Data Reduction

Grenoble, 24/25 October 2013 | Joachim Wuttke, Jülich Centre for Neutron Science

Two steps of data analysis

$N(j, i)$ neutron count histogram

↓ *raw data reduction*

$S(q, \omega)$ scattering law

↓ *data analysis, fitting, interpretation*

results (qualitative and quantitative)

Limitations of the two-step approach

- "$S(q,\omega)$" is scattering-length weighed
- spectrometer resolution $R(q,\omega)$
- q depends on 2θ and ω
- multiple scattering
- Poisson statistics described by Gaussian error bars
- systematic uncertainties

S(q,w) still contains scattering physics

$$\frac{\partial^2\sigma}{\partial\Omega\partial\omega} = \frac{k'}{k}\frac{\sigma}{4\pi}S(q,\omega)$$

with $$S(q,\omega) = \int \mathrm{d}t\, \mathrm{e}^{-i\omega t}\frac{1}{N}\sum_{jk}\langle \mathrm{e}^{iqR_j(t)}\mathrm{e}^{-iqR_k(0)}\rangle$$

holds only in few simple systems — otherwise:

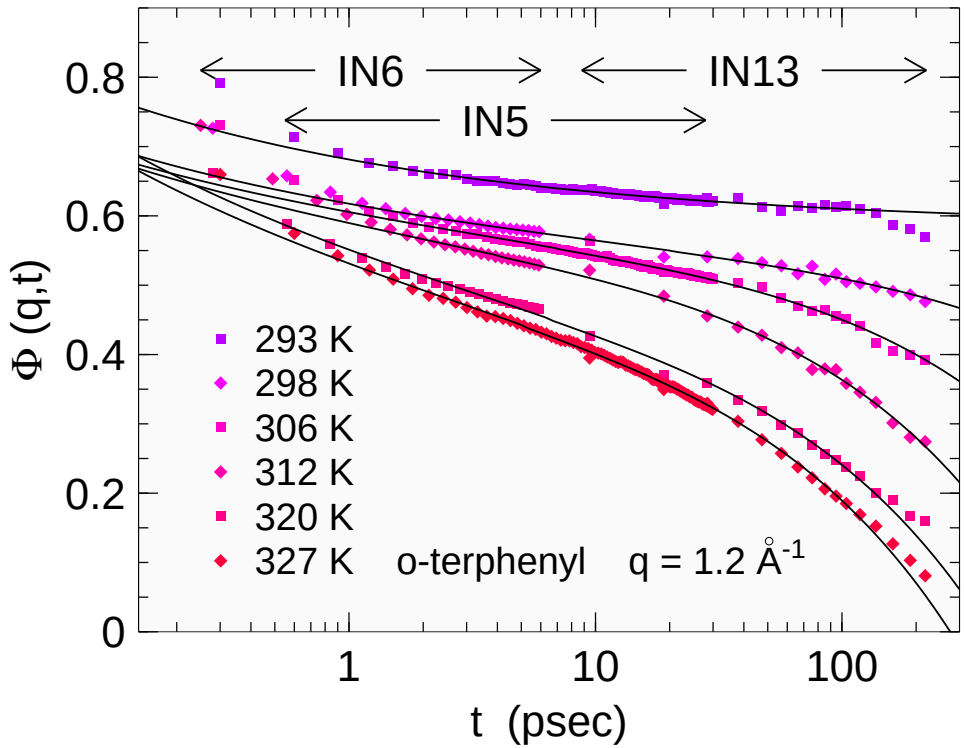
- coherent and incoherent contributions
- elements have different scattering lengths

$$\frac{\partial^2\sigma}{\partial\Omega\partial\omega} = \frac{k'}{k}\int \mathrm{d}t\, \mathrm{e}^{-i\omega t}\frac{1}{N}\sum_{jk}\langle b_j b_k^* \mathrm{e}^{iqR_j(t)}\mathrm{e}^{-iqR_k(0)}\rangle$$

