Data collection - considerations and compromises

Kay Diederichs

Universität Konstanz

This talk

- is about 1) *geometrical aspects* 2) *important error types* of the data collection experiment

- I will take some material from Zbigniew Dauter's talks (brown background)

- supplemented by facts investigated by James Holton

- finally, new aspects (mainly) for the Pilatus detector

Why collect data?

Goal of the experiment is: collect -
(quantitatively) *complete* and
(qualitatively) *accurate* data

Ideally, after data processing, the intensities of (ideally) all unique reflections of the asymmetric unit of reciprocal space should be known accurately

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Data-collection strategies

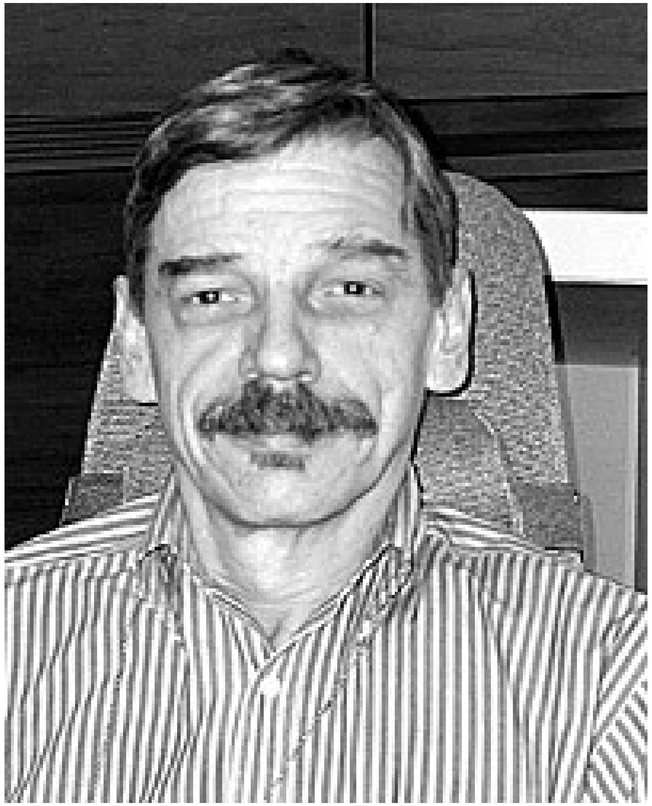
Zbigniew Dauter

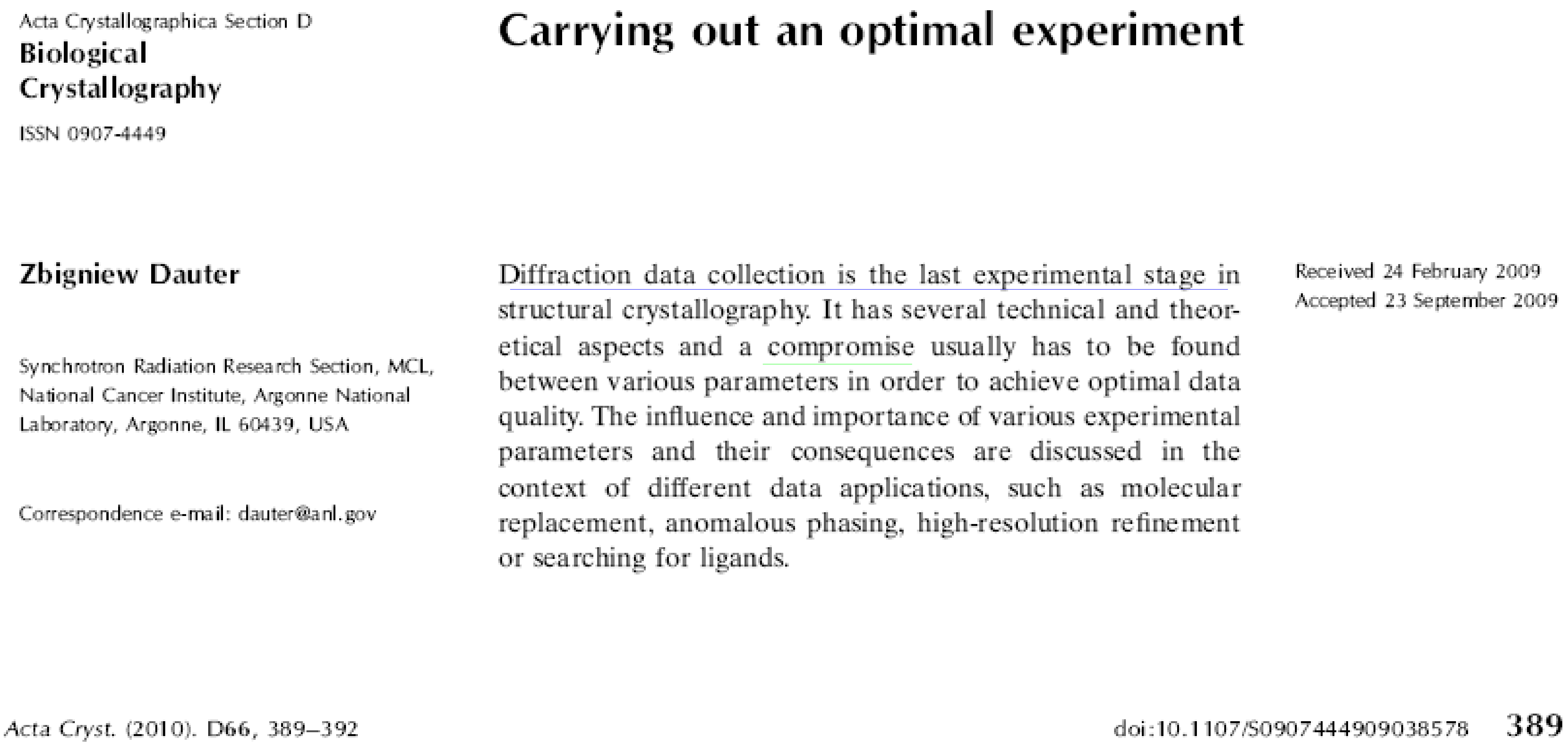
National Cancer Institute, Frederick and Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA

Correspondence e-mail: dauter@bnl.gov

The optimal strategy for collecting X-ray diffraction data from macromolecular crystals is discussed. Two kinds of factors influencing the completeness of data are considered. The first are geometric, arising from the symmetry of the reciprocal lattice and from the experimental setup; they affect quantitatively the completeness of the measured set of reflections. The second concern the quality, or information content, of the recorded intensities of these measured reflections.

Received 28 January 1999
Accepted 22 June 1999





Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Carrying out an optimal experiment

Zbigniew Dauter

Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: dauter@anl.gov

Received 24 February 2009
Accepted 23 September 2009

Diffraction data collection is the last experimental stage in structural crystallography. It has several technical and theoretical aspects and a compromise usually has to be found between various parameters in order to achieve optimal data quality. The influence and importance of various experimental parameters and their consequences are discussed in the context of different data applications, such as molecular replacement, anomalous phasing, high-resolution refinement or searching for ligands.

Acta Cryst. (2010). D66, 389–392 doi:10.1107/S0907444909038578 **389**

From: http://www.ccp4.ac.uk/schools/APS-2011/tutorials/data-peculiarities/Dauter_data.ppt

Advance preparations

1. Crystals must be prepared

2. You must be prepared

because

Anything that can go wrong,
will go wrong (Murphy, 2000 BC)

Data collection process

- Easy to screw-up in many ways
- Involves lots of technical problems
- But it is science, not technicality
- Pays off to "engage your brain"
- Last truly experimental step;
  later mostly computing (and writing-up)
  which may be repeated many times
- *good quality data make all subsequent steps much easier*

Type of experiment

Always best to have diffraction data
complete, high resolution and accurate
but of particular importance are:

- Native data for refinement
  highest resolution (multiple passes)

- Molecular replacement
  medium resolution, no overloads

- Heavy atom derivative
  medium resolution, accurate

- Anomalous (MAD, SAD)
  modest resolution (radiation damage)
  very accurate and complete at low res.

I. Completeness

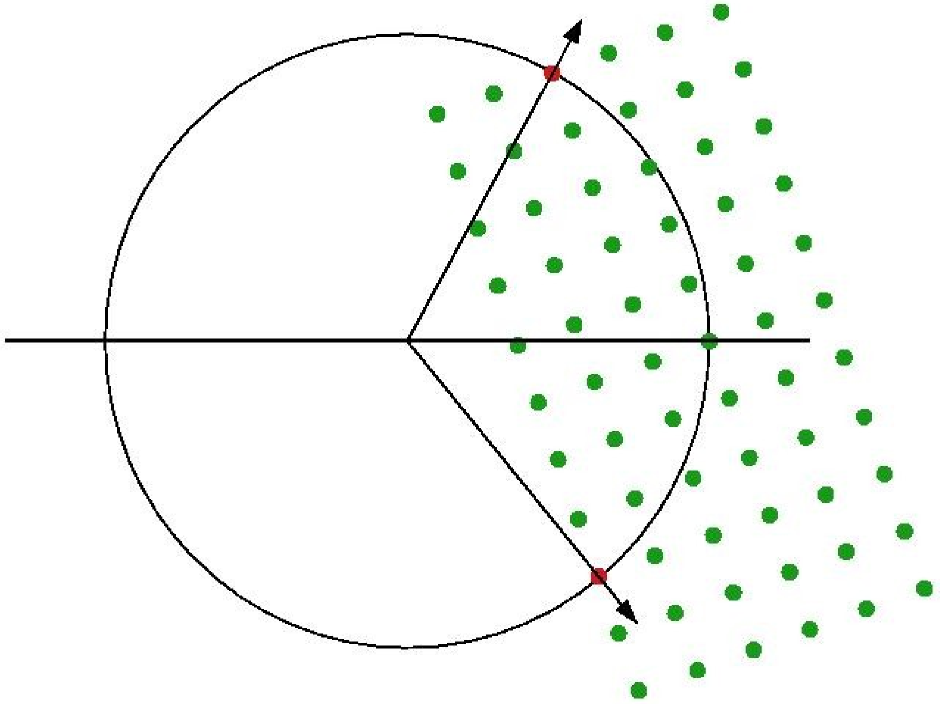
Quantitative completeness of indices

Depends entirely on the geometry
and mutual disposition of

Reciprocal lattice (crystal)
and
Ewald sphere (radiation)

