JOURNAL OF SYNCHROTRON RADIATION

Volume 27 (2020)

Supporting information for article:

Seed-skewness algorithm for X-ray diffraction signal detection in the time-resolved synchrotron Laue photocrystallography

Dariusz Szarejko, Radosław Kamiński, Piotr Łaski and Katarzyna N. Jarzembska

Orientation-matrix-less integration and original seed-skewness method

The concept allowing for data integration without the information on the orientation matrix was introduced by Kalinowski *et al.* (Kalinowski *et al.*, 2012), and originally implemented in the *LAUEUTIL* software, as a part of a wider project dedicated to small-molecule TR X-ray Laue diffraction data processing (Kalinowski *et al.*, 2011; Makal *et al.*, 2011; Coppens & Fournier, 2015). The idea was to treat the detector surface as a set of independent pixels. The signal in each pixel can then be found through the analysis of intensity changes *versus* frame number (Figure 1S). Such a procedure follows the observation that in most of X-ray diffraction experiments large number of frames at different sample orientations are collected. Consequently, the pattern recorded on a detector is different for each frame. Since the X-ray diffraction pattern is discrete in the case of single crystal samples (signals are localised at reflection positions), the majority of collected data (in terms of the available detector surface) contains background.

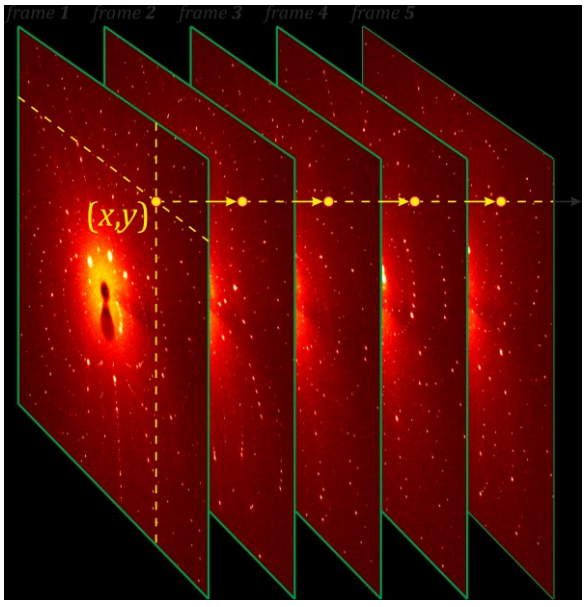

Fig. S1 The idea behind the method: intensity of a selected pixel (x, y) is analysed *vs.* frame index. The procedure is repeated for every pixel on the detector surface.

Kalinowski *et al.* (Kalinowski *et al.*, 2012) presented two methods for signal detection in 1-dimensional sets containing intensity data (the so-called vectors). The first simple approach assumed that a certain percentage (*e.g.* 20-30%) of the highest intensity values contribute to the signal (the constant-fraction method; CF). The second method, the more

sophisticated one, was based on the non-parametric Kruskal-Wallis (KW) statistical test in that matter. While the first method has its obvious drawbacks (see original article for details (Kalinowski *et al.*, 2012)), the second one yields very good results, especially in terms of the estimated background distribution. However, the KW test can be used solely for data sets containing redundant measurements (Kamiński *et al.*, 2010; Makal *et al.*, 2011). Consequently, it performs very well for time-resolved X-ray Laue diffraction data sets, in which light-ON and light-OFF frame pairs are collected multiple times for a given crystal orientation. In other cases, the method is not applicable.

The original seed-skewness (SS) method was published by Bolotovsky *et al.* (Bolotovsky *et al.*, 1995), later extended for taking into account the $K\alpha_1$-$K\alpha_2$ splitting (Bolotovsky & Coppens, 1997), and carefully evaluated (Darovsky & Kezerashvili, 1997). Since the original description of the approach is very comprehensive, only essentials will be provided here. In general, the idea behind the seed-skewness algorithm is that being given the intensity data (regardless of its dimensionality) one can determine whether the signal is present there, or not, by analysing the 3rd moment (*i.e.*, skewness) of the distribution. For example, if there is no signal in the data the skewness should be zero, or in practice, very close to that. In contrary, the presence of a signal significantly elevates the skewness of the distribution, as it contains a certain number of intensity values higher than the background. The skewness of the distribution is defined as follows:

$$\mu_3 = \frac{1}{n}\sum_{k=1}^{n}\Big(I_k - \langle I\rangle\Big)^3 \tag{S1}$$

where the index k runs from 1 to n and

$$\langle I\rangle = \frac{1}{n}\sum_{k=1}^{n} I_k \tag{S2}$$

is the mean intensity value. Assuming that the recorded pixel intensity values follow the Poisson distribution and there is no correlation between them, the standard deviation of the skewness can be expressed as:

$$\sigma(\mu_3) = \frac{3}{n}\left[\sum_{k=1}^{n}\Big(I_k - \langle I\rangle\Big)^4\left(I_k + \frac{\langle I\rangle}{n}\right)\right]^{1/2} \tag{S3}$$

The SS algorithm originally consisted of the following steps: (i) calculation of initial skewness of the selected part of the data containing intensity distribution, and decision on the signal existence in the data, (ii) construction of initial seed, as a set of intensity

values higher than that of the predefined value, (iii) seed growing (*i.e.*, enlarging its area) by moving background pixels adjacent to the seed to the signal set; procedure is monitored by the skewness value minimisation and aborted when the skewness minimum is reached. The use of 2-dimensional algorithm was originally applied to experimental charge-density distribution determination (Iversen *et al.*, 1998; Iversen *et al.*, 1999) and to TR Laue X-ray diffraction data (Kamiński *et al.*, 2010; Makal *et al.*, 2011), whereas a 3-dimensional case was tested on a monochromatic neutron data set (Peters, 2003).

