skip karabo data test	0

Failure report for: test_checksums

```
Traceback (most recent call last):
  File "/home/haufs/pycalibrate_tmp/tests/correction_base.
    print(ckfname)
NameError: name 'ckfname' is not defined
```

Basic characterization parameters

*Corrections applied by default

- Offset*
  - Dark current is negligible for MHz detectors, but important for CCDs
  - Derived from dark image
- Noise
  - Standard deviation of dark signal
  - In general non-correctable
- Bad pixel
  - Bit mask, which describes peculiarity of a given pixel
- Gain*
  - Converting factor from ADUs to number of photon
  - Derived from measurement with x-rays and injected charge using pulse capacitor and current source, depending on detector type.
  - Affected by charge transfer efficiency and memory droop, depending on detector type.

All values are evaluated per pixel, memory cell and gain stage

Gain characterization (MHz detectors)

- Both LPD and AGIPD detectors have 3 gain stages
- Gain assumed to be linear within a given stage and applied after offset correction
  - Initial assumption only

$$I' = (I - O^{x,y,m}) G^{x,y,m}$$

- To reduce number of images a high gain $G_{FF,high}$ is derived for a given pixel over all memory cells using x-ray data.
- Pulse capacitor data are used to derive relative slope of a given pixels memory cells with respect to all memory cells: $G_{PC,high} = m_h / \langle m_h \rangle_{mem}$
- Factor between high and medium (medium and low) gain is derived with the same way over all memory cells: $f_{PC} = \langle m_m \rangle_{mem} / \langle m_h \rangle_{mem}$
  - For AGIPD medium to low factor is derived from charge injection data
- For AGIPD Choice between gain stages are performed based on two thresholds T_0 and T_1

$$G^{x,y,m} = \begin{cases} G_{high}^{x,y,m} = G_{FF,high}^{x,y} * G_{PC,high}^{x,y,m} & \text{if } I > T_1 \\ G_{medium}^{x,y,m} = G_{high}^{x,y,m} * f_{PC,medium}^{x,y} & \text{if } T_0 < I < T_1 \\ G_{low}^{x,y,m} = G_{medium}^{x,y,m} * f_{PC,low}^{x,y} & \text{if } I < T_0 \end{cases}$$

  - For LPD gain is already encoded digitally

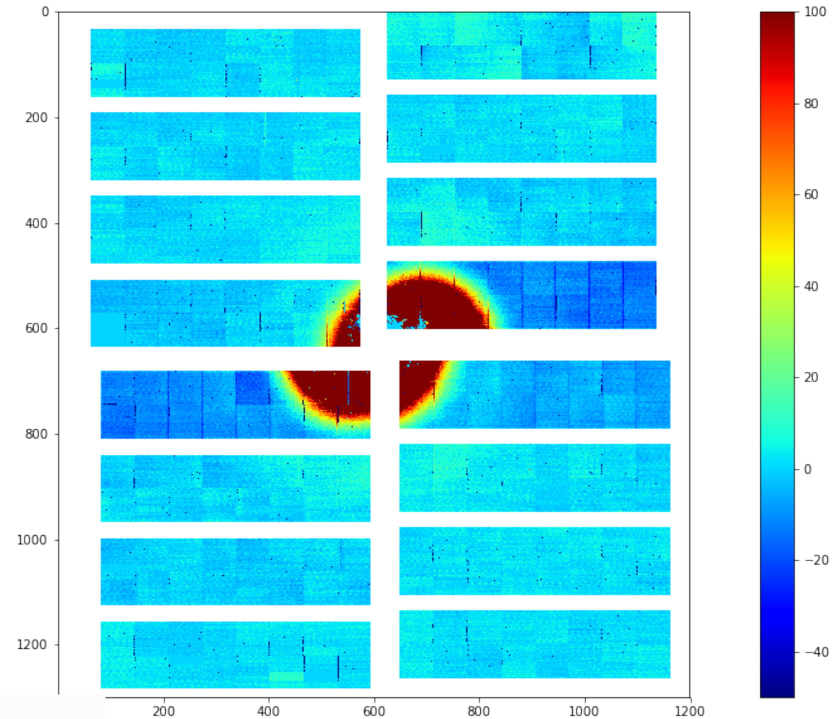
Detector-specific related effects

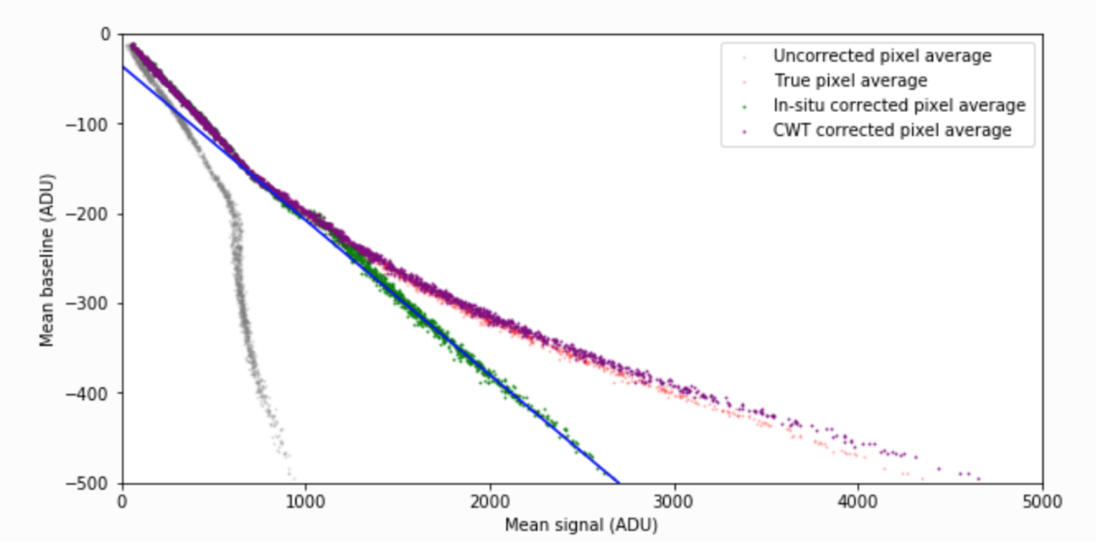
* Corrections applied by default

- AGIPD
  - Baseline shifts
  - ASIC level drop-offs
  - “snowy” pixels, i.e. pixels in gain transition region
- LPD
  - Gain non-linearity in the gain-transition region
- CCD detectors
  - Common mode
  - Charge transfer inefficiency*
  - Charge splitting*

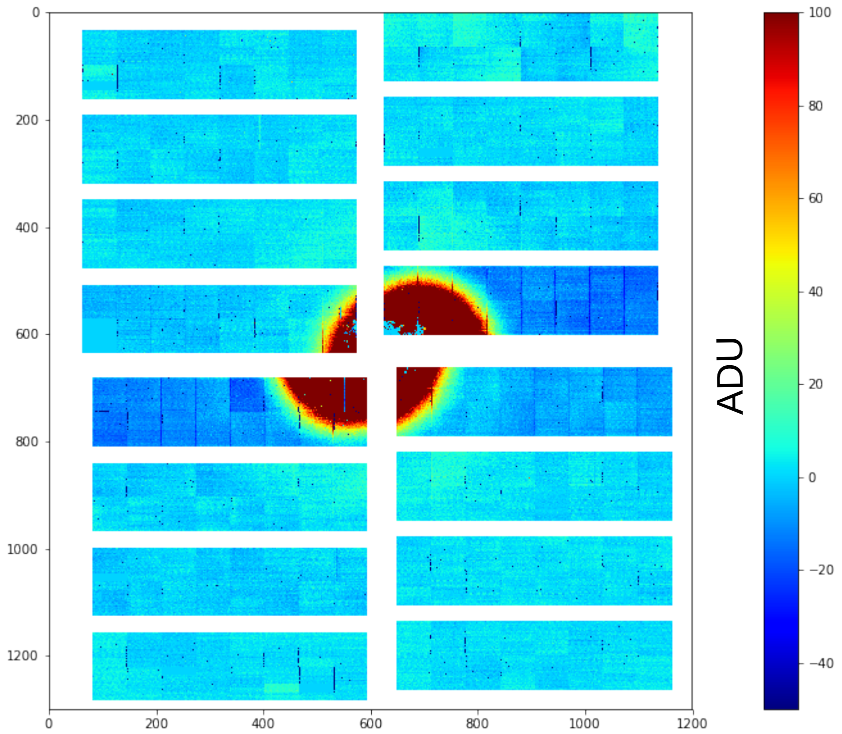
Baseline shifts in AGIPD

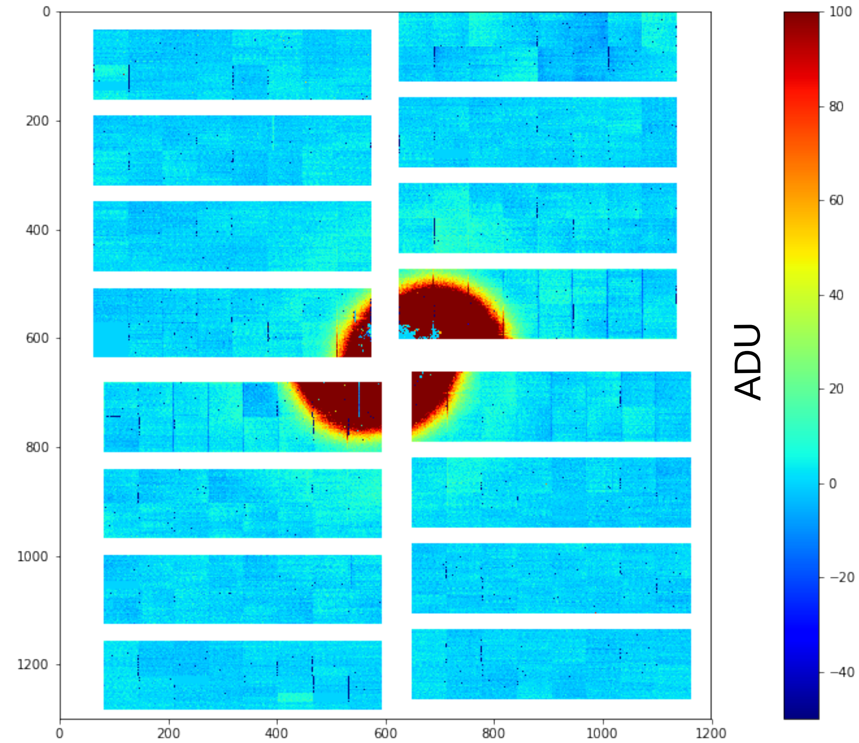
- Baseline of modules, containing a significant fraction of photons is shifted to lower values
- Dedicated characterization of baseline shifting by partially masking detector
  - Points averaged over pixel and cells
  - Functional dependency between input data and baselines shift is not trivial due to gain switching (green and gray points)
  - CWT-based noise peak evaluation method works good if a statistically relevant amount of pixels has no signal in every image





