Journal of
Materials Sciences and Applications

JScholar
Fostering Scholarly Communication

Research Article Open Access

Statistical Noise-Filtering of the X-Ray Diffraction Tomography Data: Simulation and Analysis

Chukhovskii FN*, Konarev PV, Volkov VV

A. V. Shubnikov Institute of Crystallography, Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Leninsky pr. 59, 119333 Moscow, Russian Federation

***Corresponding author:** Chukhovskii FN, A. V. Shubnikov Institute of Crystallography, Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Leninsky pr. 59, 119333 Moscow, Russian Federation, Tel: 79161792286, Email: f_chukhov@yahoo.ca

Received Date: May 23, 2022 **Accepted Date:** June 22, 2022 **Published Date:** Jun 24, 2022

Citation: Chukhovskii FN, Konarev PV, Volkov VV (2022) Statistical Noise-Filtering of the X-Ray Diffraction Tomography Data: Simulation and Analysis. J Mater sci Appl 6: 1-12

Abstract

Based on statistical analysis of generating the Gaussian-distributed pseudorandom noise data beingcome out under the experimental X-ray diffraction tomography technique, the noise-filtering issue of the 2D image patterns is analyzed. The noisy 2D image data of the crystal-lattice point defect isnumerically simulated by using the random number generator algorithm. Accordingly, to denoising,one applies the additive acquisition network of the 2D noisy data. The special attention is paid for cumulative denoising the 2D image data governed by the goodness-of-fit parameter of the recoveryof the nanoscale point defect in a crystal. In terms of the signal-to-noise ratio, improving the qualityof the 2D image pattern denoised by fusing a series of the 2D images into one has been approved.

Keywords: X-ray diffraction tomography, Coulomb-type point defects, statistical noise filtering,signal-to-noise ratio

Introduction

Substantial success in non-destructive studying the crystal-lattice defects properties is deeply connected with applying the high-resolution X-ray diffractometry (XRD), the so-called reciprocal- space mapping (RSM) method [1-4] and relevant X-ray diffraction technique (XRD) technique [5- 9] (see references in [5-9] as well). Schematic setup of the XRDT is shown in Fig. 1. The key elements of the XRDT method are the precision rotation of the sample around the diffraction vector ***h*** (angle Φ in Fig. 1), as well as the noise filtering to the reference 2D imaging pattern data recorded via the CCD detector (D) with the acquisition system AS/PC (Fig. 1). The latter collect the 2D image pictures (IPs) information recorded in the reciprocal and real space, respectively, and it is utilized to decode the 2D IP data aiming the quantitative diagnostics of nanoscaled bulk materials. Using the direct methods to decode the experimental X-ray IPs data, recovering of elastic displacement fields around the crystal-lattice defects has been carried out in [5-9].

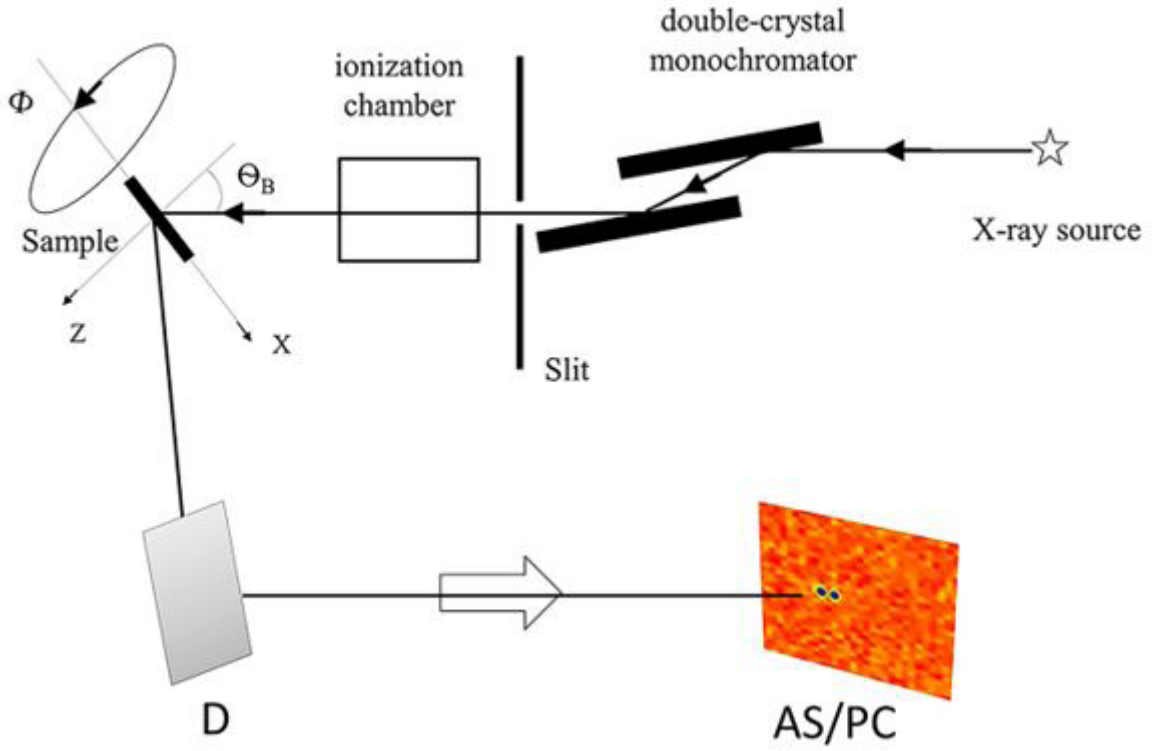

Figure 1. Schematic setup of the XRDT. Sample is rotated around the diffraction vector $h \parallel sin0X$; Φ is the rotation angle, θ is the Bragg angle. D is the CCD detector; AS/PC is the acquisitionsystem for accumulating of the 2D IP frames

