



How to extend the existing experimental dataset?

Current status:



May 27, 2022

Dataset Open Access

Reflectometry curves (XRR and NR) and corresponding fits for machine learning

Pithan, Linus; Greco, Alessandro; Hinderhofer, Alexander; Gerlach, Alexander; Kowarik, Stefan; Rußegger, Nadine; Dax, Ingrid; Schreiber, Frank

Public reflectometry data collection

- about 250 experimental XRR profiles
- sample parameters & classical fits are published together with measurements (in ML-context: “labelled data”)

L.Pithan et al. 2022 <https://doi.org/10.5281/zenodo.6497438>
<https://schreiber-lab.github.io/reflectometry-dataset/>

```
In [1]: #prepare jupyter notebook for plots shown below
%run prepare_plot.py

#imports
from silx.io.dictdump import nxdictdict #read NeXus hdf5 to python dict
import numpy as np
import pandas as pd
from IPython.display import display

#use same q_range as during fitting
q_fit = np.linspace(0.02, 0.15, 130, endpoint=False)

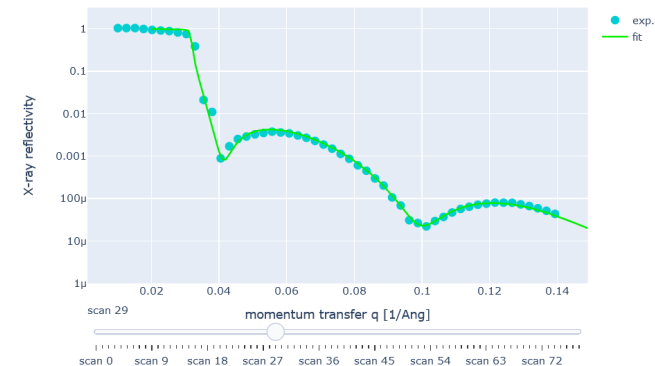
#produce plots
for key, ds in nxdictdict("xrr_dataset.h5").items():
    if "q" not in key: #skip nexus attributes
        print("Dataset: ", key)
        print("Experimentalists: ", ds["metadata"].pop("Experimentalists", "?"))
        ds["metadata"].pop("@X_class", "")
        display(pd.DataFrame.from_dict({"Dataset": [key], **ds["metadata"]}))

        fig = prepare_figure(ds, q_fit, str(ds["metadata"]["Layer_material"]) + " on SiOx")
        fig.show()
```

Dataset: DIP_1
Experimentalists: ['Kowarik, Stefan']

Dataset	Layer_CAS	Layer_formula	Layer_material	Substrate_temperature	Substrate_temperature@unit	Instrument	q_max
0	DIP_1	188-94-3	C32H16 Diindenoperylene	303	K	ESRF, ID10b	0

Diindenoperylene on SiOx: 94.45 Ang.



The screenshot displays the Reflectometry DB DRAFT web application. On the left, a sidebar contains search filters: Search (with a Clear button), Text Search, Location, Group, Type, Keywords, Start Date – End Date, and an Add Condition button. The main area shows a table of data entries under the 'Type' column, which lists 'raw' and 'derived'. Each entry has a corresponding plot thumbnail. Red circles highlight the 'raw' and 'derived' labels in the 'Type' column. Green arrows point from these labels to green text boxes explaining their meanings: 'raw' refers to experimental or processed data like XRR curves, while 'derived' refers to fit results or predictions. A third green box points to the search filters, stating that a powerful search is needed to explore the data. On the right, a yellow box partially visible says 'work Intended'.

Reflectometry DB DRAFT

Items per page: 25 | 1 – 10 of 10

Search [Clear]

Text Search

Location

Group

Type

Keywords

Start Date – End Date

+ Add Condition

Name
DIP_1
DIP_1

“raw”: Experimental or processed data (e.g., XRR curves) ... could just be references to data in facility catalogues or Zenodo.