Spectrometer resolution

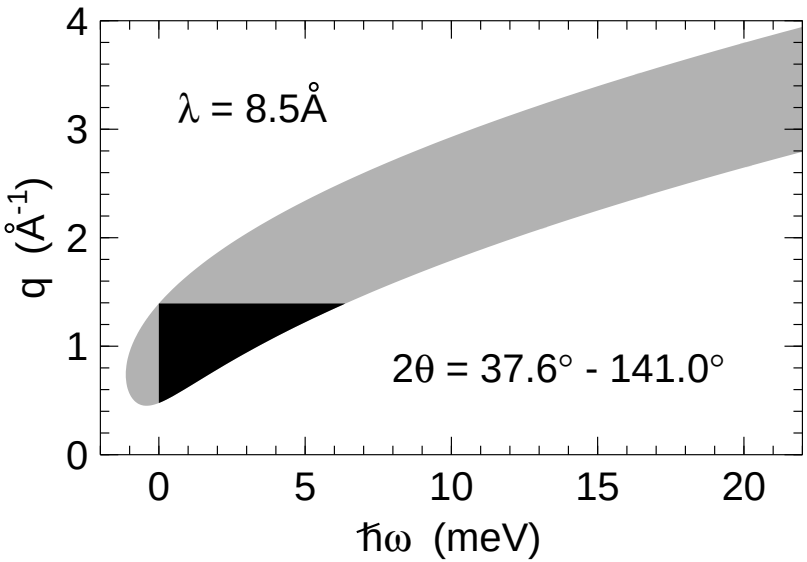
complete separation of sample physics and spectrometer physics through deconvolution:

$$S(q,t) = S_{\mathrm{exp}}(q,t)/R_{\mathrm{exp}}(q,t)$$



Dynamic range

q depends not only on 2θ but also on ω



multiple scattering is nonlinear in S, nonlocal in q and ω
$\Rightarrow$correction requires model $S(q,\omega)$ beyond measured range

Errors and Uncertainties

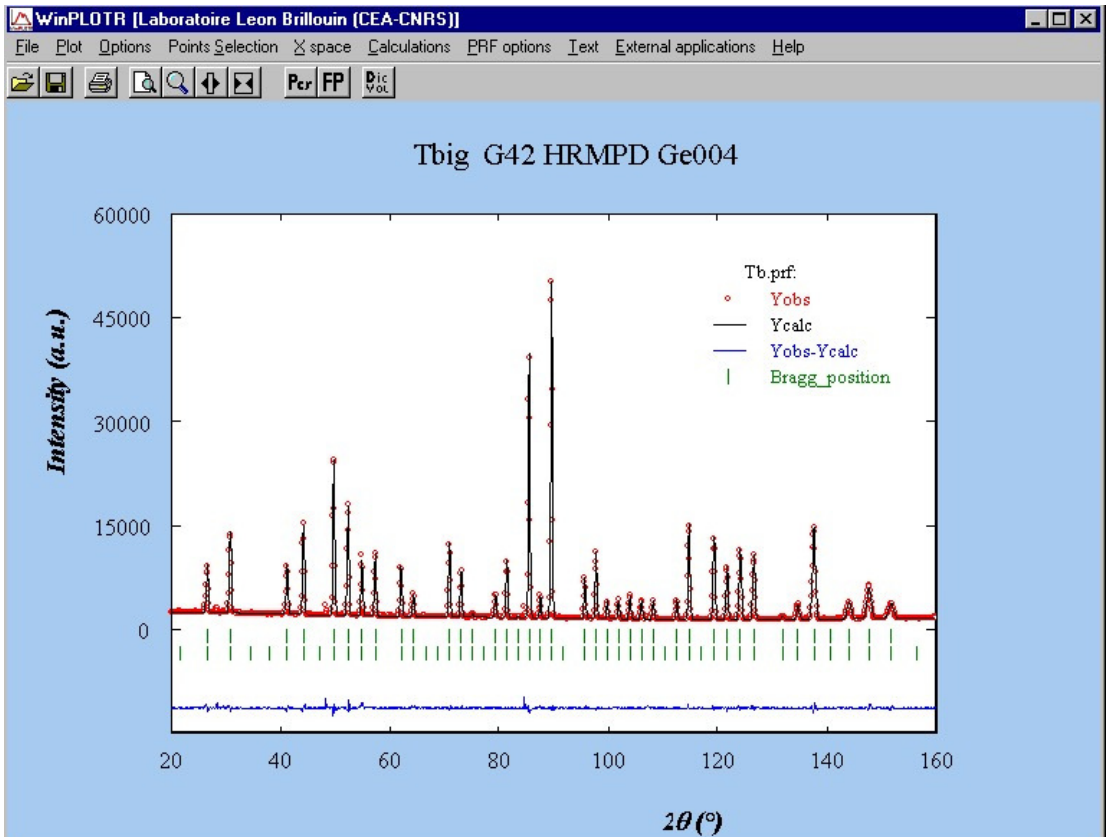
best practice:
least-squares fitting weighed with reciprocal variance
$\Leftarrow$ Gaussian error propagation
$\Leftarrow$ Gaussian approximation for Poisson distribution
$\Leftarrow$ sufficient binning (at least 5-10 counts per channel)