Ewald construction

3-D illustration of Bragg's law: $n \cdot \lambda = 2 \cdot d \cdot \sin\theta$



Ewald sphere	represents	radiation
Reciprocal lattice	represents	crystal

Asymmetric unit in reciprocal space

Asymmetric unit in reciprocal space is always a wedge bounded by rotation axes (or planes in Laue group):

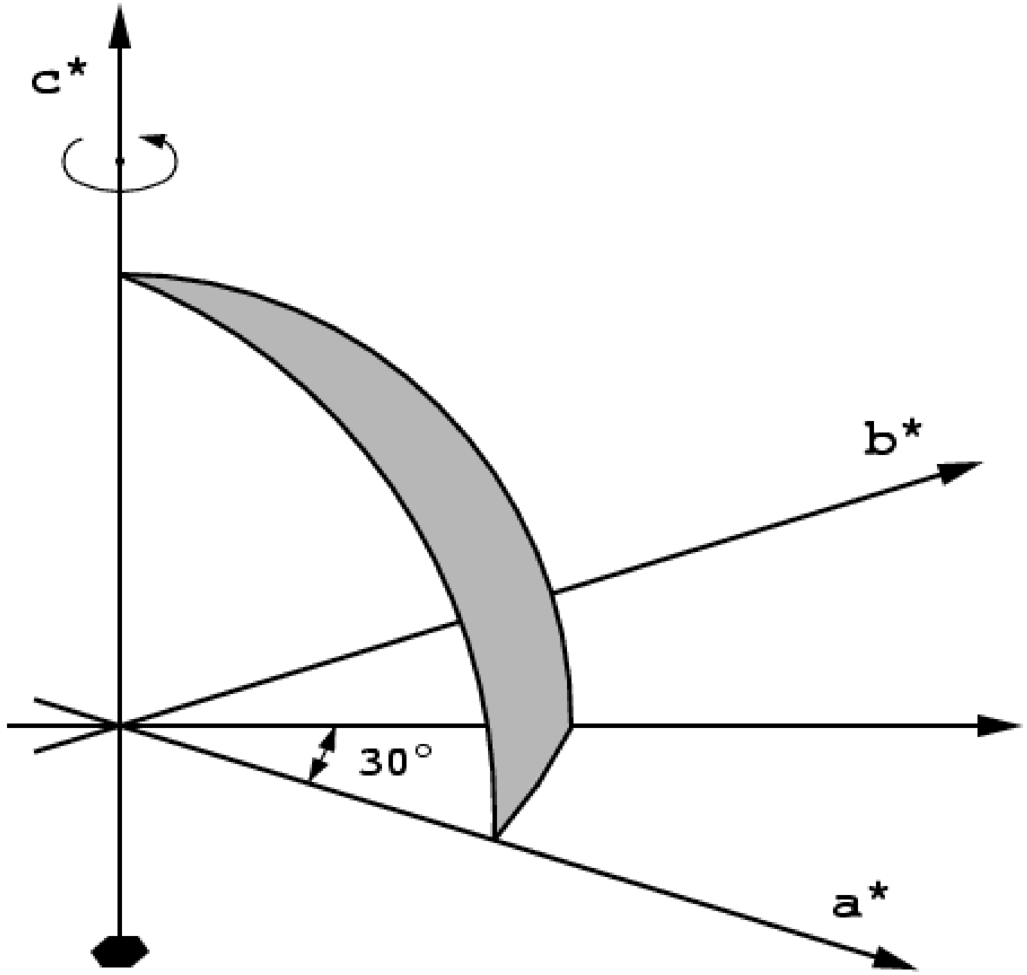
	native	anomalous
Triclinic	hemisphere	sphere
Orthorhombic	octant	quadrant

etc.

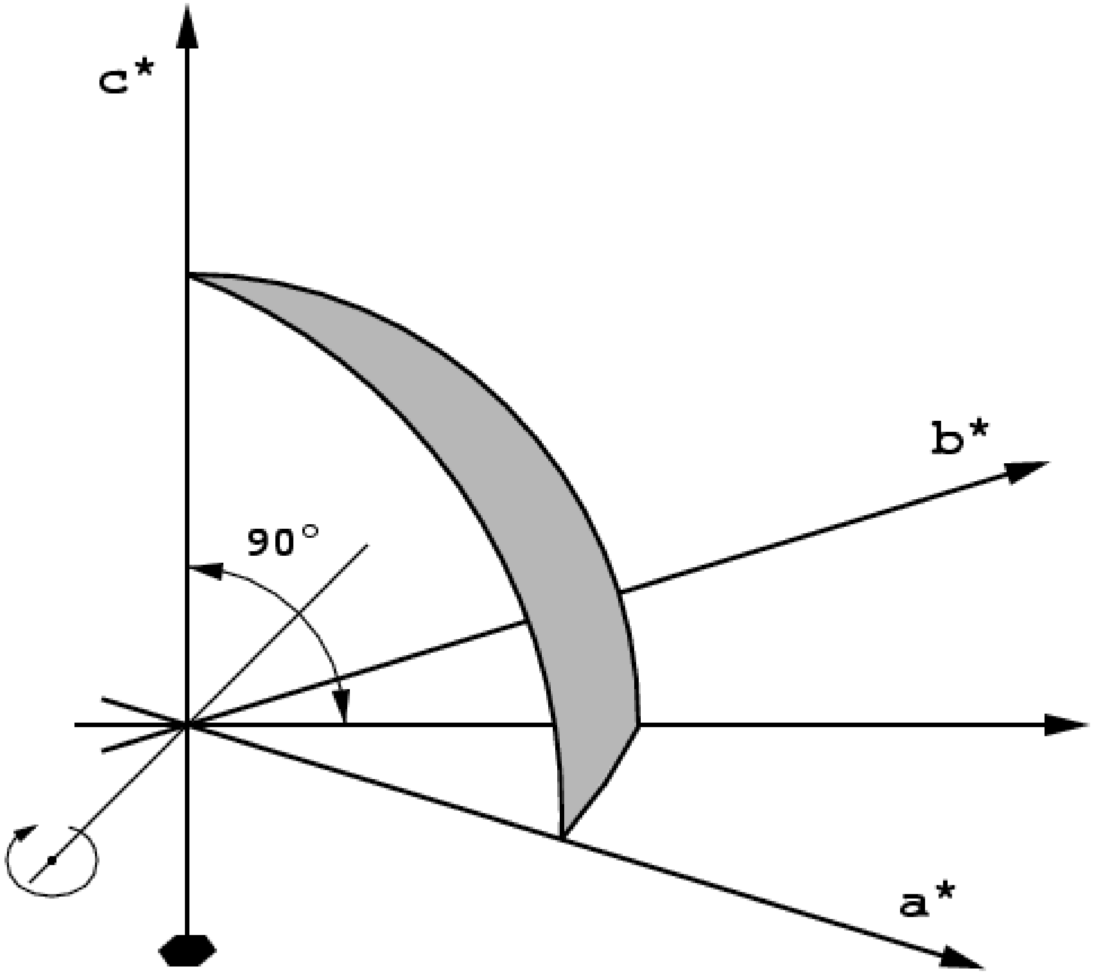
It is important to know where to start and how much to rotate the crystal

1999 !

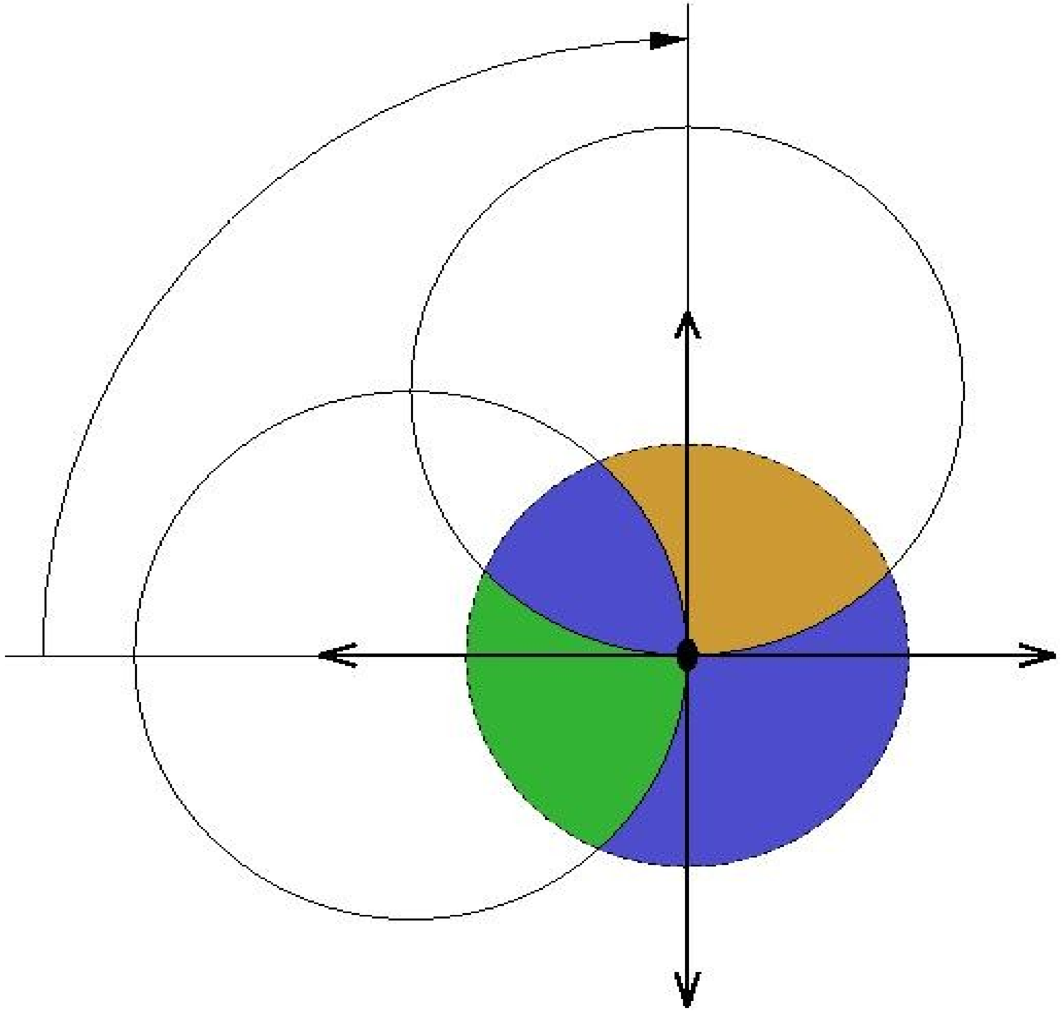
Asymmetric unit in 622 - c axis rotation



