University of Pavol Jozef Safarik
Faculty of Science

Effective representations of tomoscopy experiments suitable for machine learning

Master thesis

Bc. Daniel Moško

Field of Study:	Computer Science
Institute:	Institute of Computer Science
Tutor:	doc. RNDr. Jozef Uličný, CSc.
Counsellor:	Dr. Patrik Vagovič, PhD.

Košice 2026

Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Daniel Moško
Študijný program: informatika (jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: Informatika
Typ záverečnej práce: Diplomová práca
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Effective representations of tomoscopy experiments suitable for machine learning

Názov SK: Kódovanie 3D objektov pre výpočtovo efektívne rozoznávanie dynamiky technikami strojového učenia (tomoskopický experiment)

Cieľ:
1. Provide an overview of existing machine learning methods applied to the recognition of 3D objects from projection data, highlighting their strengths and limitations.
2. Modify and extend current machine learning approaches for predicting the temporal evolution of deformable objects and propose innovative solutions for addressing problems related to dynamic 3D objects.
3. Implement the proposed method of efficient data representation for use with existing machine learning techniques and compare the performance of the proposed solution with other available studies.

Literatúra:
[1] Koliyadu, Jayanath C. P., Mosko Daniel, Asimakopoulou Eleni M., Bellucci Valerio, Birnsteinova Sarlota, Bean Richard, Letrun Romain, Szeles Peter, Ulicny Jozef, Tokushi Sato, Vagovic Patrik et al. “Development of MHz X-Ray Phase Contrast Imaging at the European XFEL,” Accepted to Journal of Synchrotron Radiation, ref. yi5159, 2024
[2] Brunton, Steven L., and J. Nathan Kutz. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge New York, NY Port,

Anotácia: The thesis is focused on the reconstruction of a three-dimensional structure from an extremely limited number of projections available in MHz tomoscopy. MHz tomoscopy is an imaging technique that captures multiple simultaneous X-ray views on sub-microsecond time scales allowing to track fast dynamic processes such as shock waves and cavitation. The task is to first develop a physically realistic model of the dynamics of model objects observed in a tomoscopic X-ray experiment. Based on the data analysis, select a suitable data representation for the simulation of projection data on the detector. Subsequently, on the basis of known physical reality, to reconstruct spatially structured images of objects in phase contrast.

Kľúčové slová: MHz Tomoscopy, X-ray phase-contrast imaging, 3D Gaussian Splatting, sparse-view reconstruction, Fresnel propagation, coarse grain molecular dynamics

Univerzita P. J. Šafárika v Košiciach
Prírodovedecká fakulta

Vedúci:	doc. RNDr. Jozef Uličný, CSc.
Konzultant:	Dr. Patrik Vagovič, PhD.
Oponent:	doc. RNDr. Jozef Jirásek, PhD.
Ústav :	ÚINF - Ústav informatiky
Riaditeľ ústavu:	doc. RNDr. Ondrej Krídlo, PhD.

Dátum schválenia: 24.04.2026

Affidavit

Prehlasujem, že som prácu *Effective representations of tomoscopy experiments suitable for machine learning* vypracoval samostatne pod vedením doc. RNDr. Jozefa Uličného, CSc. a Dr. Patrika Vagoviča PhD. a uviedol v nej všetky použité literárne a iné odborné zdroje.

Košice 27.4.2026 . . . . . .

Acknowledgment

Here, I would like to thank my supervisor, Doc. RNDr. Jozef Uličný, CSc., for his insightful remarks and professional guidance that helped me prepare this thesis while also showing me patience and humanity. Finally, I would like to express my gratitude to everyone who helped with making this thesis, whether professionally or otherwise. Special thanks go out to Dr. Patrik Vagovič, PhD., from EuXFEL, Maxwell Cluster, and the Slovak National Super Computing Center for providing valuable computational resources.