Baseline shifts in AGIPD



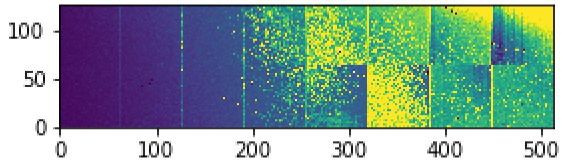


- Corrected by shifting a noise peak location to 0
  - Alternatively histogram matching between high and medium gain can be used (less stable)

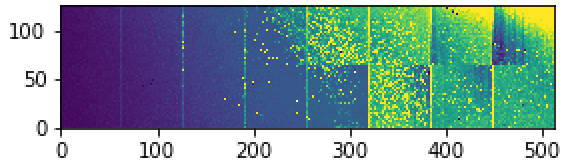
Snowy pixels in AGIPD

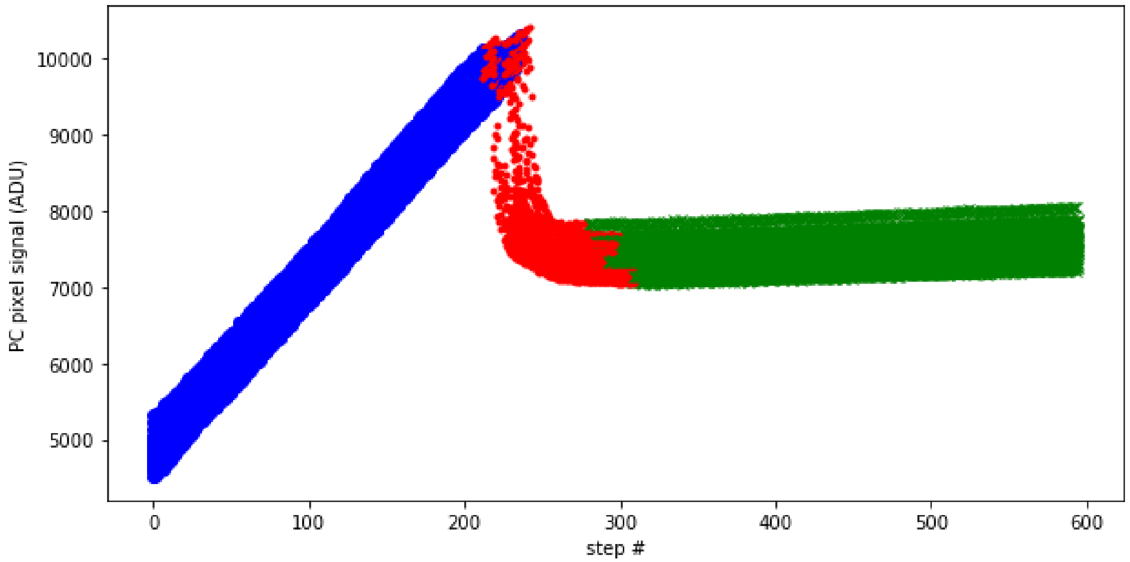
- Pixels with values in gain transition region evaluated to medium gain stage but have transitional image values between the largest high gain value and the smaller medium gain value.
- Identification is performed using Gaussian mixture clustering
  - Not all pixels can be identified
- Value set to nan or to interpolated value of neighbors
  - Identified in the bad pixel mask

Before correction



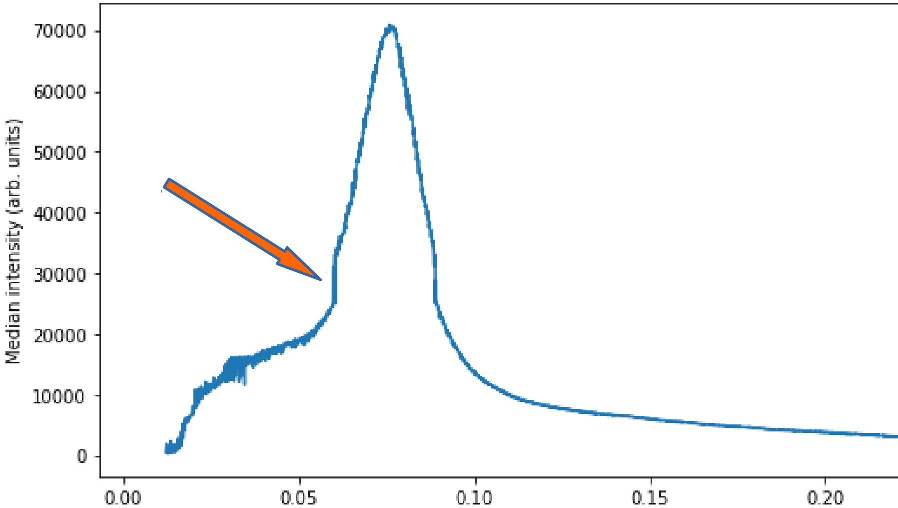
After correction

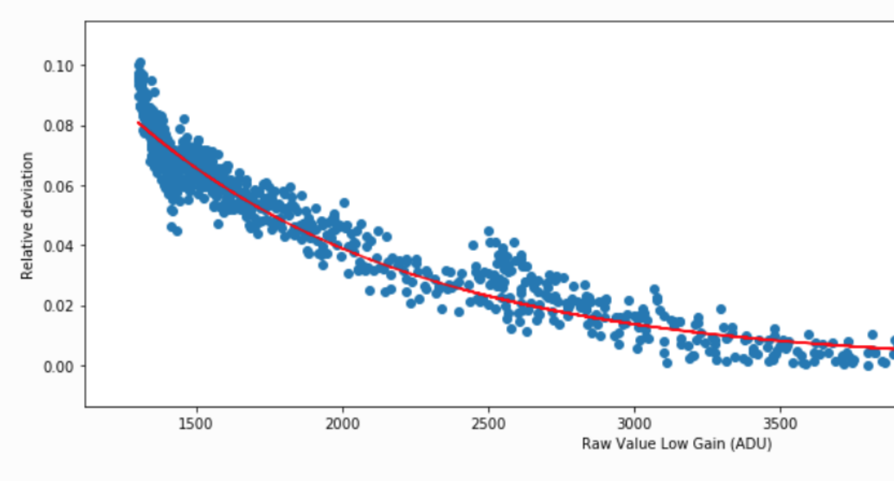




Gain non-linearity in LPD

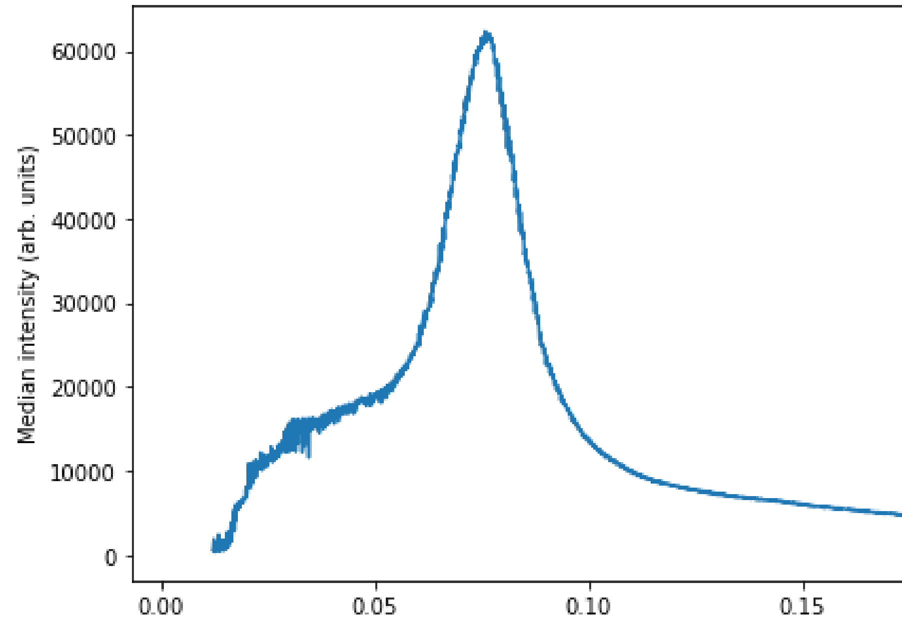
- Users observed unphysical discontinuity in radial profiles at gain switching region
  - Offline characterization based on Jupyter notebooks allowed for quick evaluation
  - Found non-linearity which was phenomenologically corrected during user beamtime