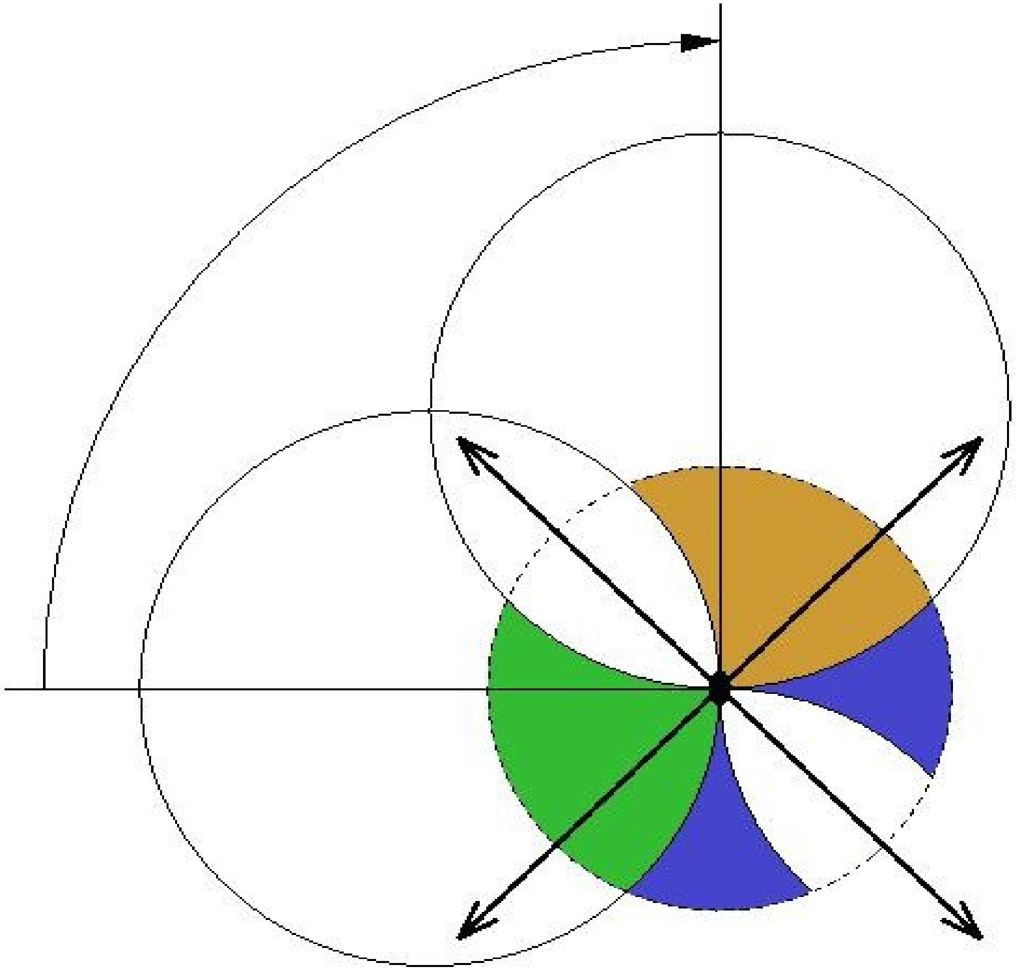
Asymmetric unit in 622 - a/b axis rotation



Asymmetric unit in 222 – 90° axial



Asymmetric unit in 222 – 90° diagonal



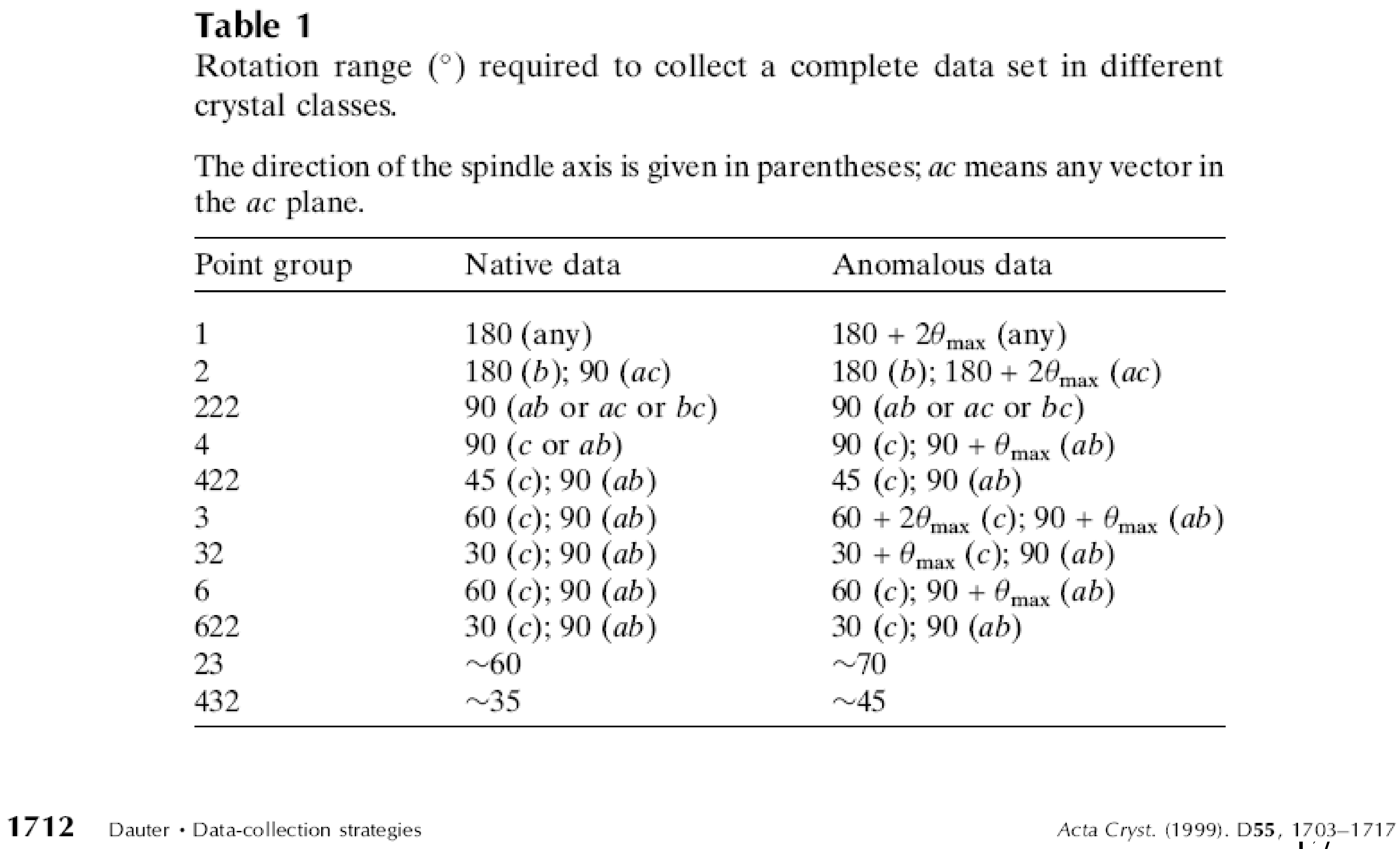


Table 1

Rotation range (°) required to collect a complete data set in different crystal classes.

The direction of the spindle axis is given in parentheses; *ac* means any vector in the *ac* plane.

