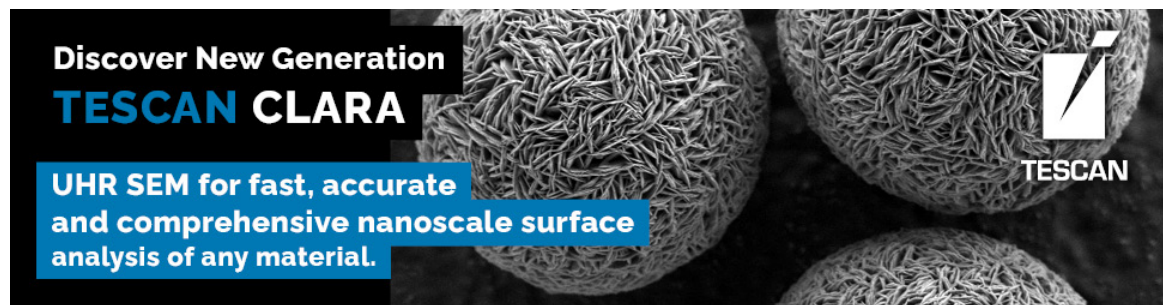




Wigner Distribution Deconvolution Adaptation for Live Ptychography Reconstruction

Arya Bangun, Paul F Baumeister, Alexander Clausen, Dieter Weber, Rafal E Dunin-Borkowski



Wigner Distribution Deconvolution Adaptation for Live Ptychography Reconstruction

Arya Bangun^{1,*} , Paul F. Baumeister², Alexander Clausen¹, Dieter Weber¹ , and Rafal E. Dunin-Borkowski¹

¹Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany

²Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

*Corresponding author: Arya Bangun, E-mail: a.bangun@fz-juelich.de

Abstract

We propose a modification of Wigner distribution deconvolution (WDD) to support live processing ptychography. Live processing allows to reconstruct and display the specimen transmission function gradually while diffraction patterns are acquired. For this purpose, we reformulate WDD and apply a dimensionality reduction technique that reduces memory consumption and increases processing speed. We show numerically that this approach maintains the reconstruction quality of specimen transfer functions as well as reduces computational complexity during acquisition processes. Although we only present the reconstruction for scanning transmission electron microscopy datasets, in general, the live processing algorithm we present in this paper can be applied to real-time ptychographic reconstruction for different fields of application.

Key words: live processing, ptychography, reconstruction, Wigner distribution deconvolution

Introduction

Four-dimensional scanning transmission electron microscopy (4D STEM) is an experimental modality where a wide range of computational methods can extract information on the specimen and reduce the acquired data for human interaction (Ophus, 2019). The acquisition schema of 4D STEM is shown in Fig. 1. Acquiring such comprehensive data and applying computational analysis and reconstruction workflows allows observation of material properties by electron microscopy that are not accessible with simple detectors and signal processing methods (Ophus, 2019). Since it generates large amounts of data (Spurgeon et al., 2021), making algorithms and implementations efficient in their use of computer memory and processing time requires special attention.

If a computational method to process the recorded data is only implemented for offline use, microscopists have to acquire data relying only on simple contrast methods or even without any feedback at all. An implementation for live processing, in contrast, allows interactive use of the microscope based on advanced computational contrast mechanisms, monitoring the acquisition process, evaluating data quality, or automatically controlling the instrument in a closed loop. This requires suitable interfaces to the microscope to receive a live data stream as well as implementations that are capable to process data gradually while they arrive (Nord et al., 2020).

Ptychography (Hoppe, 1969a, 1969b; Hoppe & Strube, 1969) can be used to extract a quantitative object transmission function for a specimen using 4D STEM data. It takes advantage of information from local overlap of the illuminated regions. Recent years have seen a widespread increase in the development of ptychography algorithms by different approaches, such as inspired by the classical alternating

projection method, i.e., PIE reconstruction algorithm (Rodenburg & Faulkner, 2004; Maiden & Rodenburg, 2009), other optimization-based approaches (Wen et al., 2012; Bostan et al., 2018), as well as direct methods. Single-side band (SSB) (Rodenburg et al., 1993; Pennycook et al., 2015; Yang et al., 2015) and Wigner distribution deconvolution (WDD) (Rodenburg & Bates, 1992; Li et al., 2014; Yang et al., 2016) are examples of direct ptychography methods that extract the relevant information in a sequential processing flow, as opposed to iterative methods that optimize the object transmission function in a loop over the input data.

With SSB as an example, Strauch et al. (2021b) showed previously that direct, linear methods are particularly suitable for live processing since the result can always be expressed as the sum of partial results from processing subsets of the input data. However, reconstruction with SSB relies on the weak phase object approximation. Compared with SSB, WDD does not have this limitation (Yang et al., 2017). At the same time, the data reduction in WDD is a linear function of the input data like in SSB, meaning it is a good candidate for live ptychography. Yang et al. (2015) showed that segmented or pixelated detector geometries transfer a majority of the possible phase contrast for STEM with a focused probe at atomic resolution with 16 channels already. Strong binning is used as dimensionality reduction method in Pelz et al. (2021) to reduce the processing time for SSB after an acquisition is completed.

Related Works

A real-time phase reconstruction approach based on integrated center of mass (COM) is proposed in Yu et al. (2021). That method does not require storing the entire 4D dataset in memory and reconstructs the phase from strongly

Received: December 2, 2022. Revised: January 22, 2023. Accepted: February 5, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Microscopy Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

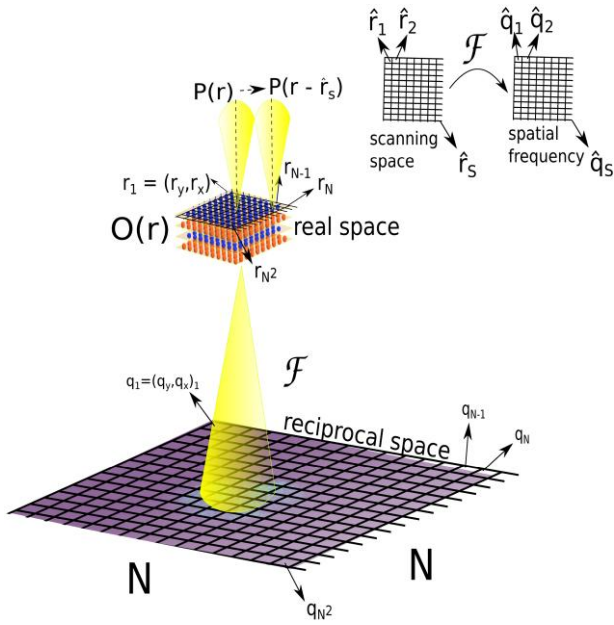


Fig. 1. Four-dimensional scanning transmission electron microscopy acquisition for a scanning point $s \in [S]$ with definition of real space grid $\mathbf{r} = (r_y, r_x) \in \mathbb{R}^2$, reciprocal space grid $\mathbf{q} = (q_y, q_x) \in \mathbb{R}^2$, scanning grid $\hat{\mathbf{r}} = (\hat{r}_y, \hat{r}_x) \in \mathbb{R}^2$ and spatial frequency grid $\hat{\mathbf{q}} = (\hat{q}_y, \hat{q}_x) \in \mathbb{R}^2$.

reduced COM information instead of full diffraction patterns. The COM can be computed efficiently from diffraction data so that this algorithm can reconstruct large-scale data. The authors coined the real-time approach as riCOM.

Summary of Contributions

Here, we demonstrate live reconstruction using the eigenfunctions of a harmonic oscillator (HO) instead of binning as a base for dimensionality reduction. In addition to data reduction, this transformation can replace the Fourier transform in several steps of the WDD method while retaining a reconstruction that is very similar to a result without dimensionality reduction. Furthermore, by building on the matrix representation of the discrete Fourier transform, we can process the intensity data and reconstruct the specimen transmission function gradually from subsets of the input data in a streaming fashion. As a benchmark, we compare the result of Live WDD with the implementation of WDD in PyPtychoSTEM (Pennycook & Hofer, 2021) as well as Live SSB (Strauch et al., 2021b).

The codes used and to reproduce the result in this paper are available at the URL below: <https://github.com/Ptychography-4-0/LiveWDD>

Notations

We provide notations used throughout this article. Vectors are written in bold small-cap letters $\mathbf{x} \in \mathbb{C}^L$ and matrices are written as a bold big-cap letter $\mathbf{A} \in \mathbb{C}^{K \times L}$ for a complex field \mathbb{C} and for a real field \mathbb{R} . A matrix can also be written by indexing its elements

$$\mathbf{A} = (a_{k\ell}), \quad \text{where } k \in [K], \ell \in [L].$$

The set of integers is written as $[N] = \{1, 2, \dots, N\}$ and calligraphic letters are used to define functions $\mathcal{A}: \mathbb{C} \rightarrow \mathbb{C}$. Specifically, we denote the discrete two-dimensional Fourier transform by \mathcal{F} . For both matrices and vectors, the

notation \circ is used to represent an element-wise product, also called Hadamard product. \mathbf{A}^T is used to denote the transpose of a matrix \mathbf{A} . The notation $\text{vec}: \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N^2}$ is an operator that vectorizes a matrix, and $\text{mat}: \mathbb{C}^{N^2} \rightarrow \mathbb{C}^{N \times N}$ constructs a matrix from a vector. The complex conjugate is indicated by a bar. For a matrix $\bar{\mathbf{X}}$, it is the element-wise conjugate. The convolution and Kronecker operator are denoted by \otimes and \boxtimes , respectively. The partial derivative is given by ∂ , and ∇ is the nabla or vector differential operator.

Wigner Distribution Deconvolution

In this section, we introduce the matrix representation of the WDD method developed from its original formulation as in Rodenburg & Bates (1992) and an Open Source implementation from Pennycook & Hofer (2021). In order to visualize the definition for different spaces used in this article, we refer to Fig. 1. The specimen transmission function is denoted by a matrix $\mathbf{O} \in \mathbb{C}^{N \times N}$ where values at the row and column indices (i, j) define the object function at a position $\mathbf{r} \in [N^2]$ in the specimen plane. Similarly, values of the illuminating probe $p(\mathbf{r})$ in the specimen plane at position \mathbf{r} without shifting can be written in matrix form as $\mathbf{P} \in \mathbb{C}^{N \times N}$.

For all shifting coordinates in the set of flattened scan position coordinates, $\hat{\mathbf{r}}_s \in \{\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_S\}$, we write the shifted matrix probe as \mathbf{P}_s for $s \in [S]$, where the $S = S_y \times S_x$ is the set of scanning points. Consequently, the matrix for a shifted matrix probe is given by

$$(p_{ij})_s = p(\mathbf{r} - \hat{\mathbf{r}}_s), \quad \text{where } i, j \in [N],$$

Here, we define the object, probe, diffraction patterns, and the scanning points on equi-spaced rectangular grids. Hence, the usual implementation of the discrete Fourier transform can be used. The intensity of the diffraction pattern at each scanning point can be written as

$$\mathbf{I}_s = |\mathcal{F}_r(\mathbf{P}_s \circ \mathbf{O})|^2 \quad \text{for } s \in [S]. \quad (1)$$

Similar to the real space coordinate, each pattern \mathbf{I}_s for scan position s is indexed by flattened reciprocal space coordinates $\mathbf{q}_s \in \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{N^2}\}$, where N^2 denotes the pair index in the two-dimensional grid in reciprocal space. It should be noted that the complete set of diffraction patterns can be written as flattened scanning points index $\mathbf{I} \in \mathbb{R}^{S \times N \times N}$ or as a 4D tensor $\mathbf{I} \in \mathbb{R}^{S_y \times S_x \times N \times N}$ indexed by the two-dimensional scan position $\hat{\mathbf{r}}$ and the two-dimensional reciprocal space coordinate \mathbf{q} , corresponding to the diffraction angle.