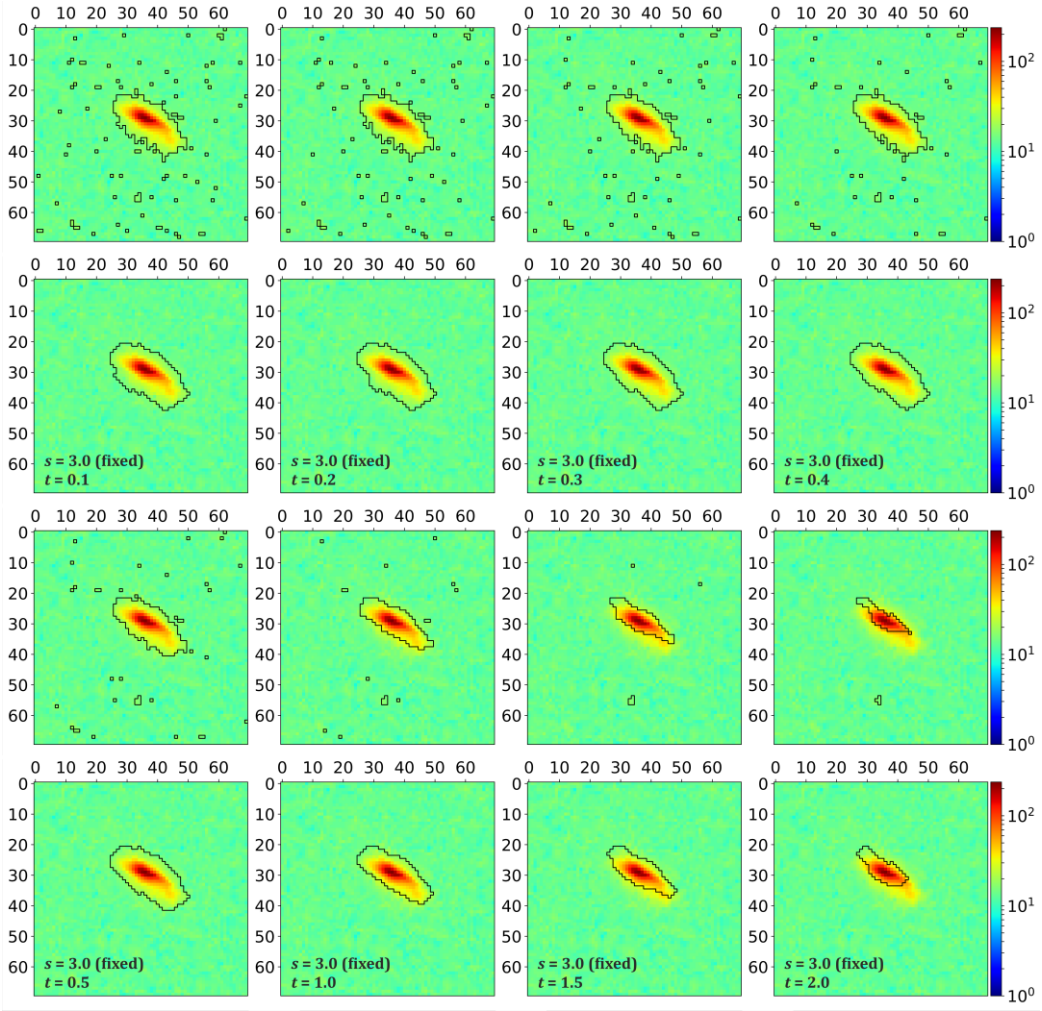

Fig. S2 Masks obtained for the *dark*-type data set with various t-parameter settings (with s = 3.0) for a strong spot (top panels – raw masks, bottom – filtered).