Abstrakt

Röntgenové žiarenie predstavuje výkonný, nedeštruktívny prostriedok na skúmanie vnútornej štruktúry nepriehľadných materiálov v troch rozmeroch. Tradičná röntgenová tomografia vytvára detailné 3D obrazy rotáciou vzorky, avšak táto sekvenčná akvizícia limituje rýchlosť, akou možno sledovať dynamické procesy. MHz-Tomoskopia tento problém obchádza súčasným zachytením viacerých projekcií pomocou röntgenového lasera, čím umožňuje zobrazovanie rýchlych dejov — šírenia šokových vĺn, kavitácie či formovania trhlín — pri frekvenciách až niekoľko MHz. Cenou za to je len niekoľko simultánnych projekcií, čo značne sťažuje kvalitnú 3D rekonštrukciu. Praktickú výzvu ďalej prehlbujú veľké objemy dát a nedokonalosti inštrumentácie, ktoré spoločne vyžadujú reprezentáciu scény dostatočne kompaktnú na efektívne ukladanie a dostatočne rýchlu na spracovanie v čo najkratšom čase. Problém ďalej komplikuje povaha biologických vzoriek: namiesto absorpcie röntgenového žiarenia ho predovšetkým fázovo posúvajú, takže klasické absorpčné zobrazovacie techniky takmer vôbec neprodukujú kontrast. Táto práca rieši uvedené výzvy prostredníctvom dvoch príspevkov. Najprv sme vyvinuli fyzikálne realistickú simulačnú pipeline, ktorá priradí röntgeno-optické parametre vypočítané z Henkeho tabuliek geometrickým modelom červených krviniek a Fresnelovým difrakčným integrálom propaguje výstupné vlnové pole na detektor, čím vznikajú syntetické fázovo-kontrastné hologramy modelujúce celé zobrazovanie. Následne sme pre röntgenové fázovo-kontrastné zobrazovanie adaptovali metódu 3D Gaussian Splatting (3DGS) ako efektívnu, diferencovateľnú reprezentáciu scény: namiesto voxelovej mriežky je scéna zložená z Gaussových primitívov, z ktorých každý nesie absorpčný index β a dekrement indexu lomu δ daného materiálu. Na rozdiel od implicitných neurónových reprezentácií 3DGS rasterizuje všetky primitívy v jednom diferencovateľnom prechode bez volaní neurónovej siete pre každý vzorkovaný bod, čo prináša výrazné výpočtové úspory oproti NeRF aj voxelovým metódam. Kľúčovými úpravami štandardného frameworku sú nahradenie perspektívnej zobrazovacej geometrie a farebného renderovania paralelnou geometriou s fyzikálnym dopredným modelom zahŕňajúcim Fresnelovo šírenie vlnového poľa a funkciu rozptylu bodu detektora, a implementovanie fázovej rekonštrukcie priamo do optimalizácie. Výsledky ukazujú, že upravený framework dokáže rekonštruovať fázovo-kontrastné zobrazenia zo 16 syntetických projekcií za približne 15 minút na jedinom GPU, čo je v rámci praktických časových možností meracích časov na XFEL. Dôležitým zistením je, že bežná metrika PSNR je pre hodnotenie holografickej rekonštrukcie nevhodná; vhodnejšími alternatívami sú SSIM a Fourierova korelácia krúžkov (FRCM). Ukázalo sa tiež, že správne nastavenie učiaceho pomeru na riedkej inicializácii prekonáva 30x zvýšenie počtu Gaussiánov, čo poukazuje na to, že hlavnou prekážkou kvalitnej rekonštrukcie nie je kapacita reprezentácie, ale dynamika optimalizácie.

Kľúčové slová: *MHz-Tomoskopia, röntgenové fázovo-kontrastné zobrazovanie, 3D Gaussian Splatting, rekonštrukcia z riedkeho počtu projekcií, Fresnelov propagátor, molekulárna dynamika*