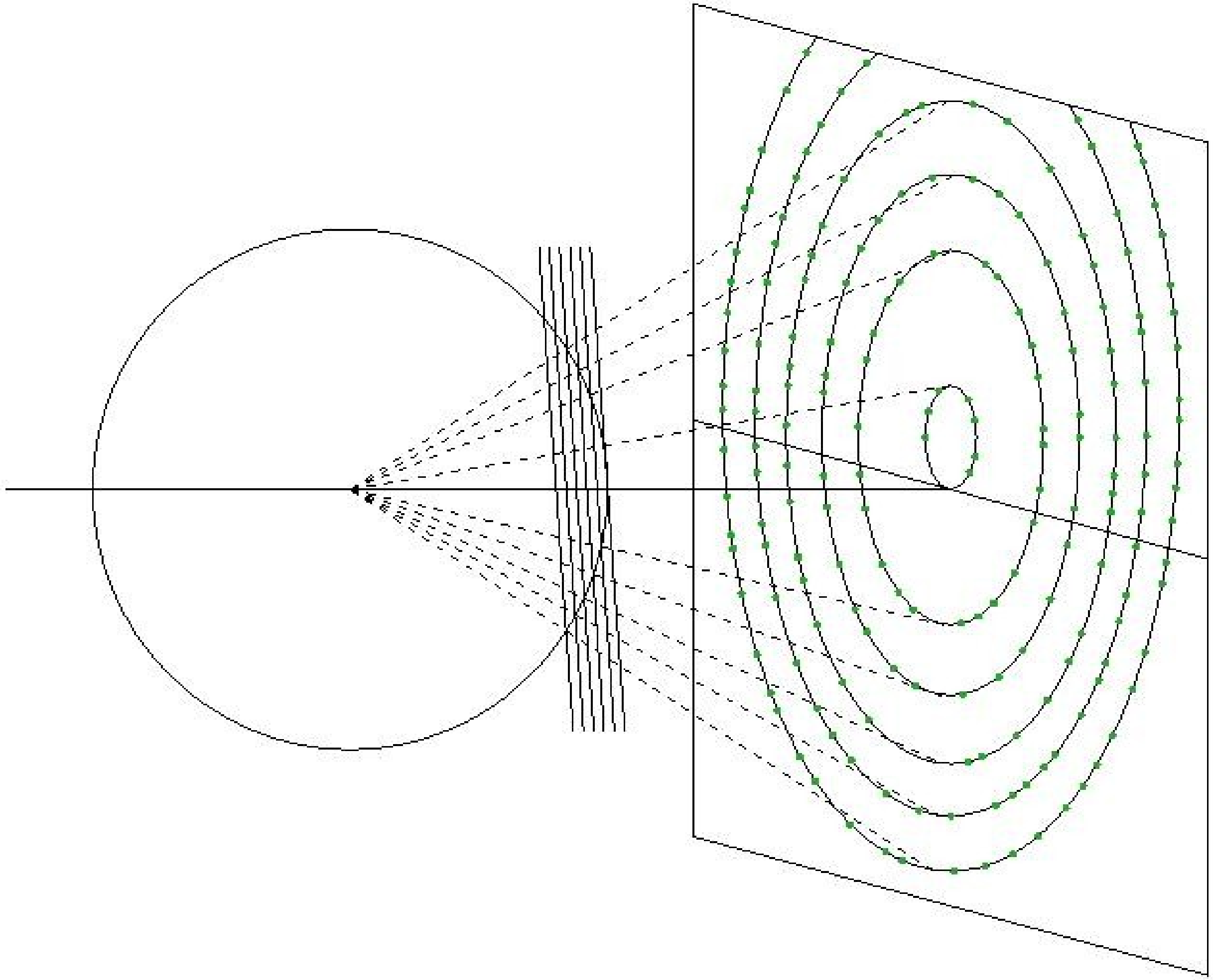
Point group	Native data	Anomalous data
1	180 (any)	$180 + 2\theta_{\max}$ (any)
2	180 (*b*); 90 (*ac*)	180 (*b*); $180 + 2\theta_{\max}$ (*ac*)
222	90 (*ab* or *ac* or *bc*)	90 (*ab* or *ac* or *bc*)
4	90 (*c* or *ab*)	90 (*c*); $90 + \theta_{\max}$ (*ab*)
422	45 (*c*); 90 (*ab*)	45 (*c*); 90 (*ab*)
3	60 (*c*); 90 (*ab*)	$60 + 2\theta_{\max}$ (*c*); $90 + \theta_{\max}$ (*ab*)
32	30 (*c*); 90 (*ab*)	$30 + \theta_{\max}$ (*c*); 90 (*ab*)
6	60 (*c*); 90 (*ab*)	60 (*c*); $90 + \theta_{\max}$ (*ab*)
622	30 (*c*); 90 (*ab*)	30 (*c*); 90 (*ab*)
23	~60	~70
432	~35	~45

II. „Oscillation“ range

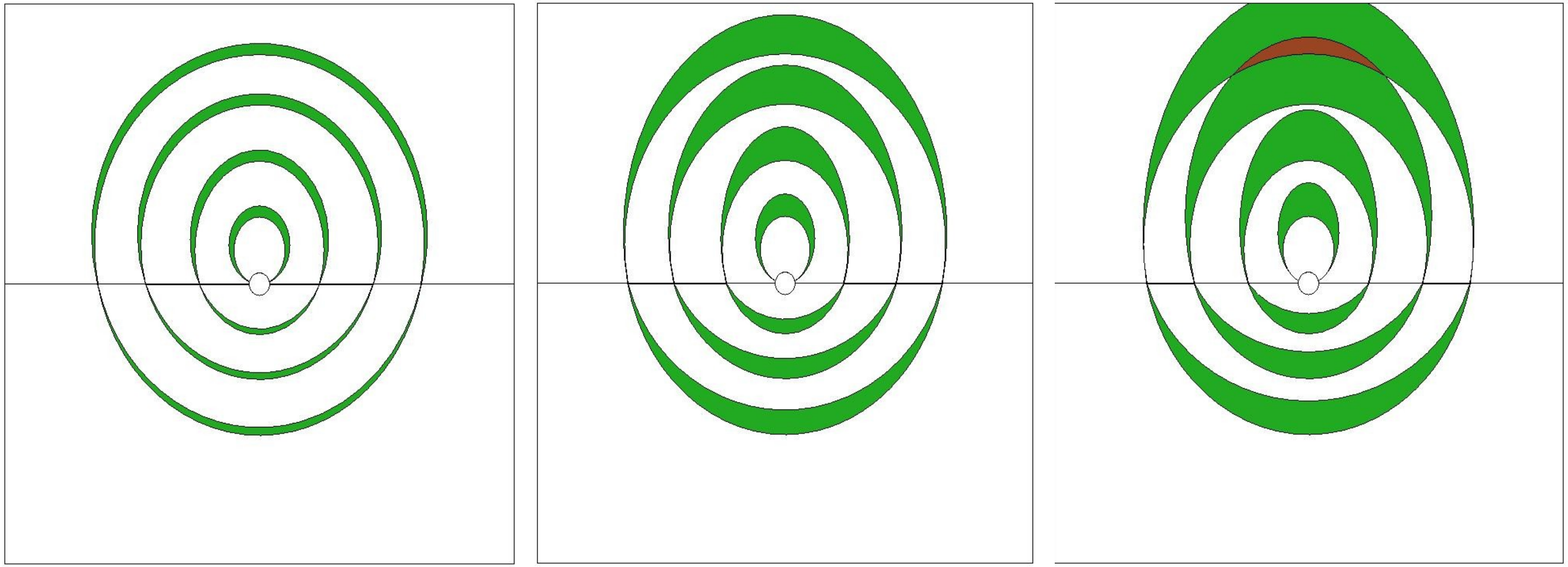
Influences:

- „partials“ versus „fullies“
- Overloads
- Background

Still image



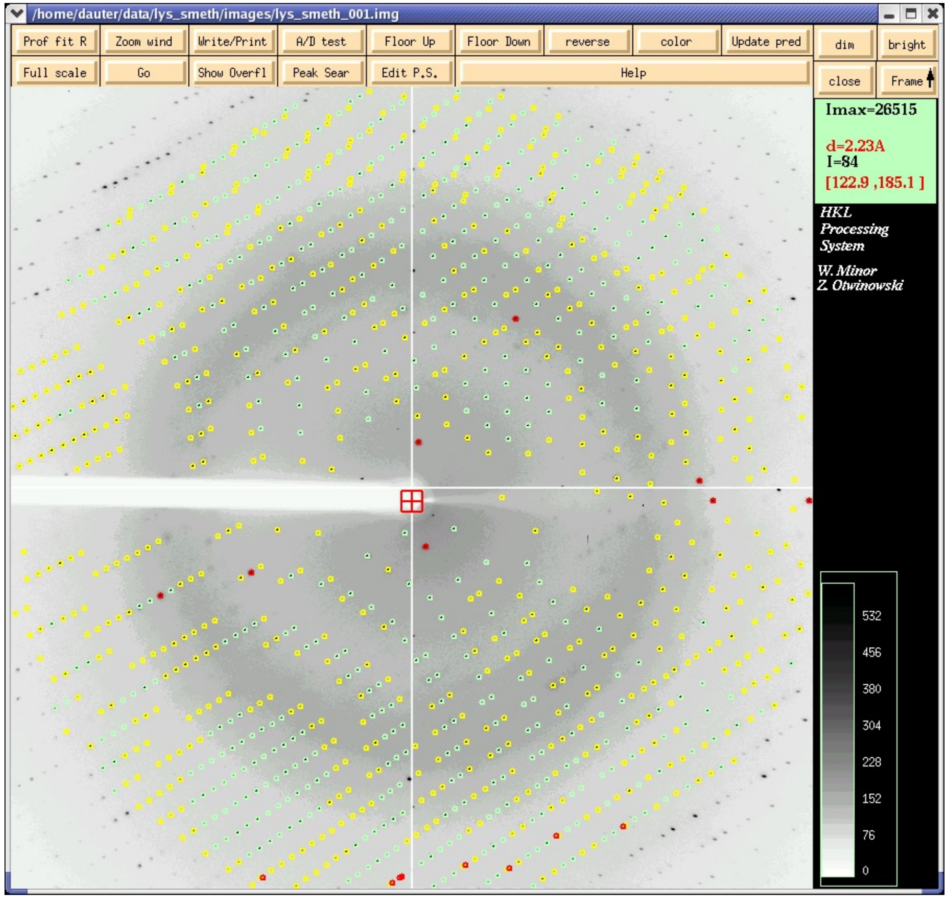
Increasing rotation



$$\Delta\varphi_{max} = \frac{180 \cdot d}{\pi \cdot a} - \eta$$

d - resolution
a - cell parameter
|| beam
η - mosaicity

Overloaded profiles

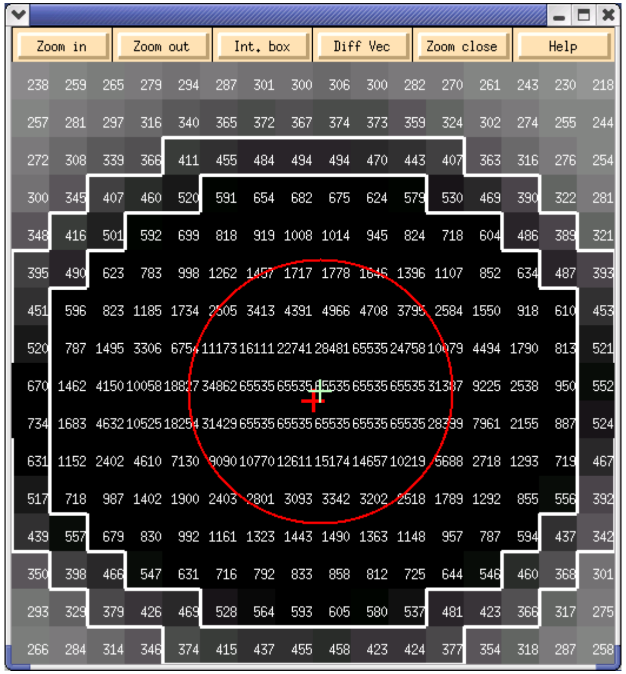


Imax=26515

d=2.23A
I=84
[122.9 ,185.1]