In practice, remaining in the modern issues frame of the materials science, one can state that plausible decoding the crystal-lattice-defects from the experimental X-ray IPs data depends on the two factors. The first is the mathematically correct solution of the inverse issue of the X-ray diffraction tomography. The second is obtaining the experimental data with good accuracy, if possible, excluding and or minimizing experimental measurement errors. In other words, it is important to use experimental data with an accuracy controlled. This means that for the experimental measurement it is necessary to achieve a level of noise that would allow reliable, at a level of significance not more than 10%, recognizing the elements of the image. Since these elements can have different sizes and shapes, it is difficult to give a general maximum limit for the noise level in all cases, we can only say that it should not exceed 10-20%. In [9], the inverse problem of the X- ray diffraction tomography has been discussed in the case of the simulated noisy 2D IPs data related to the Coulomb-types point defect in a crystal. For this, as the recovery criterion, the goodness-of- fit parameter (CP) strongly depending on the noise level of the 2D IPs data has been introduced.

In [10], authors have undertaken the endeavour to filtering the simulated noisy 2D IPs data for the Coulomb-types point defect in a crystal. It has been shown that the governed filter applied to the initial noise-levels of the order 3-10 % IPs data yields the noise levels reduction of the order of their values at

least, and significantly decreases the goodness-of-fit CPs values of the optimization procedure.

Noteworthy is the fact, in [11-13] have been pointed out a technical capacity of filtering digital signals, and quantifying the signal and noise components to improve the signal-to-noise ratio. In [12] was shown how the noise reduction by the signals average can be achieved with acquisition system of the noise signal data. Recently [14] have shown that convolutional neural networks can be a powerful tool to remove noise from reconstructed 2D images, Concerning the noise 2D IPs, hereafter the term "filtering" does mean the procedure, which leads to an improvement in the signal- to-noise ratio independent of whether the filtering algorithms are used at the post-processing stage of collected images or statistical average by combining a set of the 2D IP frames.

The 2D IPs data have been numerically simulated in the case of the incident σ-polarized plane–wave X-radiation. As is known, the Gaussian noise satisfies to the probability density function (PDF)

$$G(y,[\Lambda]) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}, \qquad (1)$$

where the dispersion r.m.s. is nothing else the noise level σ.There are some works [15-17], in which the powerful PC algorithms have been employed for modeling the noise 2D IPs data collected by the CCD area detectors in the XRD technique. All the above PC random-noise-generation algorithms are based on the inverse transform

$$\{y|n\} = inv\{\Phi(y)|n\},\ \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} d\,te^{-\frac{t^2}{2}}. \qquad (2)$$

Here Φ(y) is the PDF integral. Function Φ(y) alternates within the interval U(0, 1) with the r.m.s. Λ equal to unity; n is the current 2D IP-frame number, n = {1, 2, …, N}, N is the number of the 2D IP-frames; to be specific, the number N is large in a statistical sense.

Due to the *inverse* transform (2), the pseudorandom number set {y_k} in some interval of U(−Δ, Δ) can be generated according into some reasonable looking Gaussian function. The integer k is the pixel number of one 2D IP-frame, k = {1, 2, K}, K is the number of pixels in one 2D IP-frame. Under the Central Limit Theorem (CLT) of statistics, in the limit when $\widetilde{N} \equiv (N \times K) \to \infty$, the PDF $G_{\tilde{N}}(y_k)$ tends to the Gaussian distribution G(y_k),

$$G(y_k) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{y_k^2}{2}\right]. \qquad (3)$$

Gaussian noise, unlike Poisson's noise, can already be considered to be an admixture in the reference data.

In the present study, for simplicity, without loss of generality, we will treat an issue of the 2D noise IPs data frames in the case of frame- and pixel-independent random noise.

Appropriately, one assumes that the noisy n-th IP frame is determined by the relationship

$$\{I(k,n)\} = \{I_0(k)[1 + \sigma(y_k|n)]\}, \qquad (4)$$

Before proceeding further, it is worth to notice that there are methods to generate the Gaussian pseudorandom numbers, which do not rely on inverse transform (2). For example, good statistical properties have the maximal entropy Wallace's method [18] and the Wolfram Mathematica program package [19]. To generate the Gaussian pseudorandom numbers {y_k}, we implemented both the method based on the inverse transform (2) (see [15-17]) and the Wolfram Mathematica program package [19] when they fit our purposes. Following a general idea of reducing the signal-to-noise ratio due to the signal accumulation with the subsequent averaging [11-13], one considers the average of the noise $\{I(y_k, n)\}$ supposing that the regular IP-component $I_0(k, n)$ does not depend on the frame number n. Then, by using equation (4) averaged over N images, we have

$$\langle\{I(y_k,n)\}\rangle_N = I_0(k)\left(1 + \sigma\frac{1}{N}\sum_{n=1}^{N}\{y_k|n\}\right), \qquad (5)$$

and according to a general statistics concept, the noise part in the averaged frame $\langle\{I(y_k, n)\rangle_N$ decreases proportionally to $\sqrt{N}$ (see, *e.g.*, [11-13]).