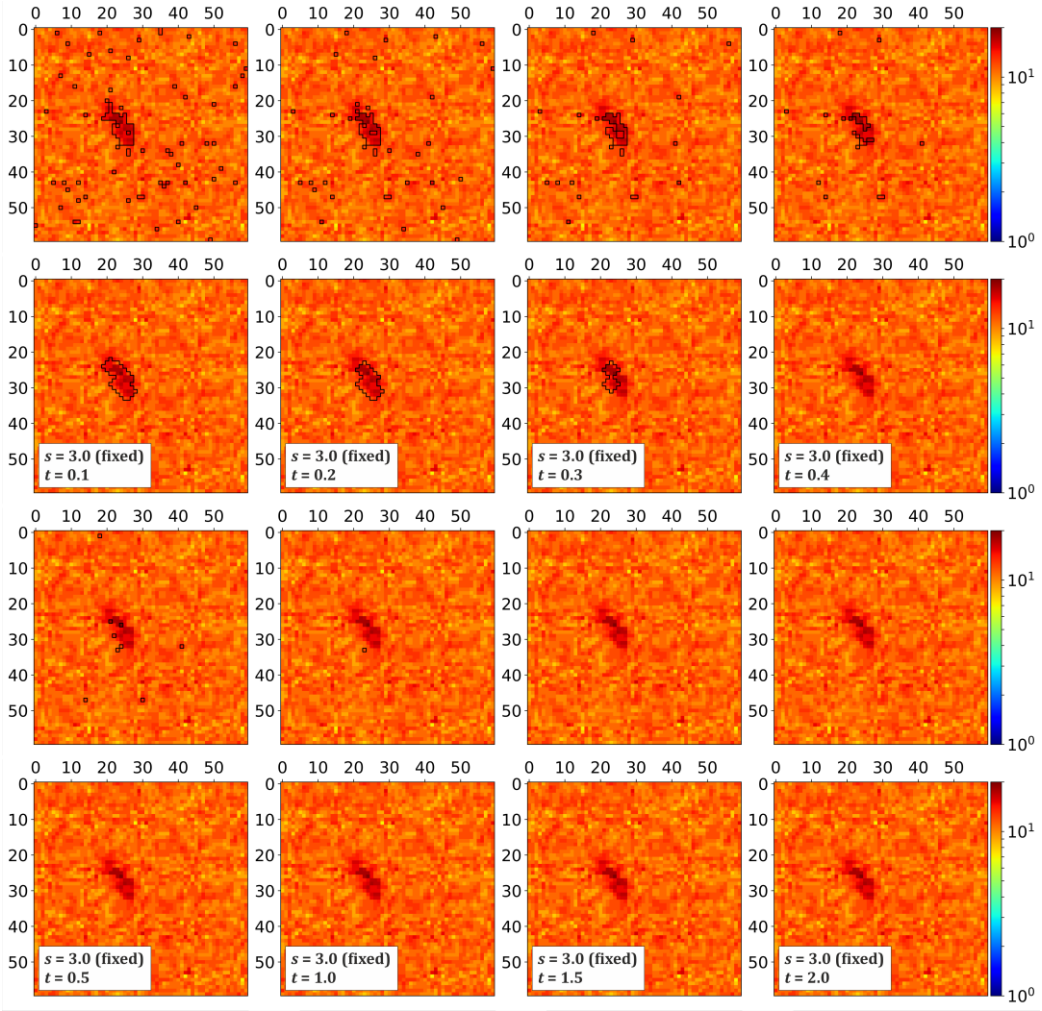

Fig. S3 Masks obtained for the *dark*-type data set with various t-parameter settings (with s = 3.0) for a weak spot (top panels – raw masks, bottom – filtered).