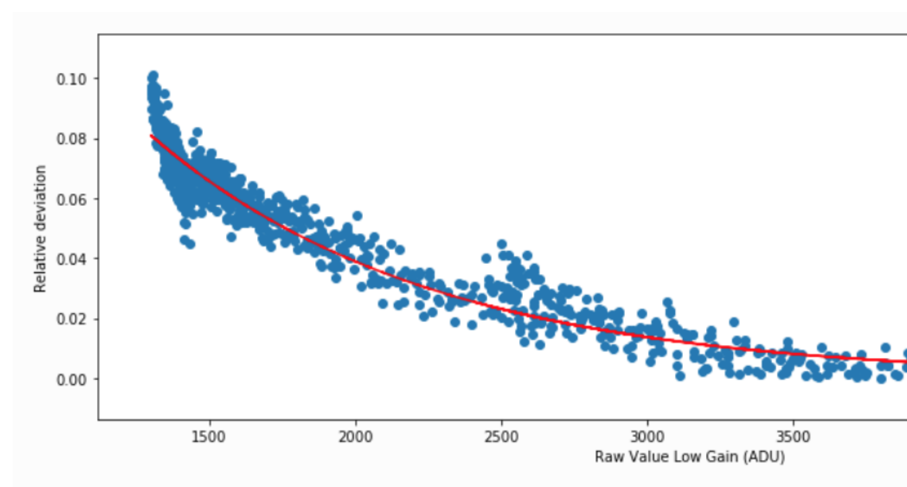




Gain non-linearity in LPD

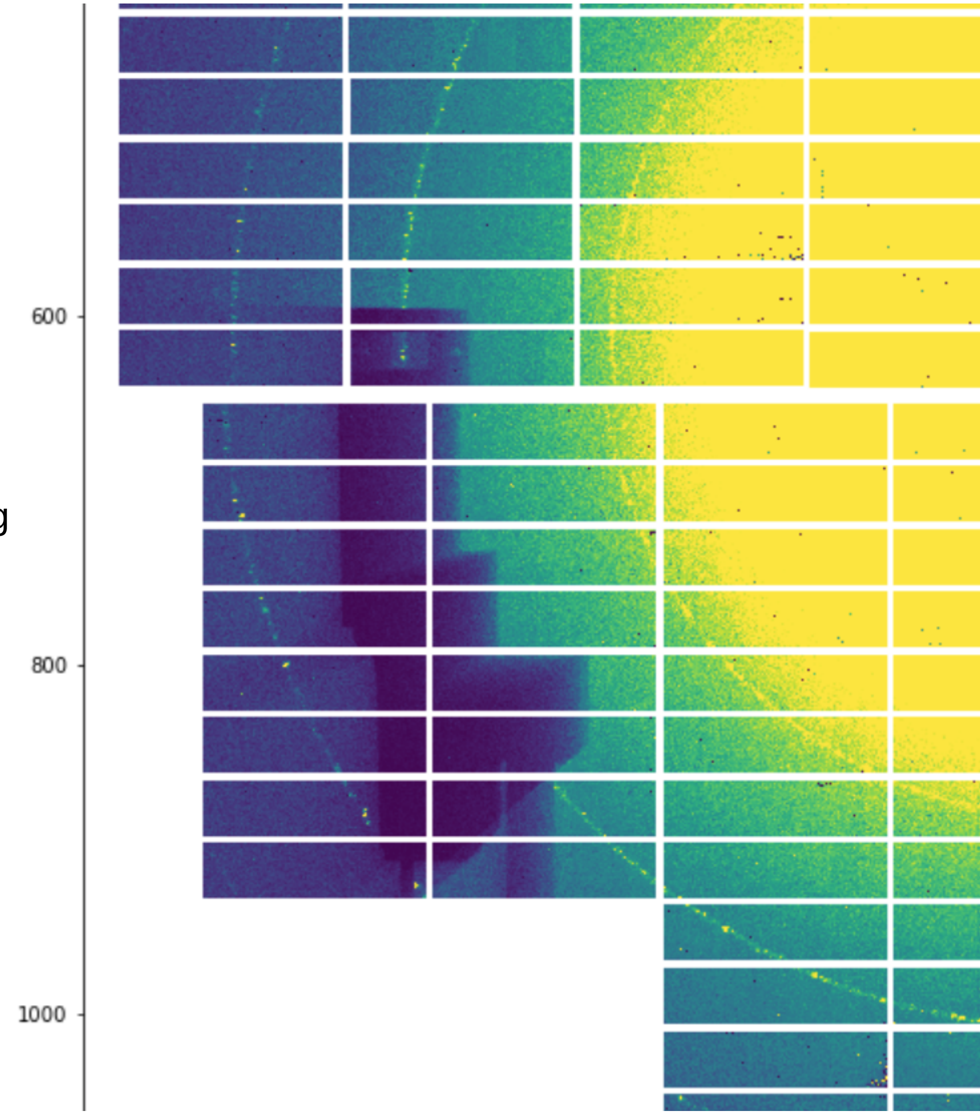
- Users observed unphysical “jump” in radial profiles at gains switching region
  - Offline calibration based on Jupyter notebooks allowed for quick evaluation
  - Found non-linearity which was phenomenologically corrected during shift





Gain non-linearity in LPD

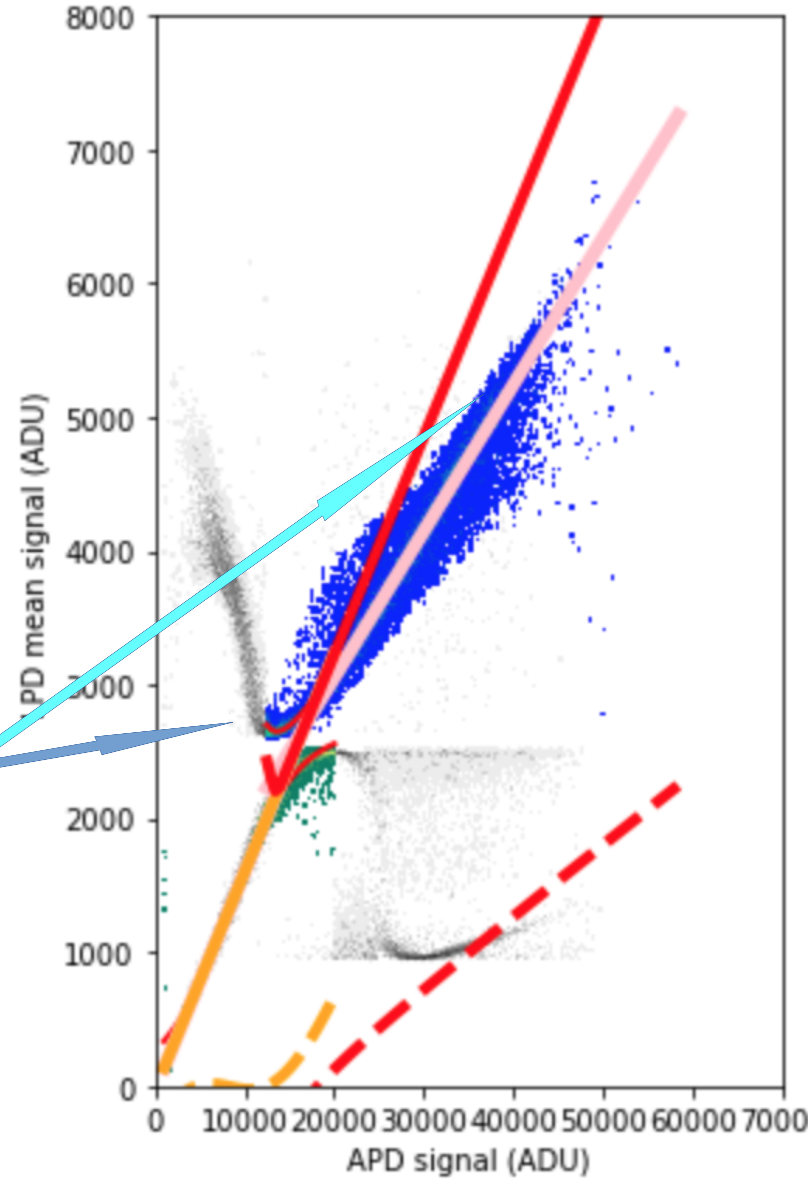
- Users observed unphysical “jump” in radial profiles at gains switching region
  - Offline calibration based on Jupyter notebooks allowed for quick evaluation
  - Found non-linearity which was phenomenologically corrected during shift
- Proper understanding required:
  - I-zero reference detector needed: used APD directly in front of LPD during commissioning beamtime
  - Compare radial average of ring region and different transmissions with APD as relative linear reference



Gain non-linearity in LPD

- Users observed unphysical “jump” in radial profiles at gains switching region
  - Offline calibration based on Jupyter notebooks allowed for quick evaluation
  - Found non-linearity which was phenomenologically corrected during shift
- Proper understanding required:
  - I-zero reference detector needed: used APD directly in front of LPD during commissioning beamtime
  - Compare radial average of ring region and different transmissions with APD as relative linear reference
  - Discontinuity in switching region
  - Different slope of high and medium gain

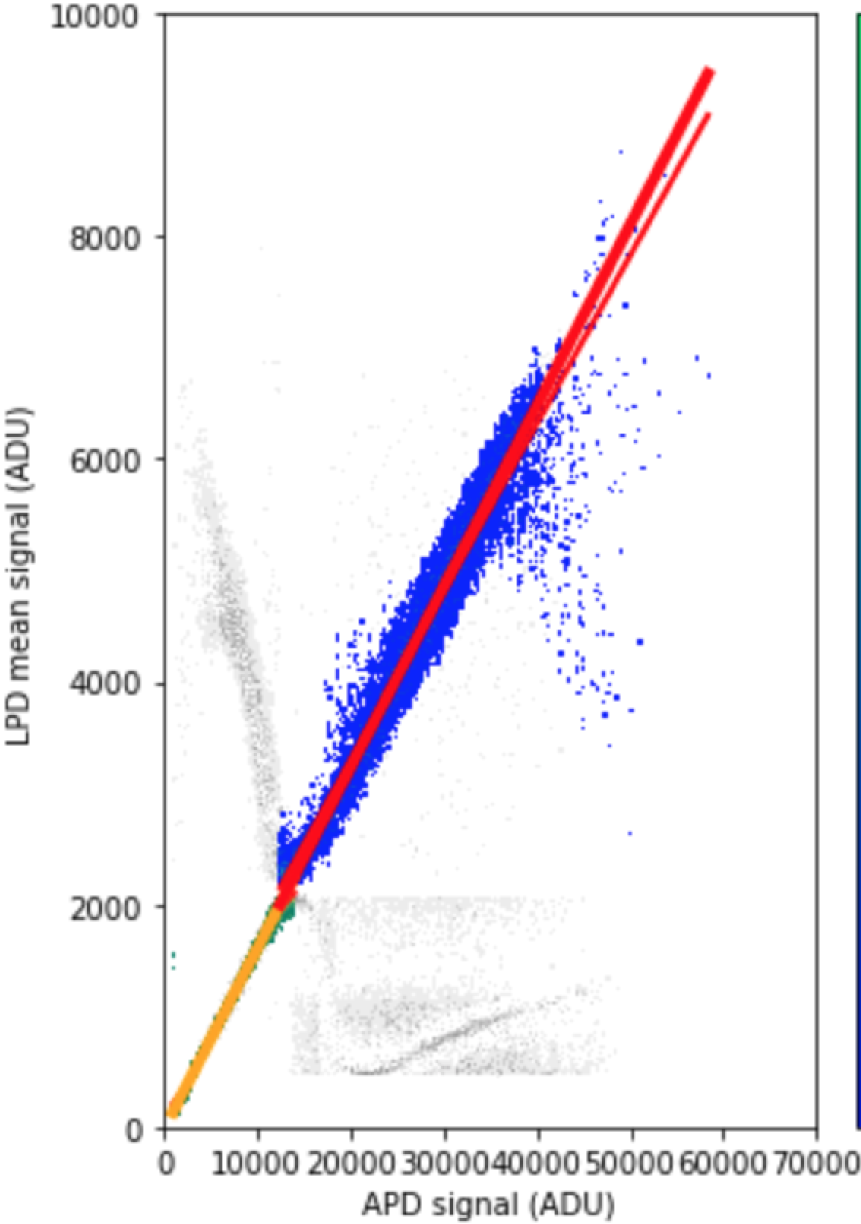
High-medium gain switching region



Gain non-linearity in LPD

- Users observed unphysical “jump” in radial profiles at gains switching region
  - Offline calibration based on Jupyter notebooks allowed for quick evaluation
  - Found non-linearity which was phenomenologically corrected during shift