In this paper, we will learn how the Gaussian statistics works on an example of the XRDT noise 2D IPs of the crystal lattice Coulomb-types defect with the 3D displacement field function $f_{Ctpd}(\mathbf{r} - \mathbf{r}_0)$ under conditions of the X-ray diffraction tomography (XRDT), as described above. Special attention is paid for cumulative reducing of the noise distortion of the simulated XRDT 2D IPs onto the effective recovery of the 3D displacement field function $f_{Ctpd}(\mathbf{r} - \mathbf{r}_0)$ (*cf.* [9]).

Hereafter, without losing a generality, the denoising problem will be considered on the example of the Gaussian noisy image pictures that is commonly taken place in the case of the X-ray diffraction tomography (XRDT). In this paper, denoising of the reference XRDT data is discussed and analyzed in the case of the 2D noisy IPs of spherical inclusion incorporated within a single crystal Si(111);

the diffraction vector h = [220], an incident linear-polarized X-ray radiation with wavelength λ = 0.0709 nm, the Bragg angle θB = 10.65°, extinction length Λ = 36.287 μm. The spherical inclusion is considered as the Coulomb-types point defect located at a point r0 with the 3D displacement field function $f_{Ctpd}(\mathbf{r}-\mathbf{r}_0)=\frac{F}{4\pi}\frac{\boldsymbol{h}(\mathbf{r}-\mathbf{r}_0)}{|\mathbf{r}-\mathbf{r}_0|^3}$, $F \cong 0.064\mu m^3$ [9].

Hereafter, the sample thickness T is chosen to be equal to the X-ray extinction length Λ, each the simulated reference 2D IP frame of the square size = Λ^2 contains of 3721 pixels. Accordingly, the linear pixel size is about 0.6 μm. Note that the linear spatial resolution of the X-ray synchrotron 2-20 keV hybrid pixel detectors used in the X-ray diffraction tomography is of the order of about 1μm. The goal of the study is to denoise the reference 2D Gaussian noise-type IPs data and to improve the quality of the digital reconstruction of the 3D function $f_{Ctpd}(\mathbf{r}-\mathbf{r}_0)$ digitally recovered.

In spite of the fact that the noise in the images obtained with CCD detectors has a Poisson distribution by nature, at high counting statistics the noise distribution is close to the Gaussian one, the case of which we will consider.

Random generation algorithm for the generating the Gaussian pseudorandom numbers

The computer algorithm code for generating the Gaussian pseudorandom numbers is employed according to the schemes, details of which may be found in the [15-19]. The results of the noise generation have been approved by applying a number of statistics criteria including the corresponding test-statistics values listed in Table 1 (*cf.* The R-Project for Statistical Computing. https://www.r-project.org and [20-27]). The *p*-value is widely used for testing the statistics hypothesis, specifically: to verify the null hypothesis (Statistics Central Limit Theorem – SCLT) that the pseudorandom numbers generated satisfy to the Gaussian distribution. If some model (*e.g.*, the null hypothesis) has been chosen, one needs to establish some significance-level α of the test statistics as the cutoff one. In the statistics data analysis, the conventional cutoff value α is taken as α = 0.05. Should the *p*-value of the test-statistics is less than 0.05, it means that the pseudorandom number data are inconsistent with the null hypothesis and therefore, the null hypothesis has to be rejected. Typically, the null hypothesis is rejected if $p < 0.05$ and accepted for $p > 0.05$.

Table 1. Test-statistics for generating pseudorandom numbers by using the RNGA algorithm code along with the tabulated PDF integral. The program R-package has been used for the various test- statistics models (The R Project for Statistical Computing. https://www.r-project.org)

events number, 961×10^N	Mean, $\times 10^{33}$	dispersion, r.m.s.	test-statistics values of P, D , A, W& the corresponding (*p*-values)			
			Pearson's,P, *p*	Kolmogorov-Smirnov's, D, *p*	Andreson-Darling's,A, *p*	Cramer - von Mises's,W, *p*
0	35	1.0476	39.68,	0.035335,	0.17516,	0.024519,
			(0.08931)	(0.1645)	(0.9244)	(0.916)
1	4.1	1.0164	82.832,	0.0077965,	0.59117,	0.096006,
			(0.3043)	(0.5776)	(0.12311)	(0.1277)
2	-0.89	1.0010	180.51,	0.0017697,	0.39168,	0.060894,
			(0.8085)	(0.9128)	(0.3791)	(0.3682)
3	0.44	1.0006	510.41,	0.0007554,	0.34016,	0.041393,
			(0.3639)	(0.618)	(0.49781)	(0.6561)
4	-0.11	1.0001	1215.2,	0.0002507,	0.34303,	0.056874,
			(0.8077)	(0.5559)	(0.4904)	(0.416)

The generated Gaussian pseudorandom numbers has been successfully passed through Pearson's test-statistics, see Table 2 [21]. Indeed, in all the cases the evaluated *p*-values exceed the level cutoff α =0.05. At the same time, it should be mentioned that the PC algorithm code elaborated is rather fast and takes the 1-2 sec to terminate the PC calculation procedure on the Intel Core i7-7820X processor.

Pearson's χ^2 test-statistics approves a null hypothesis by declaring whether the frequency distribution of the generated pseudorandom numbers is in a good consistence with the Gaussian distribution. The well-known χ_K^2-cumulative test-statistics function is determined as

$$P = \chi_K^2 = \sum_{k=1}^{K} \frac{(G_N(y_k) - G(y_k))^2}{G(y_k)}, \qquad (6)$$

where the $G_N(y_k)$ is the generated frequency distribution due to the total event number N and the $G(y_k)$ is the pure Gaussian distribution function; (*K*-1) is the intervals number of the frequency distribution histogram (see Supplementary part for details).