The two-dimensional Fourier transform with respect to the object and probe grid coordinates \mathbf{r} is written as \mathcal{F}_r , with respect to the scan coordinate $\hat{\mathbf{r}}$ as $\mathcal{F}_{\hat{r}}$, and the inverse transforms consequently as \mathcal{F}_q^{-1} and $\mathcal{F}_{\hat{q}}^{-1}$. The WDD algorithm estimates the object from the intensity of diffraction patterns and an estimate of the probe. Picking up from Rodenburg & Bates (1992, equation (8)) where a relation between the object's and probe's Wigner distributions $\mathbf{W}_O \in \mathbb{C}^{S_y \times S_x \times N \times N}$ resp. $\mathbf{W}_P \in \mathbb{C}^{S_y \times S_x \times N \times N}$ is shown:

$$\mathbf{W}_O = \frac{\mathcal{F}_q^{-1}(\mathcal{F}_{\hat{r}}(\mathbf{I})\overline{\mathbf{W}_P})}{|\mathbf{W}_P|^2 + \epsilon} \quad (2)$$

with $\mathbf{W}_P^v = \mathcal{F}_q^{-1}(\mathcal{F}_{\hat{r}}(\mathbf{P})\overline{\mathcal{F}_{\hat{r}}(\mathbf{P})})|_{(\mathbf{q}+\hat{\mathbf{q}}_v)}$, where $\mathcal{F}_{\hat{r}}(\mathbf{P})|_{(\mathbf{q}+\hat{\mathbf{q}}_v)}$ denotes a shift in reciprocal space of the Fourier-transformed probe for specific index v .

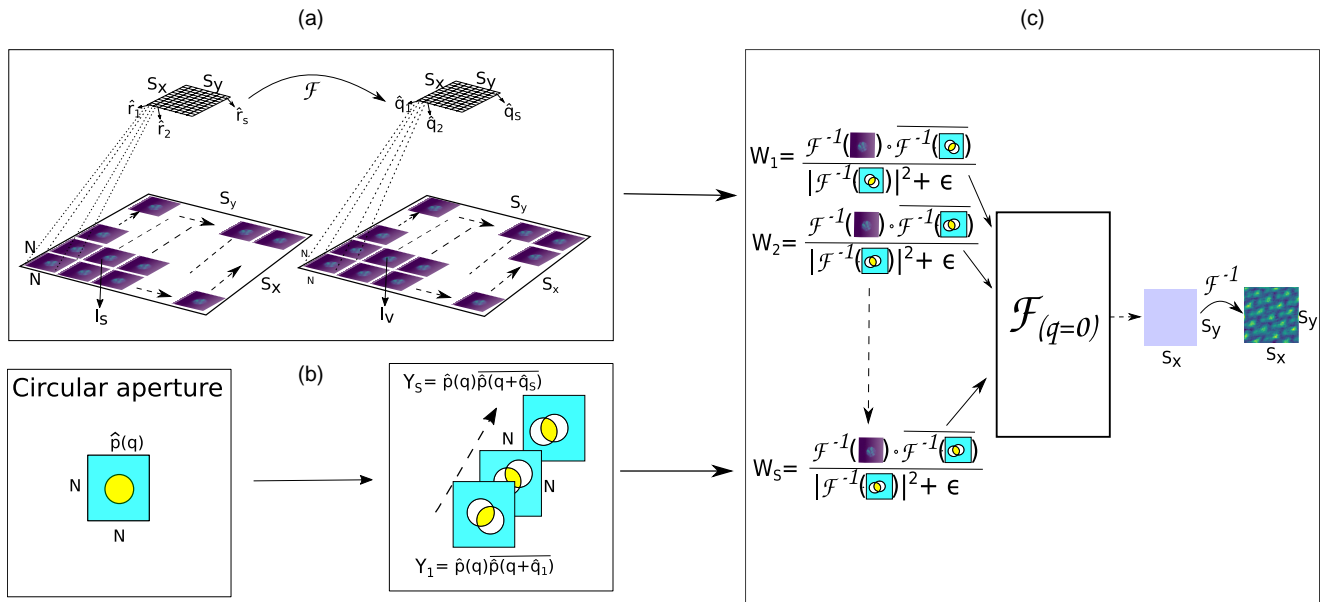


Fig. 2. The schematic diagram for classical WDD. **(a)** The calculation of spatial frequencies \hat{q}_s of a four-dimensional scanning transmission electron microscopy diffraction pattern dataset is given by applying two-dimensional Fourier transform in the real space scan coordinate \hat{r}_s , **(b)** For each spatial frequency, the autocorrelation of the circular aperture is calculated, i.e., initial probe in reciprocal space, **(c)** Apply a two-dimensional inverse Fourier transform for each diffraction pattern dataset on each spatial frequency and the probe autocorrelation and use it for the deconvolution process by using a Wiener filter. Afterward, apply Fourier transform and extract for reciprocal space $\mathbf{q} = 0$ to get an image on spatial frequency before estimating the object in real space scan coordinate with a two-dimensional inverse Fourier transform.

By inserting an estimate of the probe \tilde{P} and calculating the probe's Wigner distribution \tilde{W}_P , one can estimate the object's Wigner distribution \tilde{W}_O . Following Rodenburg & Bates (1992), an estimate for the object \tilde{O} can be obtained from \tilde{W}_O as

$$\tilde{O} = \frac{\mathcal{F}_q^{-1} \left(\mathcal{F}_r(\tilde{W}_O) \Big|_{(q=0)} \right)}{\sqrt{\mathcal{F}_r(\tilde{W}_O) \Big|_{(\hat{q}=0, q=0)}}} \in \mathbb{C}^{S_y \times S_x}, \quad (3)$$

with $\mathbf{q} = 0$ and $(\hat{q} = 0, \mathbf{q} = 0)$ denoting extraction of a subset at the specified coordinates. A visualization of the WDD algorithm is depicted in Fig. 2.

Calculating the deconvolution in (2) may consume large amounts of memory for typical 4D scanning transmission electron microscopy (STEM) data if implemented naively following the equations above since a Fourier transform of the original data, as well as \mathbf{W}_P , which has the same size as the input data, might be instantiated at the same time. Furthermore, they are complex-valued and may require higher numerical precision than the input data. Additionally, this algorithm works on entire datasets, preventing a direct implementation of live processing.

To address the computational complexity and allow live processing, we introduce a dimensionality reduction to reduce the size of the input data and \mathbf{W}_P , and rearrange the WDD algorithm. The modification allows us to process the intensity data sequentially and update the reconstruction as the scan proceeds to acquire new intensity data.

Validation

Since ptychography is a quantitative reconstruction technique, any implementation should demonstrate that it is correct, i.e.,

that it reconstructs arbitrary object functions and/or illuminations quantitatively within its inherent limitations. Simulated datasets of a crystalline specimen resemble the real-world data that ptychography is used on in electron microscopy, but they often have very high symmetry and have no natural orientation. That means errors such as a rolled, phase-reversed or inverted reconstruction, or multiplication with a factor may not be obvious. Such errors can, for example, be caused by mixing Fourier transform and inverse Fourier transform, by adding or omitting a complex conjugation, or by incorrect use of FFT shift resp. inverse FFT shift. For this reason, the implementations used in this paper were validated with a simple procedurally generated asymmetric test image. It is clearly recognizable starting at 25×25 px and contains a wrapped phase ramp at an odd angle that creates a characteristic single spot in the Fourier transform, allowing to visually confirm the correct relation between real and reciprocal space.

A test dataset was created from this test object using multiplication with a synthetic illumination rolled to the scan position, inverse FFT shift, Fourier transform and FFT shift. The illumination was calculated from a synthetic circular aperture with value 17 to catch scaling issues that may not be apparent if the value 1 was used. The illumination function in amplitude and phase as well as the calculated amplitude, phase and intensity from the forward simulation were visually inspected to conform with the expected values. In particular, the simulated diffraction patterns contain a shifted replica of the illuminating aperture at the expected position as a signature of the wrapped phase ramp. A virtual bright field image of the simulated dataset confirmed the correct spatial arrangement of the diffraction patterns. See Fig. 4 for a sample diffraction pattern and a virtual bright field image.

The WDD implementation was then confirmed to reconstruct the object quantitatively in amplitude and phase, taking

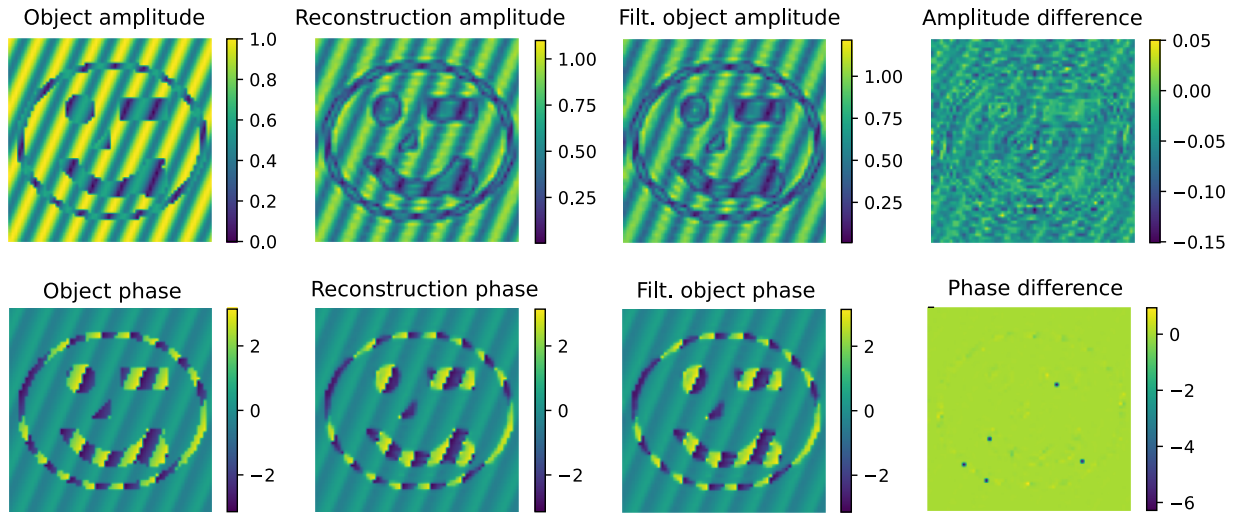


Fig. 3. Validation with a procedurally generated synthetic test dataset: comparing original object, reconstruction, approximate of the expected result calculated by bandpass-filtering the object with twice the aperture size, and difference between reconstruction and approximate expected value.

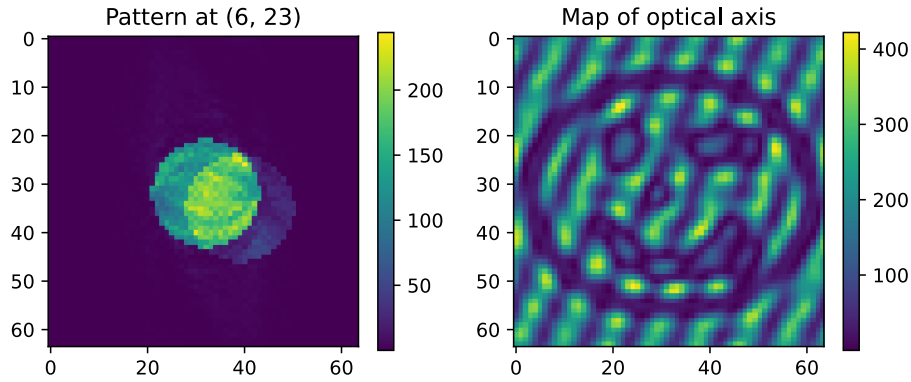


Fig. 4. Sample diffraction pattern and projection along the optical axis from the synthetic test dataset that is used for validation.

the band pass filtering of WDD with twice the semiconvergence angle of the aperture into account (Fig. 3). The correct scale for the illumination of real-world microscopy data can be derived from a vacuum reference, i.e. a scan with the same parameters but without the specimen. This enables quantitative reconstruction of both amplitude and phase of the object function.

