

Computer Programs in Physics



ePDFpy: A Python-based interactive GUI tool for electron pair distribution function analysis of amorphous materials ☆☆☆

Minhyo Kim^a, Pilsung Kim^a, Riccardo Bassiri^b, Kiran Prasai^b, Martin M. Fejer^b,
Kyung-ha Lee^{a,*}

^a Department of Physics, Sungkyunkwan University (SKKU), Suwon, 16419, Republic of Korea

^b E.L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Keywords:

ePDFpy
Amorphous material
Electron diffraction pattern
Electron pair distribution function
Short range order
Transmission electron microscope

ABSTRACT

ePDFpy is an interactive analysis program with a graphical user interface (GUI), designed to process the electron Pair Distribution Function (PDF) analysis of diffraction patterns from Transmission Electron Microscope (TEM), to identify the local atomic structure of amorphous materials. The program offers a user-friendly Python-based interface, providing a straightforward and adaptable workflow for PDF analysis. Various optimization and fitting processes were implemented to accurately reduce the electron diffraction data, including center-fitting and elliptical correction of diffraction data. An improved parameter-estimation feature is available to enhance the efficiency of the fitting process, along with an interactive GUI. ePDFpy will be freely distributed for academic purposes, with additional features, including a beam mask drawing module.

Program summary

Program Title: ePDFpy

CPC Library link to program files: <https://doi.org/10.17632/sym3sfmh7w.1>

Developer's repository link: <https://github.com/GWlab-SKKU/ePDFpy>

Licensing provisions: GNU GPLv3

Programming language: Python

Nature of problem: The general process of pair distribution function analysis consists of two major steps: image process on diffraction pattern and fitting appropriate parameters. Both of the procedures are affected by the user's proficiency, which can be responsible for producing inconsistent results and inefficiency. Thus, accurate calculation methods along with fully automated feature are required to enhance the quality of the analysis result. **Solution method:** ePDFpy offers an unbiased automated image process based on a computer vision algorithm to produce the consistent output of intensity profiles from diffraction patterns. In addition, converting the data structures into a multi-dimensional array enables efficient multi-parameter fitting features by performing parallel computation. All of these features are accomplished using various open-source libraries in the Python community, along with an interactive GUI.

1. Introduction

Interest in amorphous materials has risen significantly across various fields of science, from industrial uses such as promising media for solid-state batteries [1] to optical applications such as dielectric mirrors

with low Brownian noise for gravitational wave detectors [2]. However, unlike crystalline materials, amorphous materials lack long-range order (LRO). Instead, they exhibit localized atomic structures with short-range order (SRO) within approximately 0.5 nm length scale [3] and medium-range order (MRO) typically spanning from 0.5 nm to 5 nm

☆ The review of this paper was arranged by Prof. W. Jong.

☆☆ This paper and its associated computer program are available via the Computer Physics Communications homepage on ScienceDirect (<http://www.sciencedirect.com/science/journal/00104655>).

* Corresponding author.

E-mail address: khlee54@skku.edu (K.-h. Lee).

<https://doi.org/10.1016/j.cpc.2024.109137>

Received 19 November 2023; Received in revised form 9 February 2024; Accepted 12 February 2024

Available online 19 February 2024

0010-4655/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

scale [4]. The Pair Distribution Function (PDF) is a widely used method for characterizing the local atomic structure of amorphous materials [5–7]. The PDF represents the probability of finding a pair of atoms at a given radius and offers an intuitive means to characterize SRO. Many previous studies have utilized X-ray [8] or neutron [9] scattering experiments to extract atomic structure information about amorphous materials, for example as demonstrated in the case of zirconia-doped tantalum [10]. While these studies showed promising results, electron probes have emerged as a strong alternative for producing high-quality PDFs [11]. Compared to x-rays, electrons exhibit greater sensitivity to the electrostatic potential of matter, which makes it easier to observe the local atomic structure of thin film with higher spatial resolution, especially for low-Z materials [12,13].

Transmission electron microscopy (TEM) is an alternative tool to obtain suitable diffraction data by utilizing high energy electrons as a probe [13,14]. Using highly accelerated electrons as a probe, it can obtain a strong signal even with small-scale volumes of the sample [15], and advanced techniques, such as energy-filtering, are applied to obtain high-quality diffraction data via TEM [16]. Based on this high-quality data acquisition, electron PDF (ePDF) analysis requires a complex fitting process to achieve accurate atomic structure information. To enhance the efficiency and accuracy of ePDF analysis, ePDFpy is designed with a user-friendly graphical user interface (GUI).

1.1. Necessity for development

Several free open-source PDF analysis software tools are available for academic purposes online, including plugins in Digital Micrograph (DM) such as ePDF tools [17] and standalone tools such as SUEPDF [18] and eRDF Analyser [19]. However, for DM-based plugins, incompatibilities with other operating systems, such as Linux or Mac OS, can be a significant disadvantage. To overcome this issue, ePDFpy was developed with Python, which is an open-source and popular programming language due to its versatility and accessibility.

ePDFpy incorporates several features and workflows from eRDF Analyser, such as masking, averaging intensity, and background noise correction, which have proven to be efficient. However, key features including the image processing and parameter fitting in eRDF Analyser are partially based on an empirical approach, which requires users' expertise to select appropriate input for raw diffraction data. In contrast, ePDFpy provides a fully automated center-fitting and radial integral algorithm for the extraction of 1-dimensional azimuthally averaged intensity profiles from 2-dimensional diffraction patterns, which minimizes human intervention to prevent an inconsistent output of the intensity profile. Users can assess the fitting quality by visually confirming the real-time display of the polar coordinate transformed image, which is adapted from py4DSTEM [20]. The analysis process in ePDFpy is also significantly improved in both accuracy and efficiency by employing the advanced autofit feature, which utilizes a multi-parameter fitting algorithm. Details about this feature are discussed in forthcoming sections. In addition to data processing, organizing the output data is significant. To fulfill this task, ePDFpy includes additional tools that allow users to select specific analysis results of interest and generate average outputs for the final arrangement, as well as the variance between selected datasets to assess data quality.

ePDFpy adopts an interactive GUI, which enables an intuitive environment that allows the user to easily visualize all relevant information at each stage of the data reduction process. In addition, considering that the typical analysis involves multiple datasets from experiments, ePDFpy offers practical tools such as multiple file save/load options and batch image processing features for all loaded raw data, thereby alleviating redundancy. Furthermore, users can re-open and edit pre-saved settings such as chemical compositions and calibration factors, and essential fitting parameters from previous analysis done by ePDFpy, which enables easy revision of analysis results.

The following sections provide a detailed overview of ePDFpy and its key features: Section 2 covers the basic theory to calculate PDF, the main features of ePDFpy software are described in Section 3, and illustrative examples are shown in Section 4.

2. Theoretical background

2.1. Pair distribution function (PDF)

The determination of atomic structure information of a material involves extracting relative atomic positions or inter-atomic distance distributions. In isotropic systems, the distribution of inter-atomic distances can be obtained through the atomic pair density function [6]:

$$\rho(r) = \rho_0 g(r) = \frac{1}{4\pi N r^2} \sum_m \sum_{m \neq n} \delta(r - r_{mn}) \quad (1)$$

where ρ_0 is the average number density of the material, N is the total number of atoms inside the material, and r_{mn} is the radial distance between atom m and n . The function $g(r)$ is defined as the atomic pair distribution function and represents the probability density function of finding an atom at a distance r from a center atom [15]. The $g(r)$ function provides a useful physical quantity that acts as a map of the local density inside the material with respect to atomic distances. Meanwhile, another quantity, the reduced pair distribution function $G(r)$ is introduced [5,6], which is defined as,

$$G(r) = 4\pi r(\rho(r) - \rho_0) \quad (2)$$

$G(r)$ is more widely used than $g(r)$, since it can be directly extracted from the Fourier transform of diffraction data. Although less physically intuitive than $g(r)$, $G(r)$ is closely related to the distribution of atoms since [6]:

- (i) peak positions are aligned with $g(r)$, which indicates the dominant bond length,
- (ii) it oscillates around zero at large r ,
- (iii) it behaves like $-4\pi r \rho_0$ in the limit of $r \rightarrow 0$.

As a result, the structural information of the material can be extracted by appropriate fitting of $G(r)$ with the obtained experimental data using total scattering theory.