Furthermore, the χ_K^2 -cumulative test-statistics may be used to evaluate the value *p*, which allows to formulate the "goodness-of-fit" criterion for generating the Gaussian pseudorandom numbers by the frame of stochastic processes. And for which, the freedom-degrees number is equal to the total events number N minus the reduction coefficient *p* = s+1, where s is the number used in fitting the desired frequency distribution, for the Gaussian distribution $G(y_K)$ s=2.

Table 2. Erwartungwerten noise values $M_{\tilde{N}}[y_K]$ and Pearson's test-statistics P. String-number ñ s equal to 961×10^n, n = {2, 3}.

n	2	3
$M_{\tilde{N}}[y_K]$	0.00082090	0.00060373
P	82.832	180.51

Simulation and analysis of the noise-contamination-prone 2D IPs

Nowadays, as is shown in Fig.1, the XRDT technique allows one to collect and proceed the 2D IPs data $\{I(y_k, n)\}$ in a relevant way that contain the $\{y_k|n\}$-ensemble data. Indeed, each the *n*th frame 2D IP is nothing else the linear hybridizing of both the regular (none noise) and noise-contaminated components according to equation (5).

In terms of equation (5), the noise 2D IP frames have been calculated by using the Wolfram Mathematica program package [19] and the random number generator algorithm (RNGA, details of which are given in [15-17]), respectively. By using these algorithms for various string numbers of stochastic events, some distributions $G_{\tilde{N}}(y_k)$ of pseudorandom numbers $\{y_k\}$ have been calculated for the integer $\tilde{N} = K \times N$, *K*=961 pixels per image, *N* is the number of the 2D IP-frames under consideration. The corresponding results are presented in Fig. 2. From Fig. 2, it follows that for $N = 10^3$ the generated distribution $G_N(y_k)$ tends to the Gaussian distribution $G(y_k)$ in a good agreement with the SCLT up to high accuracy.

To be specific, to highlight a goal of the present study, we are confined ourselves by consideration of the XRDT noise 2D IP-frames referred to the crystal-lattice Coulomb-types defect when rotation angle $\Phi = 0^o$. In the case of $\Phi = 0^o$, the numerically simulated 2D IP-frames for various noise levels σ are depicted in Figs. 3a-c, particularly: (a) σ=0, (b) σ=0.001, (c) σ =0.001, respectively. Accordingly, by using the Wolfram Mathematica program package [19] for generating the pseudorandom numbers$\{y_k\}$, the straightforward procedure to averaging the noise-contamination- prone 2D IP frames has been carried out. For comparison, the corresponding no averaged 2D IP- frame with *K*= 961 and *N* = unity and the averaged 2D IP-frame with *K*= 961 and *N*=961 have been simulated. The corresponding results are shown in Fig.4, 4a, 4b, the noise level σ=0.001.

Let us introduce into consideration the Erwartungwert noise gain (*NoiseGain)* determined as the ratio of average *Mean* $N[y_k] = \frac{1}{K}\sum_{k=1}^{K} \text{Mean}_{N=1}\,[y_k]$ to average $\text{Mean}_{\tilde{N}}[Y_K] = \frac{1}{K}\sum_{k=1}^{K} Mean_{N=961}\,[y_k]$.

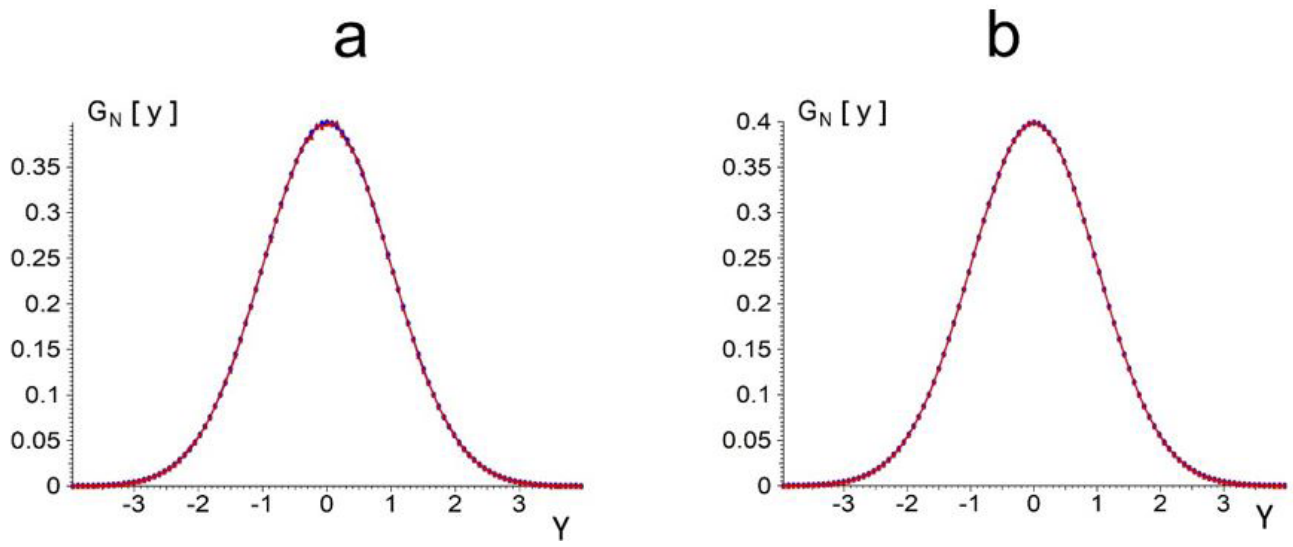

Figure 2: Generating the pseudorandom numbers *y* (red curve) *vs.* the normal distribution PDF (bluecurve) for the string numbers $\widetilde{N} = 961 \times 10^n$ for *n* equal to: a) 2, b) 3