Circular Harmonic Oscillator

In this section, we present details of the dimensionality reduction method used.

The rate-limiting step in conventional WDD is the deconvolution of a massive dataset. Therefore, one can improve the performance by projecting the dataset onto a space with lower dimension while retaining the essential information, and performing the resource-intensive steps in this reduced representation. Cropping and binning are simple examples of such projections from higher to lower dimension.

A basis of eigenfunctions of the HO has beneficial properties in this application that are detailed in the following sections. We start first by defining the HO. We provide a brief introduction to this topic and refer the interested readers to the literature (Sakurai, 1994).

One-Dimensional Harmonic Oscillator

The quantum-mechanical HO problem is described by the Hamiltonian operator

$$\hat{H}_\sigma^{[1D]} = -\frac{\partial_x^2}{2} + \frac{x^2}{2\sigma^4}. \quad (4)$$

Hartree atomic units are used in this section. Here, σ is a length scale parameter that also fixes the scale of the eigenvalues $E_n^{[1D]}(\sigma)$ of the operator $\hat{H}_\sigma^{[1D]}$, the so-called eigenenergies. The eigenvalue problem reads

$$\hat{H}^{[1D]} \psi_n^{[1D]}(x) = E_n^{[1D]} \psi_n^{[1D]}(x), \quad n \in \mathbb{N}_0. \quad (5)$$

Then, the energy eigenvalues are

$$E_n^{[1D]}(\sigma) = \sigma^{-2} \left(n + \frac{1}{2} \right) \quad (6)$$

and the corresponding harmonic oscillator (HO) eigenfunctions are

$$\psi_n^{[1D]}(x) = H_n \left(\frac{x}{\sigma} \right) \exp \left(-\frac{x^2}{2\sigma^2} \right) \quad (7)$$

with the Hermite polynomials H_n . It can be seen that these eigenfunctions are Hermite polynomials weighted by a

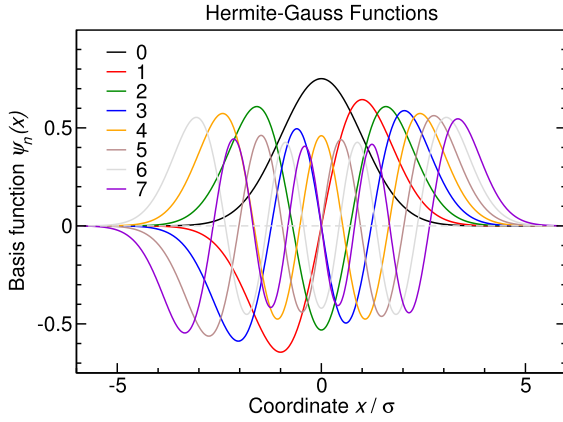


Fig. 5. Normalized Hermite–Gauss functions up to $n_{\max} = 7$. All functions are either even ($\psi(-x) = \psi(x)$) or odd ($\psi(-x) = -\psi(x)$), depending on the parity of n .

Gaussian envelope function. For the usage as a basis, a proper L_2 -normalization is necessary. The normalized Hermite polynomials for several degrees are presented in Fig. 5.

Two-Dimensional Isotropic Harmonic Oscillator

The eigenfunctions of the quantum-mechanical two-dimensional isotropic HO, in the following, we will refer to it as circular harmonic oscillator (CHO), can be written as Cartesian product of two HO eigenfunctions. The Hamiltonian

$$\hat{H}_\sigma^{[2D]} = -\frac{\nabla^2}{2} + \frac{r^2}{2\sigma^4} \quad (8)$$

has the eigenfunctions

$$\psi_{n_x n_y}^{[2D]}(x, y) = \psi_{n_x}^{[1D]}(x/\sigma) \psi_{n_y}^{[1D]}(y/\sigma) \quad (9)$$

and the eigenenergies

$$E_{n_x n_y}^{[2D]}(\sigma) = E_{n_x}^{[1D]}(\sigma) + E_{n_y}^{[1D]}(\sigma) = \sigma^{-2}(n_x + n_y + 1). \quad (10)$$

Special Property of the Basis

The HO eigenfunctions with $\sigma = 1$ are also eigenfunctions of the Fourier transform (FT) operator, i.e., the Fourier transform of $\psi_n^{[1D]}(x)$ is again a Gauss–Hermite function:

$$\mathcal{F}(\psi_n^{[1D]}(x)) = i^n \psi_n^{[1D]}(q), \quad (11)$$

for q being the reciprocal space coordinate adjoint to x . For a nonunity spread, i.e., any $\sigma > 0$, the FT produces a spread of $1/\sigma$ in reciprocal space.

Furthermore, transforming a function by this eigenfunction changes a convolution into multiplication similar to a Fourier transform, which is discussed in Glaeske (1983, Theorem 4.1). That means transforming into this basis can replace Fourier transforms in the WDD algorithm.

Dimensionality Reduction for WDD

The Hermite–Gauss functions defined in (7) can be used as a basis for dimensionality reduction of STEM datasets. We first introduce the matrix notation from sampled Hermite–Gauss function before presenting the procedure to reduce the

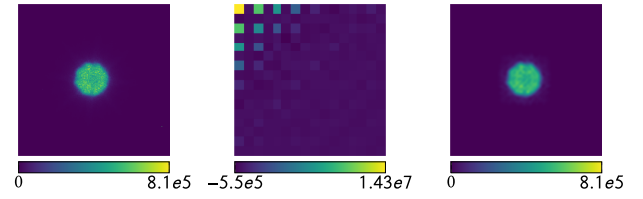


Fig. 6. Dimensionality reduction of PACBED data from Strontiumtitanat (SrTiO₃) dataset (Strauch et al., 2021a). (a) PACBED data, (b) Lower Dimension (16 × 16), and (c) Reconstructed PACBED.

dimension of the data. Second, the transformation will be presented as well as a numerical example for the transformation.

The sampling grid x and scaling factor σ in (7) should be adapted so that the basis is centered with respect to the diffraction patterns and scaled to cover the area relevant for WDD, i.e., the area illuminated by the primary beam.

We can construct a matrix from sampled Hermite–Gauss functions as presented below

$$\psi_x = \begin{pmatrix} \psi_{n_1}^{[1D]}(x_1) & \psi_{n_2}^{[1D]}(x_1) & \cdots & \psi_{n_L}^{[1D]}(x_1) \\ \psi_{n_1}^{[1D]}(x_2) & \psi_{n_2}^{[1D]}(x_2) & \cdots & \psi_{n_L}^{[1D]}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{n_1}^{[1D]}(x_N) & \psi_{n_2}^{[1D]}(x_N) & \cdots & \psi_{n_L}^{[1D]}(x_N) \end{pmatrix} \in \mathbb{R}^{N \times L},$$

where we construct a sampling grid with respect to the COM in the x direction. After introducing the matrix representation of the Hermite–Gaussian function, we can define the dimensionality reduction by using this matrix. For a $\mathbf{A} \in \mathbb{R}^{N \times N}$, the projection to the lower dimension using Hermite–Gauss functions can be expressed as

$$\mathcal{H}(\mathbf{A}) := \psi_x^T \mathbf{A}^T \psi_y \in \mathbb{R}^{L \times L}. \quad (12)$$

This function $\mathcal{H}: \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{L \times L}$ where $L \ll N$ maps from the data with dimension $N \times N$ to a lower dimension $L \times L$. In this case, if we have dimension of diffraction patterns $N \times N$, we can transform the diffraction pattern on the space of Hermite–Gauss functions with smaller dimension $L \times L$.

In contrast to the original two-dimensional discrete Fourier transformation that preserves the dimension of the dataset, this transformation allows flexibility to reduce the dimension. Hence, we can apply this dimensional reduction technique in the deconvolution step in WDD as presented below

$$\begin{aligned} \mathcal{H}(\mathbf{I}_v) &= \mathcal{H}(\mathbf{X}_v \otimes \mathbf{Y}_v) \quad \text{for } v \in [S] \\ &= \mathcal{H}(\mathbf{X}_v) \mathcal{H}(\mathbf{Y}_v), \end{aligned} \quad (13)$$

where now the dimension is reduced from $N \times N$ to $L \times L$ with $L \ll N$. The variable \mathbf{I}_v is the intensity of diffraction pattern of specific spatial frequency coordinate $v \in [S]$. Variables \mathbf{Y}_v and \mathbf{X}_v represent the autocorrelation of the illuminating probe and the autocorrelation of the specimen transmission function, respectively. In parallel with classical WDD, the $\mathcal{H}(\mathbf{Y}_v)$ and $\mathcal{H}(\mathbf{X}_v)$ represent functions at specific spatial frequency v that have similar properties to the Wigner distribution function of the probe and object in conventional WDD.

Transformation to lower dimension can be seen in Fig. 6, where we have reduced from dimension 256×256 to 16×16 . Since we apply dimensionality reduction to our 4D STEM data, we also have to apply it to the autocorrelation

of the illuminating probe to proceed with the deconvolution process, as shown in (13). The transformation from convolution into multiplication by applying the operator \mathcal{H} is derived from the properties of Hermite-Gaussian functions.

The choice of lower dimension L depends highly on the diffraction pattern from the datasets. In this section, we merely show that for the dataset given in Strauch et al. (2021a), dimension 16×16 is enough to capture the essential Hermite-Gauss functions that represent the diffraction pattern for STEM with a focused probe at atomic resolution. In general, for different datasets, the choice of dimension L can be adjusted as long as the upper left triangular structure is preserved. Numerical evaluation for different dimensions L and the effect on the reconstruction is presented in the Numerical Results Section.

Live Processing WDD

A conventional WDD implementation following the procedure in the Wigner Distribution Deconvolution Section works on an entire dataset, which is not suitable for true live processing. A WDD implementation for live processing should update the estimate for the object transmission function gradually by processing individual diffraction patterns as they arrive from the acquisition process. In WDD a Fourier transform is applied over the scan position coordinates, which is usually performed at once with an FFT in a conventional implementation, hence requiring the entire dataset.

Furthermore, live processing should be fast enough to keep up with typical detector speeds that are in the kHz range for 4D STEM. The strategy to support live processing with WDD consists of three steps:

1. Implementation of the Fourier transform over the scan position coordinates using multiplication with a partial DFT matrix instead of FFT.
2. Separation and precomputation of variables that can be computed independent of the diffraction patterns, in this case, the Wigner distribution function of the probe and Wiener filter in (2).
3. Dimensionality reduction to reduce number of individual computations.

Fourier Transform

We begin with a quick introduction of implementing a Fourier transform with a DFT matrix.