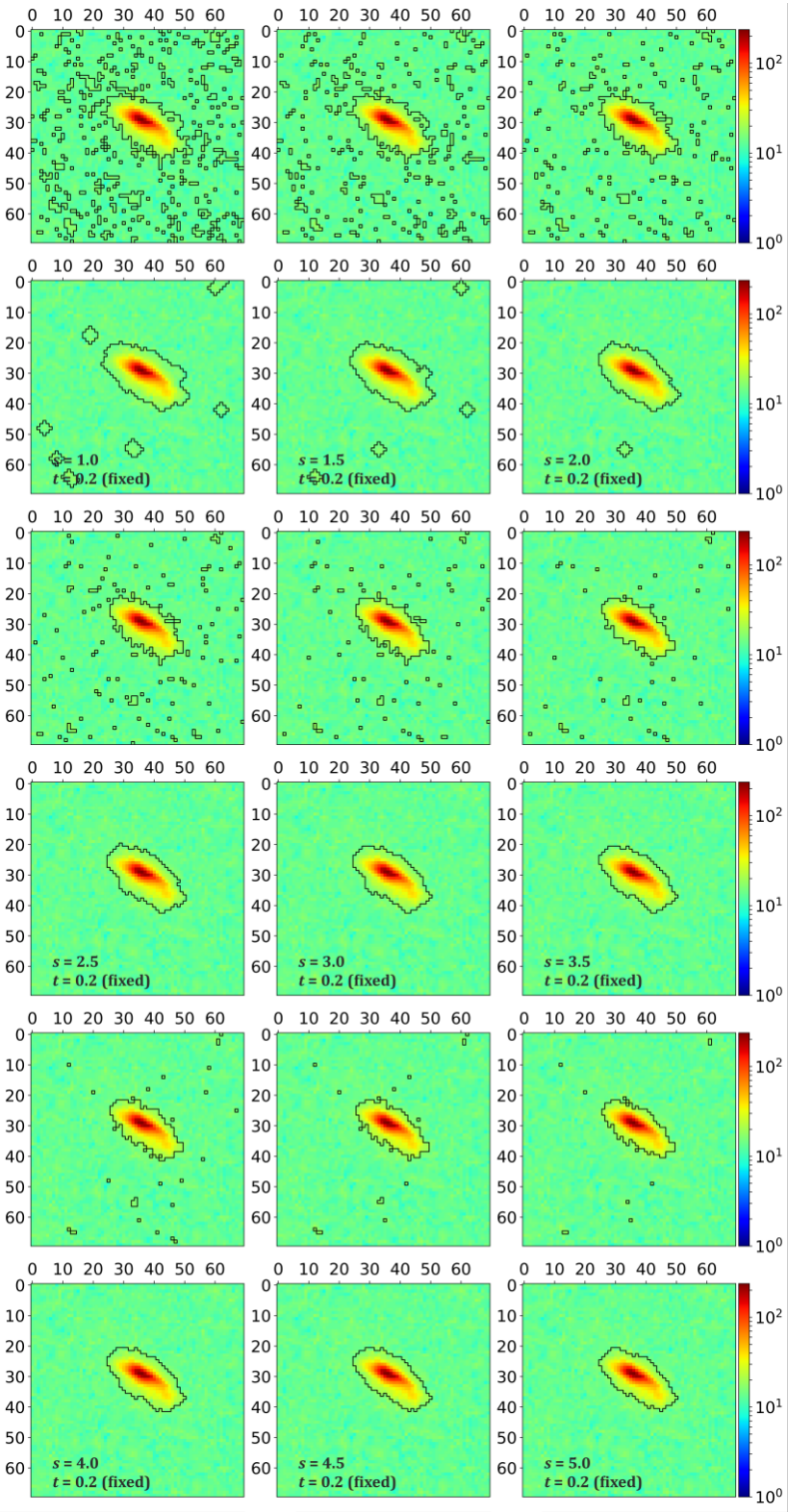

Fig. S4 Masks obtained for the *dark*-type data set with various *s*-parameter settings (with t = 0.2) for a strong spot (top panels – raw masks, bottom – filtered).

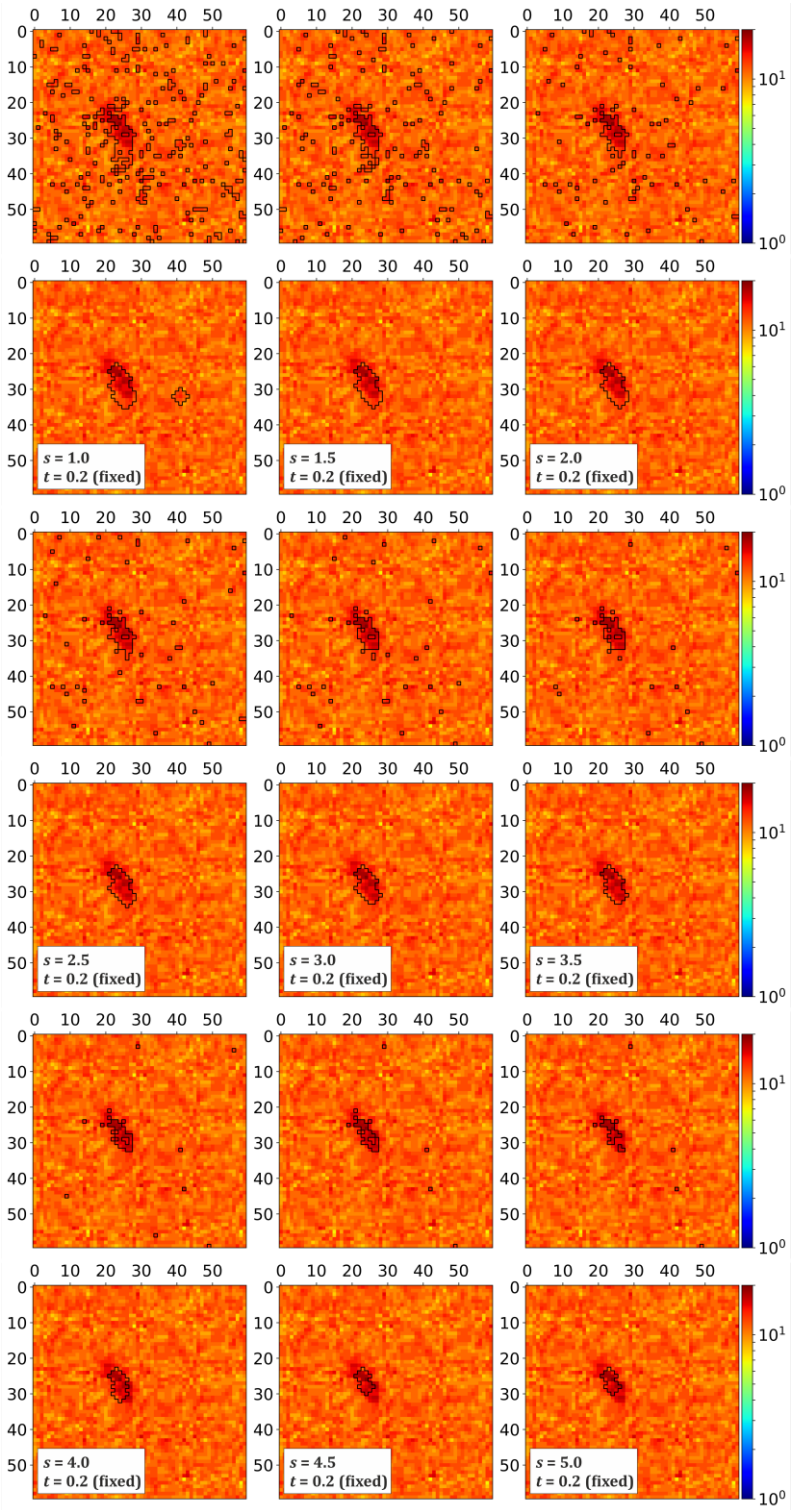

Fig. S5 Masks obtained for the *dark*-type data set with various *s*-parameter settings (with t = 0.2) for a weak spot (top panels – raw masks, bottom – filtered).