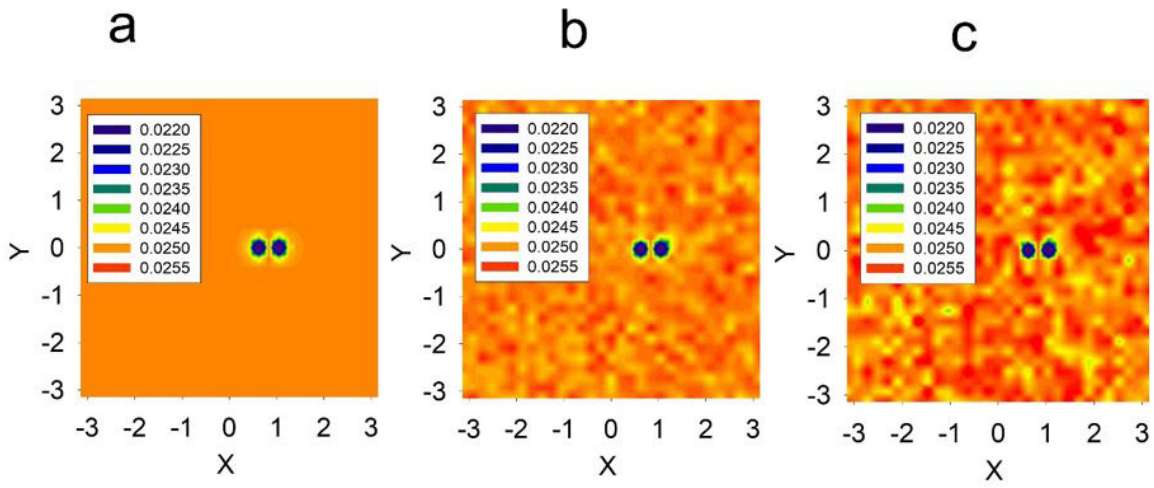

Figure 3. The XRDT 2D IPs with noise level σ equal to: (*a*) 0, (*b*) 0.01 (*c*) 0.016. The sample rotation angle $\Phi = 0°$. The total grid sizes in the 'own' crystal coordinate system (*X*, *Y*, *Z*) are equalto (61, 61, 21). Sizes of the sample voxel {ΔX, Δ*Y*, Δ*Z*}: Δ*X*= π/30, Δ*Y*= π/30, Δ*Z*= π/20, respectively. The vector $P^{(true)}$ of the 3D Coulomb-type point defect displacement function $f_{Ctpd}(r\text{-}r_0)$ assumed to be {1.50, 0.50, 1.80}