A one-dimensional Fourier transform can be implemented by constructing a Fourier matrix taken from sampled Fourier basis as follows (Jain, 1989, equation (5.44)),

$$\mathbf{F} = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-i2\pi f_1 x_1/N} & e^{-i2\pi f_1 x_2/N} & \dots & e^{-i2\pi f_1 x_N/N} \\ e^{-i2\pi f_2 x_1/N} & e^{-i2\pi f_2 x_2/N} & \dots & e^{-i2\pi f_2 x_N/N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi f_N x_1/N} & e^{-i2\pi f_N x_2/N} & \dots & e^{-i2\pi f_N x_N/N} \end{pmatrix} \in \mathbb{C}^{N \times N},$$

here the x_j for $j \in [N]$ represents the sample points on the evenly spaced scan coordinates, and f_j for $j \in [N]$ is the sample on the Fourier space. A discrete Fourier transform can be computed through matrix multiplication with \mathbf{F} . A similar approach can be done to implement a two-dimensional Fourier

transform by applying the Kronecker product of two one-dimensional Fourier matrices (Jain, 1989, equation (5.68)).

$$\mathbf{F}_{2D} = \mathbf{F} \otimes \mathbf{F} \in \mathbb{C}^{N^2 \times N^2}.$$

The Kronecker product constructs a block matrix with the total dimension of the product of the original dimension of two matrices. It should be noted that it is also possible to apply the Kronecker product even if both matrices are not square.

Suppose we have 4D STEM data $\mathbf{I}_s \in \mathbb{R}^{N \times N}$ for evenly spaced scanning points $s \in [S]$. Hence, we can vectorize our datasets into $\mathbf{I} \in \mathbb{R}^{S \times N^2}$, where the row and column space represent all scanning points and the dimension of the microscope's detector, respectively. The Fourier transform along the scan coordinates can be done by calculating the matrix product between the two-dimensional Fourier matrix and the datasets as expressed in the following:

$$\hat{\mathbf{I}} = \mathbf{F}_{2D} \mathbf{I} \in \mathbb{C}^{S \times N^2}.$$

Let us write the matrix as a collection of all vectors on the column space $\mathbf{F}_{2D} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_S) \in \mathbb{C}^{S \times S}$ and the dataset as $\mathbf{I}^T = (\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_S) \in \mathbb{R}^{N^2 \times S}$. Applying the property of matrix product, which can be expressed as the sum of outer product between column and row element of both matrices (Golub & Van Loan, 2013, Section 1.1.14, Algorithm 1.1.8), we can write the following:

$$\mathbf{F}_{2D} \mathbf{I} = \sum_{s=1}^S \mathbf{f}_s \mathbf{i}_s^T \in \mathbb{C}^{S \times N^2}, \quad (14)$$

This sum is trivial to split into partial sums for parts of the input data \mathbf{I} , allowing gradual processing with a live updating partial result. A reformulation of the WDD algorithm using this approach will be presented in the Modified WDD Section.

Precomputed Wiener Filter

In the conventional WDD as presented in the Wigner Distribution Deconvolution Section, we calculate the probe's Wigner function as the initial parameter for deconvolution. The initial guess of the illuminating probe can be generated from the acquisition settings, such as the semiconvergence angle, to compute the circular aperture in Fourier space. Therefore, the autocorrelation of the initial probe can be precomputed. The complete algorithm is presented in Algorithm 1.

The schematic diagram for the precomputed Wiener filter is presented in Fig. 7 where starting from the initial probe on the reciprocal space, i.e., circular aperture, we perform the correlation with respect to the shifted position on the spatial frequency. The results present the trotter shape yielding an intersection between both circular apertures. After calculating the correlation function, we continue with the dimensionality reduction to reduce the dimension before using the compressed correlation to calculate the Wiener filter with a small number to avoid zero division ϵ .

In typical cases for electron microscopy where the illumination is a focused convergent beam with an angular range limited by a circular aperture, we calculate the autocorrelation between shifted circular apertures. At some spatial frequencies $\hat{\mathbf{q}}_v$, we may not have an intersection at all. These frequencies do not contribute to the reconstruction and can be omitted from the calculation. An illustration is given in Fig. 8.

Algorithm 1 Precomputed Wiener filter for Live WDD**1: Initialization:**

1. Initial probe on the Fourier space $(p_{ij}) = \hat{p}(\mathbf{q})$, for $i, j \in [N]$ i.e., circular aperture, generated from the radius of diffraction patterns.
2. Choose a small number to avoid zero division ϵ

2: **for** each predefined index on the spatial frequency coordinate in the set $v \in \{1, 2, \dots, S\}$
do

- 3: Apply transformation to get physical coordinates of spatial frequency from acquisition setting $\{\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_S\}$
- 4: Construct autocorrelation matrix $\mathbf{Y}_v \in \mathbb{C}^{N \times N}$ for $v \in [S]$. The element of the matrix is $(y_{ij})_v = \hat{p}(\mathbf{q})\hat{p}(\mathbf{q} + \hat{\mathbf{q}}_v)$ for $i, j \in [N]$.
- 5: Calculate dimensionality reduction to the initial probe's autocorrelation $\mathcal{H}(\mathbf{Y}_v)$ for $v \in [S]$. Calculate the Wiener filter $\mathbf{K}_v = \frac{\mathcal{H}(\mathbf{Y}_v)}{|\mathcal{H}(\mathbf{Y}_v)|^2 + \epsilon} \in \mathbb{C}^{L \times L}$ for $v \in [S]$.

7: **end For**

The shift depends on the microscope and acquisition parameters, such as electron wavelength λ , semiconvergence angle θ , the radius of the circle in pixel R , and the scanning shift on the real space coordinates for both axes Δ_y, Δ_x . First of all, let us write the condition when the intersection occurs

$$s_y^2 + s_x^2 \leq 4R^2. \quad (15)$$

It should be noted that for spatial frequency indexes $v_y, v_x \in [S]$, we can write the transformation to physical coordinate given by

$$s_y = v_y \frac{\lambda R}{\Delta_y S \sin \theta} \quad \text{and} \quad s_x = v_x \frac{\lambda R}{\Delta_x S \sin \theta}$$

If we have the same scanning shift $\Delta_y = \Delta_x = \Delta$, we can have the condition

$$(v_y^2 + v_x^2) \frac{\lambda^2}{\Delta^2 S^2 \sin^2 \theta} \leq 4 \quad (16)$$

For all combinations of $v_y, v_x \in [S]$, we can find an upper bound $v_y^2 + v_x^2 \leq 2v^2$, where $v = \max(v_y, v_x) \in [S]$. Hence, we have

$$v \leq \hat{S} = \frac{\sqrt{2}\Delta \sin \theta}{\lambda} S. \quad (17)$$

For specific setting in the acquisition process, we can have $\sqrt{2}\Delta \sin \theta / \lambda < 1$, thereby, only required smaller intersection. To have a concrete number suppose we have a specific value of measurement settings, i.e., $\theta = 32$ mrad, $\Delta = 0.026$ nm,

$\lambda = 2.508$ pm, we have scaling factor $\hat{S} = 0.47S$, which is smaller than the total spatial frequency index S and can be used to improve the computation time of live processing WDD.

Modified WDD

Combining dimensionality reduction, gradual processing, as well as the precomputed Wiener filter, we present the modified WDD for live reconstruction in Algorithm 2. Since we

Algorithm 2 Modified WDD**1: Initialization:**

1. Precompute Wiener filter \mathbf{K}_v for $v \in [S]$ given in Algorithm 1 with the number of intersection index $\hat{S} \leq S$ with applied dimensionality reduction.
2. Initialize source for sequence of diffraction patterns $\mathbf{I}_s \in \mathbb{R}^{N \times N}$ for each scanning point $s \in [S]$.
3. Precompute DFT matrices $\mathbf{F}_y \in \mathbb{C}^{S_y \times S_y}$ and $\mathbf{F}_x \in \mathbb{C}^{S_x \times S_x}$ for efficient on-the-fly computation of subsets of \mathbf{F}_{2D} .
4. Allocate and initialize buffer for object function $\mathbf{O} \in \mathbb{C}^{S_y \times S_x}$ with zero.

2: **for** each given real space scanning point index s in the set $\{1, 2, \dots, S\}$
do

- 3: Apply dimensionality reduction $\mathcal{H}(\mathbf{I}_s) \in \mathbb{R}^{L \times L}$ and vectorization to have $\mathbf{i}_s = \text{vec}(\mathcal{H}(\mathbf{I}_s)) \in \mathbb{R}^{L^2}$.
- 4: Calculate subset of the Fourier matrix \mathbf{F}_{2D} for s from the two precomputed DFT matrices.
- 5: Calculate outer product between all column element in the subset of the Fourier matrix \mathbf{F}_{2D} and vectorized intensity \mathbf{i}_s as in (14), $\mathbf{T} = \mathbf{f}_s \mathbf{i}_s^T \in \mathbb{C}^{S \times L^2}$, where we have $\mathbf{t}_v \in \mathbb{C}^{L^2}$ for $v \in [S]$ for each row of matrix \mathbf{T} .
- 6: **For** each given spatial frequency index with nonzero intersection v in the set $\{1, 2, \dots, \hat{S}\}$ **do**
- 7: Reshape the result for each row, i.e., $\text{mat}(\mathbf{t}_v) \in \mathbb{C}^{L \times L}$.
- 8: Deconvolution process on compressed space $\mathbf{D}_v = \text{mat}(\mathbf{t}_v) \circ \mathbf{K}_v \in \mathbb{C}^{L \times L}$ for $v \in [\hat{S}]$.
- 9: Calculate zero frequency by summing all elements of matrix $\mathbf{D}_v = (d_{lk})_v$ for $l, k \in [L]$. Here, we have scalar for each spatial frequency $o_v = \sum_{l=1}^L \sum_{k=1}^L (d_{lk})_v$.
- 10: Replace the element of the nonzero index in spatial frequency with o_v .
- 11: **end for**
- 12: We have an update vector $\mathbf{o} = (o_1, o_2, \dots, o_S) \in \mathbb{C}^S$. This is added to the buffer allocation matrix for each given scanning point $\mathbf{O} = \mathbf{O} + \text{mat}(\mathbf{o}) \in \mathbb{C}^{S_y \times S_x}$.
- 13: **end for**
- 14: Applying inverse Fourier transform and complex conjugate on set of spatial frequencies to get real space coordinate of the specimen transmission function $\overline{\mathcal{F}^{-1}(\mathbf{O})}$

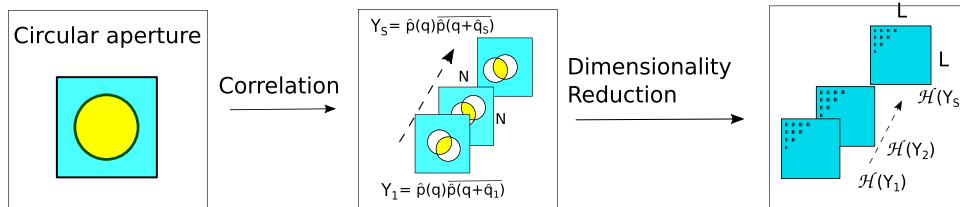


Fig. 7. Schematic diagram for precomputed Wiener filter.

compress and process the diffraction patterns per scanning point in Algorithm 2, the total reconstruction of the specimen transmission function is updated gradually, starting from a zero-initialized matrix. Furthermore, the updates from subsets of the input data can be computed independently, which allows trivial parallelization. This algorithm for Live WDD is implemented as a user defined function (UDF) for LiberTEM-live (Clausen et al., 2020). The complete schematic

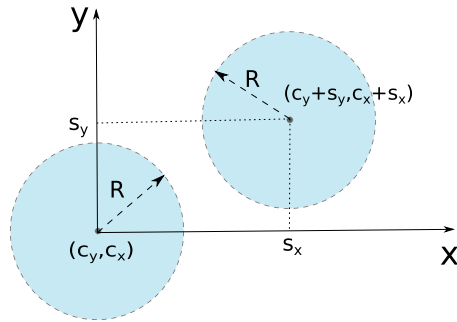


Fig. 8. Geometry of the autocorrelation of the probe in reciprocal space.

diagram for an implementation of WDD as a LiberTEM UDF is presented in Figure 9.