2.2. Total scattering theory

According to Warren [5], the scattered intensity from the material with scattering vector $\mathbf{q}$ with a modulus of $|q| = 4\pi \sin \theta / \lambda$ is expressed as,

$$I(q) = \sum_m f_m^2(q) + \sum_m f_m(q) \sum_{m \neq n} f_n(q) e^{i(\mathbf{q} \cdot \mathbf{r}_{mn})} \quad (3)$$

where $\mathbf{r}_{mn}$ is the position vector between atom m and n . $f_m(q)$ and $f_n(q)$ are the electron scattering factors ($f(q)$) of the atoms m and n respectively, which can be defined as [21,22],

$$f(q) = \frac{m_e e^2}{2\hbar^2} \left(\frac{z - f_x(q)}{q^2} \right) \quad (4)$$

where $f_x(q)$ is the x-ray scattering factor [21]. Following Cockayne's work [7], extending to a system with multiple types of atoms, the summation term that related to $m \neq n$ can be generalized by changing the sum over n to the integral of $\rho_{mj}(\mathbf{r}_{mn}) dV_n$, where $\rho_{mj}(\mathbf{r}_{mn})$ is the number density of atom of type j at $\mathbf{r}_{mn}$ within infinitesimal volume dV_n around atom m .

$$I(q) = \sum_m f_m^2(q) + \sum_j \sum_m \left[f_m(q) \times \int f_j(q) \rho_{mj}(\mathbf{r}_{mn}) e^{i(\mathbf{q} \cdot \mathbf{r}_{mn})} dV_n \right] \quad (5)$$

Introducing the number of each atom type in the material, N_i , which makes the total number of atoms $N = \sum_i N_i$, Eq. (5) can be written as,

$$I(q) = \sum_i N_i f_i^2(q) + \sum_j \sum_i \left[N_i f_i(q) \times \int f_j(q) (\rho_{mj}(\mathbf{r}_{mn}) - \rho_j) e^{i(\mathbf{q} \cdot \mathbf{r}_{mn})} dV_n \right] + \sum_j \sum_i \left[N_i f_i(q) \int f_j(q) \rho_j(\mathbf{r}_{mn}) e^{2\pi i(\mathbf{q} \cdot \mathbf{r}_{mn})} dV_n \right] \quad (6)$$

where ρ_j is the average number density of atom j . The last summation term in Eq. (6) is related to small-angle scattering, which is indistinguishable from the transmitted center beam. As the center beam is masked with a beam stopper to prevent CCD damage, this term can be neglected, as a result, $I_{\text{exp}}(q)$ is defined with the remaining terms:

$$I_{\text{exp}}(q) = \sum_i N_i f_i^2(q) + \sum_j \sum_i \left[N_i f_i(q) f_j(q) \times \int (\rho_{mj}(\mathbf{r}_{mn}) - \rho_j) e^{i(\mathbf{q} \cdot \mathbf{r}_{mn})} dV_n \right]. \quad (7)$$

Introducing the averaging $f_i(q)$ quantities over the chemical ratio of atoms as:

$$\langle f(q) \rangle = \frac{\sum_i N_i f_i(q)}{N}, \quad \langle f^2(q) \rangle = \frac{\sum_i N_i f_i^2(q)}{N} \quad (8)$$

the density function $\rho(\mathbf{r}_{mn})$ around atom m can be defined as:

$$\rho(\mathbf{r}_{mn}) \equiv \frac{\sum_j \sum_i N_j f_i(q) f_j(q) \rho_{mj}(\mathbf{r}_{mn})}{N \langle f(q) \rangle^2} \quad (9)$$

Hence, the indices i and j can be dropped in Eq. (7) to simplify the $I_{\text{exp}}(q)$ as,

$$I_{\text{exp}}(q) = N \langle f^2(q) \rangle + N \langle f(q) \rangle^2 \times \int (\rho(\mathbf{r}_{mn}) - \rho_0) e^{i(\mathbf{q} \cdot \mathbf{r}_{mn})} dV_n \quad (10)$$

Following previous studies [5,7], due to the randomized, isotropic structure of amorphous materials, there is no preferred orientation in the system. Therefore, the volume integral in Eq. (10) can be simplified by introducing a radial vector $\mathbf{r}_{mn} = \mathbf{r}$ and $r = |\mathbf{r}|$:

$$I_{\text{exp}}(q) = N \langle f^2(q) \rangle + N \langle f(q) \rangle^2 \times \int_0^\infty 4\pi r^2 (\rho(r) - \rho_0) \frac{\sin(qr)}{qr} dr. \quad (11)$$

Extracting the integral term, which contains the information on inter-atomic structure, the reduced intensity function $\phi(q)$ is defined as:

$$\phi(q) = \left[\frac{I_{\text{exp}}(q) - N \langle f^2(q) \rangle}{N \langle f(q) \rangle^2} \right] q = \int_0^\infty 4\pi r (\rho(r) - \rho_0) \sin(qr) dr. \quad (12)$$

Given the definition of $G(r)$ in Eq. (2), $\phi(q)$ can be expressed as:

$$\phi(q) = \int_0^\infty G(r) \sin(qr) dr. \quad (13)$$

Finally, $G(r)$ can be extracted using the inverse Fourier sine transform:

$$G(r) = \frac{2}{\pi} \int_0^\infty \phi(q) \sin(qr) dq. \quad (14)$$

2.3. Practical PDF analysis from experiment data

It is necessary to modify the ideal form of $\phi(q)$ and $G(r)$, since in experiments background noise such as the dark current of the detector is also recorded in the raw data. Assuming a correction factor $C(q)$ as a value of noise from the detector, it should be removed from $I_{\text{exp}}(q)$. Thus, Eq. (12) can be modified as:

$$\phi(q) = \left[\frac{I_{\text{exp}}(q) - N \langle f^2 \rangle - C(q)}{N \langle f \rangle^2} \right] q. \quad (15)$$

In addition, the effective range of scattering vector q in a given diffraction pattern is limited by various instrumental conditions, such as strong center beam, and finite detector size. Therefore, applying a precise minimum (q_{min}) and maximum (q_{max}) cut-off value of q to Eq. (14) is necessary, which corresponds to the center beam masking area and upper bound of the effective data range. However, reducing the integral range in Eq. (14) can generate a termination error in the Fourier transform, such as an unusual oscillation or unphysical peaks in $G(r)$. To compensate for this problem, applying a Gaussian function as a damping function on $\phi(q)$ is introduced [6,19] to $G(r)$ as:

$$G(r) = \frac{2}{\pi} \int_{q_{\text{min}}}^{q_{\text{max}}} \phi(q) \cdot \exp(-bq^2) \sin(qr) dq \quad (16)$$

where constant b is the damping factor. Note that scattering vector $s = 2 \sin \theta / \lambda = q / 2\pi$ is often conventionally used, which results in Eq. (16) becoming:

$$G(r) = 8\pi \int_{s_{\text{min}}}^{s_{\text{max}}} \phi(s) \cdot \exp(-bs^2) \sin(2\pi sr) ds. \quad (17)$$

3. Software description

ePDFpy is open-source software written in Python, featuring a platform-independent GUI facilitated by the PyQt5 toolkit. Users can load diffraction pattern data files, compatible with the most commonly used file formats such as TIFF, MRC, and DM3. In addition, the software itself is capable of loading multiple raw data using a batch loading process, thus users can easily navigate between files and conduct analysis.

The overview of the workflow is shown in Fig. 1. Once users load the raw data, using a center fitting feature modified from py4DSTEM software [20], the radially averaged intensity profile from the raw diffraction image can be extracted in the 'Profile extraction' tab, after applying a beam stop mask and background noise subtraction on the raw data. Users can also avoid redundant procedures by simply using batch calculation processes of center fitting and intensity averaging on entire loaded images, with fixed beam masks. After profile extraction, users can proceed to the 'PDF analysis' tab in which fitting parameters can be defined using automatic and/or manual fitting to get $G(r)$. In the final step, the output from the profile extraction and PDF analysis can be saved in various text-based file formats like csv and txt. Every workflow step can easily be revisited, so that users can check the output quality of each procedure, without starting from scratch.

3.1. Profile extraction

The aim of profile extraction is to obtain a radially averaged 1-dimensional intensity profile from raw diffraction pattern data. To efficiently facilitate this process, the profile extraction tab is composed of three main panels as shown in Fig. 2: a control panel (upper-left), an image panel (lower-left), and a plot panel (right-hand side). Users can intuitively check each calculation output in the corresponding panels.