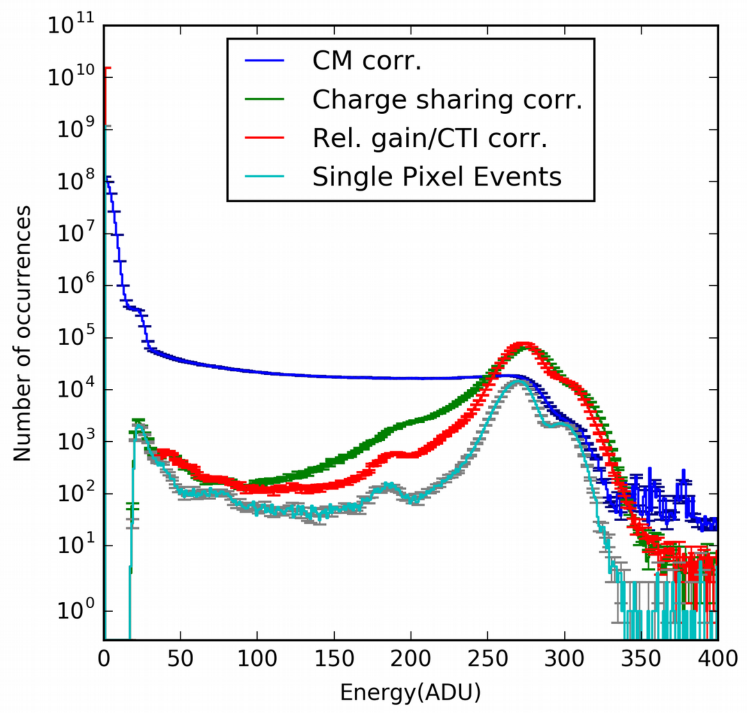
- Proper understanding required:
  - I-zero reference detector needed: used APD directly in front of LPD during commissioning beamtime
  - Compare radial average of ring region and different transmissions with APD as relative linear reference
  - Additional correction functions: reciprocal and sigmoid functions

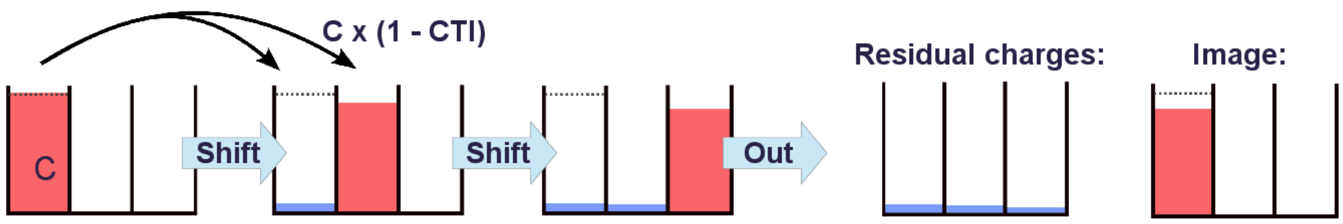


CCDs - Charge transfer inefficiency

- Related to CCD-type detectors
  - Charges are shifted through the semi-conductor sensor during read-out
  - Charge may be trapped due to defects in semi-conductor lattice
- Can be corrected separately from gain correction
  - Effect is order of 10^{-6} – 10^{-4} for the FastCCD

FastCCD, Fe^{55}

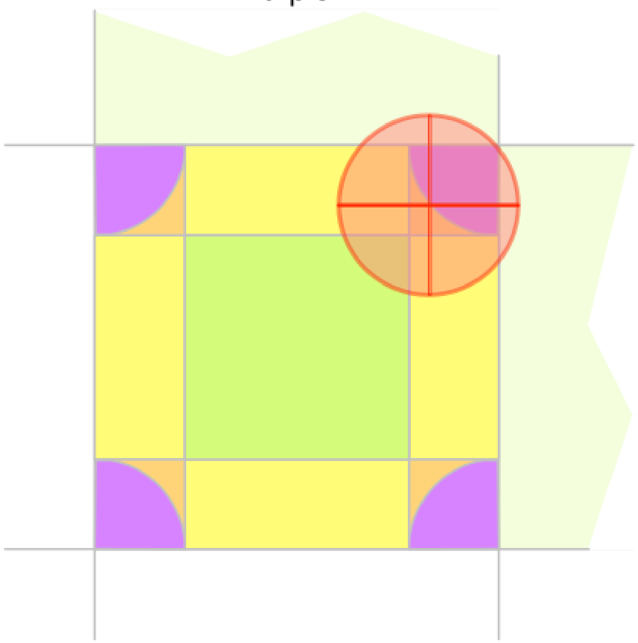




Charge splitting

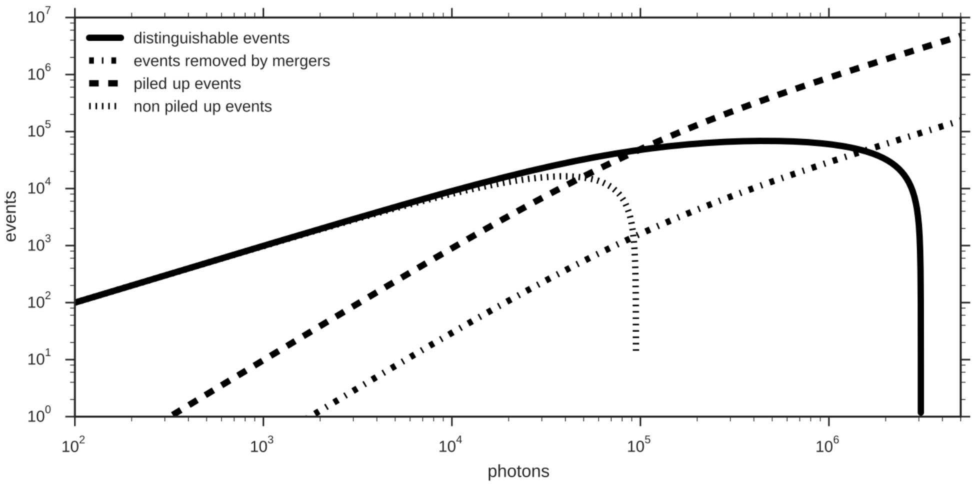
- Charge from a single photon may be present in several neighbor pixels
  - Due to diffusion of the quasi-free electrons within collection time
  - Depends on pixel size, photon energy and bias voltage, temperature
- Can be corrected in case of low photon flux
  - Split events can be identified by patterns of pixels above a noise threshold
  - Sum charge in split event partners to reconstruct original event
  - Not possible for high photon flux due to ambiguity of identification, when pattern or events overlap

triple



$$p_3 = \frac{(4-\pi)\sigma^2}{a^2}$$

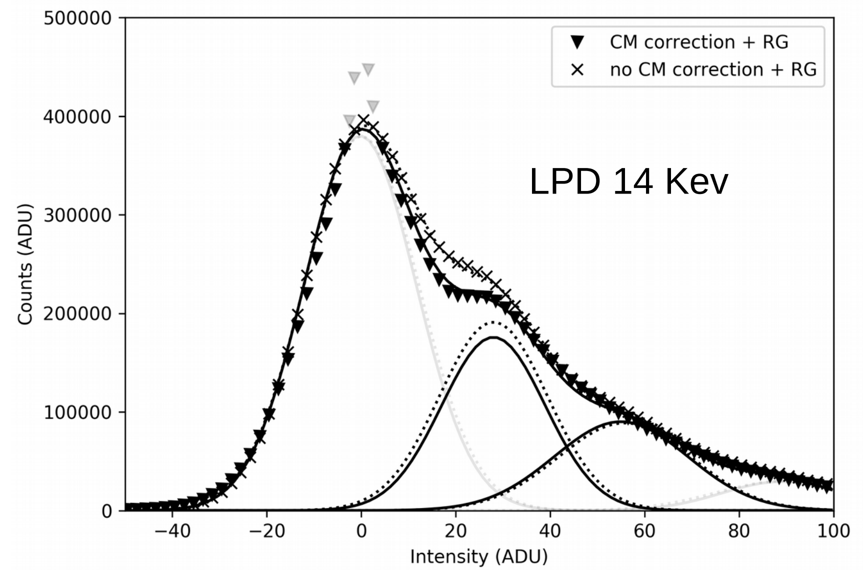
1 Mpixel detector with 200 µm pixel pitch



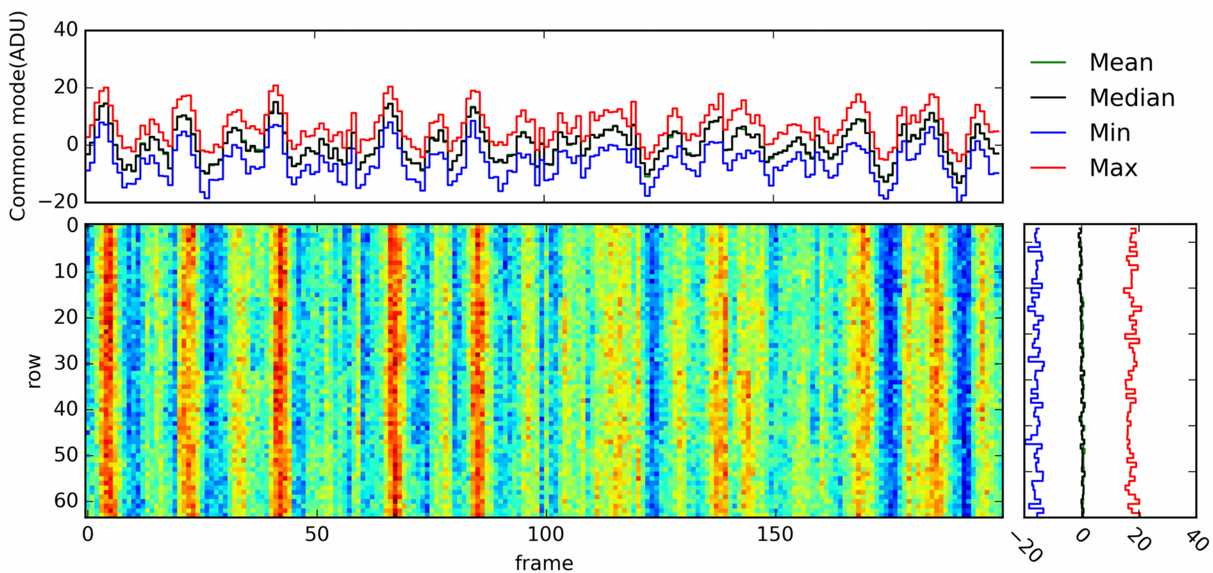
Common mode

- Variation of a signal, which affects group of channels in coherent way
  - Can be caused by common electromagnetic pick-up, noise of the supply electronics, etc
  - Time-dependent effect

- Effect can be corrected for frames with sufficient amount of non-illuminated pixels
  - Shift median to 0 for each row(column)



PnCCD, Common mode over time



Offline Calibration – Data format

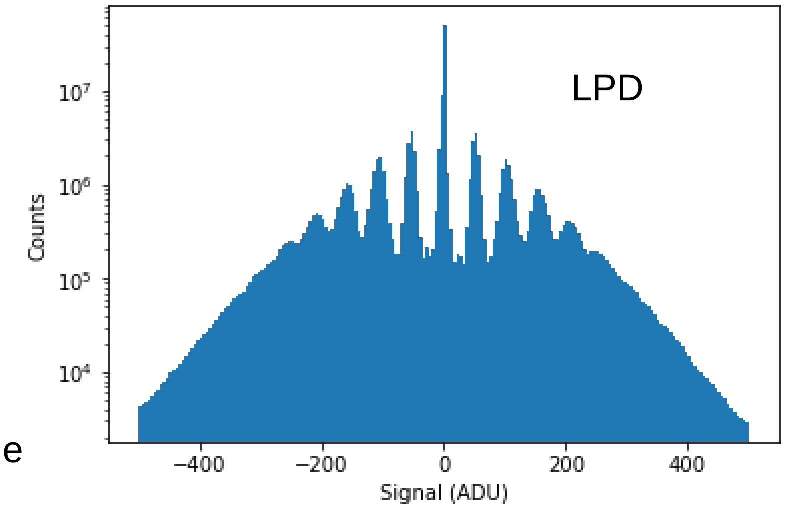
- File and directory structure from RAW files is preserved:
  - ./raw/r0012 → ./proc/r0012
  - RAW-R0039-AGIPD00-S00000.h5 → CORR-R0039-AGIPD00-S00000.h5
  - All data not touched by corrections is simply copied
- New data containers are added:
  - Image.data: the corrected image data as float32 data
  - image.gain: the digitised gain bit setting for each pixel and image where applicable
  - image.mask: the bad pixel value for each pixel and image.
  - image.cellId: memory cell id the image was stored in (if applicable).
  - image.pulseId: pulse id the image refers to.
  - Image.trainId

