

# REWRITING THE DCT PRE-PROCESSING PIPELINE

## COPIL BIGMECA AUGUST 2021

João P C Bertoldo

Materials Center @ MINES Paristech - PSL University  
ID11 @ The European Synchrotron Radiation Facility (ESRF)



26 August 2021

# DIFFRACTION CONTRAST TOMOGRAPHY (DCT)

## A SHORT INTRODUCTION

### What is it?

- Acquire 2D images from different angles
- Acquire transmitted and diffracted beam  
(notice: the acquisition area is larger than the beam)
- Identify and group diffraction spots by grain
- Reconstruct the grains individually

### How is it different than XCT?

- Reveals 3D grains orientation and shape
- Allows to build digital models of the sample
- Enables new experiments; e.g.: visualize slip bands

Figure 1: DCT setup (simulation). Credits: Wolfgang Ludwig.

↑ [click](#)

# DIFFRACTION CONTRAST TOMOGRAPHY (DCT)

## A SHORT INTRODUCTION

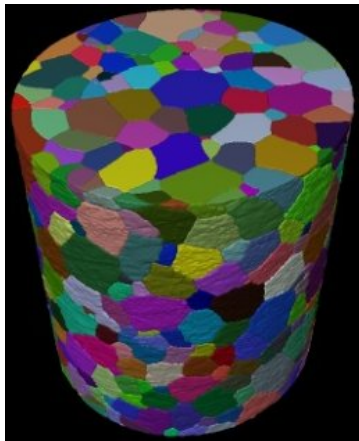


Figure 2: Reconstructed volume with individual grains segmented (different colors). Credits: Wolfgang Ludwig.

### What is it?

- Acquire 2D images from different angles
- Acquire transmitted and diffracted beam (notice: the acquisition area is larger than the beam)
- Identify and group diffraction spots by grain
- Reconstruct the grains individually

### How is it different than XCT?

- Reveals 3D grains orientation and shape
- Allows to build digital models of the sample
- Enables new experiments; e.g.: visualize slip bands

# PIPELINE: FROM ACQUISITION TO 3D RECONSTRUCTION

## DCT reconstruction pipeline

1 dataset = 3k ~ 14k images of 2048<sup>2</sup> pixels = 56GB ~ 252GB (float32)

3k ~ 100k 3D-blobs

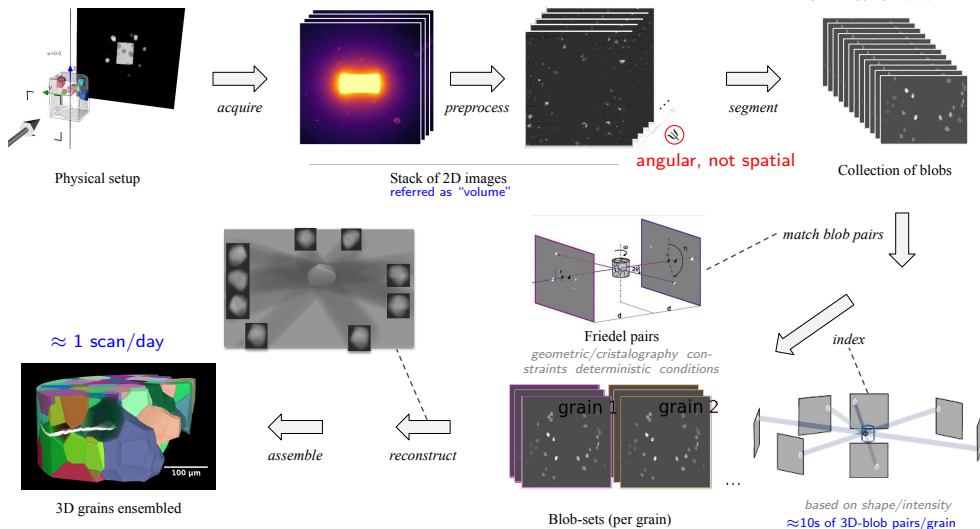
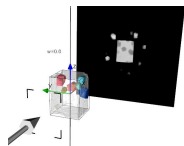