3.1.1. Control panel

The extraction of the intensity profile involves a series of automated steps. The first step is to set the mask on the beam stop region; users can select from pre-saved mask data (Fig. 3a) or draw a polygon from the mask drawing module (Fig. 3b).

If users have a noise reference image, it can be subtracted via a noise subtraction toolbox. A single blank image file can be loaded, and users can choose whether to subtract it from the raw data, or revert to the original for each loaded data. This indication of noise subtraction will

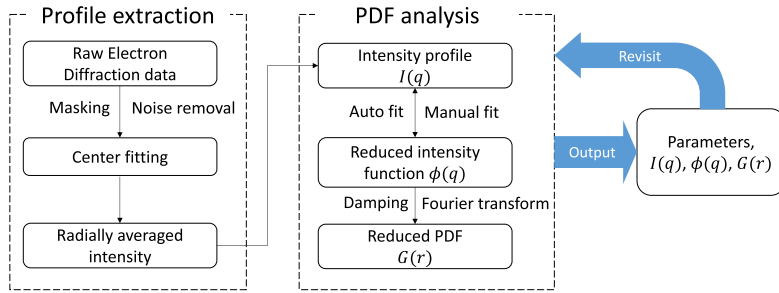


Fig. 1. Flowchart of ePDFpy analysis process. The dashed line box indicates each tab of procedures and the solid line box shows the major step of each procedure. The reversed arrow from ‘output data’ means revisiting the previous analysis.

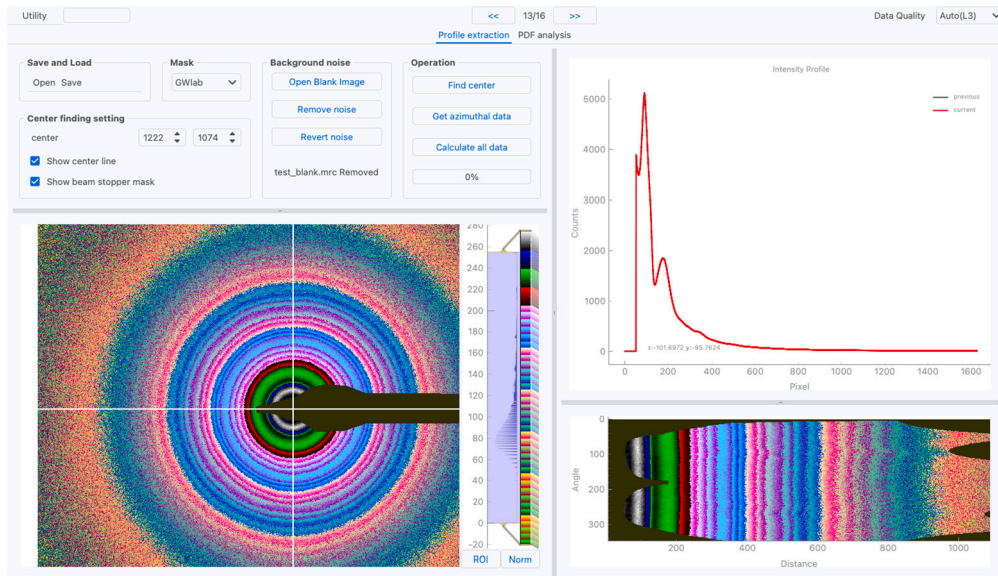


Fig. 2. GUI panel showing profile extraction. Each feature can be activated by clicking the button, and the user can adjust the parameters, such as center position and elliptical fitting parameters.

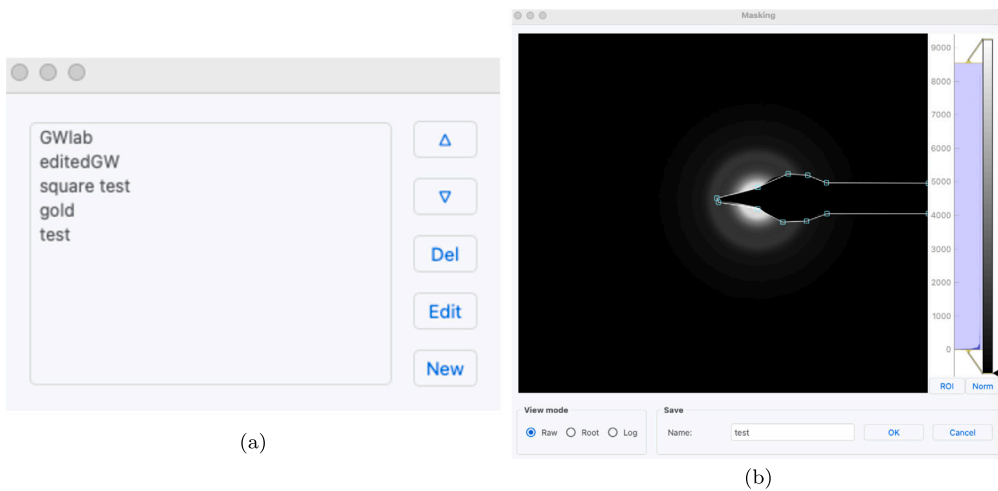


Fig. 3. ePDFpy masking GUI module. (a) Mask selection module, (b) Mask drawing module. (a) is used to select pre-saved modules and (b) can be used to draw masking area using a polygon.

be shown in the text box, and the images will be instantly updated on the corresponding panels.

After applying the mask and noise subtraction, users can find the center of the raw data with a two-step fitting algorithm. The initial center finding is done by fitting the center of mass [20] of a circle with the intensity over the threshold which is made by the flood-fill algorithm in Python's openCV library [23]. The second stage employs the gradient descent algorithm, with the center coordinates serving as a parameter in the cost function. The cost function itself represents the standard deviation of intensity along the radial axis of the polar-transformed image which is generated by the code in py4DSTEM [20]. The azimuthally averaged intensity profile $I(q)$ is extracted from the fitted center coordinates, with respect to pixel distance.

The entire procedure of these image processes can be done for all loaded data files, including both center fitting and extraction of intensity profile with fixed mask and noise data, by activating the batch process from 'Calculate all data'. Unless users manually change the center coordinates, all image processes are automated without human intervention, therefore a consistent $I(q)$ will be generated.

3.1.2. Image and plot panels

A visualization of the image processing is displayed in both panels. The image panel depicts the colormap of the log-scaled raw diffraction image to highlight the concentric diffraction rings. White perpendicular lines indicate the location of the center and the coordinates values are displayed in a spin box located above the image panel. Users can adjust the intensity scale with the scalebar on the image panel.

The plot panel consists of two plots. The first plot shows the radially averaged intensity profiles, plotted with respect to pixel distances from the center. The second plot shows the polar-transformed image of the raw diffraction image, with the distance and angle plotted on the horizontal and vertical axes, respectively, which are commonly used for multiple software, like py4DSTEM [20] and *Emilys* package (<https://github.com/ju-bar/emilys>). The transformation is done by applying the py4DSTEM's algorithm [20], with $dr = 1$ and $d\theta = 1^\circ$ step size. Users can visually check the validity of the center and angular aberration of the diffraction pattern [24,25], by confirming the straight lines formed with respect to the distance axis.

3.2. PDF analysis

$G(r)$ is calculated from the extracted $I(q)$ in the PDF analysis tab (Fig. 4). Users can adjust the fitting parameters from the control panel and immediately observe the results in the plot panels.

3.2.1. Preparation for analysis

Prior to the analysis of $I(q)$, users must enter the basic experimental parameters set up essential parameters, which can be done manually or by selecting pre-saved settings. The required parameters are the type of atomic scattering factor $f(q)$, chemical composition of the sample, and calibration factor. Users can choose between the atomic scattering factors calculated by Kirkland [21] or Lobato [26] for fitting $\phi(q)$. The chemical composition, including atom type and ratio, is used to calculate the averaged quantity of the scattering factor in Eq. (8), in which users can input up to 5 elements. The calibration factor is the value to convert the pixel distance of the extracted $I_{\text{exp}}(q)$ to a unit of scattering vector $s = q/2\pi$ in reciprocal space. Users can input the values from the metadata of raw images or calculated values directly from calibration samples such as gold or silicon. The converted $I_{\text{exp}}(q)$ is presented in the lower plot panel.