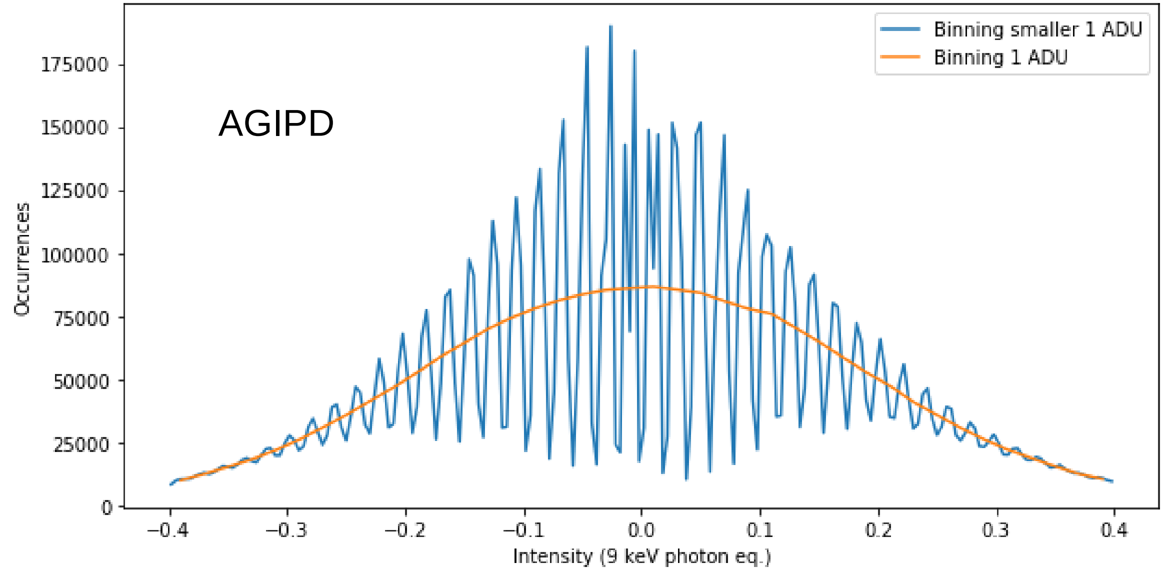
Type of bad pixel	Bit
OFFSET_OUT_OF_THRESHOLD	1
NOISE_OUT_OF_THRESHOLD	2
OFFSET_NOISE_EVAL_ERROR	3
NO_DARK_DATA	4
CI_GAIN_OF_OF_THRESHOLD	5
CI_LINEAR_DEVIATION	6
CI_EVAL_ERROR	7
FF_GAIN_EVAL_ERROR	8
FF_GAIN_DEVIATION	9
FF_NO_ENTRIES	10
CI2_EVAL_ERROR	11
VALUE_IS_NAN	12
VALUE_OUT_OF_RANGE	13
GAIN_THRESHOLDING_ERROR	14
DATA_STD_IS_ZERO	15
ASIC_STD_BELOW_NOISE	16
INTERPOLATED	17

Overbinning artifacts

- Offset correction for AGIPD (Bottom) have a half-integer resolution
  - Calculated as median of a series of a dark frame to reduce effect on outliers without multiple iterations
  - Multiplied to approx. normally distributed gain factor

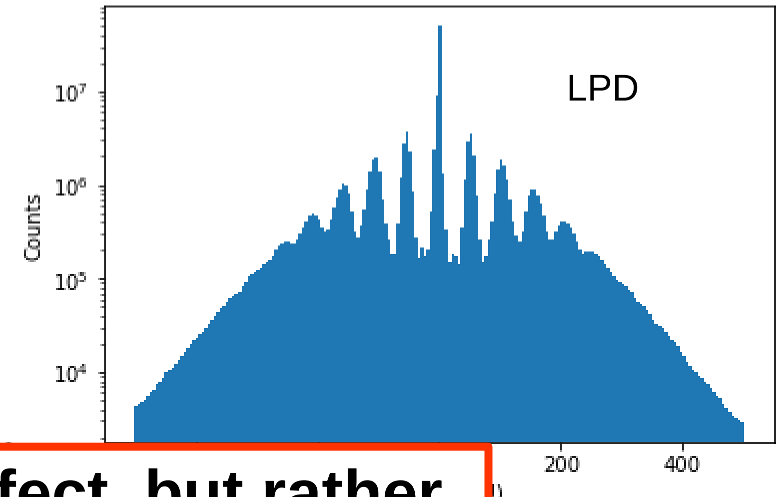
- Large gain multiplication factor for low gain mode of LPD (Right)
  - Discrete deviation of the dark signal is multiplied by the large value of gain factor
  - Effect is negligible at the signal level the low gain will be switched to.



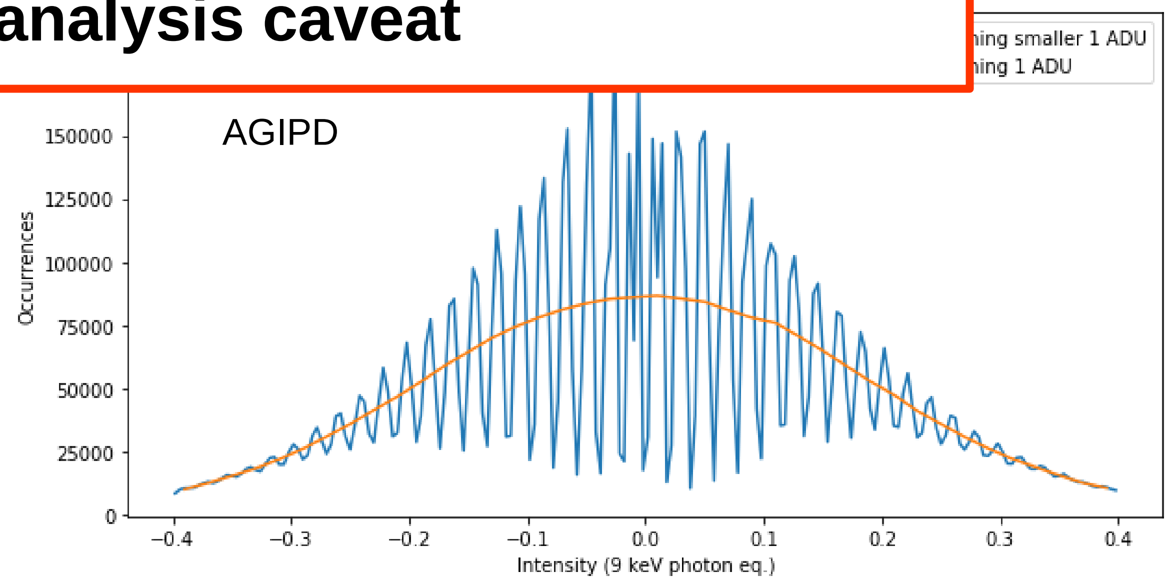


Overbinning artifacts

- Offset correction for AGIPD (Bottom) have a half-integer resolution
  - Calculated as median of a series of a dark frame to reduce effect on outliers without multiple iterations
  - Multiplied to approx. normally distributed gain factor
- Large gain multiplication factor for low gain mode of LPD (Right)
  - Discrete deviation
    large value of gain
  - Effect is negligible
    switched to.



Not a calibration effect, but rather an analysis caveat



Detector Performance over Time

- Initial version of standardized process to request tools for historic detector data evaluation is being use at the detector group

Template for request of analysis

Which detectors should be handled?:

☐ AGIPD, ☐ LPD, ☐ FastCCD, ☐ pnCCD, ☐ DSSC

Calibration constants to be evaluated:

☐ Noise ☐ Offset ☐ Gain ☐ Bad pixels

Request one or several analyses by filling functions in a table and specifying way of presentation:

Suggested functions: mean, std, median, min, max, sum

Time evaluation:

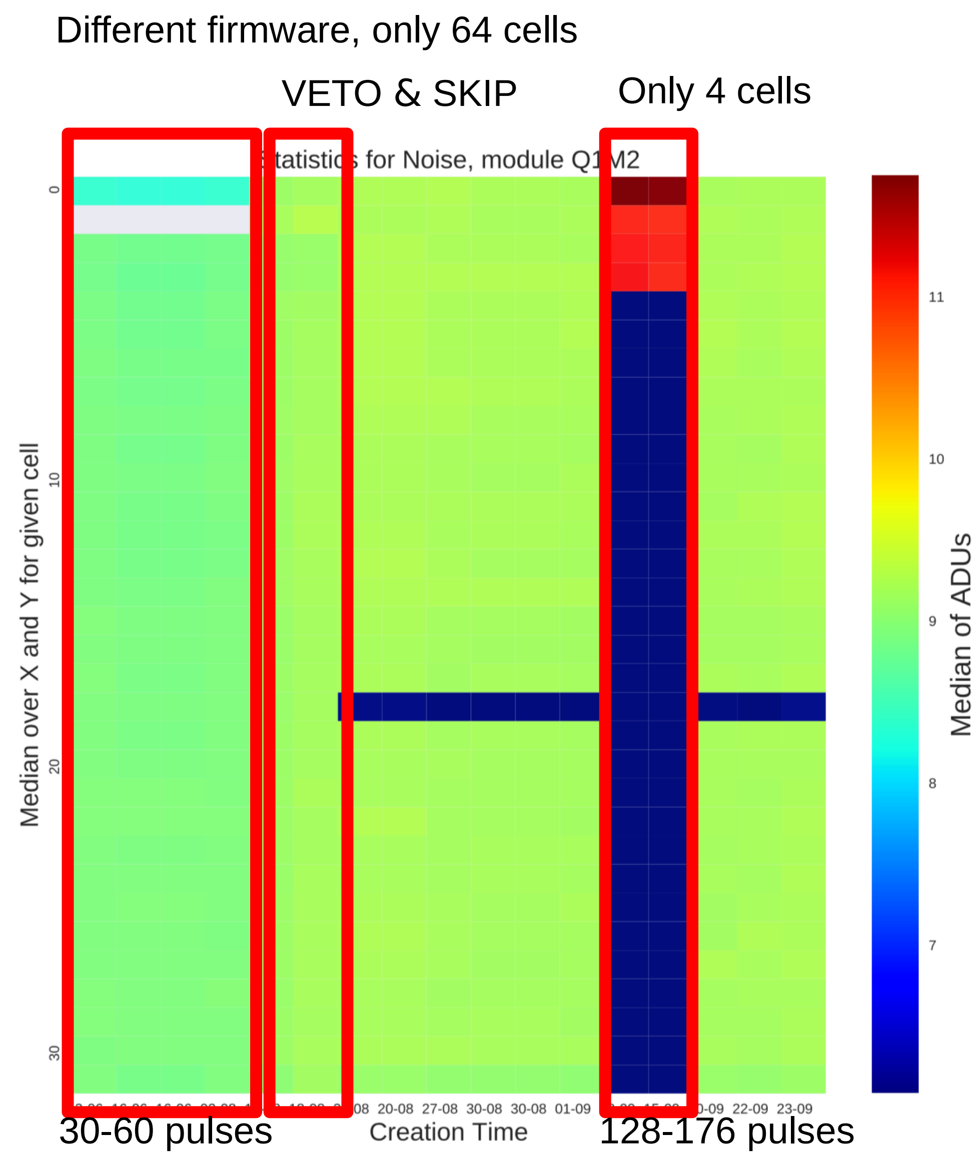
X	Y	Cell	Gain	Module

AGIPD Noise development

	Q1M4			Q2M1			Q3M4			Q4M1		
	high	med	low	high	med	low	high	med	low	high	med	low
2018-10-11	12.02	12.83	13.10	10.84	12.16	12.97	11.60	12.65	13.63	11.55	12.64	13.68
2018-12-06	9.64	10.40	10.66	8.35	9.10	9.30	9.58	10.47	10.68	9.45	10.40	10.59
2018-08-02	9.42	10.93	11.28	8.81	10.67	10.91	10.39	12.84	23.70	9.13	10.83	11.11
2018-08-16	9.44	10.86	11.11	8.98	10.64	10.81	9.31	10.76	10.97	9.22	10.75	10.94
2018-12-06	9.06	10.95	11.14	8.54	10.59	10.76	9.00	11.00	11.19	8.97	10.95	11.14

Detector Performance over Time

- Detector operating conditions and configuration can have significant impact on calibration constants
- New configurations may require re-characterization of the detector, even if they seem trivial → ask a DET expert
- Detector experts can help to chose detector configuration for optimal performance



Online Calibration

Online calibration:

- The (calibrated) streaming data can be used for near real-time analysis, providing, for example, the latest set of detector image data in the Karabo graphical user interface.
- The same data can also be sent to the first stage of further data analysis tools that give rapid feedback on whether images are interesting for scientific analysis.
- Current implementation allows to run fast algorithms on GPU cores significantly improving the performance.
- The Karabo bridge allows external data processing pipelines to connect to Karabo for real time data processing without being implemented directly within Karabo.

Gain/offset data correction.

Gain.

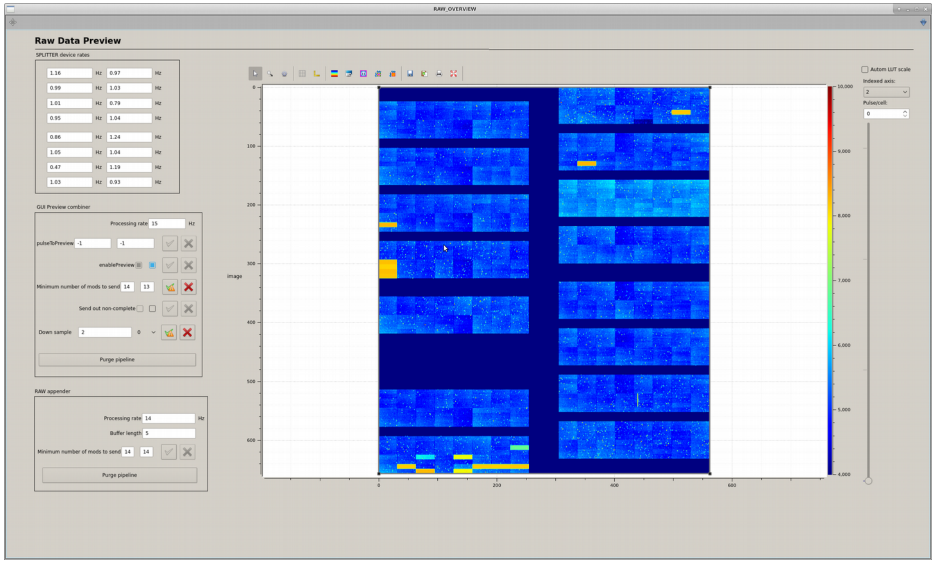
$$G_{i,j} = G_{rel.(i,j)} \cdot C_{ffield(i,j)}$$

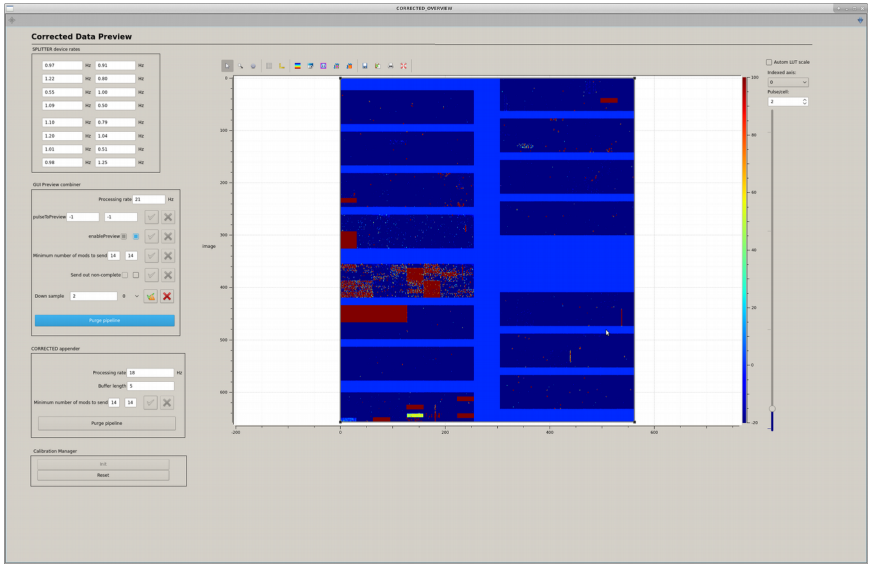
Offset (dark signal).

$$O_{dark} = C_{offset} + \langle I_{therm} \rangle$$

$$I'_{i,j} = G_{i,j} \left(I_{i,j} - O_{dark(i,j)} \right)$$

AGIPD. Raw and corrected data online preview (dark data).





Online Calibration.
Detector specific calibration.

Detector	Correction	Constants generation.

AGIPD	Offset and gain correction.	• Dark image characterization:
		  • Offset map.
		  • Gain map (from charge injection).
		• Flat field characterization.
		• Gain thresholds.

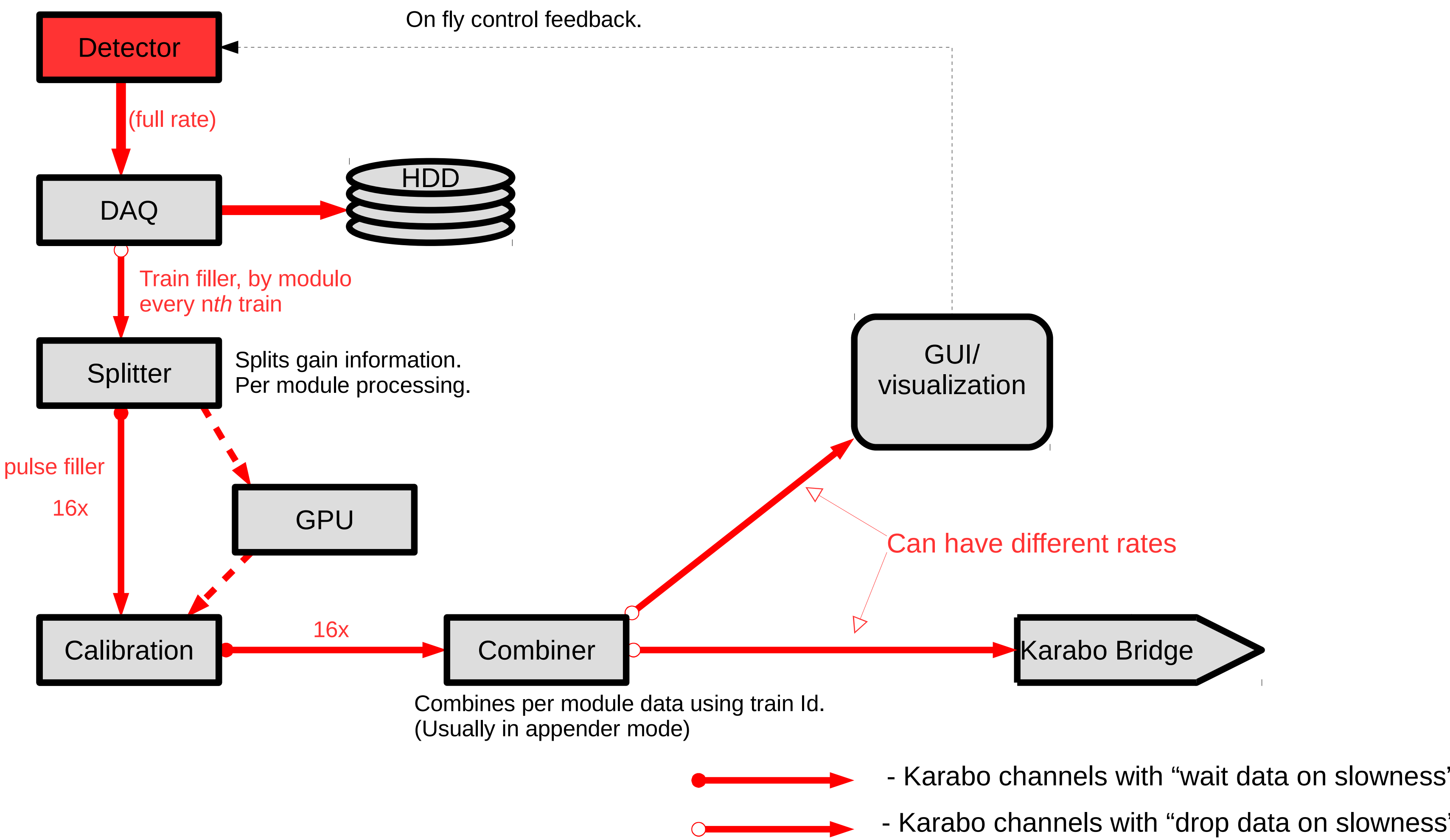
LPD	Offset and gain correction.	• Dark image characterization:
		  • Offset map.
		  • Gain map (from charge injection).
		• Flat field characterization.