HKL Processing System

W. Minor
Z. Otwinowski



238	259	265	279	294	287	301	300	306	300	282	270	261	243	230	218
257	281	297	316	340	365	372	367	374	373	359	324	302	274	255	244
272	308	339	366	411	455	484	494	494	470	443	407	363	316	276	254
300	345	407	460	520	591	654	682	675	624	579	530	469	390	322	281
348	416	501	592	699	818	919	1008	1014	945	824	718	604	486	389	321
395	490	623	783	998	1262	1457	1717	1778	1646	1396	1107	852	634	487	393
451	596	823	1185	1734	2505	3413	4391	4966	4708	3795	2584	1550	918	610	453
520	787	1495	3306	6754	11173	16111	22741	28481	65535	24758	10079	4494	1790	813	521
670	1462	4150	10058	18827	34862	65535	65535	65535	65535	65535	31387	9225	2538	950	552
734	1683	4632	10525	18254	31429	65535	65535	65535	65535	65535	28399	7961	2155	887	524
631	1152	2402	4610	7130	9090	10770	12611	15174	14657	10219	5688	2718	1293	719	467
517	718	987	1402	1900	2403	2801	3093	3342	3202	2518	1789	1292	855	556	392
439	557	679	830	992	1161	1323	1443	1490	1363	1148	957	787	594	437	342
350	398	466	547	631	716	792	833	858	812	725	644	546	460	368	301
293	329	379	426	469	528	564	593	605	580	537	481	423	366	317	275
266	284	314	346	374	415	437	455	458	423	424	377	354	318	287	258

Best, strongest reflections – very important for Fourier maps, Pattersons, direct methods, phasing

III. Multiplicity

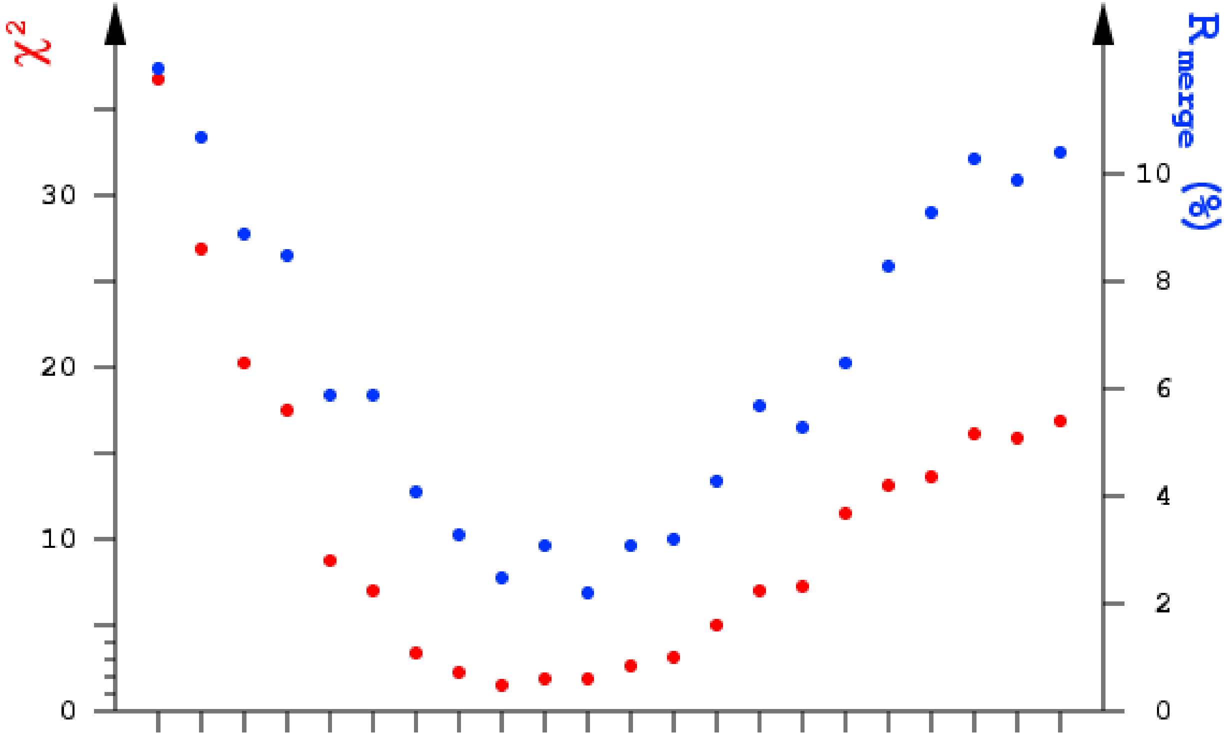
More measurements of equivalent reflections

lead to more accurate average and σ estimation

Also scaling and merging is more effective

But beware of radiation damage

IV. Radiation damage



Typical syndrome of radiation damage –
first and last data do not agree with average

Radiation damage

Actually, multiplicity enables us to

- discover radiation damage
- quantify radiation damage: R_d-plot
- (at least partially) compensate for it: zero-dose extrapolation (XSCALE)

Structure solution would be easy – without experimental error

Error = random + systematic

Random = counting + detector

Multiplicity of n reduces the random error by √n

Systematic = Radiation damage + absorption + non-linearities + vibrations + instabilities + …

Multiplicity does *not* necessarily reduce the systematic error!

“If you don’t have good data, then you have no data at all.” -Sung-Hou Kim

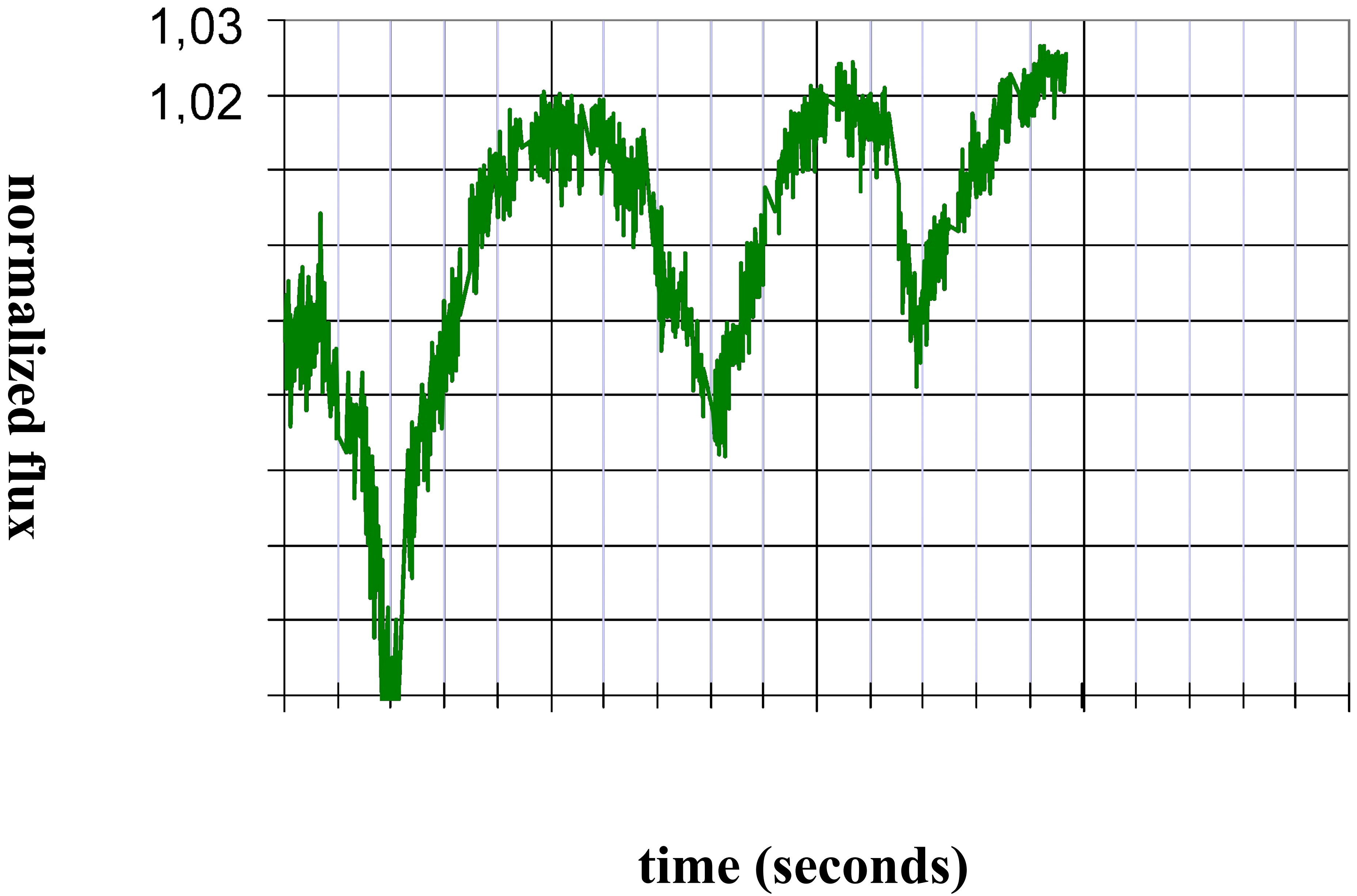
“If you don’t have good data, then you must learn statistics.” - James Holton