Time and Space Complexity

Here, we discuss the analysis of the Live WDD algorithm and compare it with conventional WDD. For both approaches, we derive the time and space complexity required to perform the steps specified by the algorithm. In this section, we denote the total scanning points as S for raster position on both x, y axis, i.e., $S = S_x \times S_y$. In addition, the dimension of the detector is denoted as $N \times N$. In Table 1, we provide a summary for the time and space complexity analysis for both algorithms.

For space complexity, it appears that the Live WDD scales better compared with conventional WDD since $L \ll N$.

Table 1. Computational Complexity of Conventional WDD and Live WDD.

Complexity	Conventional WDD	Live WDD
Time	$\mathcal{O}(SN^2 \log SN^2)$	$\max(\mathcal{O}(SN^2 L), \mathcal{O}(S^2 L^2))$
Space	$\mathcal{O}(SN^2)$	$\mathcal{O}(SL^2)$

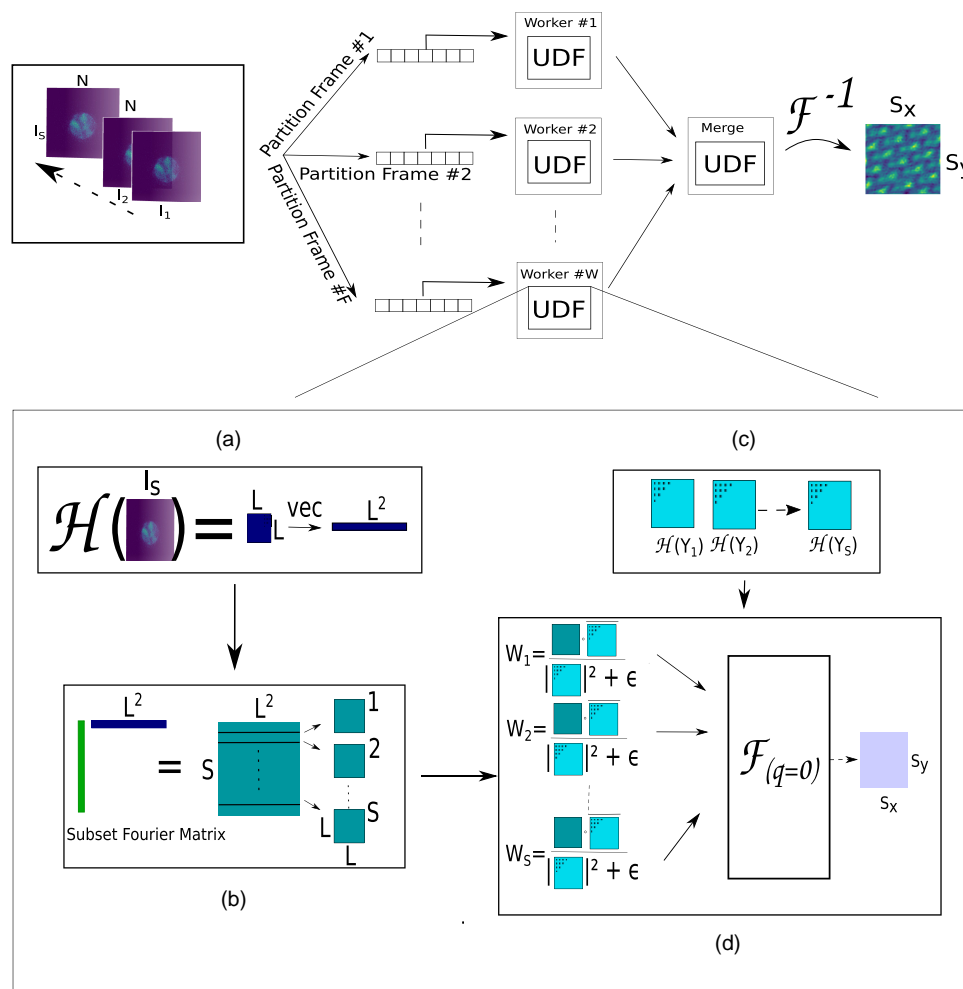


Fig. 9. The schematic diagram for running the Live WDD as a UDF in LiberTEM-live. The data stream from the detector is split into partitions, which are streamed to worker processes via queues. Each diffraction pattern is processed in the UDF as follows: (a) Applying dimensionality reduction and vectorization, (b) Processing the compressed frame with a subset of the Fourier matrix, (c) Accessing the precomputed Wiener filter to perform a single reconstruction as presented in (d). The merge function sums up the contributions from each partition, and the final reconstruction in real space is determined by the two-dimensional inverse Fourier transform.

However, for time complexity, it highly depends on the total scanning points S and the logarithmic factor $\log(SN^2)$ compared with the low dimension L , here we use $L = 16$. The derivation of complexity analysis is provided in the Conventional WDD and Live WDD Sections, respectively. It should be noted that complexity analysis does not predict the run time of the code that is implemented in the computer. It only describes the scaling behavior between the number of operations and the size of input data towards infinite size. However, we also provide the run-time evaluation of our implementation that utilizes LiberTEM Live in the Numerical Results Section.

Conventional WDD

First of all, conventional WDD performs a Fourier transform along the scanning position in real space to obtain the spatial frequencies. The computational complexity of a fast Fourier transform for all scanning points is $\mathcal{O}(SN^2 \log S)$. For each spatial frequency, we calculate the autocorrelation of the probe, which gives us computation time $\mathcal{O}(N^2)$. Afterward, we calculate the Wigner distribution function by applying inverse Fourier transform on the autocorrelation as well as the intensity of diffraction patterns, for each taking $\mathcal{O}(N^2 \log N^2)$. The process is followed by applying Wiener filtering or deconvolution that requires $\mathcal{O}(N^2)$. Since we have to calculate all spatial frequencies, the computation is therefore $S(\mathcal{O}(N^2) + \mathcal{O}(N^2 \log N^2))$, which gives total for processing all spatial frequencies $\mathcal{O}(SN^2 \log N^2)$. The next step is to calculate the Fourier transform of all deconvolution data before taking only zero reciprocal space, i.e., $\mathbf{q} = 0$, hence we perform operation $\mathcal{O}(SN^2 \log N^2)$. In the last step, we apply an inverse Fourier transform for estimated object which requires $\mathcal{O}(S \log S)$. The total time computation for conventional WDD then $\mathcal{O}(SN^2 \log(S)) + \mathcal{O}(SN^2 \log N^2) + \mathcal{O}(S \log S)$. Simplification gives us time complexity of conventional WDD as $\mathcal{O}(SN^2 \log SN^2)$.

For space complexity, we start with total memory allocation to store all 4D datasets to perform a Fourier transform on scanning points on the real space, i.e., to obtain the spatial frequencies, which requires $\mathcal{O}(SN^2)$. For each spatial frequency, we have to process the data with dimension $\mathcal{O}(N^2)$. In the last process to calculate the estimated object, we have space complexity $\mathcal{O}(S)$. Thereby, total space complexity of conventional WDD is $\mathcal{O}(SN^2) + \mathcal{O}(N^2) + \mathcal{O}(S)$, which then we have scaling for space complexity $\mathcal{O}(SN^2)$.

Live WDD

The Live WDD presented here includes a compression step where the dimension of the compressed data is $L \times L$, much smaller than detector size ($L \ll N$). For the time complexity of Live WDD, we separate the processing into the computation of the Wiener filter in Algorithm 1 and the actual data processing. The Wiener filter requires $\mathcal{O}(N^2)$ for computing the autocorrelation function. Since we apply a dimensionality reduction, which is performed with three matrix multiplications, we have $\mathcal{O}(LN^2 + L^2N)$. At the end of the process, we compute the element-wise division for the Wiener filter with complexity $\mathcal{O}(L^2)$ and incorporate the process for all spatial frequencies S . Therefore, we have time complexity for the precomputed Wiener filter as $\mathcal{O}(SN^2L)$. For processing each scanning point, we calculate the dimensionality reduction for each diffraction pattern with complexity $\mathcal{O}(L^2N + N^2L)$. Afterward, the computation of the outer product between each column of Fourier matrix with a compressed

diffraction pattern requires $\mathcal{O}(SL^2)$. The next process is the deconvolution with the precomputed Wiener filter for all spatial frequencies which takes $\mathcal{O}(SL^2)$. Getting zero frequencies and the update data have complexity $\mathcal{O}(SL^2)$ and $\mathcal{O}(S)$, respectively. Thereby, processing with all real space scanning points is given by $\mathcal{O}(SL^2N + SN^2L) + \mathcal{O}(S^2L^2) + \mathcal{O}(S^2)$. In the last step, to get the estimated object on the real space with dimension S , we apply an inverse Fourier transform that requires $\mathcal{O}(S \log S)$.

In the end, for Live WDD we have total time complexity $\mathcal{O}(SL^2N + SN^2L) + \mathcal{O}(S^2L^2) + \mathcal{O}(S^2) + \mathcal{O}(S \log S)$, which then gives us $\max(\mathcal{O}(SN^2L), \mathcal{O}(S^2L^2))$. The result highly depends on the dimension of N and S , where for a larger detector dimension than total spatial frequency, we have time complexity $\mathcal{O}(SN^2L)$. On the contrary, if we have a large field of view, there is a possibility then we have time complexity $\mathcal{O}(S^2L^2)$.

Regarding the space complexity, in Live WDD, we can process the data per frame, i.e., per diffraction pattern, the size of which corresponds to the detector's dimension. In comparison, conventional WDD stores the complete 4D dataset to compute the spatial frequencies. In addition, we apply a dimensional reduction technique that only requires $\mathcal{O}(L^2)$. For each scanning point on the real space as well as the nonzero intersection on the spatial frequency, the algorithm only requires $\mathcal{O}(L^2)$ except for computing the estimated object that has dimension $\mathcal{O}(S)$. For the precomputed Wiener filter, we have to store $\hat{S} \leq S$ nonzero intersection spatial frequencies for the autocorrelation process with compressed dimension $L \times L$. Hence we have $\mathcal{O}(SL^2)$ space complexity of the precomputed Wiener filter. In total, we have $\mathcal{O}(SL^2) + \mathcal{O}(S) + \mathcal{O}(L^2)$ and give us complexity $\mathcal{O}(SL^2)$, which still better than conventional WDD.

In Numerical Results Section, we compare real-world time and space use of the proposed Live WDD and conventional WDD on various datasets.

Simulation and Experimental Datasets

The information of datasets and parameter settings to evaluate Live WDD are given in Pennycook (2021) for simulated graphene. The specimen has a hexagonal lattice structure, shown in Fig. 10. The 4D STEM data are generated by parameter settings presented in Table 2.

We can also add the effect of Poissonian noise to the simulated diffraction patterns data. Suppose we have dose level per pixel represented by variable v that has a unit $e^-/\text{\AA}^2$. The model used to generate a noisy dataset for each intensity of

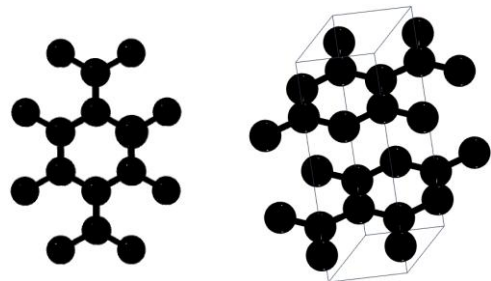
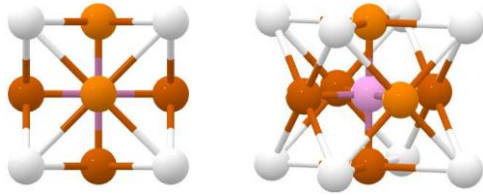


Fig. 10. Structure of graphene. The structure is downloaded from materialsproject.org (Jain et al., 2013).