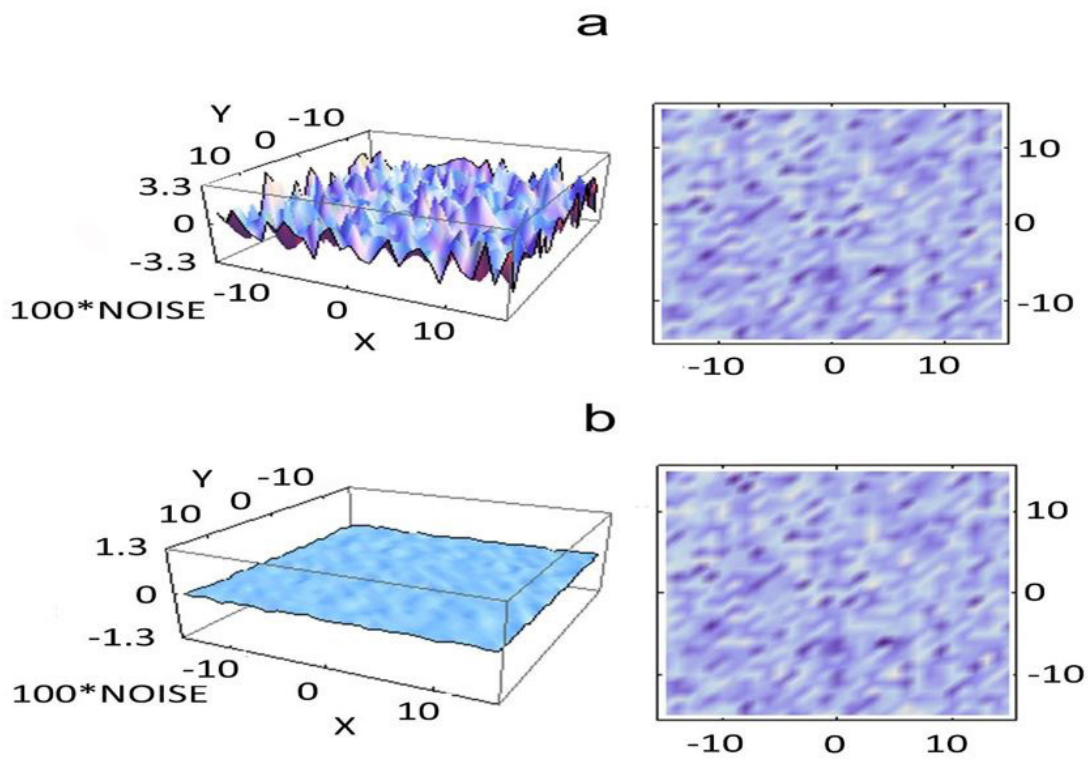

Figure 4. Statistical quit noise-filtering of the 2D IP frames. Noise level σ=0.01. Rotation angle Φ= 0°. *K* is the 2D IP-frame pixel number, *K*= 31 × 31. *N* is the 2D IP-frames number. (a) no averagednoise 2D IP-frame, the string number, ñ=K, *K*=961, *N*=1; (b) the averaged noise 2D IP-frame in terms of equation (5), the string number $\widetilde{N}$ = KXN, *K*=961, *N* = 961. The grid sizes in the 'own' crystal coordinate system (*X*, *Y*, *Z*) are equal to (31, 31, 21). The grid voxel sizes {Δ*X*, Δ*Y*, Δ*Z*}: Δ*X*= π/30, Δ*Y*= π/30, Δ*Z*= π/20. Wolfram program code [19] has been applied for generating the pseudorandom numbers{y_k}

Accordingly, in our calculations $Mean_{961}$ = −0.0226414 and $Mean_{961\times961}$ = −0.000911784 and thus, one obtains *Noise Gain* = 24.832 (remind that according to the SCLT the *Erwartungwert* noise values of $Mean_{\tilde{N}}[y_k]$ tends to zero for ñ→ ∞). In the case of the values $\widetilde{N}$ = 961 × 10^n (*n* = 0, 1, 2), one applies the random number generator algorithm [15-17] (RNGA) algorithm package for generating the noise-contamination-prone 2D IP- frames. The corresponding *Erwartungwert* noise values of $Mean_{\tilde{N}}[y_k]$ are listed in Table 2, they allow one to state that the procedure of averaging some noise 2D IP-frames can be effective tool to achieve the statistical noise filtering of the initial XRDT 2D IPs. The noise-filtered 2D IPs data, which should be incorporated in the optimization recovery procedure allow to obtaining more stringent information about the nanoscaled crystal-lattice defects. Loosely speaking, the use of a large number of the noise 2D IP frames opens a way to enhance the precision of the XRDT technique.

To illustrate how equation (5) works in the case of the noise-contamination-prone 2D IPs, the calculated results of the 3D function $f_{Ctpd}(\mathbf{r}-\mathbf{r}_0)$ recovery are listed in Table 3. Having an aim to minimize the XRDT 2D IPs target function, all the calculations have been carried out by using the combined iterative quasi-Newton-Levenberg-Marquardt – Simulated Annealing algorithm (qNLMSA) (*cf.* [9]). All the calculation have been calculated for total grid crystal sizes along to the dimensionless coordinates (*X*, *Y*, *Z*)-coordinates are equal to (61, 61, 21); the voxel sizes along to (*X*, *Y*, *Z*)-coordinates are equal to: Δ*X*= π/30, Δ*Y*= π/30 and Δ*Z*= π/20 in the units of Δ/π, the dimensionless thickness T of the sample Si(111) is chosen to be equal to π. From Table 3 it follows an assertion that optimized processing of the noise 2D IPs in terms of equations (5), (6) allows to decode the noise XRDT 2D IPs data up to the enhancing accuracy.

On the other hand, based on the calculation results listed in Table 3, one concludes that the averaging technique proposed is approved to work and improve a signal-to-noise ratio of the XRDT 2D IPs observed.

Table 3. Coulomb-types point defect in a crystal. Rotation angle $\Phi = 0°$. Recovery of the true 3D displacement field function $f_{Ctpd}(\mathbf{r} - \mathbf{r}_0)$ under retrieval: true vector $p^{(true)}$ = {1.50, 0.50, 1.80}, start vector $p^{(start)}$ is {1.12, 0.554, 2.40}. Parameter $CP^{(end)}$ is the goodness-of-fit, for the noise-freeIP-frame $CP^{(end)} = 6.5*10^{-11}$. Total grid crystal sizes along to coordinates (*X*, *Y*, *Z*)-coordinates areequal to (61, 61, 21); the voxel sizes along to (*X*, *Y*, *Z*)-coordinates are equal to: $\Delta X = \pi/30$, $\Delta Y = \pi/30$ and $\Delta Z = \pi/20$

Noise level, %	Number of averaged frames	$M_N[y]\times10^3$	r.m.s.	Vector $p^{(end)}$	Target function $\mathcal{F}\{p^{(end)}\}\times10^5$	Goodness-of-fit, $CP^{(end)}$
3	1	37	1.000	(1.55;0.49;1.83)	1.5	0.271
3	1000	-0.42	0.031	(1.51,0.50,1.81)	$1.5*10^{-3}$	0.009
3	10000	0.13	0.011	(1.50,0.50,1.80)	$1.4*10^{-4}$	0.0006
5	1	34	1.000	(1.83;0.52;2.40)	4.2	0.822
5	1000	0.43	0.031	(1.50,0.50,1.82)	$4.0*10^{-3}$	0.014
5	10000	-0.10	0.011	(1.50,0.50,1.80)	$4.1*10^{-4}$	0.0011

Discussion

A goal of the paper is to propose a statistical processing of the XRDT 2D IPs data to obtain quantitative information about the nanoscaled crystal-lattice defects in the case when a noise composes some part of the 2D IPs data observed.