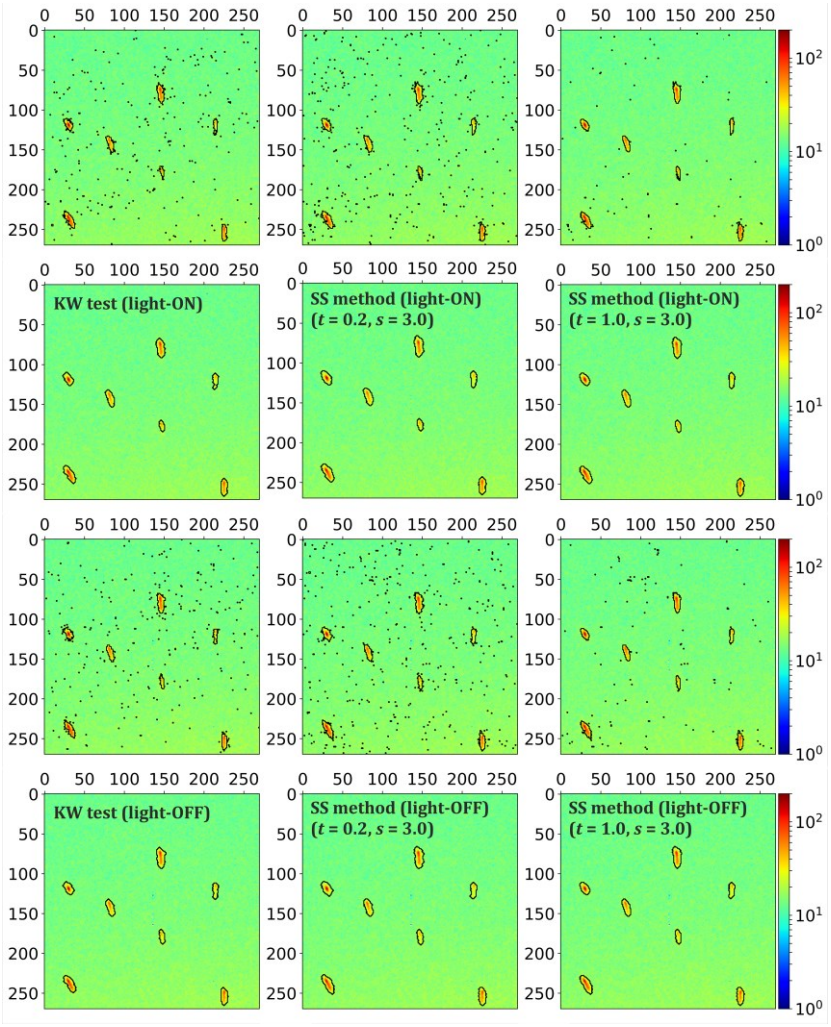

Fig. S6 Raw (top) and filtered (bottom) 2-dimensional masks obtained for a fragment of a selected frame based on the analysis of full TR Laue *laser*-type data set from ESRF; two methods were tested: KW test and SS method with two different t-parameter settings. Masks' outline is marked in black.