Table 2. Parameters for Simulated Graphene (Pennycook, 2021).

Parameters	Graphene
Rotation (deg)	0.0
Semiconv. angle (mrad)	30
Accel. voltage (keV)	60
Scanning step size (nm)	0.02
Scanning points (Sy, Sx)	(64, 65)
Detector size (pixel)	(256, 256)

**Fig. 11.** Structure of material SrTiO₃. The structure is downloaded from materialsproject.org (Jain et al., 2013).

diffraction patterns $\mathbf{I}_s \in \mathbb{R}^{N \times N}$ for $s \in [S]$ is given as follows:

$$\text{Poisson}(\tilde{\mathbf{I}}_s) \in \mathbb{R}^{N \times N},$$

where the $\tilde{\mathbf{I}}_s$ is normalization of the diffraction pattern for each scanning point and Poisson is a function to generate Poisson distribution applied to our dataset. It should be noted that this function preserves the dimension of the data.

Additionally, we also used an experimental datasets of SrTiO₃ specimen acquired using Medipix Merlin EM detector (Strauch et al., 2021a), where the structure is presented in Fig. 11. The parameters setting of this dataset is presented in Table 3. Complete information about this dataset can be directly seen (Strauch et al., 2021a).

Numerical Results

We present numerical evaluations of the Live WDD in terms of reconstruction, computation time, and memory allocation. For Live WDD, we also present how the live reconstruction evolves from partial results. As a comparison to the existing WDD implementation, we refer to the implementation in Yang et al. (2016) as the reference to check the dynamic range of the phase, where the source code is available in <https://gitlab.com/ptychoSTEM/ptychoSTEM> and <https://gitlab.com/PyPtychoSTEM/PyPtychoSTEM>. We use the same reconstruction parameters in WDD as given in the source code, for instance, the small constant ϵ for the Wiener filter.

The evaluation is performed individually on the same workstation with AMD EPYC 7543P with 32 CPU with 64 threads operating at a base frequency of 2.80 GHz, and 512 GB DDR4 RAM with an operating frequency of 3.2 GHz. It should be noted that the simulation for each algorithm is performed without a noise background, i.e., no other processes or algorithms were running during the evaluation.

Effect of Dimensionality Reduction

In this section, we also evaluate the effect of dimensionality reduction, represented by $L \in \{8, 16, 32, 64, 128\}$, to the

Table 3. Parameters for Experimental Dataset SrTiO₃ (Strauch et al., 2021a).

Parameters	SrTiO ₃
Rotation (deg)	88
Semiconv. angle (mrad)	22.13
Accel. voltage (keV)	300
Scanning step size (nm)	0.0127
Scanning points (Sy, Sx)	(128, 128)
Detector size (pixel)	(256, 256)

diffraction pattern from the graphene dataset (Pennycook, 2021) as well as the phase and error reconstructions. It should be noted that, since the ground truth is not available, we measure the error with respect to the complex value and phase reconstruction from PyPtychoSTEM.

First, we present the transformation of PACBED of graphene datasets to the space of Hermite–Gauss functions by applying matrix product in (12). Figure 12 shows the structure of Hermite–Gauss coefficients for several dimensions, where the lower dimension $L = 16$ suffices to preserve the important information of diffraction patterns. Hence, it can be used to justify the choice of dimension for graphene datasets. However, for different datasets, the lower dimension L should be adapted. The errors of complex and phase reconstruction of specimen transmission function for different L are presented in Table 4, where we use relative mean square error to measure the difference, as follows:

$$\frac{\sum_{i=1}^{S_y} \sum_{j=1}^{S_x} |o_{ij} - \hat{o}_{ij}|^2}{\sum_{i=1}^{S_y} \sum_{j=1}^{S_x} |o_{ij}|^2},$$

where o_{ij} and \hat{o}_{ij} are element of matrices from complex or phase reconstruction for both PyPtychoSTEM and Live WDD, respectively. It can be seen that the error reconstruction improves starting from $L = 16$. The rest of numerical section are proceeded with dimension $L = 16$.

Reconstruction

We performed numerical comparisons of the Live WDD and the conventional WDD given in the PyPtychoSTEM, with a similar setting for both algorithms, e.g., $\epsilon = 0.01$. In the first part, we focus on the evaluation of noise-free conditions for graphene datasets. The second part covers the performance of algorithms when applied to data that is degraded by Poissonian noise corresponding to different dose levels. Since the goal is to show the specimen transmission function, we present the final phase reconstruction from the Live WDD.

Figure 13 shows the reconstruction of the graphene dataset for both the PyPtychoSTEM implementation of WDD and the Live WDD for noise-free conditions. It can be seen that the Live WDD can reconstruct the specimen transmission function with the correct orientation of the atom represented by the phase and similar dynamic range. In the next evaluation, we conduct numerical evaluation for different dose levels, namely for $v \in \{10^2, 10^3, 10^4, 10^5\} \text{ e}^-/\text{\AA}^2$. The reconstruction for noisy setting is depicted in Fig. 14.

Compared with the conventional WDD, which performs faithful reconstruction starting from dose $v = 10^5 \text{ e}^-/\text{\AA}^2$, Live WDD yields a recognizable reconstruction starting from $v = 10^4 \text{ e}^-/\text{\AA}^2$. Hence, it can be seen that the Live WDD is more robust against Poissonian noise. Besides the

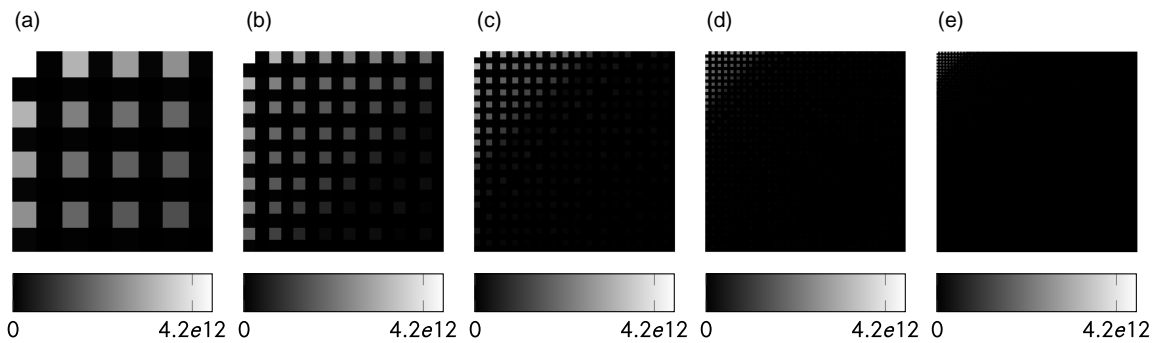


Fig. 12. The lower dimension after dimensionality reduction of PACBED from a graphene dataset (Pennycook, 2021) with dimension: (a) $L = 8$, (b) $L = 16$, (c) $L = 32$, (d) $L = 64$, and (e) $L = 128$.

Table 4. Error Reconstruction in Terms of Dimensionality Reduction.

L	8	16	32	64	128
Complex	$1.1e-4$	$2.3e-5$	$2.1e-5$	$2.1e-5$	$2.2e-5$
Phase	0.12	0.025	0.022	0.022	0.023

property of flexible dimensional reduction, the arbitrary order of the Hermite–Gauss functions can be seen as a low-pass filter, where the width σ is optimized so that the any pixel noise is strongly suppressed by the dimensionality reduction, as in Fig. 5.

A line scan through the reconstruction at dose level $10^4 \text{ e}^-/\text{\AA}^2$ is presented in Fig. 15, which shows that the noise affects the phase reconstruction of both algorithms. It can be seen that the Live WDD has better noise suppression than PyPtychoSTEM, where the latter requires more dose to reliably find atom positions in the phase reconstruction. For infinite dose, both algorithms present the same atom positions with a different value range for Live WDD compared with PyPtychoSTEM. A difference is to be expected because Live WDD reduces the dimensionality using Hermite–Gauss functions, while PyPtychoSTEM uses Fourier space without dimensionality reduction. That means the two methods are not numerically equivalent.

Computation Time

The numerical computation time for Live WDD compared with the conventional WDD given in Yang et al. (2016) is discussed. The evaluation is presented in Tables 5 and 6, where we measure the median, as well as the standard deviation of computation time for live and conventional WDD for different dimension of datasets generated from Graphene in Pennycook (2021). To investigate the effect of both increasing scanning points and detector dimensions on the computation time, we also observe both parameter settings, where we use the convention S_y , S_x as the number of scanning points for both axes in the raster scan. Dimension of detector is given by N_y , N_x . Here, we evaluate by increasing the dimension of graphene datasets, where the larger dimension of scanning points is generated by replicating the datasets. In addition, for the detector dimension, we use zero padding techniques. The focus is merely to provide the conceptual study on the effect of larger datasets on the memory allocation and run time of the proposed algorithm in the code implementation.

We perform 10 trials to measure the numerical computation time for both algorithms. From these measurements, we show

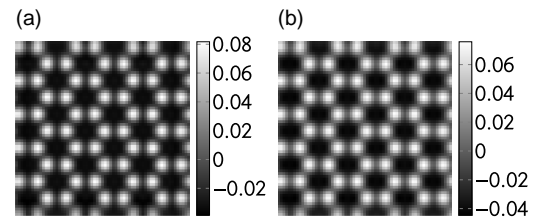


Fig. 13. Phase reconstruction of simulated graphene with (a) PyPtychoSTEM and (b) Live Processing WDD.

the median computation time. It can be seen that PyPtychoSTEM requires more memory than available to accomplish the reconstruction. As discussed in the Time and Space Complexity Section, accommodating an entire dataset requires a large memory allocation for PyPtychoSTEM and impinges on the computation performance in general. In all cases, Live WDD performs faster numerical computation than conventional WDD implemented in PyPtychoSTEM. However, when the same dimension for both scanning points and detector size is evaluated, i.e., (1,024, 1,024), we observe the computation time increases approximately a 100-fold due to the quadratic scaling in the time complexity, as discussed in the Time and Space Complexity Section.

Memory Allocation

Apart from the numerical computation time, we are also interested in independently observing the memory allocation for both algorithms during the reconstruction process. For this reason, we record the memory usage every 0.2 s. The evaluation for memory allocation is also performed independently of the investigation of numerical computation time in the previous section and separately for each algorithm. Similar to the numerical computation time, we present the evaluation for both increasing dimensions of scanning points and detector.