FastCCD	Offset, gain, temperature drift and split charge* correction.	• Dark image characterization:
		  • Offset map.
		  • Noise map.
		• Flat field characterization.
		• Temperature drift is compensated.

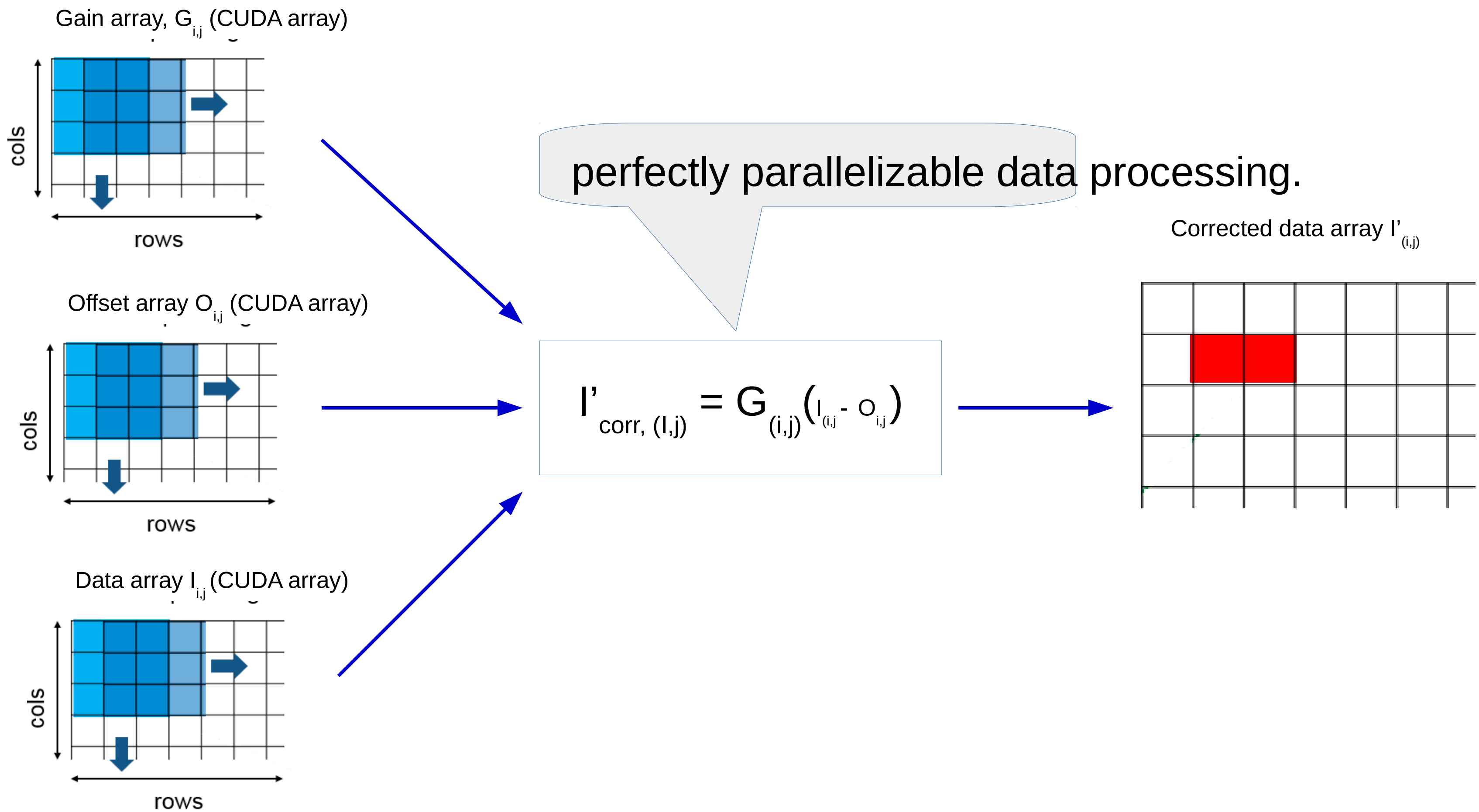
JUNGFRAU	Offset, gain and split charge* correction.	• Dark image characterization:
		  • Offset map.
		  • Noise map.
		• Flat field characterization.

* - not currently implemented in production use

Online Calibration.
Data flow pipeline.



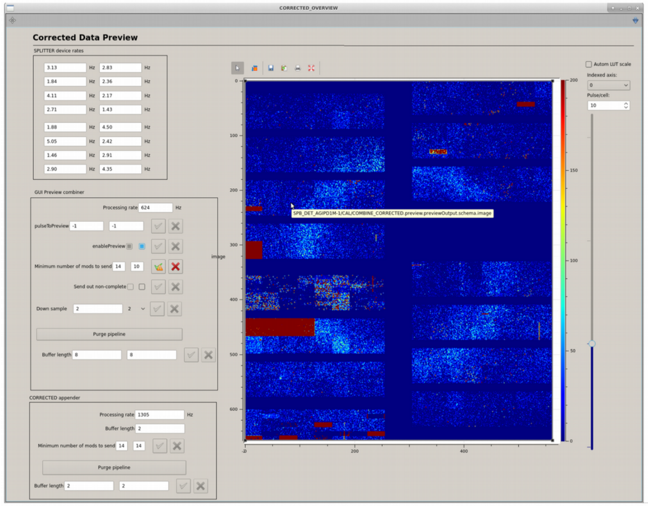
Online Calibration.
Offset/Gain correction on GPU.



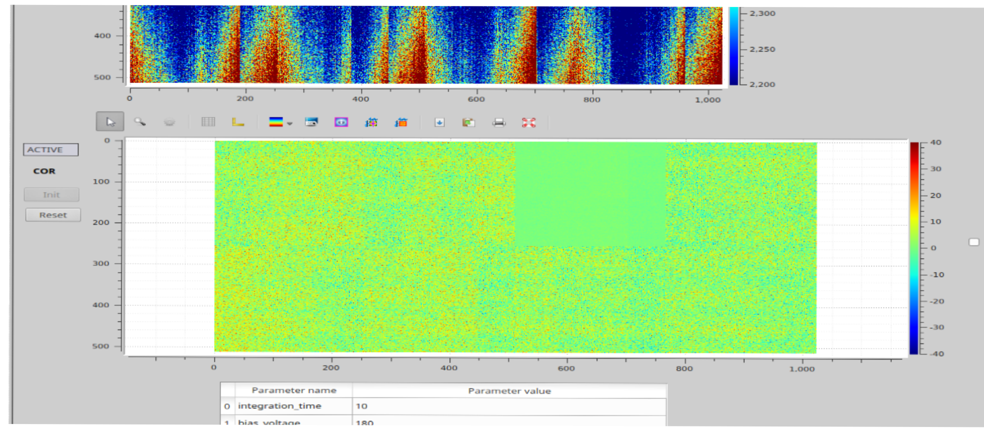
Online Calibration.

GUI.

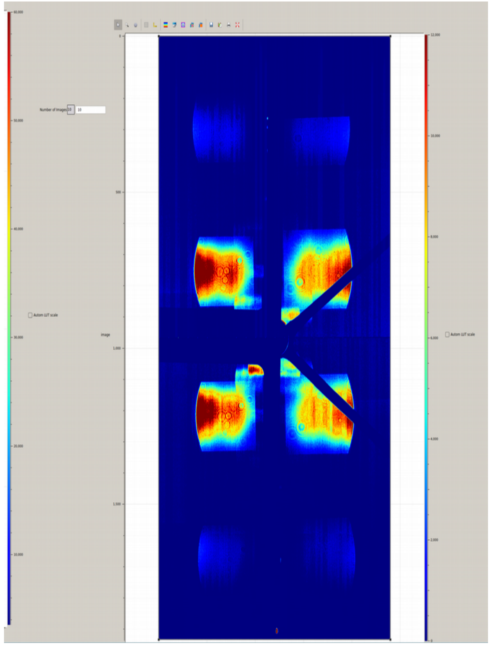
AGIPD 1M



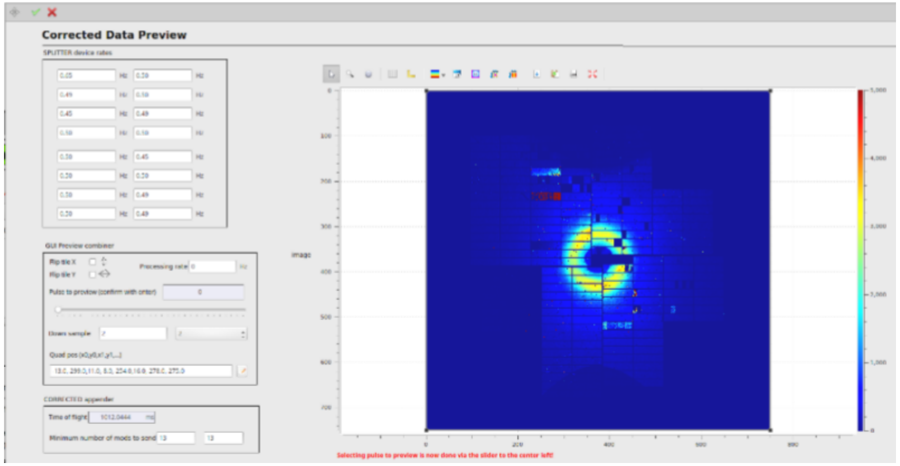
JUNGFRAU



FastCCD



LPD 1M



Online Calibration.
Performance.

- MHz rate detectors produce up to 5080 Megapixel images per second, and DSSC will produce up to 8000. This requires the data acquisition (DAQ) system to digest rates between 10−16 Gigabytes/sec.
- In online processing data rates of up to 128 corrected Megapixel images/sec, with a latency to user processing of below 2 sec are routinely achieved, and 256 images have been stably processed at 1Hz, and 3-5 Hz rate at 128 images.
- Connects to CAS-provided Karabo bridge.

Applications so far:

- ONDA.
- CASS.
- Hummingbird.
- karaboFAI.

Online Calibration.

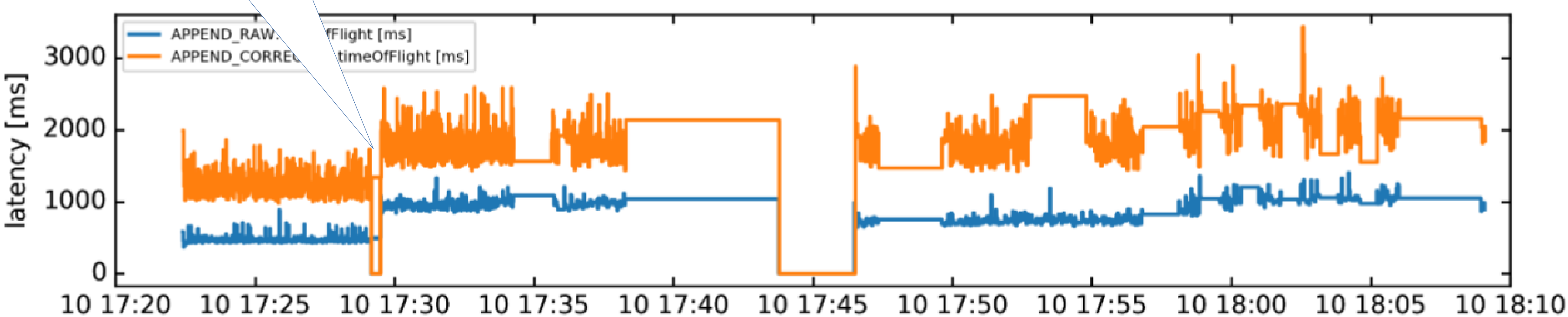
Perfomance (latency).

