An overview of offline data analysis at different PETRA beamlines (and its relation to online data analysis)

Ch. Rosemann

DESY FS-EC

PETRA III/Photon Science Users Meeting
Thursday, January 24, 2018

Overview

PS@DESY is a very heterogeneous thing

by beamline

(which I know of) e.g. P03, P06, P10 (and P07, P08, P09)

by scientific topic, e.g.

- diffraction (including SAXS/WAXS)
- tomography/ptychography
- GISAXS
- "simple" spectra

by detector

frame rates, output format, compression, (dimensions)

... and there a lot of tools and data formats around

Basic unit of processing

Single spectrum

Resonance scans, flourescence

Single image, no order

masking, background subtraction, transformation

Singular image(s) as sequence

dependency on monotonically varying parameter

Image stack

processing of bulk data

Higher view

More Heterogenity

- large variance in data rate and computational complexity
- different processing chains:
  some strictly sequential
  some heavily parallelizable
- combinations occur:
  a parallel problem as a part of a sequence

No standards

- no standard data format
- no standard tool
- no standard toolchain

Ex 1: Tomography reconstruction – back projection

Three distinct steps

1. data taking
2. azimuthal integration
3. tomo reconstruction

Azimuthal integration is key

for every point in a projection

- open the image & mask it (takes a lot of time)
- polar transformation
- integration along azimuthal direction

Quick first step – quasi Online

- do transformation and integration asap
- collect spectra by scan position to create 3D data

Ex 2: Correlation spectroscopy

In principle: time-series analysis of image stack

Three distinct steps

1. data taking: a complete stack
2. factorial analysis of:
   - regions of images (single pixel to subregions of image)
   - time intervals: single frame, two frames, four frames, ...

Parallelizable code

- question of processing chain
- currently MATLAB implementation

Ex 3: Sequence processing – GISAXS sputtering

From image to spectrum

record image and create projective cut along line

Extract feature from spectrum

curve fitting plus data processing

Time evolution of feature

time is equal to scan number

Special processing tool

DPDAK

Important problem part data compression

depends mainly on detector

hep solution

don't compress data

Photon Science?

we need a good processing scheme and infrastructure:

- factorize steps for easy scaling
- decompression $\otimes$ processing/data reduction
- key question for scaling ability: fast/short communication

Soon

bigger data fans are already approaching, e.g. Lambda 9M

Online and offline merge

every BM I approached wants live data processing

- **live** can have very different meanings
- usually as informative tool for decision making:
  don't waste time
  don't waste space

code maintenance

impractical to waste resources for two versions

Challenges

My experience so far: it is a completely fragmented picture

$$\text{detector} \otimes \text{processing} \otimes \text{tools}$$

$$\text{detector} \approx \text{rate} \otimes \text{compression} \otimes \text{fileformat}$$

$$\text{processing} \approx \text{parallel} \otimes \text{sequential} \otimes \text{single/bulk}$$

$$\text{tools} \approx \text{OS} \otimes \text{custom}$$

... and more challenges

$$\text{language}(\textit{topic}, \textit{beamline}, \textit{application})$$

+ wildly varying experience and knowledge basis