Figure 16 shows that for detector dimension (1,024, 1,024) PyPtychoSTEM requires more memory than available to complete the reconstruction. Live WDD only requires a constant amount of memory around 10^4 MB independent of detector size. The maximum memory allocation to complete each algorithm for increasing detector dimension is presented in Table 7. In addition, we also investigate the effect of increasing scanning points dimension or the dimension of the field of view, as presented in Fig. 17. PyPtychoSTEM requires more

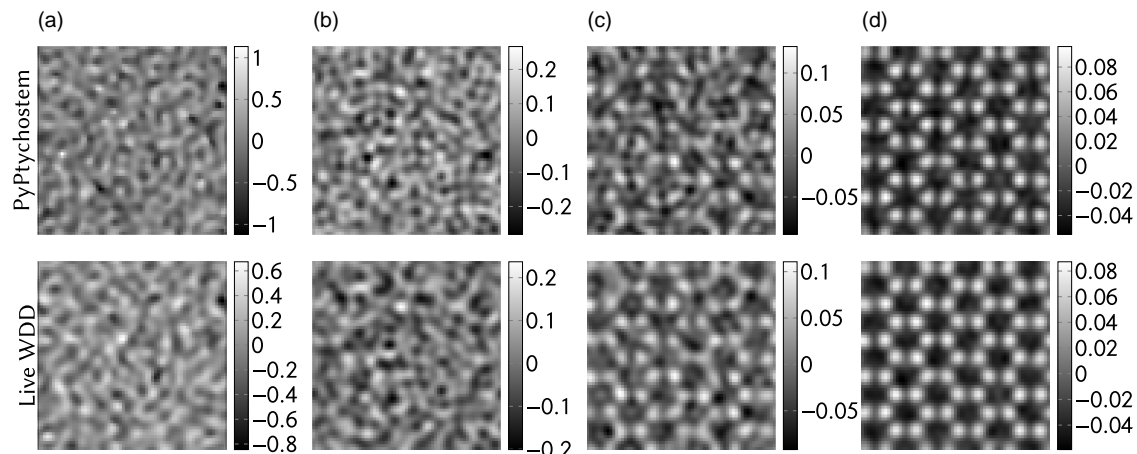


Fig. 14. Phase reconstruction of simulated graphene with different dose levels (a) $10^2 \text{ e}^-/\text{\AA}^2$, (b) $10^3 \text{ e}^-/\text{\AA}^2$, (c) $10^4 \text{ e}^-/\text{\AA}^2$, and (d) $10^5 \text{ e}^-/\text{\AA}^2$.

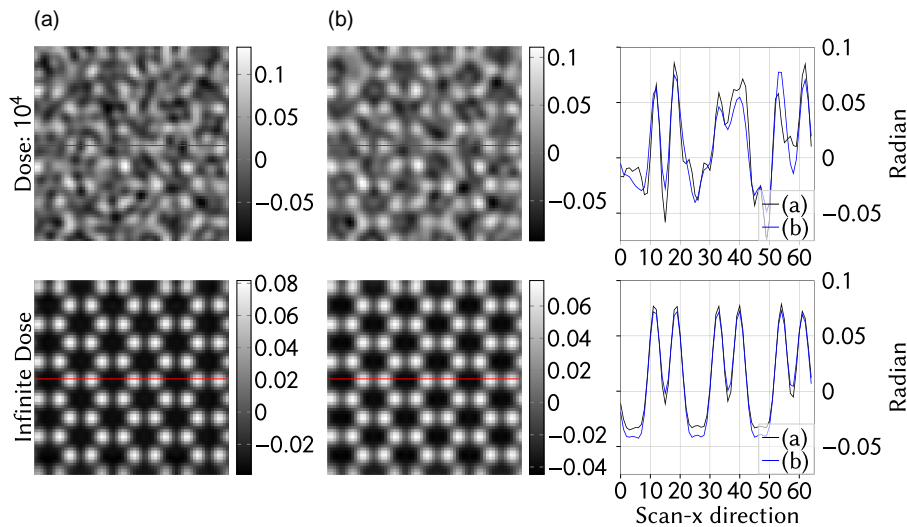


Fig. 15. Phase reconstruction of simulated graphene with dose level $10^4 \text{ e}^-/\text{\AA}^2$ and infinite dose for both algorithms (a) PyPtychoSTEM and (b) Live WDD. The line scan reconstruction is also presented where the scan location on Y-direction: 32.

Table 5. Numerical Median Computation Time in Seconds for PyPtychoSTEM and Live WDD for Fixed Dimension Scanning Points $S_y = S_x = 128$ and Increasing Detector $N_y = N_x$. Here, we use $L = 16$.

Dimension	PyPtychoSTEM	Live WDD
128	48.76 ± 0.14	1.28 ± 0.07
256	145.51 ± 2.16	2.21 ± 0.069
512	631.71 ± 9.05	12.60 ± 0.65
1,024	–	37.42 ± 0.85

Table 6. Numerical Median Computation Time in Seconds for PyPtychoSTEM and Live WDD for Fixed Dimension Detector $N_y = N_x = 128$ and Increasing Scanning Points $S_y = S_x$. Here, we use $L = 16$.

Dimension	PyPtychoSTEM	Live WDD
128	48.76 ± 0.14	1.28 ± 0.07
256	199.12 ± 0.60	14.82 ± 0.05
512	824.04 ± 7.45	228.59 ± 0.18
1,024	–	3048.48 ± 5.94

memory than available to complete the reconstruction for scanning points (1, 024, 1, 024). Also, in this scenario, Live WDD uses memory more efficiently than PyPtychoSTEM. The maximum memory allocation for increasing scanning points is given in Table 8.

Live Processing Evaluation

In this section, we demonstrate that the performance of Live WDD is sufficient for live acquisition and reconstruction with real-world 4D STEM detectors. In 4D STEM, illustrated

in Fig. 1, the acquisition time per scanning point is usually limited by the detector frame rate. The specification for different detectors, namely Merlin MedipixEM,¹ Dectris Quadro,² and Dectris Arina,³ are given in Table 9. The maximum frame rate may depend on the chosen bit depth and

¹ <https://quantumdetectors.com/wp-content/uploads/2022/01/MerlinEM-app-notes.pdf>

² <https://www.dectris.com/detectors/electron-detectors/for-materials-science/quadro/>

³ <https://www.dectris.com/detectors/electron-detectors/for-materials-science/arina/>

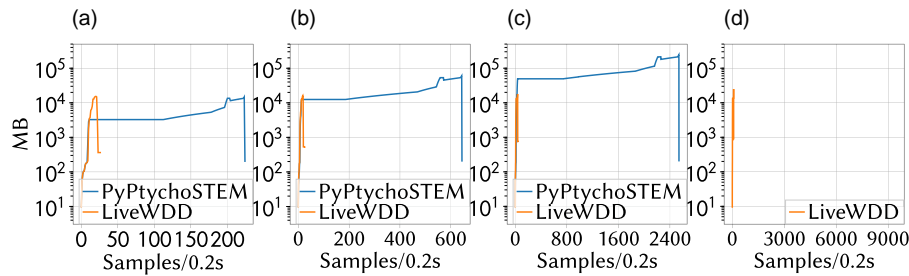


Fig. 16. Memory allocation for conventional WDD and Live WDD ($L = 16$) for different detector sizes: (a) (128,128,128,128), (b) (128,128,256,256), (c) (128,128,512,512), and (d) (128,128,1,024,1,024).

Table 7. Maximum Memory Allocation in MB for PyPtychoSTEM and Live WDD for Fixed Scanning Points Dimension $S_y = S_x = 128$ and Increasing Detector Size $N_y = N_x$. Here, we use $L = 16$.

Dimension	PyPtychoSTEM	Live WDD
128	14,725	15,343
256	60,505	16,136
512	244,557	17,889
1,024	—	24,887

readout area for a given detector. To accomplish a continuous reconstruction, data processing time per detector frame needs to be faster than the STEM dwell time. Following the considerations on computational complexity, this highly depends on the number of scanning points, the detector size, and the number of nonzero entries in the Wiener filter. Since the processing is parallelized using a UDF and LiberTEM-live, the number and speed of CPU cores is a major factor as well.

To illustrate the scalability of Live WDD, we show the scaling behavior of the computation time as a function of number of CPU cores for dimension (128, 128, 128, 128), as presented in Fig. 18. The scaling is nearly linear for up to eight cores and tapers off after that.

Based on the performance data from our 32 core CPU in Tables 5 and 6, we can therefore support live reconstruction up to the frames per second (fps) as presented in Table 10. Therefore, a reconstruction using Live WDD can keep up with Merlin Medipix and Dectris Quadro without ROI up to a dimension of (256, 256, 128, 128) when used with the given setup and settings.

Figure 19 shows simulated live reconstruction for different stages of scanning progress for Live WDD and Live SSB from (Strauch et al., 2021b), where the update is added gradually until completing all scanning points. Here, the dimension of the 4D STEM data is (128, 128, 256, 256), as described in Table 3.

Discussion

Since we only perform the object reconstruction using a synthetic probe initialization, a potential next step for Live WDD could be to factor in the microscope alignment and update the probe. After estimating the object, we can swap the deconvolution process to reconstruct the probe. This process is straightforward, but the additional computation and update for the Wiener filter would affect the performance. Therefore, an efficient and on-fly computation should be implemented to overcome this issue.

The time complexity of Live WDD for a large field of view is scaled quadratically with total scanning points and reduced dimension, respectively. Although the Live WDD implementation can complete the reconstruction, our numerical observation shows that it takes approximately one hour with dimension (1,024, 1,024, 128, 128), which is inefficient. This could be overcome by subdividing the field of view into smaller patches that are reconstructed independently and subsequently merged. In this case, the time complexity can be reduced to a linear scale of the number of subsets.

Another strategy to optimize the performance of Live WDD could be to choose an optimal scan step based on the intersection of the probe's auto correlation in reciprocal space, as presented in (17). Making sure that the scan grid is not unnecessarily fine can reduce the number of scanning points to process. It is also thinkable to adapt WDD and Live WDD for scan patterns that are not on an equispaced grid. In that case, a matrix for a nonuniform discrete Fourier transform should be used to match the scanning points.

In this article, we focus on the application of live WDD on STEM with a focused probe at atomic resolution using datasets from the literature, i.e., graphene and SrTiO₃. It will be interesting to observe more complicated specimen structures, simulations with different parameters, and the effects of a probe with defocus or aberrations.

We will defer such possible improvements to future works.

Summary

As an evolution of the classical WDD algorithm, we demonstrated Live WDD that can reconstruct in a streaming fashion while acquiring diffraction patterns to support real-time reconstruction. Our investigation shows that Live WDD produces object reconstructions that approximate the conventional result. The algorithm uses less memory and runs faster than the classical WDD for typical parameters. As a side effect of dimensionality reduction, we also observe that it acts as a filter for Poissonian noise to attain a more robust reconstruction from low-dose diffraction patterns. We compare the numerical computation time of the proposed algorithm with the dwell time of Merlin Medipix and Dectris Quadro detectors, where we can perform live continuous reconstruction with a field of view up to (256, 256) on a system with 32 CPU cores.

Acknowledgments

Arya Bangun, Alexander Clausen, Dieter Weber, Rafal E. Dunin-Borkowski acknowledge support from Helmholtz Association under contract no. ZT-I-0025 (Ptychography

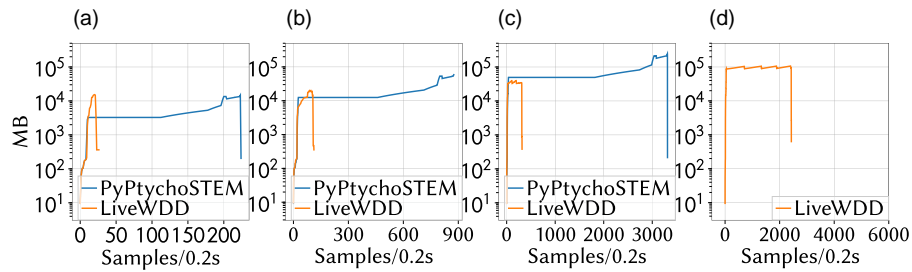


Fig. 17. Memory allocation for conventional WDD and Live WDD ($L = 16$) for different scanning point dimensions: (a) (128, 128, 128, 128), (b) (256, 256, 128, 128), (c) (512, 512, 128, 128), and (d) (1,024, 1,024, 128, 128).