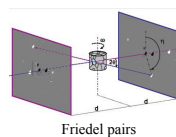
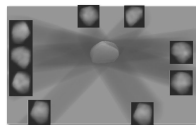
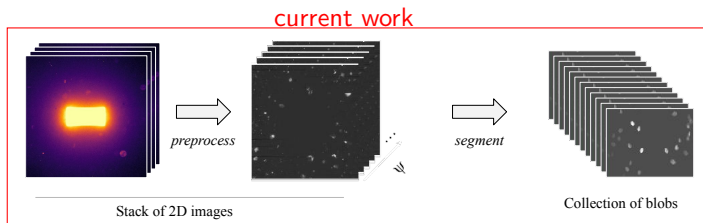
Figure 3: Main steps in DCT's reconstruction pipeline

# PIPELINE: FROM ACQUISITION TO 3D RECONSTRUCTION

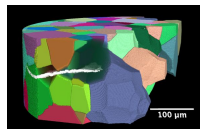
## DCT reconstruction pipeline



acquire



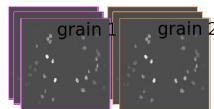
match blob pairs



3D grains ensemble

assemble

reconstruct



Blob-sets (per grain)

index

Figure 3: Main steps in DCT's reconstruction pipeline

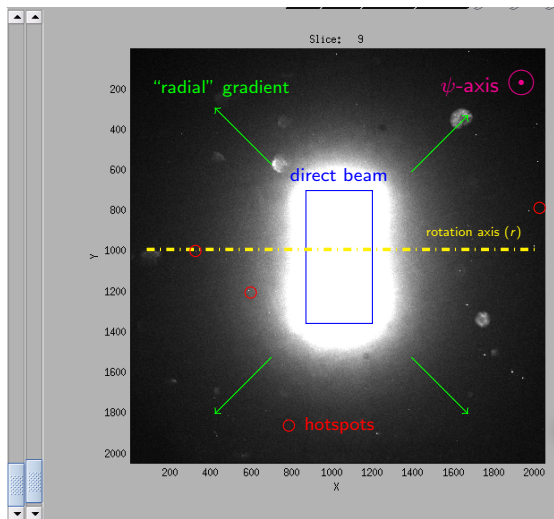


Figure 4: Raw acquisition

## Image characteristics

- 4 sensors grid
- source intensity oscillates  
i.e. beam intensity variation over  $\psi$
- beam/blobs brightness 2 scales apart
- blobs closer to  $r$  are  $\psi$ -longer

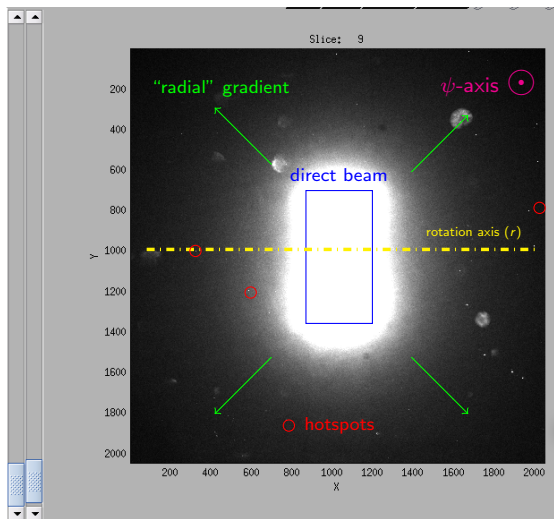


Figure 4: Raw acquisition

## Image characteristics

- 4 sensors grid
- source intensity oscillates  
i.e. beam intensity variation over  $\psi$
- beam/blobs brightness 2 scales apart
- blobs closer to  $r$  are  $\psi$ -longer

## Preprocessing steps

- subtract the offset
- (optional) normalize by the margin's average
- subtract a per-pixel  $\psi$ -wise moving median  
i.e. 1D moving median on  $2048^2$  positions
- 2D median filter (per frame)