Based on the simulated noise XRDT 2D IPs data, a statistical imperative noise-filtering procedure has been developed. It allows one to conclude that in particular, the statistical noise-averaging technique of the XRDT 2D IPs acquired secures retrieving of plausible information about small- size defects in a crystal with rather high accuracy. Both the RNGA algorithm and Wolfram Mathematica program have been employed for the computer simulating and processing of the noiseXRDT 2D IPs. As can be seen from Table 3, the goodness-of-fit parameter can be significantly improved by noise-filtering procedure from high discrepancy values of more than 27% (obtained without noise-averaging technique) to values of less than 1% (it corresponds to an almost perfect reconstruction of the 3D displacement field function of Coulomb-type point defect). Based on the qNLMSA algorithm applied for solving the inverse XRDT issue, one has justified that the XRDT method may be a powerful tool for obtaining quantitative information about small-size crystal-lattice-defects, *e.g.*, clusters, small dislocation loops *etc*.

As to a general solution of the inverse XRDT problem, a question of how the above statistical approach will work in a proper way for other kinds of crystal-lattice defects is still open for research.This is a good topic for future work. Staying on the study presented in the paper, we would like onlyto state that the elaborated statistical formalism works well and pushes ahead of the XRDT techniqueby accumulating the noise-filtered 2D-IPs data.

It is clear, the question of the easy-of-access XRDT 2D IPs rate is important. As is shown in the paper, statistical combining of the noise 2D IPs frames may effectively to improve the signal-to- noise ratio. In some cases, statistical processing of the 2D IP fragments may to reveal low-intensitydetails masked by a noise.

In present, the 2D IP-frame rate of the CCD detectors is quite high, and it can be altered within a wide range from 1 kHz to tens of MHz, only governed by reasons of the sufficient X-ray intensity per pixel [28]. Fast detectors would have reduced the time of the experiment, but in practice the flux of quanta after a highly absorbing sample turns out to be too small and the exposure time must be increased toachieve an acceptable relative level of noise in the image. Assuming the typical 2D IP-frame exposure time being about 10-20 sec [8], one can design a set of 100 frames under conditions when each of them contains, at least, 100 photons per the detector pixel in the 2D IP fragments needed.

Conclusion

In concluding, it is worth to mention that in the paper, the statistical filtering of the 2D IPs really relates to the experimental XRDT conditions in the two aspects as the measurement duration of the2D IP frames from several milliseconds and so the Gaussian amplitude noise values. They correspond to the incident X-ray beam intensity of the order of 10^{14} photons/sec that is typical for most of the X-ray experimental facilities being in the modern synchrotrons of the 3rd - 4th generations.

All the above allows to make an assertion that the statistical noise filtering of the 2D IPs data proposed is a good way to improve a signal-to-noise ratio, at least, for investigating and analyzing the small-size defects by using the XRDT technique. Checking the conformity of the noise distribution in the frame combined from several frames against the normal law allows the experimenter to achieve a quality image by calculating the necessary number of individual measurements and conducting the necessary series of experiments.

Acknowledgement

Special gratitudes to Prof. V.E. Asadchikov, Dr. D.A. Zolotov for their permanent interest and fruitful discussions concerning the present study. This work was supported by the Ministry of Science and Higher Education of the Russian Federation within the State assignment FSRC «Crystallography and Photonics» RAS in part of developing methods of structural analysis using X rays and synchrotron radiation.

References

1. Bowen DK, Tanner BK (1998) High Resolution X-ray Diffractometry and Topography. New York: Taylor & Francis.

2. Pavlov KM, Punegov VI (2000) Statistical dynamical theory of X-ray diffraction in the Bragg case: application to triple-crystal diffractometry. Acta Cryst A 56: 227–234.

3. Authier A (2001) Dynamical Theory of X-ray Diffraction. New York: Oxford Univ.

4. Benediktovitch A, Feranchuk I, Ulyanenkov A (2013) Theoretical Concepts of X-ray Nanoscale Analysis. Heidelberg: Springer.

5. Ludwig W, Cloetens P, Härtwig J, Baruchel J, Hamelin B, et al. (2001) Three-dimensional imaging of crystal defects by 'topo-tomography'. J Appl Cryst 34: 602-607.

6. Danilewsky AN, Wittge J, Croell A, Allen D, McNally P, Vagovič P, et al. (2011) Dislocation dynamics and slip band formation in silicon: In-situ study by X-ray diffraction imaging. J Cryst Growth 318: 1157-1163.

7. Hänschke D, Danilewsky A, Helfen L, Hamann E, Baumbach T (2017) Correlated three-dimensional imaging of dislocations: Insights into the onset of thermal slip in semiconductor wafers. Phys Rev Lett 119: 215504.

8. Asadchikov V, Buzmakov A, Chukhovskii F, Dyachkova I, Zolotov D, et al. (2018) X-ray topo- tomography studies of linear dislocations in silicon single crystals, J. Appl. Cryst. 51: 1616-1622.

9. Chukhovskii FN, Konarev PV, Volkov VV (2020) Towards a solution of the inverse X-ray diffraction tomography challenge: theory and iterative algorithm for recovering the 3D displacement field function of Coulomb-type point defects in a crystal. Acta Cryst A 76: 16-25.

10. Bondarenko VI, Konarev PV, Chukhovskii FN (2020) On the theory of reducing the level of statistical noise and filtering of 2D images of diffraction tomography. Cryst Rep 65: 821-826.

11. Hudgings DW, Dragt AJ (1972) Signal averaging at modest cost. Am J Phys 40: 1206-1211. 12. Kraftmakher Y (2006) Noise reduction by signal accumulation. The physics teacher 44: 528-530.