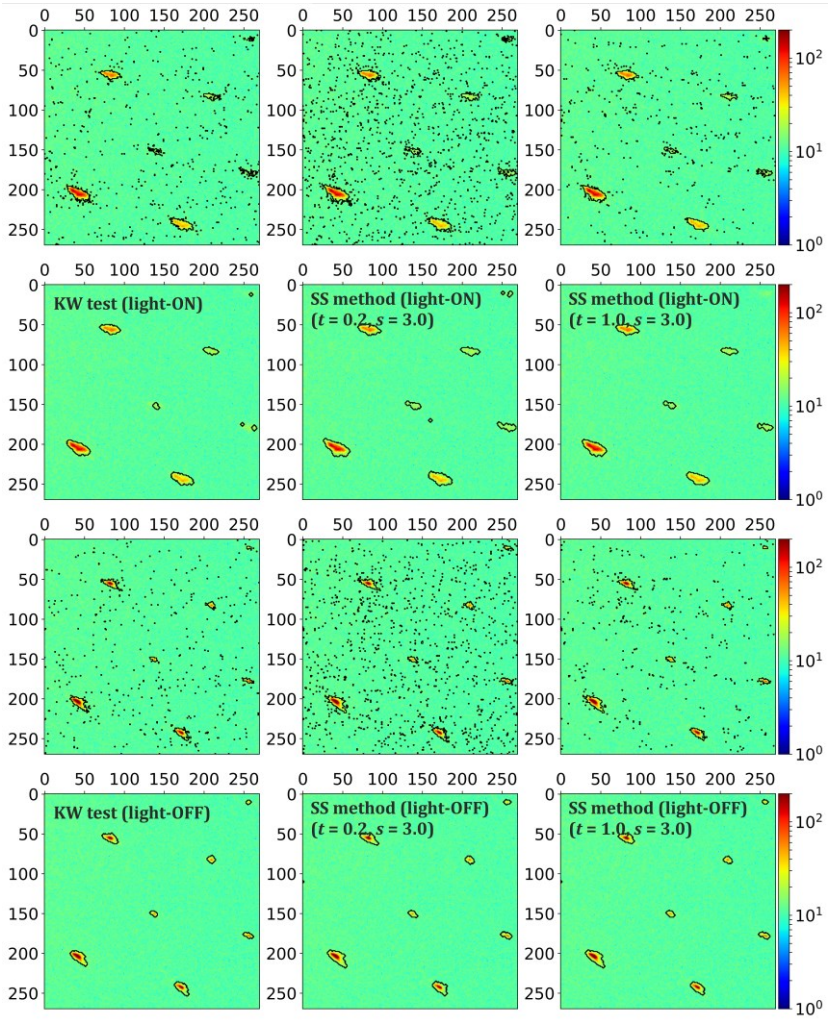

Fig. S7 Raw (top) and filtered (bottom) 2-dimensional masks obtained for a fragment of a selected frame based on the analysis of full TR Laue *laser*-type data set from APS; two methods were tested: KW test and SS method with two different t-parameter settings. Masks' outline is marked in black.

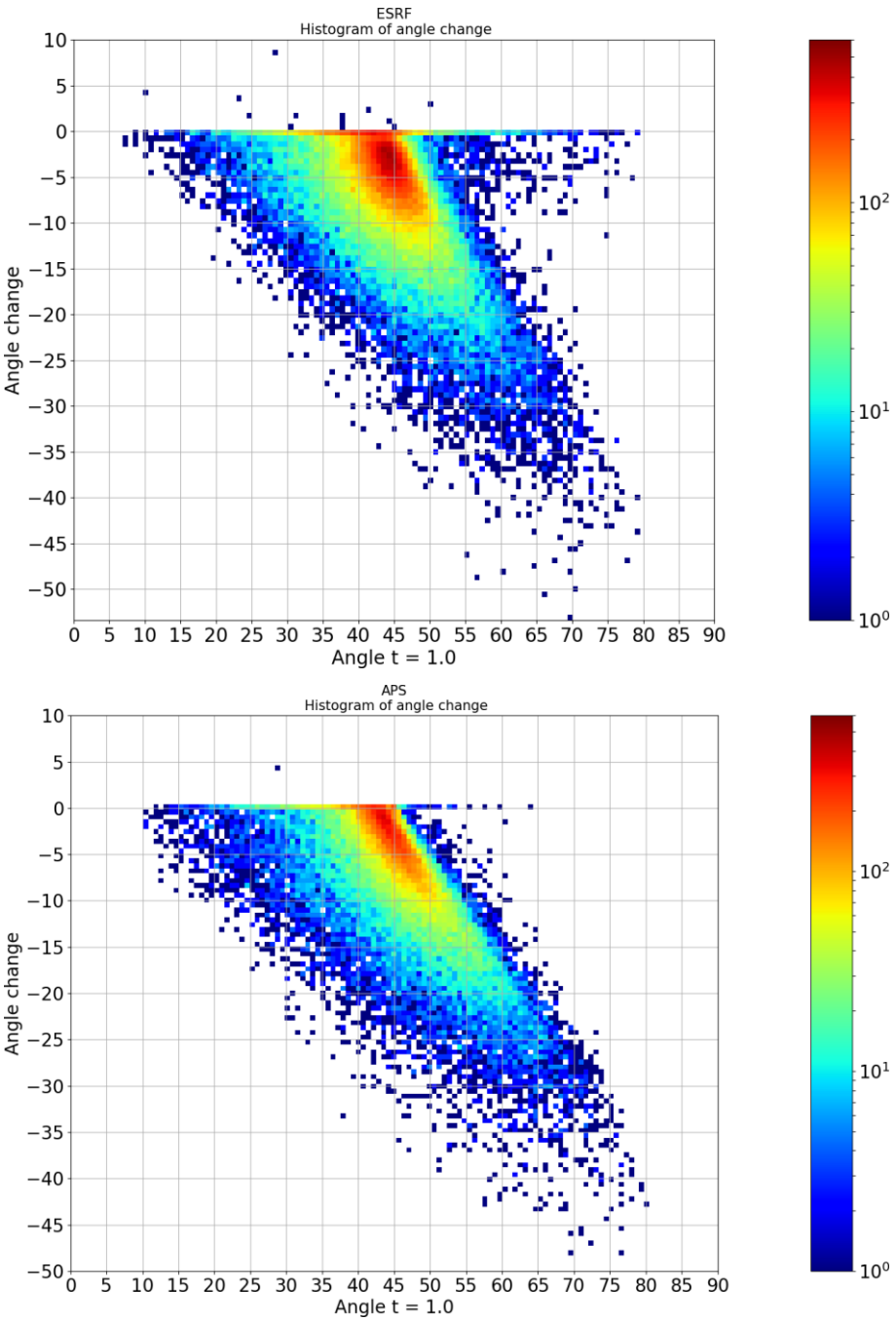

Fig. S8 Histograms indicating the relationship of the spot angle (in degrees) on the correlation plot in Figure 6 from the SS-processing with t = 1.0 (horizontal axis), and its change when the SS t parameter is set to 0.2 (vertical axis). Colours indicate the density of spots. Top panel – ESRF data set, bottom panel – APS data set. Standard (counterclockwise) mathematical definition of the angle is used in respect to the origin of the correlation plot.