3.2.2. Fitting $\phi(q)$ and calculating $G(r)$

After setting the prerequisite parameters, users can start the analysis by adjusting the fitting parameters, which are used to calculate $\phi(q)$ and $G(r)$ based on Eq. (15) and (17), from the obtained $I_{\text{exp}}(q)$. Each calculated result is presented in the top right ($\phi(q)$) and bottom right

($G(r)$) plot panels, and users can intuitively visualize changes in the results to find the optimum values. The list of the major parameters is q range, q_k , and N , which are required to optimize $\phi(q)$.

(i) Users can easily set the **q range** by adjusting the highlighted ranges in the $I_{\text{exp}}(q)$ plot panel or by manually changing the numbers in the text box. Both methods will change q_{min} and q_{max} , which will determine the integration range in Eq. (17). In addition, the software automatically fixes q_{min} at the first local minimum point in $I_{\text{exp}}(q)$, which is expected to be beyond the effect of the beam stop mask. If necessary, users can also modify q_{min} from the 'select' menu (Fig. 5).

(ii) **Cut-off scattering vector q_k** is used to set the correction factor $C(q)$ in Eq. (15). From Anstis's work [27], the effects of inelastic scattering signals from multiple scattering can be ignored for thin samples, hence ePDFpy assumes the correction factor as a constant $C(q_k)$, where q_k is the intersection point between $N\langle f^2(q) \rangle + C(q)$ and the experimental signal $I_{\text{exp}}(q)$. Thus, q_k specifies $C(q_k)$ with following condition [19]:

$$C(q_k) = I_{\text{exp}}(q_k) - N\langle f^2(q_k) \rangle \quad (18)$$

The default setting in ePDFpy fixes the q_k value as the same as q_{max} , but users can change it by adjusting values in 'fit at q' text box.

(iii) **Coordination number N** is the total number of atoms in the illuminated area in the sample. The optimum N value minimizes the difference between $I_{\text{exp}}(q)$ and $N\langle f^2(q) \rangle + C(q_k)$. Users can adjust its value manually in the 'N' text box or use the 'Autofit' feature, which optimizes N using least-square fitting. The corresponding $N\langle f^2(q) \rangle + C(q_k)$ is represented by the green line in the $I_{\text{exp}}(q)$ plot panel.

While these 3 parameters can be manually adjusted by users, an 'Advanced fitting' feature is also available to optimize all of these parameters at once with unbiased values, which will be described in the next subsection. Once the $\phi(q)$ is set, users can customize additional factors required in the calculation of $G(r)$. Firstly, the **damping factor b** in Eq. (17), which has a default value of 0.15, can be adjusted, and the corresponding $\phi(q)$ is displayed in the $\phi(q)$ plot panel as a green line. The **r range** and **resolution** of the Fourier transform can be adjusted by modifying the values of 'r(max)' and 'dr'. These parameters are crucial for capturing the short-range order structure in $G(r)$, and the default values are 10 Å and 0.01 Å, respectively. Each of the parameters can be adjusted manually by changing the values in the corresponding spinboxes. Users can choose to update plots instantaneously with each parameter change, allowing for efficient manual optimization of parameters, by enabling the 'instant update' option.

3.2.3. Advanced fitting

To predict the optimum parameters of $\phi(q)$, an automated fitting feature was developed to fit all 3 factors simultaneously. To initiate this multi-parameter fitting process, users first need to set the parameter searching range, which will involve q_{max} , q_k in the control panel ('Fitting Setting') section in the GUI panel (Fig. 6). For convenience, the search range for q_{max} is shown as the corresponding pixel range box in the control panel. In addition, users can adjust the search range and step size of the parameters to achieve the desired level of accuracy. The other two control boxes adjust the threshold condition of fitting and the number of best-selected results. After setting up all initial variables, users can start the fitting process by activating the 'Autofit' button to conduct the multi-parameter calculation. The fitted results are presented in the table (fitting parameters) and in graph panels ($\phi(q)$ and $G(r)$).

In multi-parameter fitting, 3-dimensional array structures (datacube) are formed to stack up all possible cases of $\phi(q)$ and calculated $G(r)$ within the input parameter searching ranges, as the total calculation process is described in Figure 7. $\phi(q)$ 1-dimensional array is built by applying different q range, q_k and N value, forming each row in $\phi(q)$ datacube. For the q range, users only need to set q_{max} , as q_{min} is determined automatically in the 'pdf analysis' tab. The N range is set within ± 5 values from the result of least-square fitting, which is done within

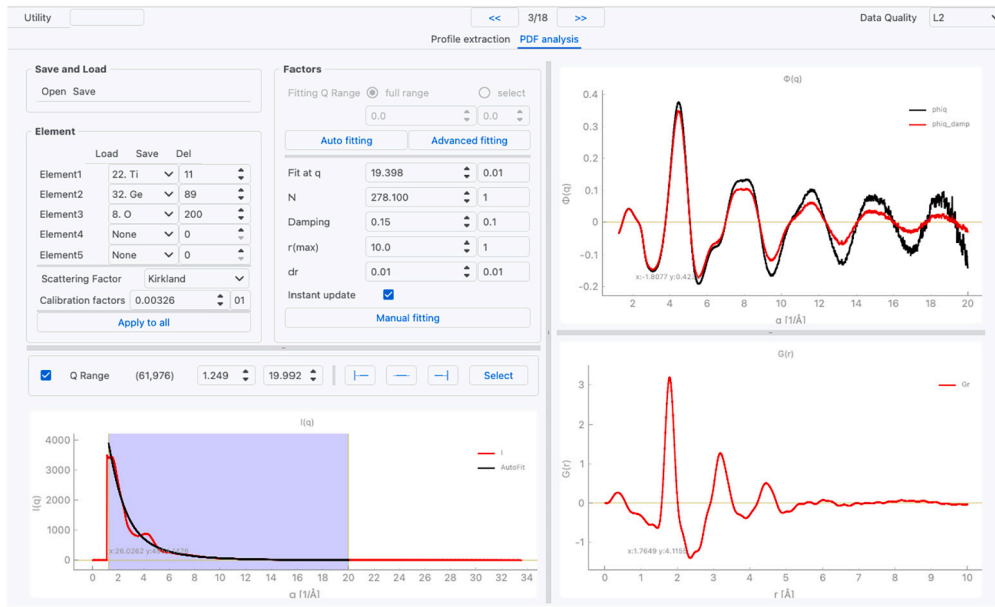


Fig. 4. PDF analysis GUI panel. The top left corner is the control panel and the rest are the plot panels of $I(q)$ (bottom-left), $\phi(q)$ (top-right), and $G(r)$ (bottom-right).

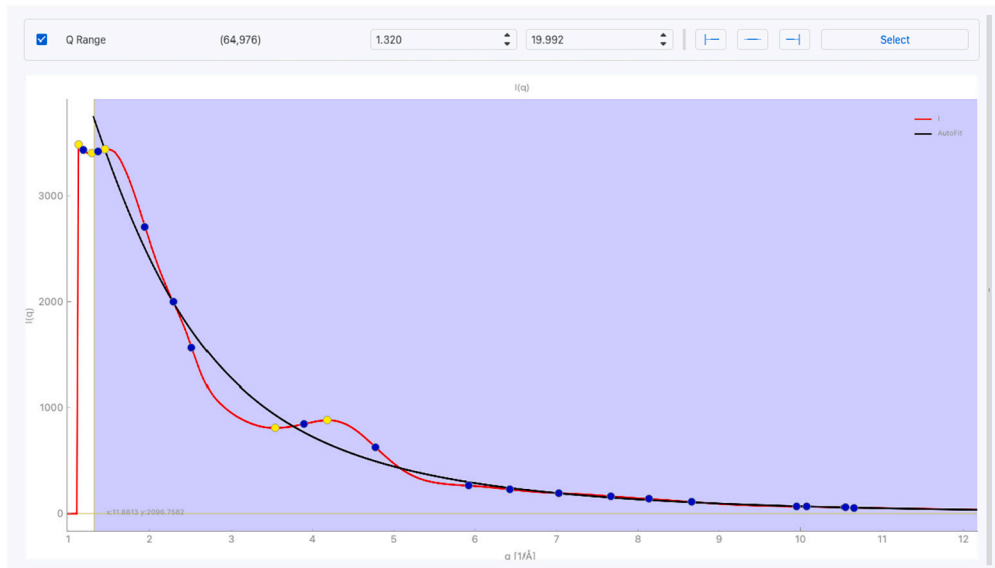


Fig. 5. $q_{\min}$ selection GUI panel. Yellow dots represent the 1st derivative of $I(q)$ to be zero, and blue dots correspond to the points where the 2nd derivative of $I(q)$ is zero. (For interpretation of the figure(s), the reader is referred to the web version of this article.)

the q range from 2π to $q_{\max}$. This fitting range is determined based on the relation between reciprocal space distance q and real space distance r , corresponding to an $r < 1$ Å range. Some of the individual $\phi(q)$ rows are defined with different lengths in q , due to different $q_{\max}$, which makes it impossible to stack up the $\phi(q)$ datacube. Therefore, those arrays contain manually added NaN (Not a Number) values to match the length of every row array.