# PREPROCESSED IMAGE

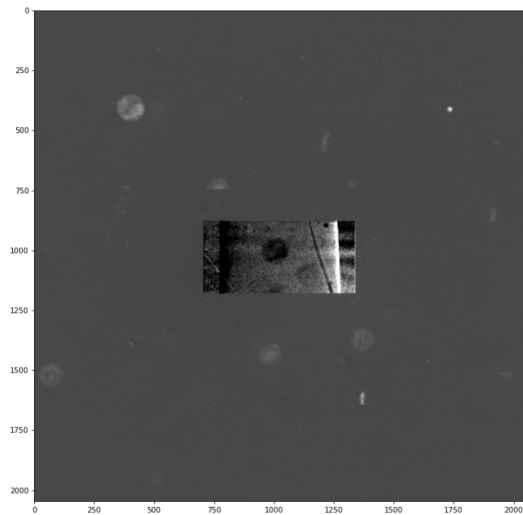


Figure 5: Preprocessed image



# SEGMENTATION

## DOUBLE THRESHOLD IN 3D

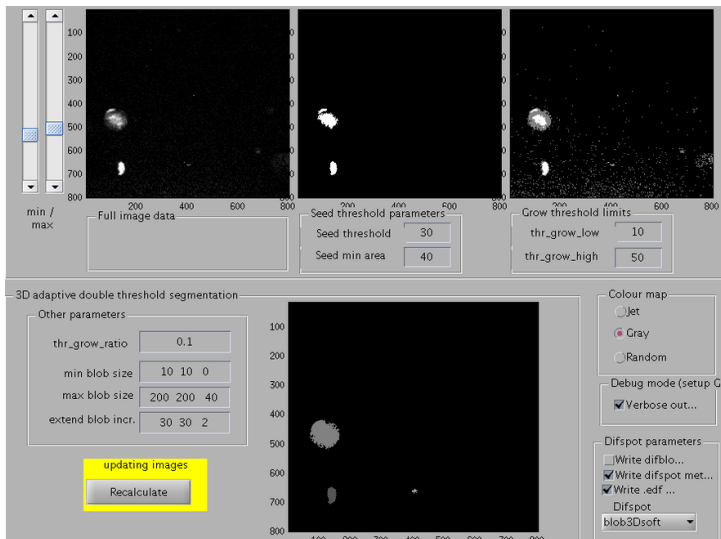


Figure 6: Double threshold interface

- threshold the image
  - find all connected regions
- for each connected region
- find the local maximum value  $M$
  - the local threshold is  $\tau M$   
 $\tau \in [0, 1]$  is the tolerance parameter
  - from  $M$ 's location, get a  $BB^a$  of size  $2 \times bb_{max}$   
 $bb_{max} \in \mathbb{R}^3$  is the maximum BB parameter
  - threshold the BB with  $\tau M$  and get its connected component  $c$
  - keep it if  $bb_{min} \leq c \leq bb_{max}$