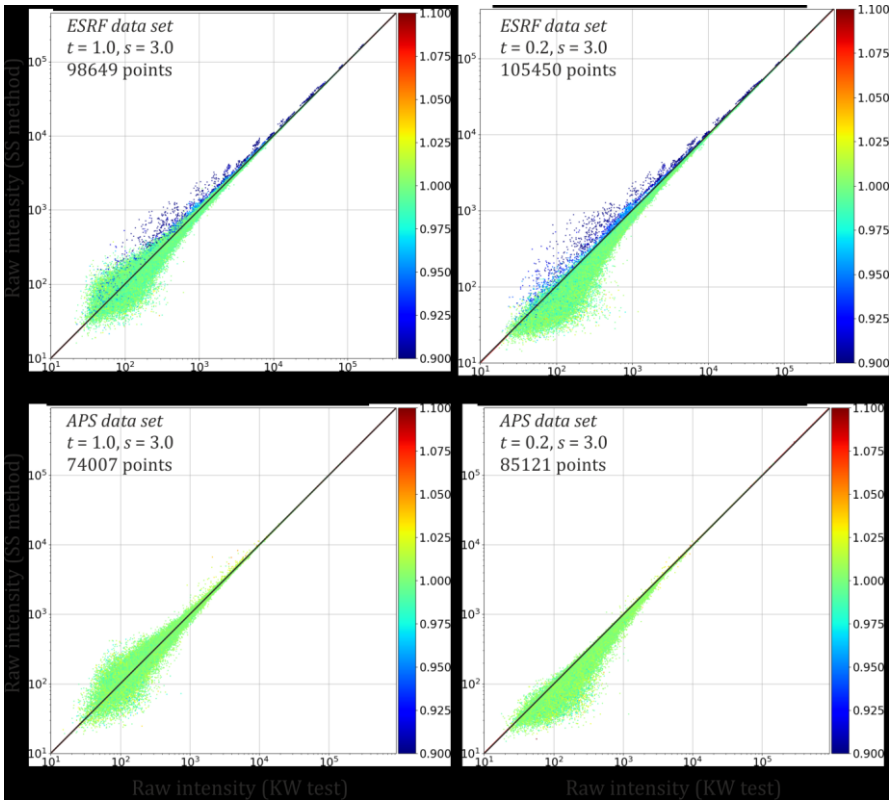

Fig. S9 Correlation plots of the SS-integrated spot intensities (vertical axes) *vs.* KW-integrated ones (horizontal axes). Two SS integration schemes were used with t = 1.0 (left panels) and t = 0.2 (right panels) for two *laser*-type data sets: from ESRF (top panels) and APS (bottom panels). Plots are coloured with mean reflection background ratios (SS- over KW-reflection mean background ratios), $\langle B\rangle_{l,\mathrm{SS}}/\langle B\rangle_{l,\mathrm{KW}}$ where for l-th reflection $\langle B\rangle_l = \frac{1}{n_l}\sum_{(i,j)}\langle B\rangle_{ij}$, n_l – number of pixels for a given reflection, $\langle B\rangle_{ij}$ – mean background in a given pixel (i,j), summation goes over reflection area; intensities are given in ADU; logarithmic scale is used; black solid straight line is the reference line of $I_{\mathrm{SS}} = I_{\mathrm{KW}}$.