“derived”: fit results or predictions of
There can be many derived datasets towards the same “raw / source”
In future this would allow meta studies of different models.

Evidently, we need a powerful search to explore the data.

work
Intended

Open Refl
An international, open effort to improve
<https://www.reflectometry.org/>

work in progress –
Intended to be proposed to



ORSO

Open Reflectometry Standards Organisation

An international, open effort to improve the scientific techniques of neutron and X-ray reflectometry

<https://www.reflectometry.org/>



Display of raw and processed data

ORSC DAPHNE Reflectometry DB DRAFT

Datasets / XRRDB/1ae74349-3240-429d-b67a-ed61ec41cbce /

Details Datafiles

General Information

Name DIP_1

PID XRRDB/1ae74349-3240-429d-b67a-ed61ec41cbce

Type RAW

Creation Time 2022-11-19 22:30

Keywords

Creator Information

Owner xrr

Principal Investigator Kowark, Stefan

Contact Email someone@somewhere.de

Owner Group xrr

Access Groups xrr

File Information

Source Folder http://xrrdb.sytes.net/download/DIP_1.hdf5

Related Documents

Creation Location ESRF, ID10b

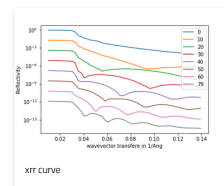
Scientific Metadata

Search x

Experimentalists

Layer_CAS	188-94-3
Layer_formula	C32H16
Layer_material	Diindenoperylene
Substrate_temperature	303
Substrate_temperature@unit	K
Instrument	ESRF, ID10b

Some identifiers,
could be DOI...



Graphical preview of the dataset

Relevant download info. For XRR the file
format could be ORSO .txt or ORSO .hdf5

All relevant metadata to find the dataset and
the studied sample (materials, expected layer
structure, thicknesses...)

For XRR relatively easy: follow ORSO
definitions and extend them later

ORSC

Universal Descriptive Model Language

Motivation

When building a physical model of samples for analysis of reflectivity, there is no standard naming convention for some shared concepts (e.g. a single layer or multilayer)

Implementations

Two routes have been discussed for implementing a shared model language:

Simple model language to use for sample description prior to analysis

A draft for a computer readable sample description that is quite universal and can be included in the ORSO file header was introduced by the file formats working group: [Simple Model Language Specification](#)

Edit this page on GitHub by clicking here
© ORSO - Powered by Jekyll & whiteglass - Subscribe via RSS

https://www.reflectometry.org/projects/model_language/



Display of analysed data

Reflectometry DB DRAFT

Datasets / XRRDB/d233342e-768f-45a1-a56e-cd2ebfa86ad0 /

Details Datafiles Lifecycle

General Information

Name DIP_1

PID XRRDB/d233342e-768f-45a1-a56e-cd2ebfa86ad0

Type derived

Creation Time 2022-11-19 22:30

Keywords

Creator Information

Owner xrr

Investigator Pithan, Linus and Greco, Alessandro

Contact Email someone@somewhere.de

Owner Group xrr

Access Groups xrr

File Information

Source Folder http://xrrdb.sytes.net/download/DIP_1.hdf5

Related Documents

Input Datasets XRRDB/1ae74349-3240-429d-b57a-ed51ec41cbce

Derived Data

Software Used GenX

Scientific Metadata

Search x

Air_sld 0

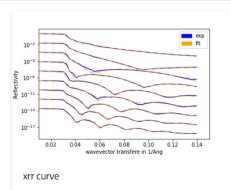
Film_roughness

Film_roughness@unit Ang

Film_thickness

0 0.3362124901223951

1 6.167671247444737



Graphical preview of analysis & fit results

Download link for analysed data

Link to raw dataset

Software used for analysis / fitting

Searchable metadata describing the analysis.
This should rather contain a machine-readable descriptive summary of the analysis than the fit itself...

Why do we care about this?

- Collect open data for machine learning efforts
- Systematic comparison of different analysis tools
- Long-term: Build reproducible processing chains