---

<sup>a</sup>bounding box

# METAMORPHOSIS

BEFORE

```
Activities Terminal 25 août 17:05 ssh -X rnice8
joaopcbertoldo@joaopcbertoldo-pc:~/projects/mytex/projects/dct-slides/src ssh -X rnice8

gtDefFwdProjGvdn2NW          gtGeoP1LabFromThetaEta      gtMathsGradient              gtTaperUpdateGrains
gtDefFwdProjGvdn2UVP         gtGeoPointProjectionsOnDetector gtMathsIsPointInPolyhedron   gtTest_gt60UpdateDualDetector_c
gtDefFwdProjGvdn2UVM        gtGeoProjectedSampleEnvelope gtMathsLaplacian             gtTest_gt60UpdateDualL1_c
gtDefFwdProjGvdn2W          gtGeoProjForReconstruction  gtMathsLinePairsIntersections gtTest_gt60UpdatePrimal_c
gtDefGvdn2Dvovl             gtGeoRecDefaultParameters    gtMathsLinePlanesIntersections gtTestExceptionHandling
gtDefModifyDevot            gtGeoSam2Lab                  gtMathsLinePolyhedronIntersections gtThinCrack
gtDefShapeFunctionsCreateNW  gtGeoSam2Saa                  gtMathsLinesDists            gtThinCrack2
gtDefShapeFunctionsCreateW   gtGeoSamDefaultParameters    gtMathsLinesIntersection      gtTIFInfoReader
gtDefShapeFunctionsFwdProj   gtGeoSamEnvFromAcq           gtMathsMatrixProduct         gtTIFVolReader
gtDefShapeFunctionsNW2N      gtGeoSamEnvInSampleRef       gtMathsMirrorPointsOnAxis     gtTIFVolReaderWithInfo
gtDefShapeFunctionsNW2UV     gtGeoScattVecFromDirFFVec    gtMathsMisorientation         gtTIFVolWriter
gtDefShapeFunctionsNW2W      gtGeoSin                       gtMathsNonNegative           gtTriggerUpdateCallback
gtDefShapeFunctionsNW2VM     gtGeoThetaFromDirFFVec       gtMathsNorm_l1                gtTwinCalculateVariants
gtDefSyntheticGrainCreate    gtGeoTwoThetaLimits          gtMathsNorm_l2                gtTwinOrientations
gtDetectorTwoTheta          gtGetAllCellValues           gtMathsNormalizeVectorsList   gtUpdateSpotGrain
gtDilateGrains              gtGetAMSIColour              gtMathsOriMat2AngleAxis      gtUpdateBinFile
gtDisorientation            gtGetDirFImage               gtMathsOriMat2Euler           gtVector2Matrix
gtDrawArrow3                gtGetExtImage                gtMathsOriMat2Rod            gtVectorCryst2Lab
gtDrawAxis3d                gtGetFullImage               gtMathsOutLiers              gtVectorLab2Cryst
gtDrawGrainInInitCells      gtGetFunctionDepes           gtMathsPointLineDists         gtVectorOrientationColor
gtDrawPoly3D                gtGetGeometry                gtMathsProjectPointsOnPlane   gtVeri_FyDFHeader
gtDrawSampleGeometry        gtGetHST_FileNamesAndType    gtMathsRod2Euler             gtVersion
gtDriftsEndOfScan           gtGetHTMLColour              gtMathsRod2OriMat            gtVGStudioLUT
gtEBSDLoadMapEulerCSVFile   gtGetID10Depes               gtMathsRod2RodInFundZone     gtVolRead
gtEBSDLoadMapEulerCTFFile   gtGetLastDirName             gtMathsRodSum                 gtVolReadSingle
gtEBSDLoadMapEulerTSVFile   gtGetMaxDisplacement          gtMathsRotationMatrixComp     gtVolumeGetSlice
gtEBSDLoadMapIndexingTSVFile gtGetMaxExtImage              gtMathsRotationTensor         gtVolumeMaskBorder
gtError                      gtGetMeanExtImage            gtMathsRotationTensorComposite gtVTXConvertTypeMatlab2VTK
gtESF2LSF                   gtGetMeanFullImage           gtMathsRotationTranslationTensor gtVTXConvertTypeVTK2Matlab
gtEstimateErrors            gtGetOOFFromDevot            gtMathsSegmentConstrainedLevelset gtVTXMeshReader
gtEvalSignalAutocorrelation gtGetOOFFromDevot            gtMathsSIAFiveCell           gtVTXMeshWriter
gtEvaluateCalibrationQuali  gtGetParentFigure            gtMathsSquareNorm            gtVTXStructGridWriter
gtEvaluateCalibrationScan    gtGetRawROI                  gtMathsSumCellVolumes        gtWriteBlobToHDF5
gtText                       gtGetReflections             gtMathsSumNDVol              gtWriteDirStackASTRA_roi
gtExternalCompileFunctions  gtGetSignalAutocorrelation    gtMathsUniqueTol             gtXMLLookupTableWriter
gtExtract                   gtGetSignalEnergy            gtMathsSuppplVolume          gtZsetCup
gtFedActiveVolumeVoxels     gtGetSignalNoiseRatio        gtMathsVectorsAngles         gtZsetMeshReader
gtFedApplyField4            gtGetSignalPower             gtMathsVectorsAnglesRad     gtZsetMeshWriter
gtFedApplyFieldBalance      gtGetSignalSpectralDensityPower gtMatrix2Vector
gtFedApplyFieldFlow         gtGetSinglesValues           gtMATVolReader
gtFedApplyFieldFlowPolyChr
>> gtPr
gtPrepDefaultParameters     gtPreprocessing              gtPrintException
>> gtPr
gtPrepDefaultParameters     gtPreprocessing              gtPrintException
>> gtPr
gtPreprocessing
Recompile functions for OAR? Not normally needed! [y/n]: [n] n
Do you want to reset all the parameters for preprocessing? [y/n]: [n] n
Do you want to verify/change the (default) preprocessing parameters? [y/n]: [n] n
Do you want to redefine the direct beam BoundingBox (currently 1709 881 632 304)? [y/n]: [n] n
Do you want to (re)-calculate the Rotation Axis position (currently 1022,8) [y/n]: [n] n
Do you want to (re)-define the Sample BoundingBox (currently 1732 881 583 304)? [y/n]: [n] n
Do you want to (re)-define the Active Area of the detector? [y/n]: [n] n
```