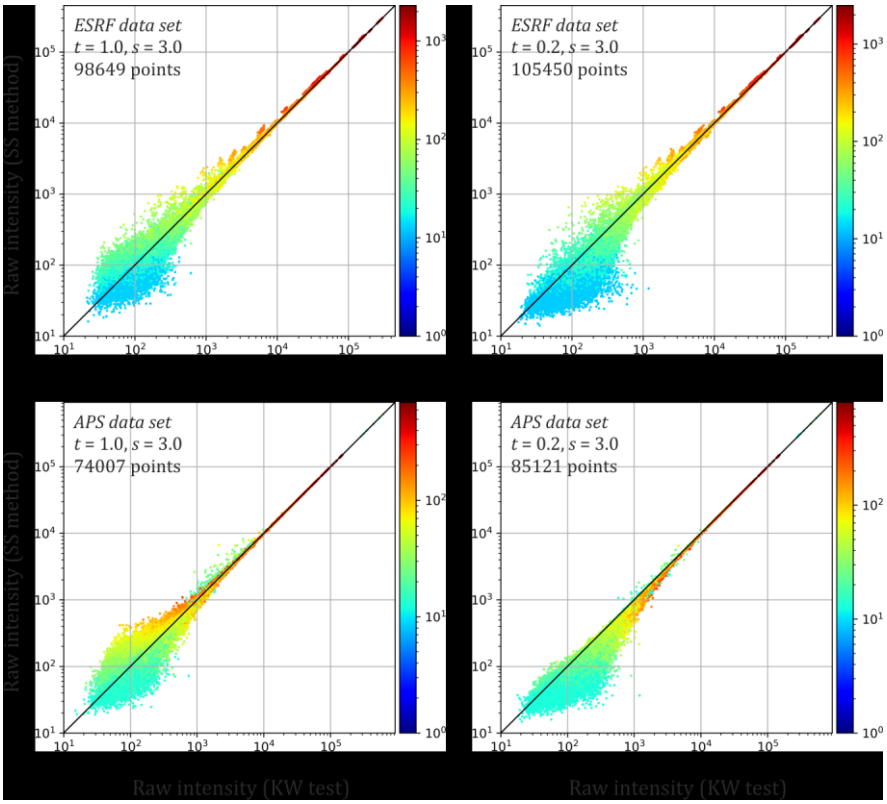

Fig. S10 Correlation plots of the SS-integrated spot intensities (vertical axes) *vs.* KW-integrated ones (horizontal axes). Two SS integration schemes were used with t = 1.0 (left panels) and t = 0.2 (right panels) for two *laser*-type data sets: from ESRF (top panels) and APS (bottom panels). Plots are coloured with reflection areas (given in pixels); intensities are given in ADU; logarithmic scale is used; black solid straight line is the reference line of $I_{SS} = I_{KW}$.

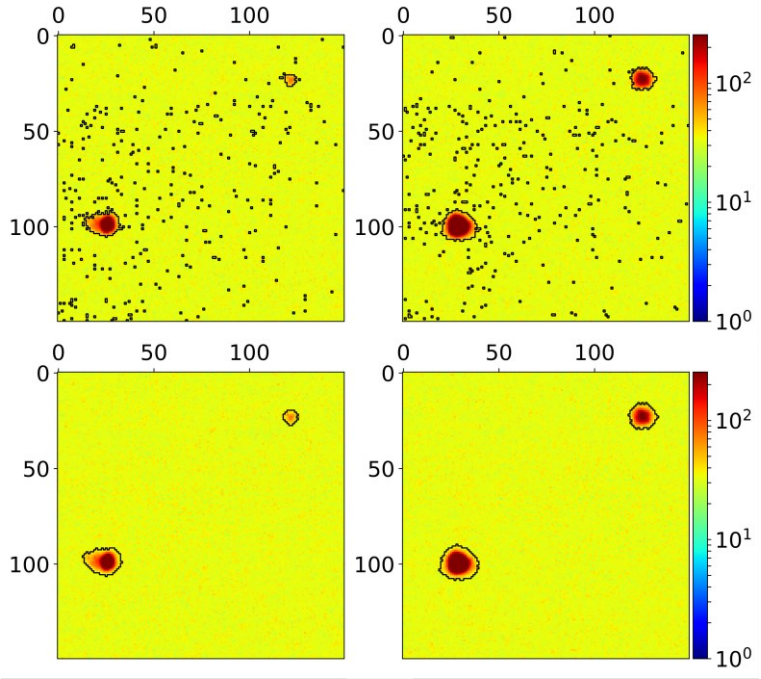

Fig. S11 Raw (top) and filtered (bottom) 2-dimensional masks obtained for selected fragments of two consecutive frames (left & right) for monochromatic in-house data for the K(UMP) single crystal (SS algorithm with t = 0.2 and s = 3.0). Masks' outlines are marked in black; all panels are drawn using the same intensity scale.

Table S1. Extended Table 4 with selected structural and refinement parameters computed for data integrated with three different methods (see Table 3).

Parameter	*KW test*	*SS method* (t = 1.0, s = 3.0)	*SS method* (t = 0.2, s = 3.0)	*QM/MM results*
$d_{\mathrm{Ag1\cdots Ag2}}$ / Å	2.65(2)	2.59(2)	2.70(2)	2.749
$d_{\mathrm{Ag1\cdots Cu2}}$ / Å	2.85(2)	2.80(3)	2.95(2)	3.114
$d_{\mathrm{Ag2\cdots Cu2}}$ / Å	2.69(2)	2.66(3)	2.71(2)	2.822
Pop.% (set no. 1)	0.47(4)	0.37(3)	0.54(4)	–
Pop.% (set no. 2)	1.11(6)	0.87(4)	1.21(7)	–
Pop.% (set no. 3)	0.77(4)	0.61(3)	0.89(5)	–
Pop.% (set no. 4)	0.95(5)	0.75(4)	1.04(6)	–
$R[R]$ (set no. 1)	1.70%	1.81%	1.70%	
No. of reflections (set no. 1)	2650	2831	2727	
$R[R]$ (set no. 2)	2.25%	2.36%	2.36%	
No. of reflections (set no. 2)	2354	2714	2589	
$R[R]$ (set no. 3)	2.78%	2.91%	2.72%	
No. of reflections (set no. 3)	3982	4528	4217	
$R[R]$ (set no. 4)	2.47%	2.54%	2.56%	
No. of reflections (set no. 4)	2811	3123	2975	
$R[R]$ (all data)	2.34%	2.46%	2.37%	–
No. of reflections (all data)	11797	13196	12508	–

R-factor calculated on intensity ratios (Coppens *et al.*, 2010): $R[R] = \sum_i \left| R_{\mathrm{o},i} - R_{\mathrm{c},i} \right| / \sum_i R_{\mathrm{o},i}$. QM/MM results are taken from the original reference of Jarzembska *et al.* (Jarzembska *et al.*, 2014).