Using a parallel computing algorithm, the Fourier transform based on Eq. (16) (or Eq. (17)) of this 3-dimensional $\phi(q)$ array is done, which in results, forms the $G(r)$ datacube, where each row corresponds to each $\phi(q)$ row with different fitting parameter. To find the optimum parameter sets, $G(r)$ results are filtered based on the following two conditions:

(i) The peak height of $G(r)$ within the distance r less than 1 Å should be minimized, since it indicates atoms existing within 1 Å length scale, which is physically impossible. Therefore, users can set the lower percentage threshold for selecting the minimum noise peak value. The

default setting for this is to choose the lowest 1% results among the entire sets, and its level can be adjusted by changing the threshold in 'Noise peak cut (%)'.

(ii) The existence of a local maximum in $G(r)$ in the range between 2 to 3 Å. This peak position corresponds to the second nearest bond length, which is likely the oxygen-oxygen bond length in most oxide materials.

Based on the nature of $G(r)$, it should exhibit a linear trend of $-4\pi\rho_0r$. Therefore, the least-squares method is applied to determine the best-fitted $-4\pi\rho_0r$ for each possible $G(r)$. From this fitting, grading values are defined as the sum of the squared differences between the fitted line and $G(r)$ for $r < 1$ Å, in order to identify the best-fitted $G(r)$ that satisfies the filtering conditions. Some of the top best-fitting results are shown in the bottom-left table and corresponding plots on the right-side plot panels. Among those results, users can choose the parameter sets that will be used for further manual fitting.

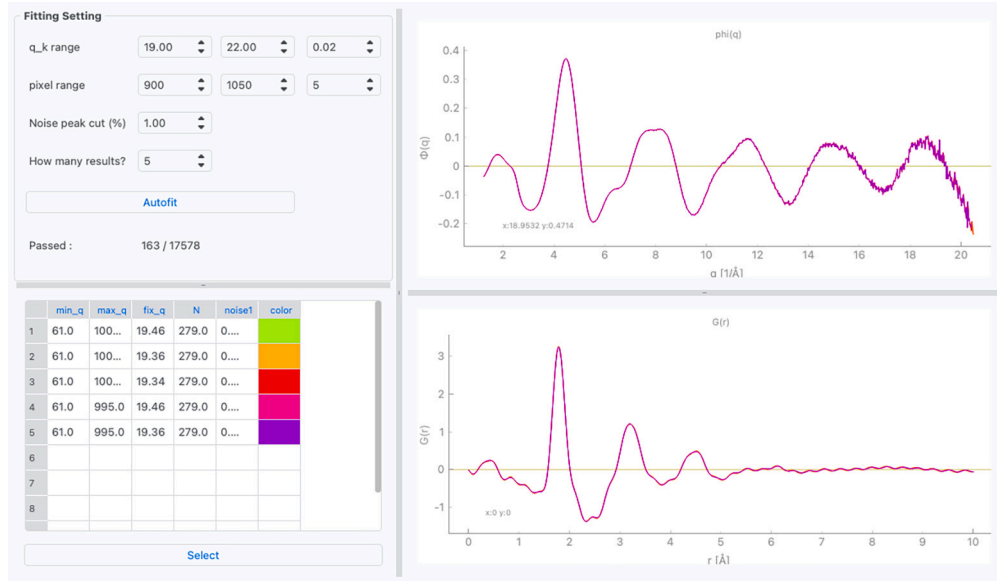


Fig. 6. Advanced fitting module GUI panel. The control panel is in the top-left corner, which controls the parameter searching ranges ($q_{\max}$ and q_k) and threshold conditions. The $\phi(q)$ and $G(r)$ in graph panels are corresponding to each row in the table panels.

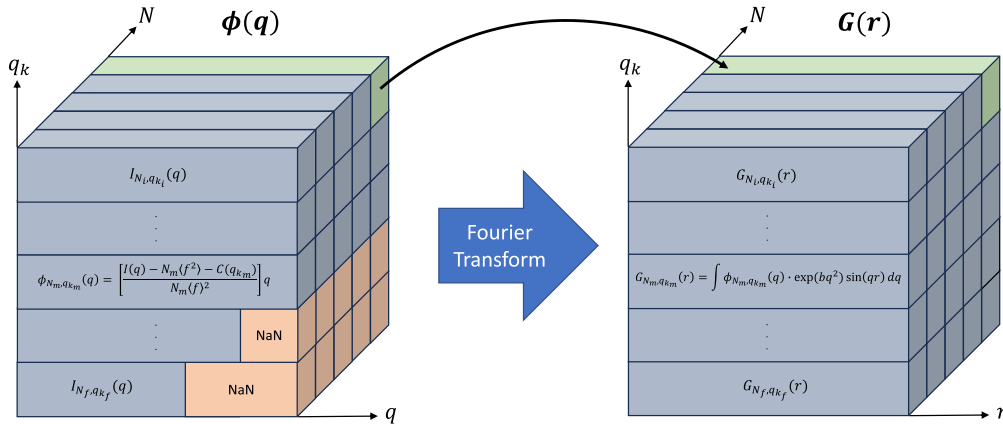


Fig. 7. Graphical representation of the 3-dimensional array (datacube) calculation used in the autofit process. The lefthand array consists of $\phi(q)$ values and the righthand array consists of $G(r)$ values, calculated from $\phi(q)$ datacube. Each axis shows the parameter space, and subscripts represent the different fitting parameters. The black arrow describes the relation between each row (green) in $\phi(q)$ and $G(r)$. The orange blocks show manually added NaN values.

Table 1
Descriptions of the output file from ePDFpy software.

File format	Description
.q.csv	Effective q range, intensity profile, $\phi(q)$
.r.csv	Atomic distance, $G(r)$
.jpg	Screenshot of PDF analysis panel
.preset.json	All parameters required for analysis

3.3. Output data

ePDFpy provides various output files to save the final analysis results, as listed in Table 1. These output files enable users to revisit previous analysis results from where they left off.

In addition, ePDFpy allows users to load multiple specific output files for comparison in a single plot panel or to create averaged values for selected results. Fig. 8 demonstrates the selection analysis on multiple test analysis result files with a Ta_2O_5 sample as an example.

The white line in the plot represents the averaged value of $G(r)$ with a selected list of datasets. Users can save the selected list of data or only the averaged data in single csv files, along with the txt file containing the directory information of selected data files.

4. Example and discussion

To demonstrate standardized analysis procedures of ePDFpy, two illustrative examples are presented in this section. The first example shows the extraction of a calibration factor from the calibration standard sample (polycrystalline gold), while the second demonstrates a PDF analysis of an amorphous material sample.

4.1. Polycrystalline Au

The first example is polycrystalline gold (Fig. 9), which is obtained using the FEI Tecnai G2 F20 X-TWIN TEM in Stanford Nano Shared Facility with acceleration voltage of 200 kV, condenser aperture (C2) of 100 μm , a camera length of 75 mm and no selected-area aperture (SA) applied. The image was taken with total exposure time with 0.1 s with 1 frame image. Fig. 9a shows polar-transformed images of the raw data with center-fitting conducted by ePDFpy, along with off-centered transformations varied by ± 5 pixels. The graph confirms that ePDFpy's center-fitting results are notably superior, characterized by straighter lines in the polar-transformed images, particularly within the near-center region where intensity is most pronounced. In contrast, the off-centered results exhibit curved lines in a polar-transformed image.