# METAMORPHOSIS

## BEFORE

- MATLAB
  - paid
  - window-based application (ssh -X...)
  - pipeline “spread” over several files/functions
  - user types commands
  - parameters in .m file
  - compiled
  - command line interface + windowed interfaces
  - backend vs. fronted
  - dependency on file/folder naming convention

## NOW

- Python
  - free, open-source
  - IPython/JupyterLab-based application
  - pipeline in Jupyter Notebooks
  - notebook guides the user
  - parameters in YAML files (human-readable)
  - interpreted
  - notebooks with functions and widgets
  - code as software
  - user-configurable URLs (path + hdf5 link)

# METAMORPHOSIS

## AFTER

The screenshot shows a JupyterLab environment with the following components:

- Code Editor (Left):** Contains Python code for loading data and visualizing it. The code includes comments and function calls like `load_data` and `diffraction pipeline`.
- Terminal (Right):** Displays system logs and output from the code execution, including file paths and system information.
- Figure (Center):** A 2D diffraction pattern visualization showing a bright central spot and surrounding rings, with axes ranging from 0 to 2000.
- Environment (Bottom):** Shows the current environment as `base (conda)` and the active kernel as `Python 3.8.10 (base)`.

# METAMORPHOSIS

## AFTER

The screenshot shows the JupyterLab environment with a code editor and an output view.

**Code Editor Content:**

```

confirm_execution
If False, the 'confirm_execution' calls will be ignored.
...
@autoreload 2
@import pydicom.common
w = pydicom.common.get_confirm_execution_widget(parameters)
ParameterFile[2021-08-19_devrun04.preprocess.yml]::key='confirm execution' loaded from parameters file value=False
...
skip_visualization
If True, the visualization controls will not be displayed.
...
ParameterFile[2021-08-19_devrun04.preprocess.yml]::key='skip visualization' loaded from parameters file value=True
...
io parameters
This is where you point to the input files and where you want the outputs to be saved.
...
dark input format
Is the dark stage in 'hdfs' or 'sftp'?
...
ParameterFile[2021-08-19_devrun04.preprocess.yml]::key='dark input format' loaded from parameters file value=InputFormat.HDFS:
HDFS>
dark_input_format hdfs
...
dark filepath
If you prefer to set it manually, uncomment the line below and fill the variable in
...
@ dark_filepath = "/path/to/file.h5" # or "/path/to/file.esf"
...
@autoreload 2
@import pydicom.preprocess
From pydicom.preprocess import PreprocessParameterKeys as PK
try:
    parameters[PK.dark_filepath] = Path(dark_filepath)
    print(f'Manually selected: {dark_filepath}')
except NameError:
    pydicom.preprocess.get_dark_filepath_widget(parameters)
finally:
    del PK
...
dark_filepath
/home/ll1/bera04/128524/11/img_11/img_11_img_002.h5
...
img_11_img_002.h5
img_11_img_002.h5
img_11_img_003.h5
img_11_img_004.h5
img_11_img_005.h5
img_11_img_006.h5
...
ParameterFile[2021-08-19_devrun04.preprocess.yml] updating 'dark_filepath'.
ParameterFile[2021-08-19_devrun04.preprocess.yml]

```

**Output View Content:**

Parameter	Value
confirm_execution	False
dark_input_format	hdfs
dark_remove_reduction	True
PreprocessParameterKeys.dark_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002.h5
dark_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002.h5
data_input_format	InputFormat.HDFS
dark_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002.h5
dark_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002.h5
preprocessed_save	True
preprocessed_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002_preprocessed.h5
backgrounds_save	True
output_filepath	/home/ll1/bera04/128524/11/img_11/img_11_img_002_preprocessed.h5
skip_visualization	False
PreprocessParameterKeys.dark_remove_reduction	DarkDataRemovalOptions.mean
dark_remove_reduction	DarkDataRemovalOptions.mean
normalization	NormalizationOptions.none
margin_thresholding_box_size	(1, 100, 100, 1, 100, 100)
median_votability	100
median_window	500
median_iter_x	3
median_iter_y	3
save_backgrounds	False

**Code Editor Footer:** Working completed

# METAMORPHOSIS

## BEFORE

- EDF files
  - ESRF-specific format
  - 1 volume = 3600 files
  - metadata in yet other files

## NOW

- HDF5 files
  - generic format
  - 1 volume = 1 file
  - metadata in the same file