The latency is measured from acquisition of a train on the detector to the output of the Karabo bridge and includes:

- Data acquisition and formatting on DAQ
- Data formatting to pipelines
- Data selection at pipelines entry points:
  - Every n*th* train
  - Cells containing FEL pulses
- Offset and gain corrections
- train-matched combination of the 16 independent data streams for the detector modules into a single array
- Data advertising on Karabo Bridge

Reconfiguration (number of images)

Latency for providing raw (blue) and corrected (orange) LPD Megapixel images, measured as the time from acquisition on the detector to output to the Karabo bridge. At 256 image/s the total data rate processed by the pipeline is 1.792 GB/s.



Summary & Outlook

- Detector calibration and characterization tools are established in common framework
  - User operation and good performance have been demonstrated
  - Common tools and approaches across detectors
  - Continuous feedback from instrument groups
  - Evaluation of correction factors on the best knowledge of detector experts
    - Non-linear effects of correction factors
    - Studies of time-evolution

- Future plans
  - Interface to meta-data catalog
  - Make (re-) calibration more user friendly

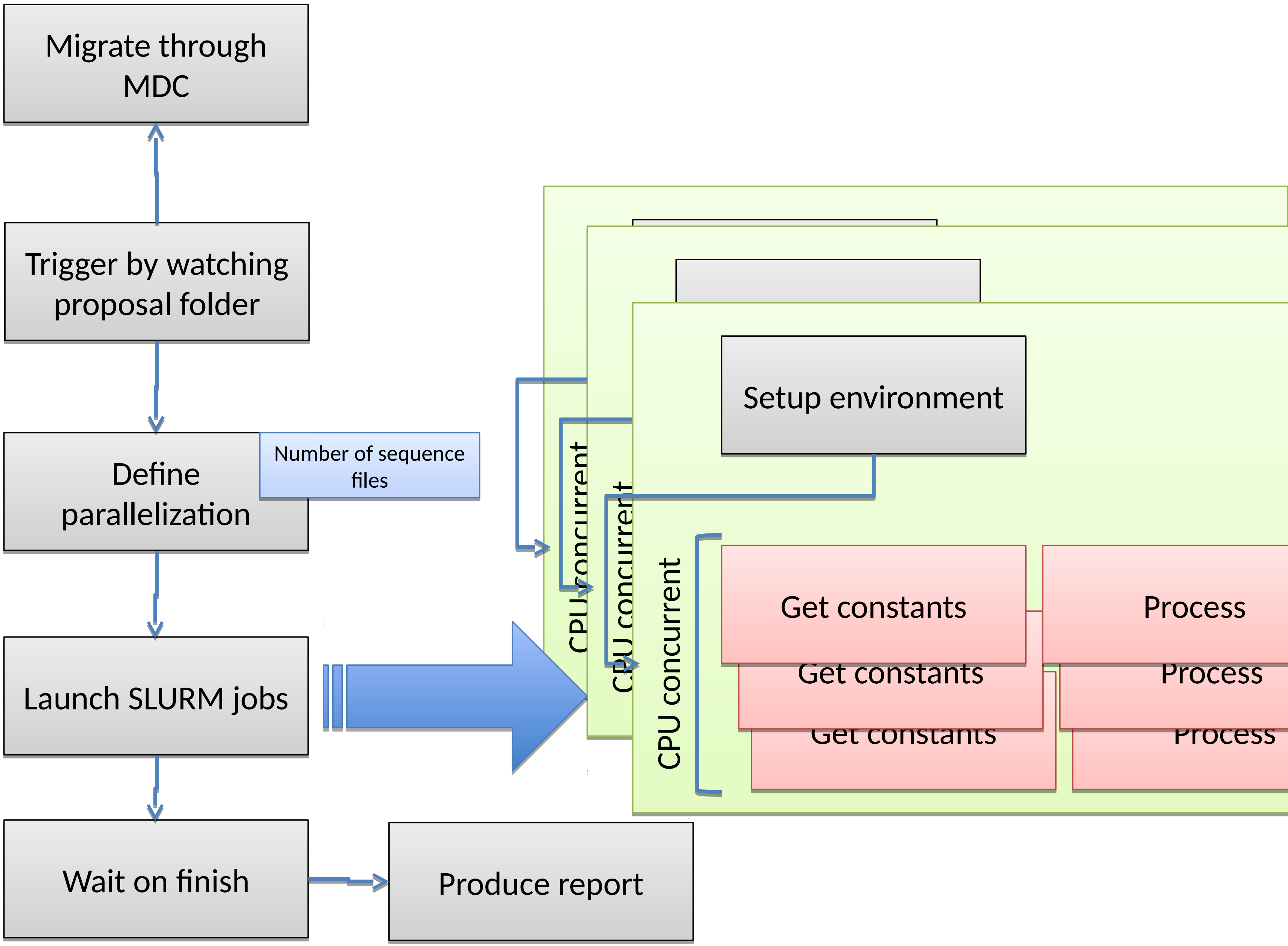
Documentation:

- **Detector documentation**: https://in.xfel.eu/readthedocs/docs/detector-documentation/en/latest/
- **Calibration:** https://in.xfel.eu/readthedocs/docs/detector-documentation/en/latest/calibration/detector_specific.html
- **Interface to calibration DB:** https://in.xfel.eu/readthedocs/docs/icalibrationdb/en/latest/
- **Libraries for detector related analysis:** https://in.xfel.eu/readthedocs/docs/pydetlib/en/latest/
- **Offline calibration:** https://in.xfel.eu/readthedocs/docs/european-xfel-offline-calibration/en/latest/
- **Online calibration:** https://in.xfel.eu/readthedocs/docs/online-calibration-tools/en/latest/

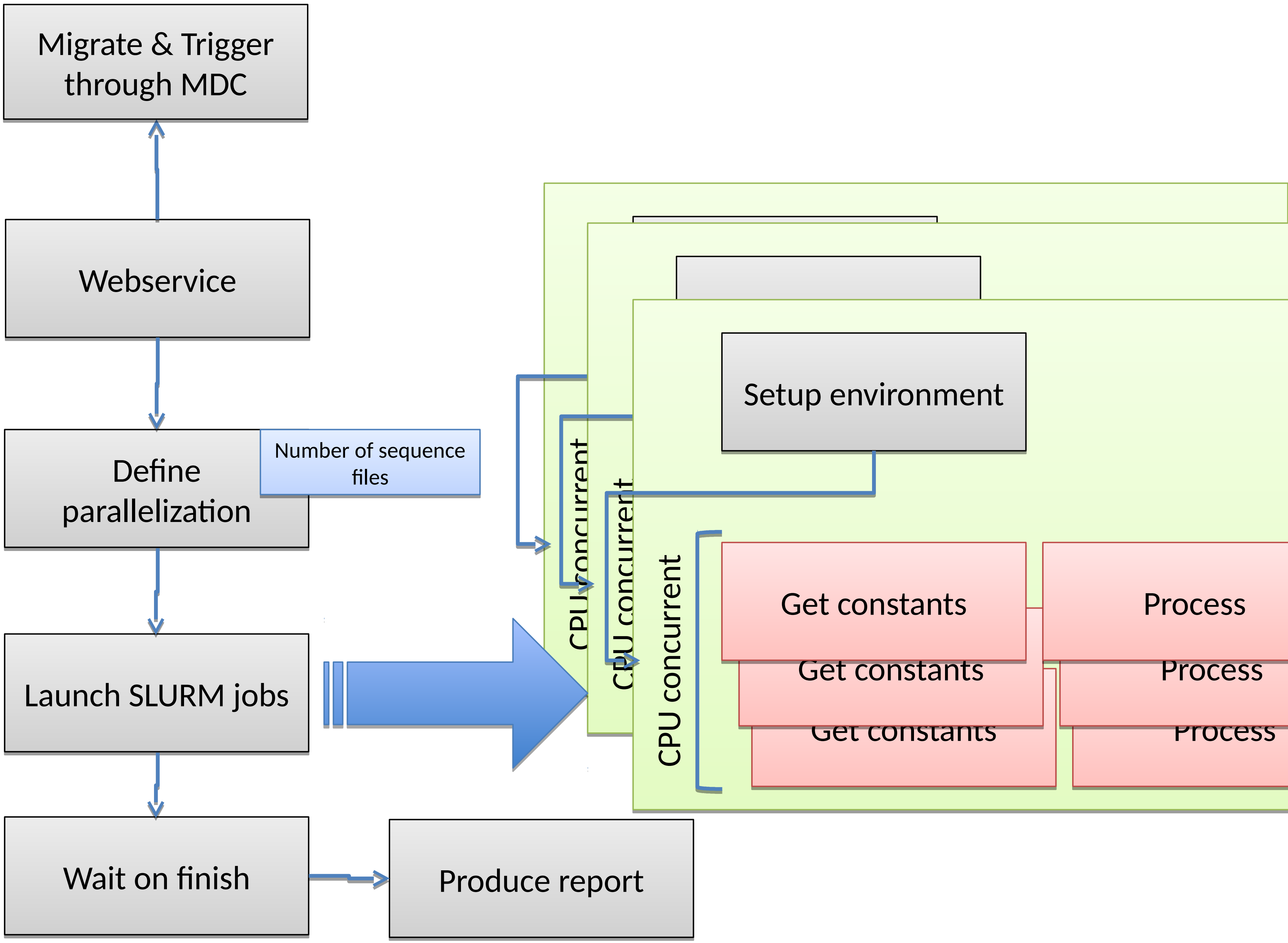
  https://in.xfel.eu/readthedocs/docs/online-calibration-base-classes/en/latest/

  https://in.xfel.eu/readthedocs/docs/online-pipeline-splitter-and-combiner-devices/en/latest/

Backup Slides



Migrate through MDC
Trigger by watching proposal folder
Define parallelization
Number of sequence files
Launch SLURM jobs
Wait on finish
Produce report
CPU concurrent
CPU concurrent
CPU concurrent
Setup environment
Get constants
Get constants
Get constants
Process
Process
Process


Migrate & Trigger through MDC
Webservice
Define parallelization
Number of sequence files
Launch SLURM jobs
Wait on finish
Produce report
Setup environment
CPU concurrent
CPU concurrent
CPU concurrent
Get constants
Get constants
Get constants
Process
Process
Process

DSSC per module automated trimming pipeline.