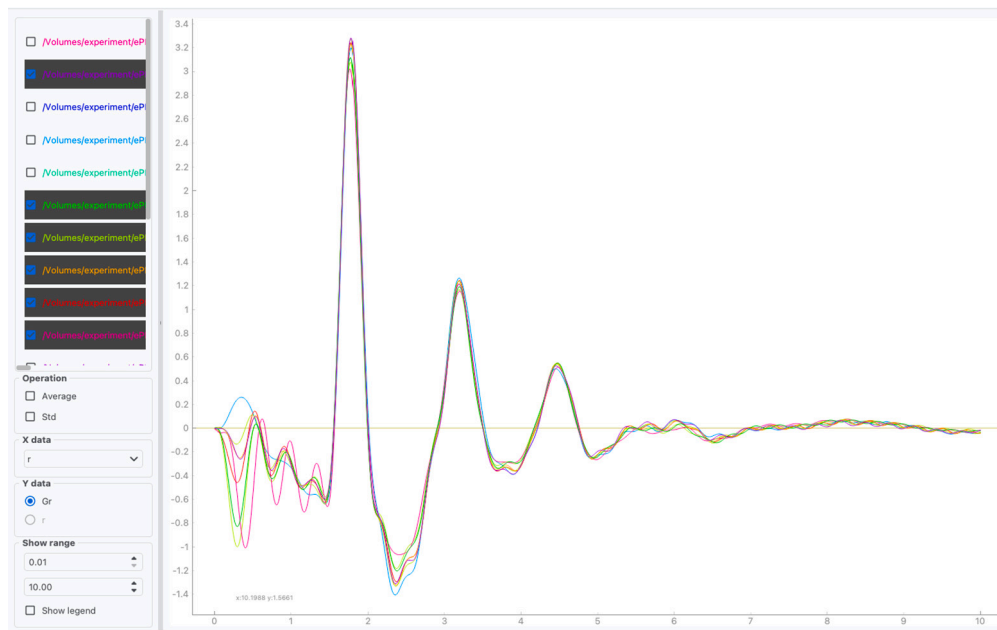


Fig. 8. Data selection process for pre-analyzed amorphous Ta_2O_5 data. Poor quality data are excluded from the analysis, which can be confirmed in the legends and checkbox.

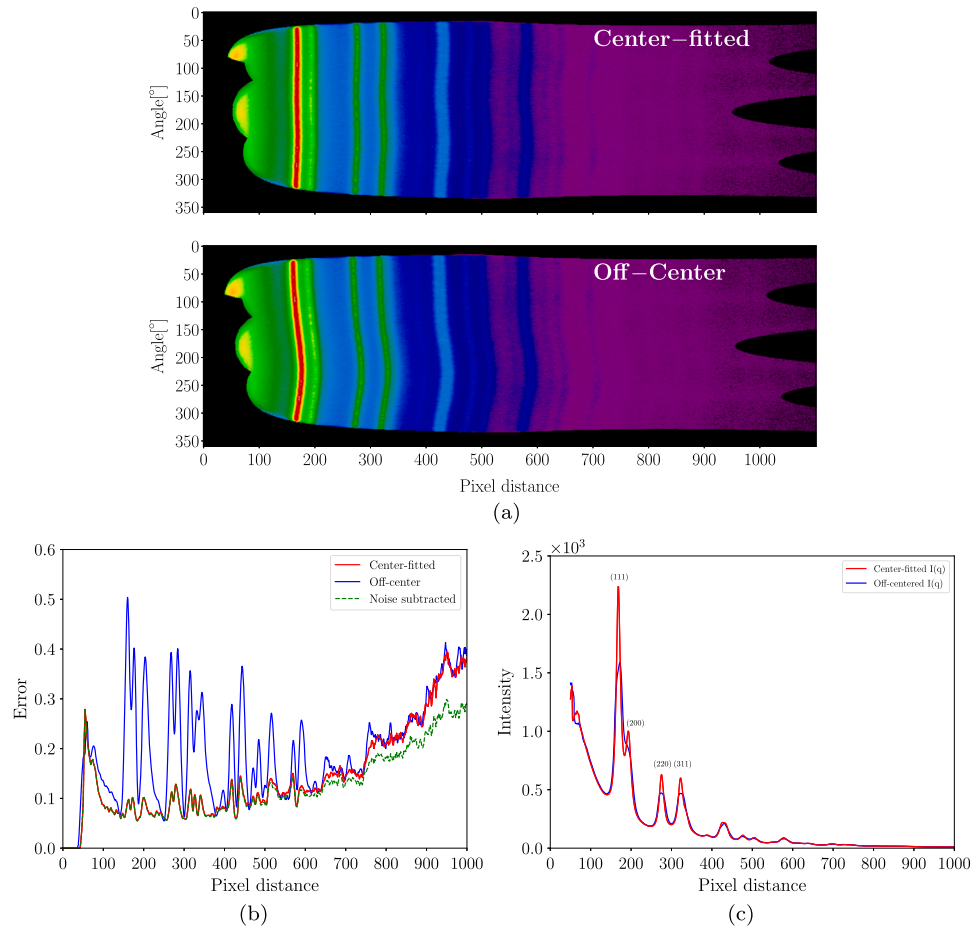
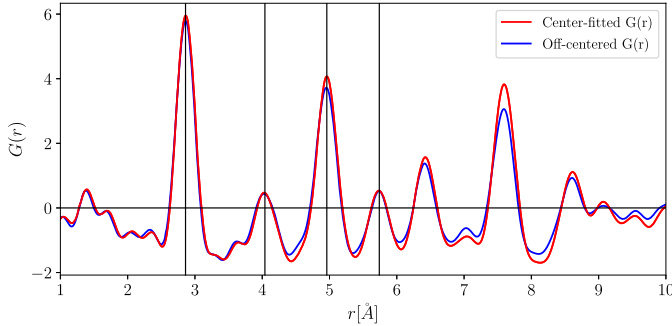


Fig. 9. Image process results of polycrystalline gold (Au) data. (a) Polar-transformed images of the raw data with center-fitted (top) and off-centered by 5 pixels from the optimal center along both the x and y positions (bottom). (b) Error comparison which is calculated along the angle axis in the polar-transformed data as a function of distance. Background subtracted data is in a dashed line to distinguish itself from center-fitted data in overlapped areas. (c) Comparison of $I(q)$. The numbers (hkl) show the indices of corresponding lattice planes.

Table 2

Calculated calibration factor from extracted intensity profile.

(hkl)	(111)	(002)	(022)	(113)
d	2.35 Å	2.05 Å	1.44 Å	1.23 Å
D	168	196	275	322
ds	0.00253	0.00254	0.00252	0.00253

**Fig. 10.** Fitted $G(r)$ comparison between center-fitted and off-center $I(q)$ results. The vertical lines indicate the first 4 peak positions: 2.87 Å, 4.06 Å, 4.96 Å, and 5.73 Å.

This comparison confirms that users can intuitively check if the center was correctly defined in the analysis process, even as a small offset (near 5 pixels) can be visually inspected very easily. In Fig. 9b, the error due to aberration is defined as a normalized variance with respect to the pixel distance, which is calculated along with the angle axis in Fig. 9a and normalized by the average intensity calculated in the same manner. A comparison between these error values and the center-fitted results demonstrates a significant reduction in errors, particularly in the near-center region. Moreover, the analysis of other error values derived from noise-subtracted data indicates an improvement in data quality in regions further from the center, owing to a decrease in aberration effects when compared to center-fitted data. As a result, Fig. 9c presents the intensity profile comparison between noise-subtracted center-fitted raw data and off-centered data, which exhibits better-defined peaks. Furthermore, certain peaks, such as those corresponding to the (200) plane, become obscured by background noise in the off-centered transformations. The extracted intensity profiles of the standard sample can be used for the calculation of the calibration factor (ds) to convert the pixel distance of the CCD detector to scattering vector s . This conversion can be achieved using the equation: $ds = (1/d)/D$, where d represents the d-spacing corresponding to the distance between atomic planes, and D is the pixel distance of the detector. Referencing the known d-spacing value of face-centered cubic (fcc) gold calculated from the lattice parameter 4.08 Å, the calibration factor is extracted as the following Table 2. In this particular data, the average value of the calibration factor is calculated as 0.00253, which can be used in the calculation of $G(r)$.