13. Hassan U, Anwar MS (2010) Reducing noise by repetition: introduction to signal averaging, Eur J Phys 31: 453-465.

14. Hendriksen AA, Bührer M, Leone L, Merlini M, Vigano N, et al. (2021) Deep denoising for multi-dimensional synchrotron X-ray tomography without high- quality reference data. Scientific Reports 11: 11895.

15. Strecok A (1968) On the calculation of the inverse of the error function. Math Comp 22: 144-158.

16. Blair JM, Edwards CA, Johnson JH (1976) Rational Chebyshev approximations for the inverse of the error function. Math Comp 30: 827-830.

17. Marsaglia G (1984) A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. SIAM J Sci Stat Comput 5: 349-359.

18. Wallace CS (1996) Fast pseudorandom generators for normal and exponential variates. ACM Trans Math Soft 22: 119-127.

19. Wolfram S (1999) Mathematica Version 4, Wolfram Media, Cambridge University Press.

20. Chernoff H, Lehmann EL (1954) The use of maximum likelihood estimates in $\chi 2$ tests for goodness of fit. Ann Math Statist 25: 579-586.

21. Plackett RL (1983) Karl Pearson and the Chi-Squared Test, International Statistical Review. International Statistical Institute (ISI) 51: 59-72.

22. Kolmogorov A (1933) Sulla determinazione empirica di una lgge di distribuzione. G Ist Ital Attuari 4: 83-91.

23. Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions, Annals of Mathematical Statistics 19: 279-281.

24. T. W. Anderson, D. A. Darling, Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes, Annals of Mathematical Statistics 23 (1952) 193–212.

25. T.W. Anderson, D.A. Darling, A test of goodness of fit, Journal of the American Statistical Association 49 (1954) 765–769.

26. Cramér H (1928) On the composition of elementary errors. Scandinavian Actuarial Journal 1: 13-74.

27. Von Mises RE (1928) Wahrscheinlichkeit, Statistik und Wahrheit. Julius Springer.

28. Olbinado MP, Just X, Gelet JL, Lhuissier P, Scheel M, et al. (2017) Mancuso, J. Morse, A. Rack, MHz frame rate hard X-ray phase-contrast imaging using synchrotron radiation. Optics Express 25: 13857-13871.

Supplementary part

Pearson's chi-squared test description

Pearson's chi-squared test (χ^2) [S1] is a statistical test applied to sets of categorical data to evaluatehow likely it is that any observed difference between the sets arose by chance. It is the most widelyused of many chi-squared tests (e.g., Yates, likelihood ratio, portmanteau test in time series, etc.) – statistical procedures whose results are evaluated by reference to the chi-squared distribution. In contexts where it is important to improve a distinction between the test statistic and its distribution, names similar to Pearson χ-squared test or statistic are used.

It tests a null hypothesis stating that the frequency distribution of certain events observed in a sampleis consistent with a particular theoretical distribution. The events considered must be mutually exclusive and have total probability 1. A common case for this is where the events each cover an outcome of a categorical variable.

Pearson's chi-squared test is used to assess three types of comparison: goodness of fit, homogeneity,and independence. A test of goodness of fit establishes whether an observed frequency distribution differs from a theoretical distribution. A test of homogeneity compares the distribution of counts for two or more groups using the same categorical variable A test of independence assesses whether observations consisting of measures on two variables, expressed in a contingency table, are independent of each other.

For all three tests, the computational procedure includes the following steps:

(1) Calculate the chi-square test statistic, which is the sum of squares of deviations between observedand theoretical frequencies, normalized to estimates of standard deviations.

(2) Determine the degrees of freedom, df, of that statistic.

(3) For a test of goodness-of-fit, df = Cats − Parms, where Cats is the number of observation detailsrecognized by the model, and Parms is the number of parameters in the model adjusted to make themodel best fit the observations: The number of details is reduced by the number of fitted parametersin the distribution.

(4) For a test of homogeneity, df = (Rows − 1)×(Cols − 1), where Rows corresponds to the number of details (i.e. rows in the associated contingency table), and Cols corresponds to the number of independent groups (i.e. columns in the associated contingency table).

(5) For a test of independence, df = (Rows − 1)×(Cols − 1), where Rows corresponds here to the number of details associated with one variable, and Cols corresponds to the number of details in thesecond variable.

(6) Select a desired level of confidence (significance level, p-value) for the result of the test.

(7) Compare (χ^2) to the critical value from the chi-squared distribution with df degrees of freedom and the selected confidence level (one-sided, since the test is only in one direction, i.e. "is the test value greater than the critical value?"), which in many cases gives a good approximation of the distribution of (χ^2).

(8) Sustain or reject the null hypothesis that the observed frequency distribution is the same as the theoretical distribution based on whether the test statistic exceeds the critical value of (χ^2). If the test statistic exceeds the critical value of (χ^2), the null hypothesis (H_0 = there is no difference between the distributions) can be rejected, and the alternative hypothesis (H_1 = there is a differencebetween the distributions) can be accepted,

both with the selected level of confidence. If the test statistic falls below the threshold (χ^2) value, then no clear conclusion can be reached, and the null hypothesis is sustained (we fail to reject the null hypothesis), though not necessarily accepted.

This method is an effective test of whether the theoretical observations from the constructed modelcorrespond to the experimental data, since an incomplete or incorrect model leads to violations of the expected law of the distribution of the differences between them.

References

[S1] Pearson, Karl (1900) On the criterion that a given system of deviations from the probable inthe case of a correlated system of variables is such that it can be reasonably supposed to have arisenfrom random sampling. Philosophical Magazine. Series *5*: 50, 157–175.