# METAMORPHOSIS

## BEFORE

- EDF files
  - ESRF-specific format
  - 1 volume = 3600 files
  - metadata in yet other files
- Parallelized on several machines
  - OAR (job config files, management)
  - network communication
  - repeated io-operations

## NOW

- HDF5 files
  - generic format
  - 1 volume = 1 file
  - metadata in the same file
- Parallelized on a single machine's cores
  - Python's native library `multiprocessing`
  - no network transfer
  - shared memory space



# METAMORPHOSIS

## BEFORE

- EDF files
  - ESRF-specific format
  - 1 volume = 3600 files
  - metadata in yet other files
- Parallelized on several machines
  - OAR (job config files, management)
  - network communication
  - repeated io-operations
- Performance
  - $\approx 30$  minutes
  - 8 machines

## NOW

- HDF5 files
  - generic format
  - 1 volume = 1 file
  - metadata in the same file
- Parallelized on a single machine's cores
  - Python's native library `multiprocessing`
  - no network transfer
  - shared memory space
- Performance
  - 10~15 minutes
  - single 30-core machine

# METAMORPHOSIS

## BEFORE

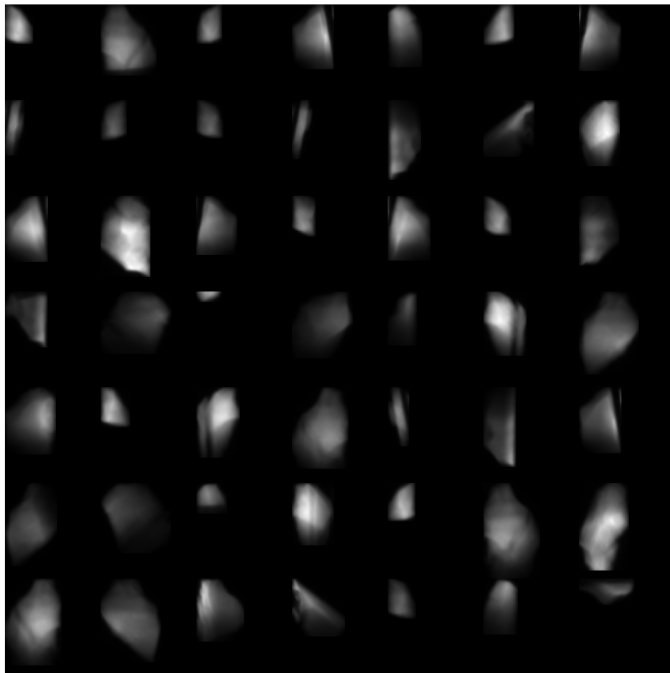
- EDF files
  - ESRF-specific format
  - 1 volume = 3600 files
  - metadata in yet other files
- Parallelized on several machines
  - OAR (job config files, management)
  - network communication
  - repeated io-operations
- Performance
  - $\approx 30$  minutes
  - 8 machines

## NOW

- HDF5 files
  - generic format
  - 1 volume = 1 file
  - metadata in the same file
- Parallelized on a single machine's cores
  - Python's native library `multiprocessing`
  - no network transfer
  - shared memory space
- Performance
  - 10~15 minutes
  - single 30-core machine

## RESULTS

- Installed and used in Psyché during an experiment
- Kenza's material (AD730) processed with `pydct`



# DIFSPOTS FROM AD730

The screenshot shows the Silx viewer interface. The left pane displays a tree view of files, with '000000.h5' selected. The right pane shows the metadata for this file, categorized into 'HDF5 File', 'Path info', and 'Attributes'.

Name	Description	Type	Shape	Link
000000.h5				
Area	2151	int64	scalar	
BoundingBox	Compressed...	int64	2 × 3	
BoundingBoxOriginSize	Compressed...	int64	2 × 3	
BoundingBoxXorigin	762	int64	scalar	
BoundingBoxXsize	45	int64	scalar	
BoundingBoxYorigin	1094	int64	scalar	
BoundingBoxYsize	63	int64	scalar	
CentroidImage	1.37985	float64	scalar	
CentroidIndex	Compressed...	float64	3	
CentroidX	18.9391	float64	scalar	
CentroidY	35.5055	float64	scalar	
Data	Compressed...	float32	6 × 63 × 45	
DataMasked	Compressed...	float32	6 × 63 × 45	
difspotID	0	int64	scalar	
EndImage	1350	int64	scalar	
Integral	2.79796e+07	float32	scalar	
Mask	Compressed...	int8	6 × 63 × 45	
MaxImage	1	int64	scalar	
MaxIntensityIndex	Compressed...	int64	3	
Projection	Compressed...	float32	63 × 45	
StartImage	1345	int64	scalar	