Table 8. Maximum Memory Allocation in MB for PyPtychoSTEM and Live WDD for Fixed Dimension Detector $N_y = N_x = 128$ and Increasing Scanning Points $S_y = S_x$. Here, we use $L = 16$.

Dimension	PyPtychoSTEM	Live WDD
128	14,725	15,343
256	61,521	19,952
512	244,121	39,577
1,024	–	106,585

Table 9. Specification of Different Detectors That Support Experimental Acquisition.

Detectors	Frame Rate (kHz)
MedipixEM	18.8 (1-bit), 3.2 (6-bit), or 1.6 (12-bit)
Dectris Quadro	2.25 (16-bit), 4.5 (8-bit), ROI 9 (16-bit), ROI 18 (8-bit)
Dectris Arina	120 (12-bit)

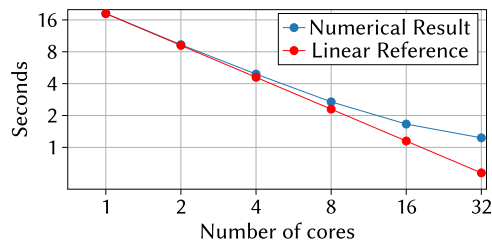


Fig. 18. Computation time depending on the number of cores for a dataset dimension (128, 128, 128, 128). Despite using the CPU bandwidth as a shared resource the algorithm exhibits a 15 × speedup on 32 cores.

Table 10. Processing Speed in Frames per Second for Live WDD for Fixed Dimension Detector $N_y = N_x = 128$ and Increasing Scanning Points $S_y = S_x$. Here, we use $L = 16$.

Dimension	Average (s)	Frame/s
128	1.32	12,412
256	14.82	4,422
512	228.65	1,146
1,024	3048.66	343

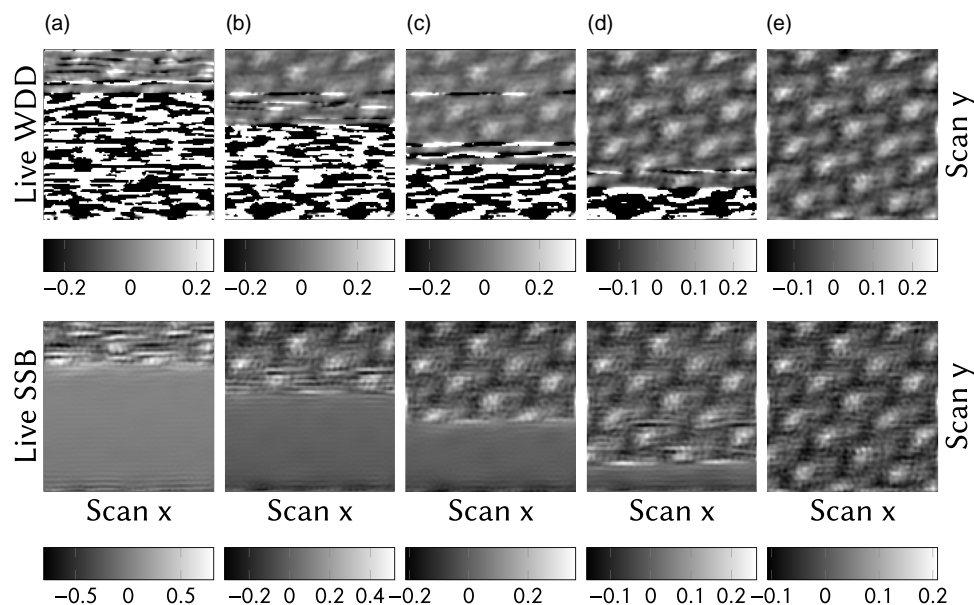


Fig. 19. Snapshots of simulated live ptychography (phase) using Live WDD (top) and Live SSB (bottom) method at different stages of the scan for SrTiO_3 datasets for different buffer processes: (a) 10, (b) 26, (c) 41, (d) 57, and (e) 72. Note how the results don't have to be merged in any particular order.

4.0) and JL-MDMC (Joint Lab on Model and Data-Driven Material Characterization). Paul F. Baumeister acknowledges the support from SiVeGCS (Sicherstellung der weiteren Verfügbarkeit der Supercomputing-Ressourcen des GCS).

Conflict of Interest

The authors declare that they have no competing interest.

References

- Bostan E, Soltanolkotabi M, Ren D & Waller L (2018). Accelerated Wirtinger flow for multiplexed fourier ptychographic microscopy. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018, pp. 3823–3827. IEEE. <https://doi.org/10.1109/ICIP.2018.8451437>
- Clausen A, Weber D, Ruzaeva K, Migunov V, Baburajan A, Bahuleyan A, Caron J, Chandra R, Halder S, Nord M, Müller-Caspary K & Dunin-Borkowski R (2020). LiberTEM: Software platform for scalable multidimensional data processing in transmission electron microscopy. *J Open Source Softw* 5, 2006. <https://doi.org/10.21105/joss>
- Glaeske H (1983). On a convolution structure of a generalized Hermite transformation. *Serdica Bulgaricae Math Publ* 9, 223–229.
- Golub GH & Van Loan CF (2013). *Matrix Computations*. Baltimore, MD: JHU Press.
- Hoppe W (1969b). Beugung im inhomogenen Primärstrahlwellenfeld. III. Amplituden- und Phasenbestimmung bei unperiodischen Objekten. *Acta Crystallogr A* 25, 508–514. <https://doi.org/10.1107/S0567739469001069>
- Hoppe W (1969a). Beugung im inhomogenen Primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeugungsinterferenzen. *Acta Crystallogr A* 25, 495–501. <https://doi.org/10.1107/S0567739469001045>
- Hoppe W & Strube G (1969). Beugung in inhomogenen Primärstrahlenwellenfeld. II. Lichtoptische Analogieversuche zur Phasenmessung von Gitterinterferenzen. *Acta Crystallogr A* 25, 502–507. <https://doi.org/10.1107/S0567739469001057>
- Jain AK (1989). *Fundamentals of Digital Image Processing*. Hoboken, NJ: Prentice-Hall, Inc.
- Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G & Persson KA (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater* 1, 011002. <https://doi.org/10.1063/1.4812323>
- Li P, Edo TB & Rodenburg JM (2014). Ptychographic inversion via Wigner distribution deconvolution: Noise suppression and probe design. *Ultramicroscopy* 147, 106–113. <https://doi.org/10.1016/j.ultramic.2014.07.004>
- Maiden AM & Rodenburg JM (2009). An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* 109, 1256–1262. <https://doi.org/10.1016/j.ultramic.2009.05.012>
- Nord M, Webster RW, Paton KA, McVitie S, McGrouther D, MacLaren I & Paterson GW (2020). Fast pixelated detectors in scanning transmission electron microscopy. Part I: Data acquisition, live processing, and storage. *Microsc Microanal* 26, 653–666. <https://doi.org/10.1017/S1431927620001713>
- Ophus C (2019). Four-dimensional scanning transmission electron microscopy (4D-STEM): From scanning nanodiffraction to ptychography and beyond. *Microsc Microanal* 25, 563–582. <https://doi.org/10.1017/S1431927619000497>
- Pelz PM, Johnson I, Ophus C, Ercius P & Scott MC (2021). Real-time interactive 4D-STEM phase-contrast imaging from electron event representation data: Less computation with the right representation. *IEEE Signal Process Mag* 39, 25–31. <https://doi.org/10.1109/MSP.2021.3120981>
- Pennycook T (2021). Graphene simulated dataset. *Zenodo*.
- Pennycook T & Hofer C (2021). PyPtychoSTEM, Online. Available at <https://gitlab.com/pyptychostem/pyptychostem/>.
- Pennycook TJ, Lupini AR, Yang H, Murfitt MF, Jones L & Nellist PD (2015). Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: Experimental demonstration at atomic resolution. *Ultramicroscopy* 151, 160–167. <https://doi.org/10.1016/j.ultramic.2014.09.013>
- Rodenburg J & Bates R (1992). The theory of super-resolution electron microscopy via Wigner-distribution deconvolution. *Phil Trans R Soc Lond A* 339, 521–553. <https://doi.org/10.1098/rsta.1992.0050>
- Rodenburg JM & Faulkner HM (2004). A phase retrieval algorithm for shifting illumination. *Appl Phys Lett* 85, 4795–4797. <https://doi.org/10.1063/1.1823034>
- Rodenburg J, McCallum B & Nellist P (1993). Experimental tests on double-resolution coherent imaging via STEM. *Ultramicroscopy* 48, 304–314. [https://doi.org/10.1016/0304-3991\(93\)90105-7](https://doi.org/10.1016/0304-3991(93)90105-7)
- Sakurai JJ (1994). *Modern Quantum Mechanics*, rev. ed. Reading, MA: Addison-Wesley.
- Spurgeon SR, Ophus C, Jones L, Petford-Long A, Kalinin SV, Olszta MJ, Dunin-Borkowski RE, Salmon N, Hattar K, Yang W-CD, Sharma R, Du Y, Chiamonti A, Zheng H, Buck EC, Kovarik L, Penn RL, Li D, Zhang X, Murayama M & Taheri ML (2021). Towards data-driven next-generation transmission electron microscopy. *Nat Mater* 20, 274–279. <https://doi.org/10.1038/s41563-020-00833-z>
- Strauch A, Clausen A, Müller-Caspary K & Weber D (2021a). High-resolution 4D STEM dataset of SrTiO₃ along the [1 0 0] axis at high magnification, Tech. Rep., Physik Nanoskaliger Systeme.
- Strauch A, Weber D, Clausen A, Lesnichaia A, Bangun A, März B, Lyu FJ, Chen Q, Rosenauer A, Dunin-Borkowski R & Müller-Caspary K (2021b). Live processing of momentum-resolved STEM data for first moment imaging and ptychography. *Microsc Microanal* 27, 1–15. <https://doi.org/10.1017/S1431927621012423>
- Wen Z, Yang C, Liu X & Marchesini S (2012). Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Probl* 28, 115010. <https://doi.org/10.1088/0266-5611/28/11/115010>
- Yang H, MacLaren I, Jones L, Martinez GT, Simson M, Huth M, Ryll H, Soltau H, Sagawa R, Kondo Y, Ophus C, Ercius P, Jin L, Kovács A & Nellist PD (2017). Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution. *Ultramicroscopy* 180, 173–179. <https://doi.org/10.1016/j.ultramic.2017.02.006>
- Yang H, Pennycook TJ & Nellist PD (2015). Efficient phase contrast imaging in STEM using a pixelated detector. Part II: Optimisation of imaging conditions. *Ultramicroscopy* 151, 232–239. <https://doi.org/10.1016/j.ultramic.2014.10.013>
- Yang H, Rutte R, Jones L, Simson M, Sagawa R, Ryll H, Huth M, Pennycook T, Green M, Soltau H, Kondo Y, Davis BG & Nellist PD (2016). Simultaneous atomic-resolution electron ptychography and z-contrast imaging of light and heavy elements in complex nanostructures. *Nat Commun* 7, 1–8.
- Yu CP, Friedrich T, Jannis D, Van Aert S & Verbeeck J (2021). Real-time integration center of mass (RICOM) reconstruction for 4D STEM. *Microsc Microanal* 28, 1–12. <https://doi.org/10.1017/S1431927622000617>



TESCAN TENSOR

Integrated, Precession-Assisted,
Analytical 4D-STEM



Visit us and learn more
about our TESCAN TENSOR

info.tescan.com/stem