Systematic errors investigated by

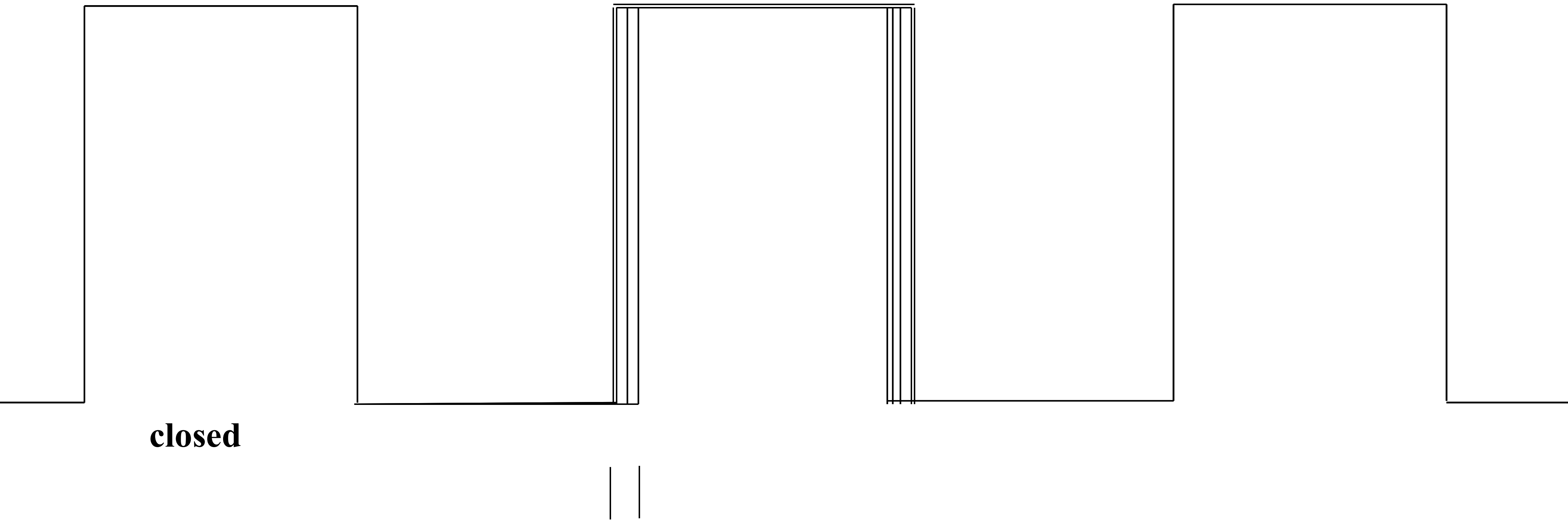
James Holton – see

http://bl831.als.lbl.gov/~jamesh/powerpoint/

Beam Flicker



Shutter Jitter

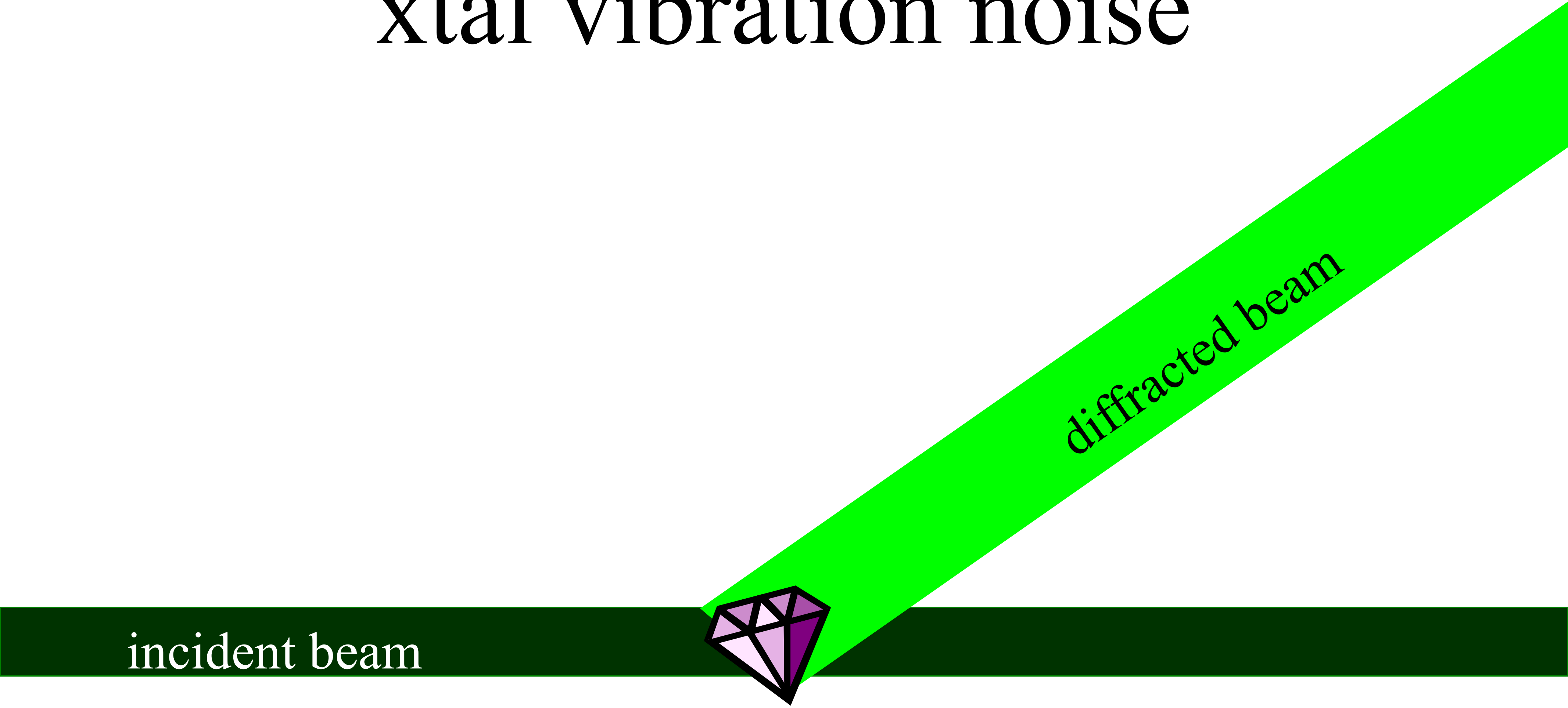


open

closed

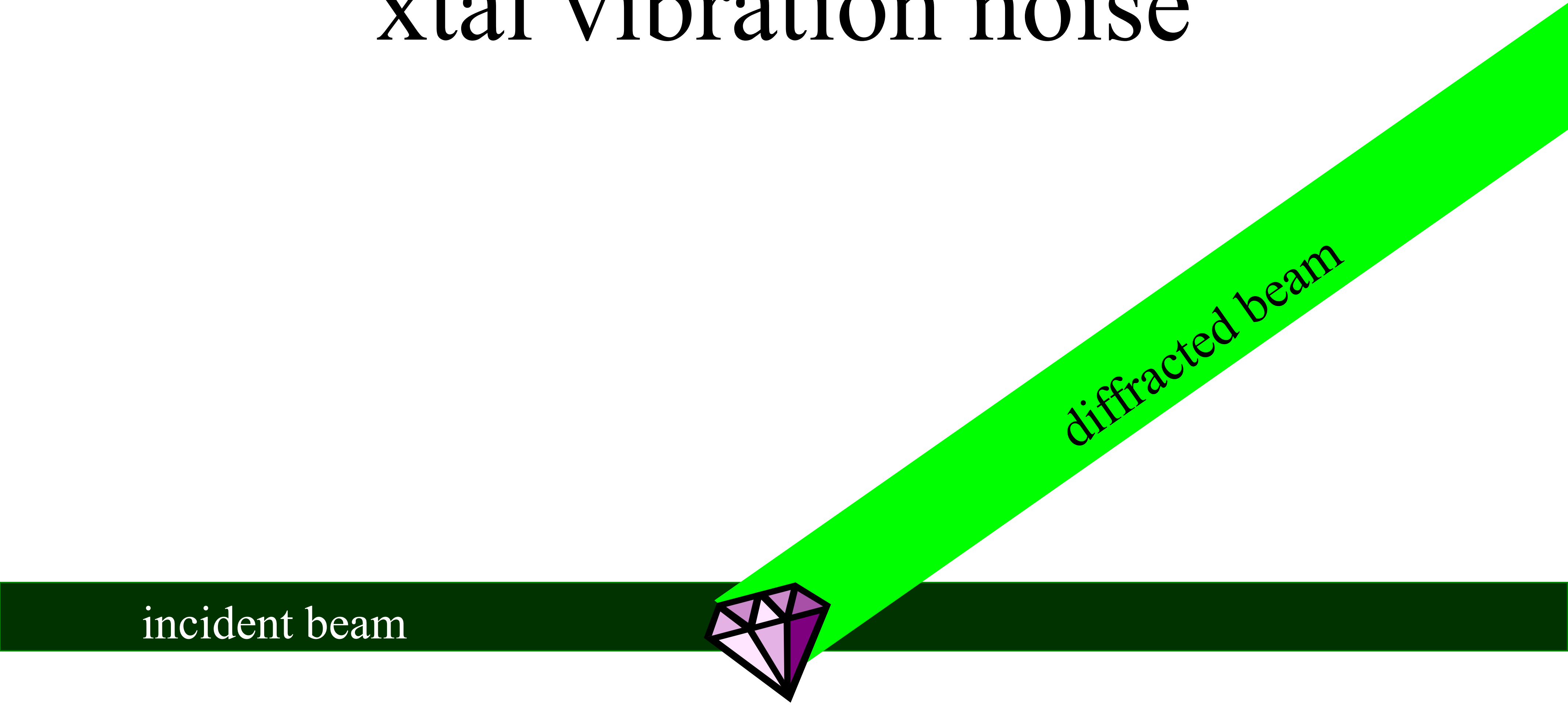
shutter jitter

xtal vibration noise