Stochastic errors say nothing about systematic uncertainties

Beyond data reduction

Long-term perspective:
replace “data reduction” by “virtual instrument” fitting?

realized e. g. for diffraction by Fullprof



Two-step data flow

$N(j, i)$ neutron count histogram

↓ *raw data reduction*

$S(q, \omega)$ scattering law

↓ *data analysis*

results

Two-step data flow

$N(j, i)$ instrument-specific format $\rightarrow$ NeXus?

$\downarrow$ instrument-specific program $\rightarrow$ generic framework?

$S(q, \omega)$ various data formats

$\downarrow$ various data-analysis programs

results

Rank

aka dimensionality

rank 3:

$S(q, \omega; T)$ inelastic temperature scan
$S(q, \omega; t)$ other inelastic time scan

rank 2:

$S(q, \omega)$ regular scan
$S(\omega; t)$ q-averaged/selected time scan
$S(q, 0; t)$ or $S(q, 0; T)$ elastic time/temperature scan

rank 1:

$S(\omega)$ q-averaged/selected spectrum
$S(q, 0)$ elastic intensity

New data reduction software

why not adapt SQW / INX ?

- support new instruments
- support time-dependent inelastic scans (rank $+= 1$)
- support various output formats
- different ways of control: dialog, GUI, scripts, programs

SLAW

SPHERES data → scattering law

software architecture

- implemented to specification (man page)
- controlled by `Slawfile`
- 1000 lines Ruby (`Slawfile` → low-level script)
- 2000 lines C++ (raw data → $S(q, \omega)$)

typical input script (`Slawfile`)

```
ip /sraw/T28/
tp T
Tp Tsam
bc *.4
dd 0-4,14-15
bd 0
4u3-28  of Cell300
5u7-    of Coff004
6u6-    of Coff200 Tp Treg
```

SLAW: lessons learned

observations

- used by 98% of instrument users
- creating a `Slawfile`:
  - nobody reads man page
  - nobody writes `Slawfile` from scratch
  - inherited `Slawfile` gets slowly messed up
  - writing in/out-rules too repetitive
  - only few have analyzed rank 3 experiments
- instrument responsibles rather break than improve sw.

next SLAW version

for TOFTOF

immediate decisions

- more verbous commands
- plus fixed comments (as in `Doxyfile`, `/etc/...`)?

make Slawfile foolproof?

- automatic regeneration of `Slawfile`??
- semi-automatic generation of in/out-rules (← catalogue)??

or make it Turing-complete?

- replace home-brewed syntax by Python?

mid/long-term aims

- GUI for parametrization
- integration into data visualization?
- only run on server under our control?