From the acquired $I(q)$ and ds value, $G(r)$ is calculated for both cases to confirm the center-fitting effect. Fig. 10 shows that the center-fitted $I(q)$ has a more well-defined peak in every region. Both results were analyzed with the advanced fitting of ePDFpy, which produced the fitting parameters for center-fitted $I(q)$ as $q_{\max} = 20.381 \text{ Å}^{-1}$, $q_k = 19.140 \text{ Å}^{-1}$, $N = 59$. Compared with the known interatomic distances of the fcc gold's coordination shell, the first four peak positions indicated with black vertical lines in Fig. 10, show well-matched results within the offset of $\pm 0.02 \text{ Å}$ in the 1st and 2nd coordination shell, and $\pm 0.05 \text{ Å}$ in 3rd and 4th coordination shell. These results suggest that although ePDFpy is designed to analyze the diffraction pattern of amorphous materials, it is also capable of analyzing the ring pattern of polycrystalline materials.

Table 3Parameters from each fitted method of a-Ta₂O₅ sample.

Methods	$q_{\max}$	q_k	N
Manual fit	20.074 Å ⁻¹	19.807 Å ⁻¹	181.250
Advanced fit	20.483 Å ⁻¹	19.254 Å ⁻¹	181

4.2. Amorphous tantalum

The diffraction image from an amorphous Ta₂O₅ (a-Ta₂O₅) thin film sample is used for the second illustrative example. Measurement was performed by FEI Themis Z TEM located in DGIST, South Korea, operated with 300 kV acceleration voltage, C2 aperture of 10 μm, a camera length of 75 mm and no Selected-area Aperture (SA) conditions. The illumination area on the sample is 150 nm. The image was acquired by a CCD camera with an exposure time of 50 s and integrated among 20 frame images. The image process on the raw data is presented in Fig. 11: polar transformed images that show the effect of the offset from the best-fitted center (11a) and the errors calculated from them (11b) as defined in Section 4.1. Similar to the example of polycrystalline gold, a minor shift in the center coordinates produced noticeable deviations that can be visually observed. Furthermore, the error curve for the center-fitted data demonstrates improved results by reducing distinct peaks in the near-center region, which are present in the off-centered data. In addition, the noise-subtracted data demonstrate reduced error in regions further from the center. This noise reduction consequently enhances the signal-to-noise ratio, especially in the high- q range of the acquired diffraction data, leading to improved analytical results for $G(r)$.

Fig. 12 and Table 3 show a comparison between the optimized parameter results from both manual and advanced fitting of $\phi(q)$, fixing $q_{\min} = 1.209 \text{ Å}^{-1}$ and using a $q_{\max}$ value of approximately 20 Å^{-1} along with a calibration factor of 0.00326. Fig. 12a displays the comparison of the reduced intensity function $\phi(q)$ for both methods, with a damping factor of $b = 0.15$. For the calculation of $G(r)$ in Fig. 12b, the Fourier transformation parameters were set to a maximum $r = 10 \text{ Å}$, and $\Delta r = 0.01 \text{ Å}$.

Overall, the results obtained through advanced fitting exhibit similar trends to those obtained through manual fitting. The primary difference lies in the oscillations observed in the advanced fitting results beyond 5 Å, which may arise from the longer tail in $\phi(q)$. Nevertheless, advanced fitting produces a $G(r)$ profile that closely aligns with that of manual fitting. Furthermore, a quantitative comparison of values, as presented in Table 3, demonstrates that advanced fitting can provide a reliable initial estimate for optimizing $G(r)$.

In Fig. 12b, vertical lines indicate well-defined peak positions after 1 Å, as each value corresponds to 1.96 Å, 2.79 Å, 3.34 Å, 3.74 Å. The first peak at 1.97 Å, which is reported as Ta-O bond length, matches up with the previous x-ray study [28] within $\pm 0.03 \text{ Å}$. Likewise, other peaks are in good agreement with those published by Shaym et al. [29], as the differences are also in the range of $\pm 0.02 \text{ Å}$. The second peak (2.79 Å) corresponds to the O-O bond, which is better defined than the previous x-ray PDF results [30].

5. Conclusion

ePDFpy is a useful tool for PDF analysis of electron diffraction data, with an interactive GUI environment. The package is based on Python, whose versatility, accessibility, and open-source environment allow users to customize the software to their requirements. Implementing various techniques using computer vision, ePDFpy provides not only an advanced image processor that allows users to extract intensity profiles with enhanced accuracy and consistency but also an advanced multi-parameter fitting to increase the efficiency of the PDF analysis. In addition, features such as selection analysis, masking module, and the

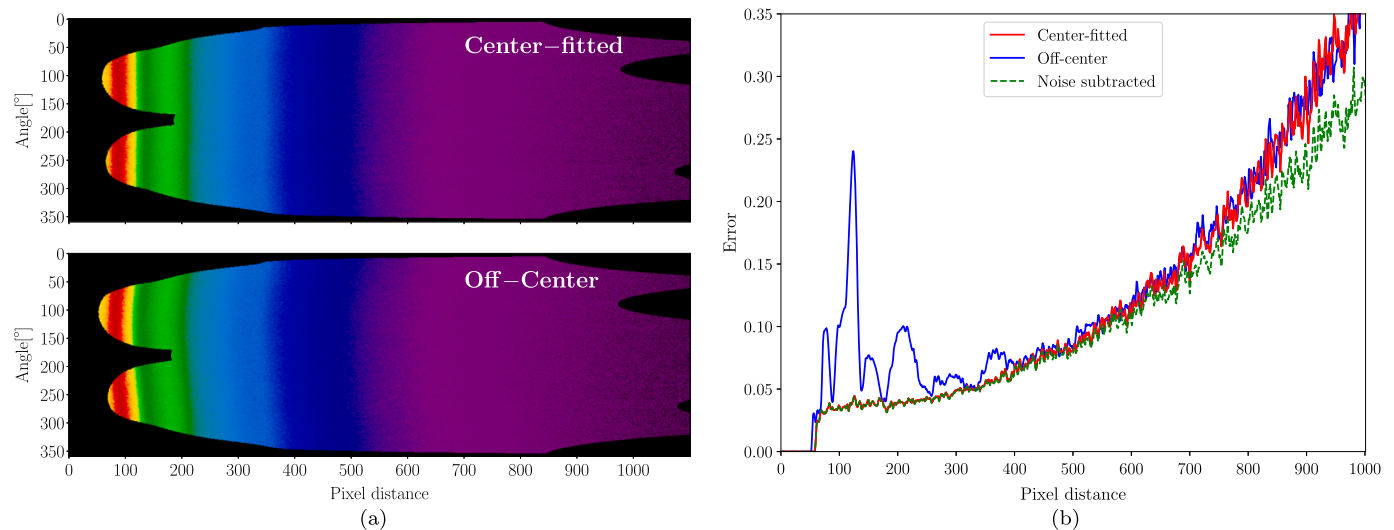


Fig. 11. Results of image processing on a-Ta₂O₅ diffraction pattern. (a) Polar-transformed image of center-fitted (top) and off-center (bottom). (b) The calculated error of polar-transformed images with respect to distance. For convenience, the error curve is smoothed by averaging the value in the nearest 3 data points. The dashed line is used to distinguish data in overlapped areas.

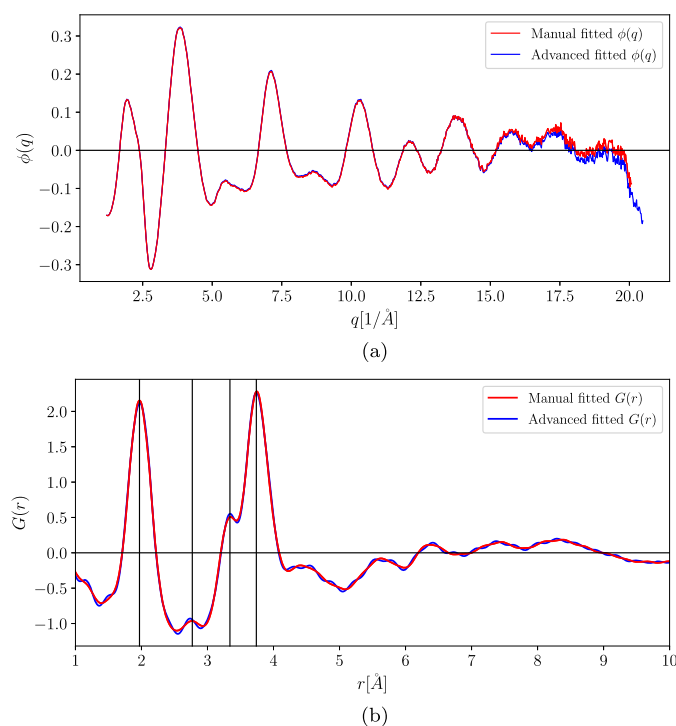


Fig. 12. Comparison between manual fitting and advanced fitting of a-Ta₂O₅ analysis. Calibration factor is extracted from different standard sample data. (a) Reduced intensity function $\phi(q)$ (b) Reduced PDF $G(r)$ of a-Ta₂O₅ sample. All results were calculated from noise-subtracted images, and the analysis was done with both manual fitting and advanced fitting for comparison.

reproducible file save-load system enhance the user's productivity and accuracy in analysis. The obtained $G(r)$ results of ePDFpy are expected to be widely used in further analysis of local atomic structure studies on disordered materials, based on electron diffraction experiments.