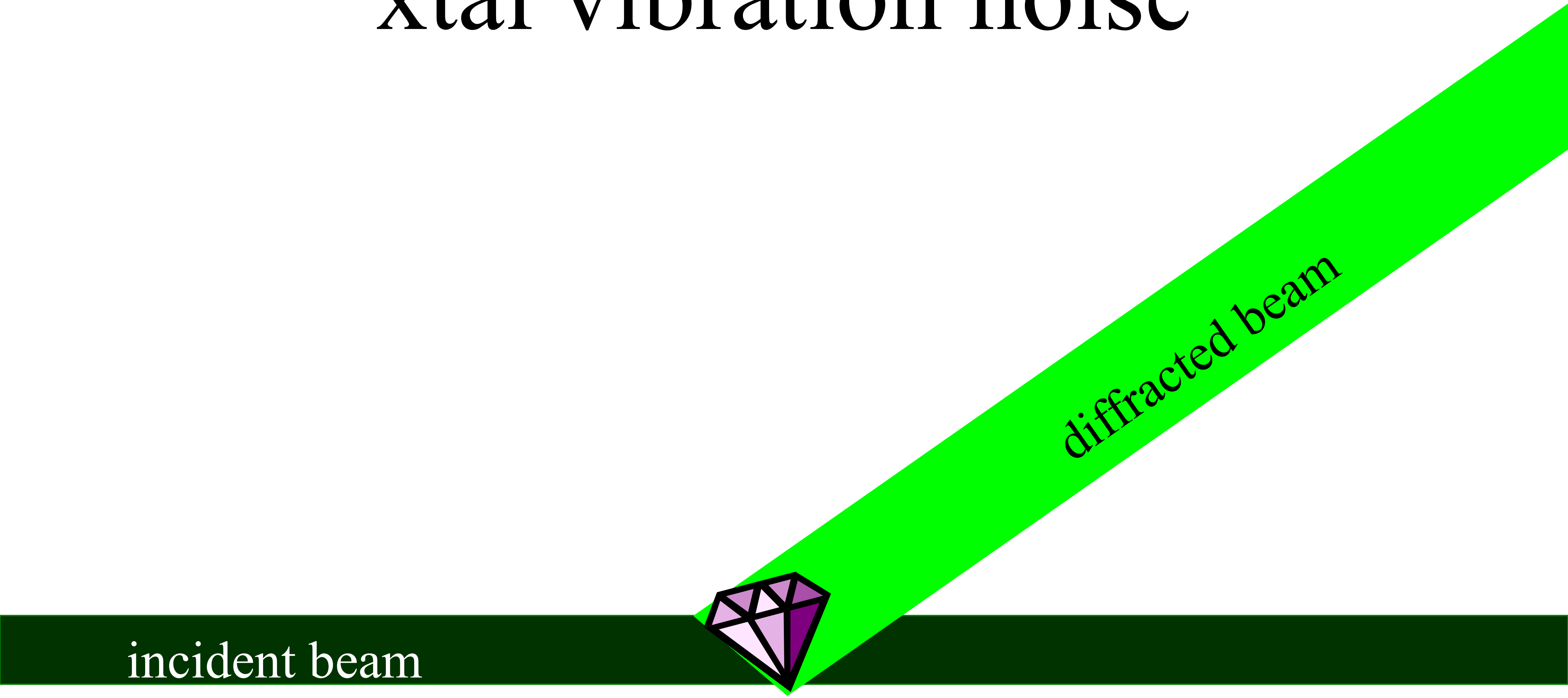
see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



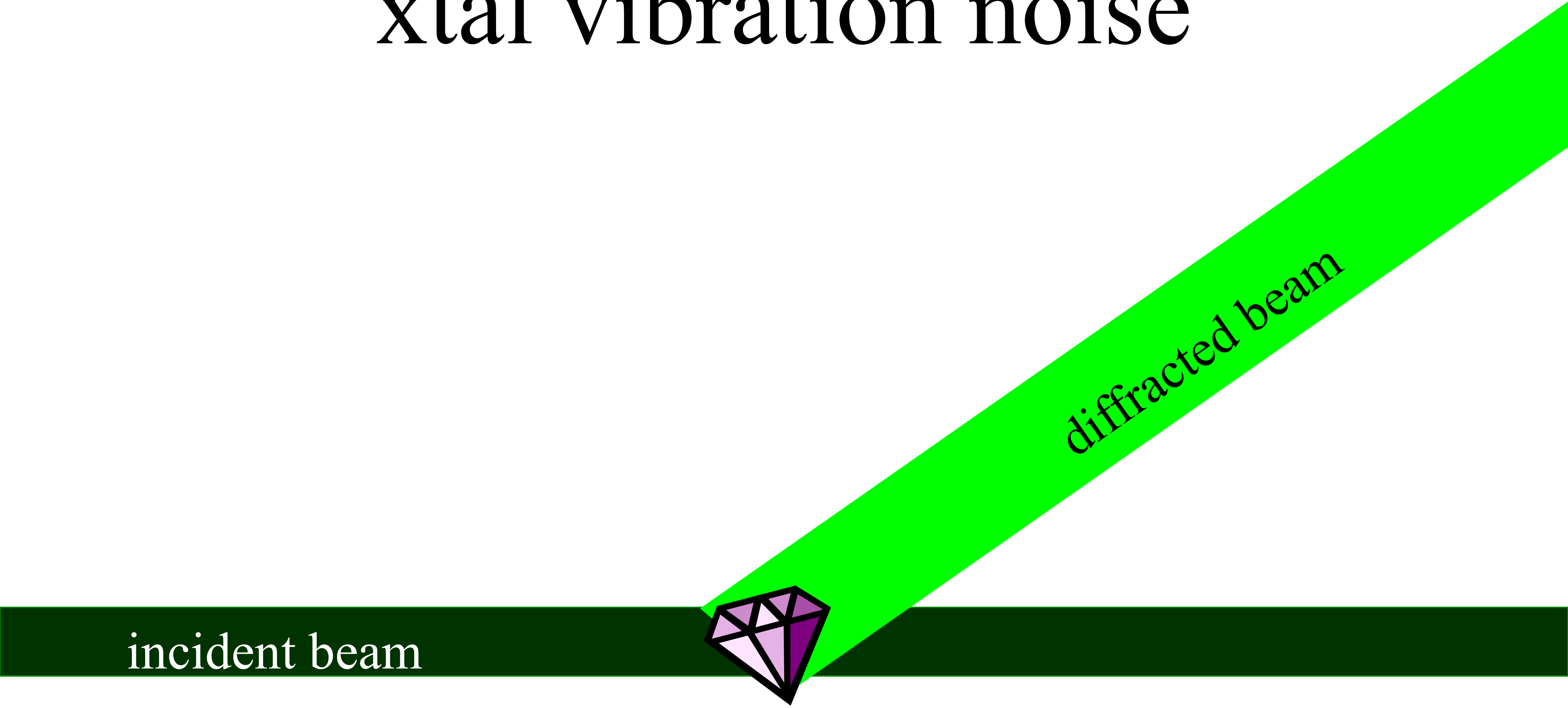
see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



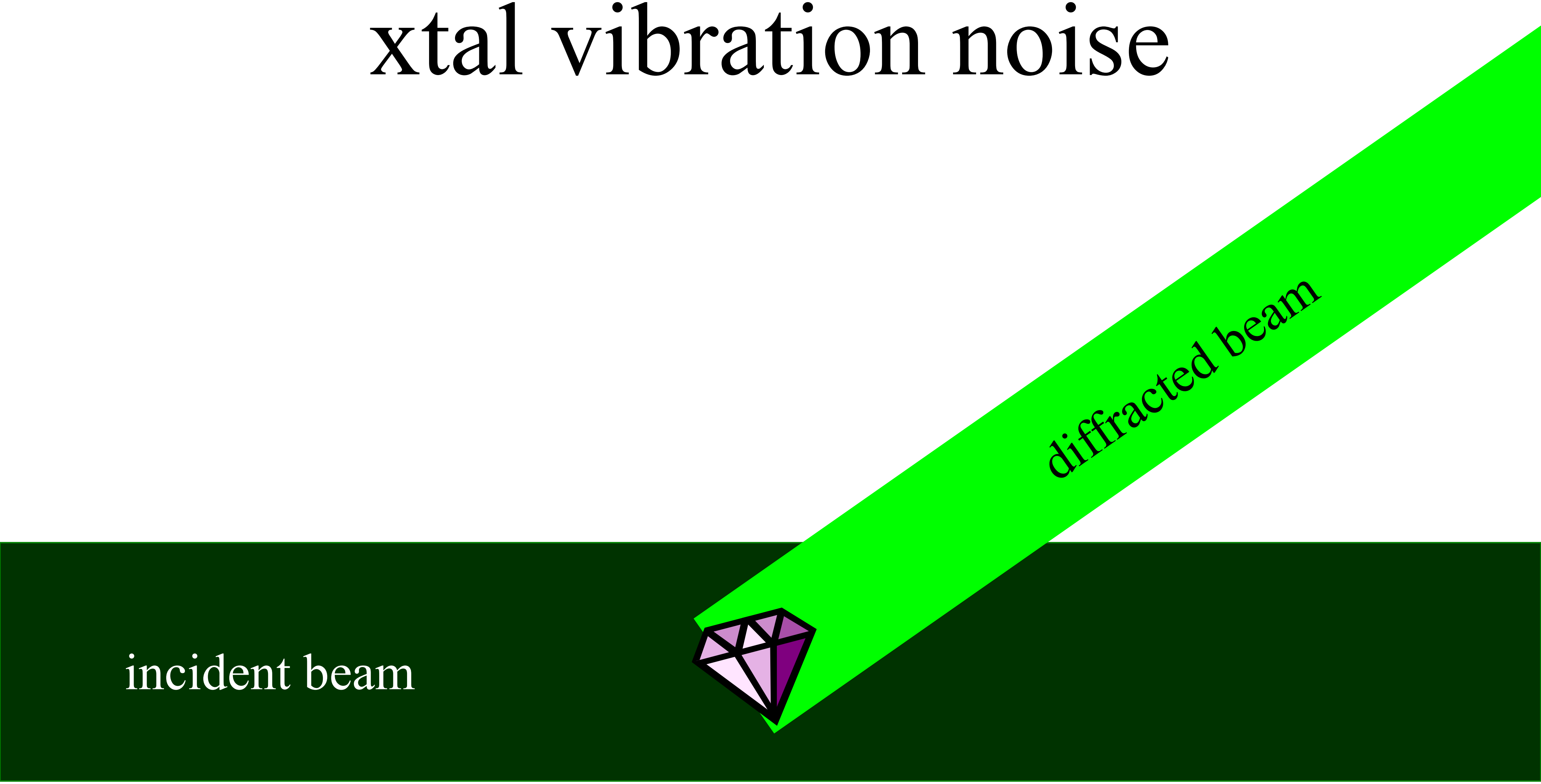
see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