**HDF5 File**

**Path info**

Basename	
Name	/
Local	/home/joaopcbertoldo/Downloads/difspots/000000.h5-:/
Physical	/home/joaopcbertoldo/Downloads/difspots/000000.h5-:/

**Attributes**

interpretation	"difspot"
notation_axes	"zyx"
notation_indexing	"starts with 1 (MATLAB)"

HDF5

# DIFSPOTS FROM AD730

Slix viewer

File Options Views Help

Name	Description	Type	Shape	Link
000000.h5				
• Area	2151	int64	scalar	
■ BoundingBox	Compressed...	int64	2 × 3	
■ BoundingBoxOriginSize	Compressed...	int64	2 × 3	
• BoundingBoxXorigin	762	int64	scalar	
• BoundingBoxXsize	45	int64	scalar	
• BoundingBoxYorigin	1094	int64	scalar	
• BoundingBoxYsize	63	int64	scalar	
• CentroidImage	1.37985	float64	scalar	
~ CentroidIndex	Compressed...	int64	3	
• CentroidX	18.9391	float64	scalar	
• CentroidY	35.5055	float64	scalar	
■ Data	Compressed...	float32	6 × 63 × 45	
■ DataMasked	Compressed...	float32	6 × 63 × 45	
• difspotID	0	int64	scalar	
• EndImage	1350	int64	scalar	
• Integral	2.79796e+07	float32	scalar	
■ Mask	Compressed...	int8	6 × 63 × 45	
• MaxImage	1	int64	scalar	
~ MaxIntensityIndex	Compressed...	int64	3	
■ Projection	Compressed...	float32	63 × 45	
• StartImage	1345	int64	scalar	

/home/joapcbertoldo/Downloads/difspots/000000.h5::Data

X: -6.256841 Y: 18.52393 Data: - Dims: 45x63

Axis selection

Dimension 0  limits: 0, 5

Dimension 1 y

Dimension 2 x

HDF5 Curve Image Raw Image stack

# DIFSPOTS FROM AD730

Sixx viewer

File Options Views Help

Name	Description	Type	Shape	Link
000000.h5				
• Area	2151	int64	scalar	
■ BoundingBox	Compressed...	int64	2 × 3	
■ BoundingBoxOriginSize	Compressed...	int64	2 × 3	
• BoundingBoxXorigin	762	int64	scalar	
• BoundingBoxXsize	45	int64	scalar	
• BoundingBoxYorigin	1094	int64	scalar	
• BoundingBoxYsize	63	int64	scalar	
• CentroidImage	1.37985	float64	scalar	
~ CentroidIndex	Compressed...	float64	3	
• CentroidX	18.9391	float64	scalar	
• CentroidY	35.5055	float64	scalar	
• Data	Compressed...	float32	6 × 63 × 45	
• DataMasked	Compressed...	float32	6 × 63 × 45	
• difspotID	0	int64	scalar	
• EndImage	1350	int64	scalar	
• Integral	2.79796e+07	float32	scalar	
• Mask	Compressed...	int8	6 × 63 × 45	
• MaxImage	1	int64	scalar	
~ MaxIntensityIndex	Compressed...	int64	3	
■ Projection	Compressed...	float32	63 × 45	
• StartImage	1345	int64	scalar	

000001.h5  
000002.h5  
000003.h5  
000004.h5  
000005.h5  
000006.h5  
000007.h5

/home/foapc Bertoldo/Downloads/difspots/000000.h5::DataMasked

X: 4.385429 Y: 0.1384615 Data: 0 Dims: 45x63

Axis selection

Dimension 0  limits: 0, 5

Dimension 1 y

Dimension 2 x

HDF5 Curve Image Raw Image stack

# THANK YOU!

Joao P C Bertoldo

Materials Center @ MINES Paristech - PSL University  
ID11 @ The European Synchrotron Radiation Facility (ESRF)



26 August 2021

# SEGMENTATION

## DOUBLE THRESHOLD IN 1D

Figure 7: double threshold illustration