CRediT authorship contribution statement

Minhyo Kim: Writing – original draft, Software, Methodology. **Pil-sung Kim:** Software. **Riccardo Bassiri:** Writing – review & editing. **Kiran Prasai:** Writing – review & editing, Conceptualization. **Martin M.**

Fejer: Writing – review & editing. **Kyung-ha Lee:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kyung-ha Lee reports financial support was provided by National Research Foundation of Korea (grant number 2021R1C1C1009248). Martin Fejer, Riccardo Bassiri, Kiran Prasai reports financial support was provided by National Science Foundation (grant number PHY-2011571 and PHY-2011706). Martin Fejer, Riccardo Bassiri, Kiran Prasai reports financial support was provided by Gordon and Betty Moore Foundation (grant number 6793). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1C1C1009248).

Bassiri, Prasai, and Fejer acknowledge the support of the LSC Center for Coatings Research, jointly funded by the National Science Foundation (NSF) and the Gordon and Betty Moore Foundation (GBMF); through NSF awards PHY-2011571 and PHY-2011706, and GBMF Grant No. 6793.

Appendix A. Installation of ePDFpy

ePDFpy is based on the PyQt5 GUI toolkit, with various supporting libraries such as Hyperspy, and Mrcfile for loading multiple diffraction data. Users can install via PyPI distribution (`pip install epdfpy`), where the source code can be downloaded from the GitHub (<https://github.com/GWlab-SKKU/ePDFpy>). The detailed information on the installation, such as the required libraries to set up the working Python environment, is described in the readme file in the source code. In addition, the detailed user guide markdown file is also available in the above GitHub.

References

- [1] J. Sakabe, N. Ohta, T. Ohnishi, K. Mitsuishi, K. Takada, Porous amorphous silicon film anodes for high-capacity and stable all-solid-state lithium batteries, *Commun. Chem.* 1 (1) (2018) 24.
- [2] J. Aasi, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, et al., Advanced ligo, *Class. Quantum Gravity* 32 (7) (2015) 074001.
- [3] J. Cowley, Short-range order and long-range order parameters, *Phys. Rev.* 138 (5A) (1965) A1384.
- [4] P. Voyles, D. Muller, Fluctuation microscopy in the stem, *Ultramicroscopy* 93 (2) (2002) 147–159.
- [5] B.E. Warren, X-Ray Diffraction, Courier Corporation, 1990.
- [6] T. Egami, S.J. Billinge, Underneath the Bragg Peaks: Structural Analysis of Complex Materials, Elsevier, 2003.
- [7] D. Cockayne, D. McKenzie, Electron diffraction analysis of polycrystalline and amorphous thin films, *Acta Crystallogr., Sect. A, Found. Crystallogr.* 44 (6) (1988) 870–878.
- [8] S.J. Billinge, The rise of the x-ray atomic pair distribution function method: a series of fortunate events, *Philos. Trans. R. Soc. A* 377 (2147) (2019) 20180413.
- [9] D.I. Grimley, A.C. Wright, R.N. Sinclair, Neutron scattering from vitreous silica IV. Time-of-flight diffraction, *J. Non-Cryst. Solids* 119 (1) (1990) 49–64.
- [10] K. Prasai, J. Jiang, A. Mishkin, B. Shyam, S. Angelova, R. Birney, D. Drabold, M. Fazio, E. Gustafson, G. Harry, et al., High precision detection of change in intermediate range order of amorphous zirconia-doped tantalum thin films due to annealing, *Phys. Rev. Lett.* 123 (4) (2019) 045501.
- [11] A.M. Abeykoon, C.D. Malliakas, P. Juhás, E.S. Bozin, M.G. Kanatzidis, S.J. Billinge, Quantitative nanostructure characterization using atomic pair distribution functions obtained from laboratory electron microscopes, 2012.
- [12] Y. Liao, Practical Electron Microscopy and Database, An Online Book, 2006.
- [13] J.B.S. Junior, G.R. Schleder, J. Bettini, L.C. Nogueira, A. Fazzio, E.R. Leite, Pair distribution function obtained from electron diffraction: an advanced real-space structural characterization tool, *Mater* 4 (2) (2021) 441–460.
- [14] J.M. Cowley, Electron Diffraction Techniques, vol. 2, Oxford University Press, 1992.
- [15] D.J. Cockayne, The study of nanovolumes of amorphous materials using electron scattering, *Annu. Rev. Mater. Res.* 37 (2007) 159–187.
- [16] Y. Hirotsu, T. Ohkubo, I.-T. Bae, M. Ishimaru, Electron diffraction structure analysis for amorphous materials, *Mater. Chem. Phys.* 81 (2–3) (2003) 360–363.
- [17] H. Shi, M. Luo, W. Wang, epdf tools, a processing and analysis package of the atomic pair distribution function for electron diffraction, *Comput. Phys. Commun.* 238 (2019) 295–301.
- [18] D.T. Tran, G. Svensson, C.-W. Tai, Suepdf: a program to obtain quantitative pair distribution functions from electron diffraction data, *J. Appl. Crystallogr.* 50 (1) (2017) 304–312.
- [19] J. Shanmugam, K.B. Borisenko, Y.-J. Chou, A.I. Kirkland, erdf analyser: an interactive gui for electron reduced density function analysis, *SoftwareX* 6 (2017) 185–192.
- [20] B.H. Savitzky, S.E. Zeltmann, L.A. Hughes, H.G. Brown, S. Zhao, P.M. Pelz, T.C. Pekin, E.S. Barnard, J. Donohue, L.R. DaCosta, et al., py4dstem: a software package for four-dimensional scanning transmission electron microscopy data analysis, *Microsc. Microanal.* 27 (4) (2021) 712–743.
- [21] E.J. Kirkland, *Advanced Computing in Electron Microscopy*, vol. 12, Springer, 1998.
- [22] P. t. Doyle, P. Turner, Relativistic Hartree–Fock x-ray and electron scattering factors, *Acta Crystallogr., Sect. A Cryst. Phys. Diff. Theor. Gen. Crystallogr.* 24 (3) (1968) 390–397.
- [23] G. Bradski, The OpenCV library, Dr. Dobb's J. Softw. Tools (2000).
- [24] E. Bøjesen, T.C. Petersen, A. Martin, M. Weyland, A.C. Liu, Statistical measures of angular correlations in amorphous materials from electron nano-diffraction in the scanning/transmission electron microscope, *J. Phys. Mater.* 3 (4) (2020) 044002.
- [25] V.D.-H. Hou, D. Li, A method to correct elliptical distortion of diffraction patterns in tem, *Microsc. Microanal.* 14 (S2) (2008) 1126–1127.
- [26] I. Lobato, D. Van Dyck, An accurate parameterization for scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints, *Acta Crystallogr. A, Found. Adv.* 70 (6) (2014) 636–649.
- [27] G. Anstis, Z. Liu, M. Lake, Investigation of amorphous materials by electron diffraction—the effects of multiple scattering, *Ultramicroscopy* 26 (1–2) (1988) 65–69.
- [28] R. Bassiri, K. Borisenko, D. Cockayne, J. Hough, I. MacLaren, S. Rowan, Probing the atomic structure of amorphous ta₂o₅ coatings, *Appl. Phys. Lett.* 98 (3) (2011) 031904.
- [29] B. Shyam, K.H. Stone, R. Bassiri, M.M. Fejer, M.F. Toney, A. Mehta, Measurement and modeling of short and medium range order in amorphous ta₂o₅ thin films, *Sci. Rep.* 6 (1) (2016) 1–7.
- [30] M. Hussain, K. Niihara, K. Fukumi, Structure determination of silica tantalum soda glasses by x-ray diffraction analysis, *Mater. Lett.* 24 (1–3) (1995) 69–73.