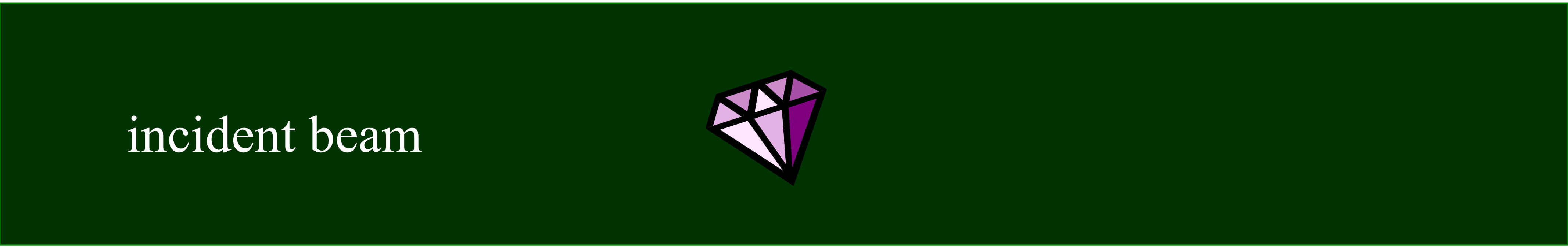
see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

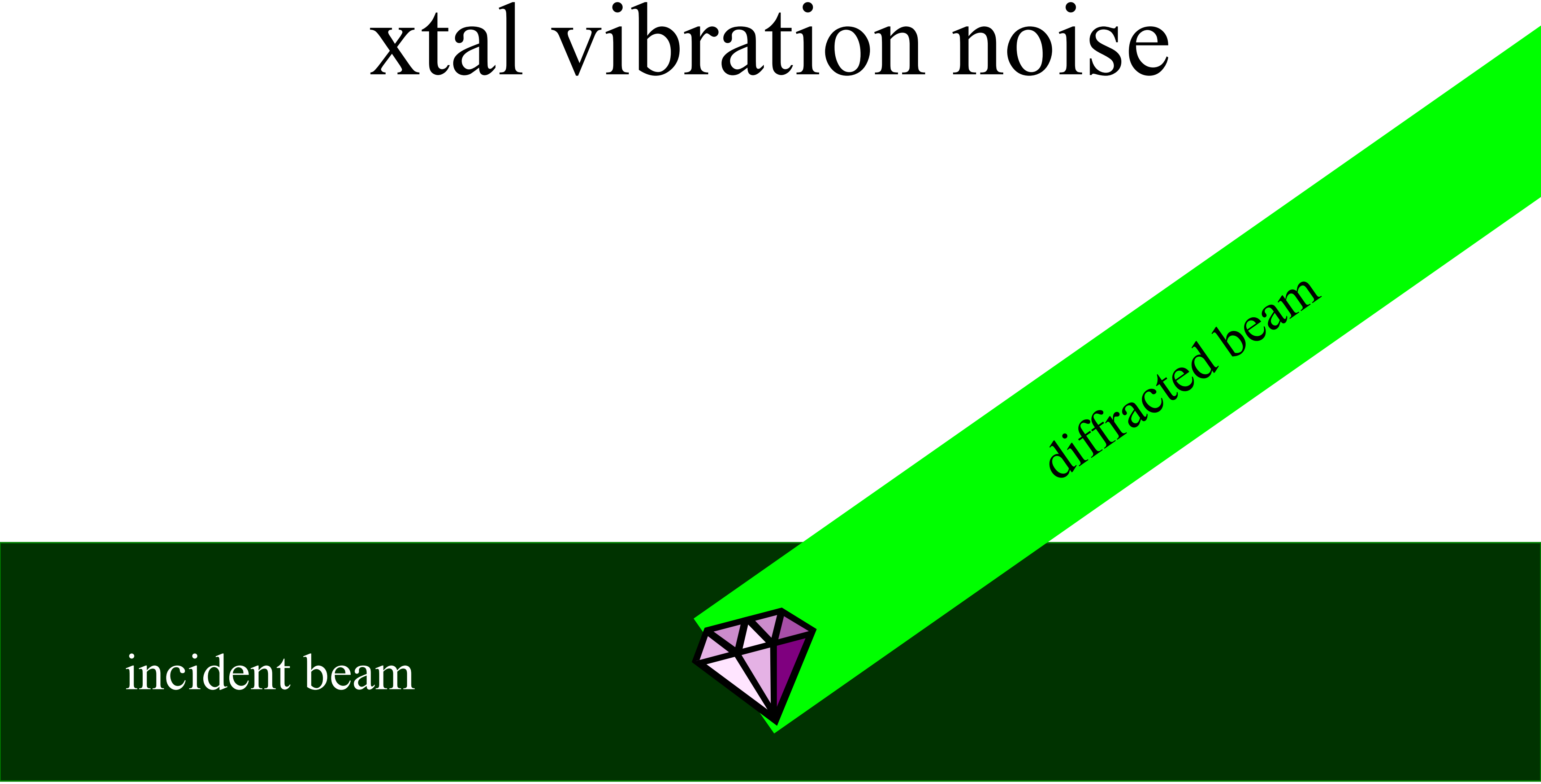
xtal vibration noise



incident beam

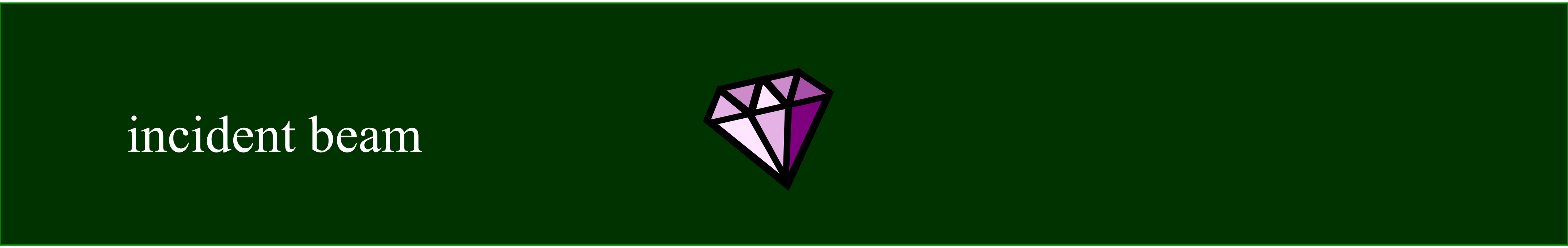
see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

xtal vibration noise



see: Alkire *et al.* (2008)."Is your cold-stream working for you or against you? An in-depth look at temperature and sample motion", *J. Appl. Cryst.* **41**, 1122-1133.

Solution to vibration (beam flicker, shutter jitter, crystal):

attenu-wait!

- reduce flux
- increase exposure time (compensating the flux reduction)

OR: improve the hardware!

Compromises

short exposure time (or strong attenuation) reduces radiation damage and avoids overloads,

long exposure time (little attenuation) improves the signal-to-noise

small oscillation better samples the reflection profiles and reduces background,

large oscillation saves readout time and minimizes the damage from shutter flicker

short crystal-detector distance ensures that even highest resolution recorded,

long distance avoids reflection overlap and increases signal-to-noise

How to find „best“ compromise?

- some parameters are ill-defined because they involve non-proven concepts: what is „high enough completeness“, „too much radiation damage“, „too weak data“, „too low resolution“?
- choice based on past experience – but is it still relevant and up-to-date?
- choice based on strategy programs: BEST
- best: build up your own experience
- in practice, a good experiment is better than a good theory

Features of Pilatus detector

- pixel-array detector (PAD): each pixel is a detector with electronics
- counts each photon as it hits the detector
- can count up to 20 bits (>1.000.000)
- noise-free readout, no intrinsic background
- fast readout (ms)
- Point spread function: one photon affects only one pixel (if the photon hits the detector at right angle)

This gives us freedom!

e.g.
- to adapt the oscillation range to the mosaicity
- to slice the tolerable dose into many low-dose frames such that we obtain more meaningful partially complete datasets from microcrystals or at RT

this changes the rules

- lack of intrinsic and read-out noise improves signal-to-noise ratio
- very low counts (0,1,2,...) are possible: this avoids overloads
- ideal for fine slicing: less background
- enables shutterless (i.e. continuous) data collection: no shutter jitter
- for the same signal-to-noise, one can expose less: this means less radiation damage, higher multiplicity
- multiple passes not required

… and the way we can do the experiment:

Conventional way	*Pilatus way*
“It is important to know where to start and how much to rotate the crystal“	Start anywhere and collect 180° (or more, as long as xtal survives)
Collect 1° oscillation	Collect 0.1° oscillation
Expose such that reflections can be seen visually	Expose weakly and rather increase multiplicity
Multiple passes	Single pass

Conclusions

X-ray data collection (with 2D detectors)

- scientific process, not technicality

- irreversible consequences (often)

- even more important due to progress in automation, phasing, refinement etc.

Always involves a compromise between time, redundancy, completeness etc.

- but it should be a wise compromise

My summary

Technical details can be understood, but questions remain:

How much completeness is enough?
How much radiation damage can be tolerated?
How good do the data have to be, to be able to solve a structure?
Anisotropy - should data be elliptically truncated?

My recommendation: always
1) use a test crystal
2) don't rely on BEST and similar „strategy“ programs. It is better to process data on site and adjust parameters for the next crystal, based on the results. But it's important to look at the right indicators!

Thank you!

If you would like to obtain a PDF of the slides, send email to kay.diederichs@uni-konstanz.de