Abstract

X-rays provide a powerful, non-destructive means of probing the internal structure of opaque materials in three dimensions. Conventional X-ray tomography builds detailed 3D images by rotating the sample through a full angular range, but this sequential acquisition limits how fast dynamic processes can be observed. MHz-Tomoscopy overcomes this by capturing multiple projections simultaneously with an X-ray laser, enabling the imaging of rapid phenomena — shock waves, cavitation, crack formation — at repetition rates of up to several MHz. The trade-off is a severely limited number of simultaneous views, making high-quality 3D reconstruction challenging. The practical difficulty is further compounded by the large data volumes and instrumentation imperfections inherent to MHz-rate acquisition, which together demand a scene representation compact enough for efficient storage and fast enough for near-real-time post-processing. The problem is additionally complicated by the nature of soft biological specimens: rather than absorbing X-rays, they predominantly shift their phase, so conventional absorption-based methods produce almost no contrast. This thesis addresses these challenges through two main contributions. First, we developed a physically realistic simulation pipeline that assigns X-ray optical parameters computed from the Henke tables to geometric models of red blood cells and propagates the exit wavefield to the detector via the Fresnel diffraction integral, producing synthetic phase-contrast holograms that model the full imaging chain. Second, we adapted 3D Gaussian Splatting (3DGS) as an efficient, differentiable scene representation for X-ray phase-contrast imaging: rather than a dense voxel grid, the scene is decomposed into Gaussian primitives each carrying the absorption index β and refractive-index decrement δ of the local material. Unlike implicit neural representations, 3DGS rasterises all primitives in a single differentiable forward pass without per-sample neural network queries, yielding significant computational savings over both NeRF and voxel-based methods. The key modifications to the standard framework are the replacement of perspective geometry and colour rendering by a parallel-beam physical forward model incorporating Fresnel free-space propagation and detector point-spread function, and the embedding of a phase-retrieval step within the scene optimisation loop. The results show that the adapted framework can reconstruct phase-contrast images from 16 projections in approximately 15 minutes on a single GPU, well within the practical timescales of beamtime experiments. A key finding is that the standard PSNR metric is unsuitable for evaluating holographic reconstruction; SSIM and the Fourier Ring Correlation Metric (FRCM) are more appropriate alternatives. It also emerged that careful learning-rate tuning on a sparse initialisation outperforms a 30x increase in Gaussian count, indicating that the bottleneck for reconstruction quality is optimisation dynamics rather than representational capacity.

Keywords: *MHz-Tomoscopy, X-ray phase-contrast imaging, 3D Gaussian Splatting, sparse-view reconstruction, Fresnel propagation, molecular dynamics*

Acronyms

MHz-Tomoscopy — MHz rate mulTiple prOjection X-ray MicrOSCOPY
PCI — Phase-contrast imaging
3DGS — 3D Gaussian Splatting
FBP — Filtered Back Projection
SIRT — Simultaneous Iterative Reconstruction Technique
NeRF — Neural Radiance Fields
ONIX — Optimised Neural Implicit X-ray imaging
XFEL — X-ray free electron laser
RBC — Red Blood Cells
SSIM — Structural Similarity Index Measure
FRC — Fourier Ring Correlation
FRCM — Fourier Ring Correlation Metric
MD — Molecular Dynamics

Contents

1 Introduction

X-rays provide a powerful, non-destructive means of exploring the internal structure of materials due to their high penetration capability. Their interactions with matter, manifesting as variations in absorption and phase shifts, enable the reconstruction of volumetric structures with high spatial resolution. This capability forms the basis of several advanced imaging modalities, such as X-ray computed tomography (CT) and MHz-Tomoscopy [1], also referred to as X-ray Multi-Projection Imaging (XMPI) [2].

While conventional X-ray tomography serves as the gold standard for 3D imaging, it fundamentally relies on temporal averaging. This makes it unsuitable for imaging fast phenomena, as data acquisition requires sequential rotations over a 180° range. MHz-Tomoscopy presents a compelling alternative, offering the ability to capture simultaneous projections from multiple angles at sampling frequencies up to several MHz. This enables the visualization of dynamic processes in almost real-time, including shock wave propagation, cavitation, crack formation, combustion, and acoustic turbulence—processes often inaccessible to optical or slower imaging techniques [2].

There are multiple challenges in MHz-Tomoscopy. At the data level, MHz-rate acquisition produces extremely large data streams from instrumentation that is inevitably imperfect, requiring both effective data filtering and an internal scene representation compact enough for efficient storage and fast enough to keep pace with the experimental acquisition rate. At the reconstruction level, the severely limited number of simultaneous views renders the inverse problem underdetermined [3], and the recovery of phase objects is further complicated by the absence of direct phase measurements and the nonlinearity of the imaging model. To tackle this, we propose a particle-based representation that inherently captures abrupt changes in optical density while reducing memory demands and training complexity. Unlike conventional voxel-based methods that tend to be computationally intensive and require large neural networks, our representation focuses on essential features of the sample. This approach promises to overcome the constraints arising from sparsity and ill-posedness, while also addressing the data-volume demands of online tomoscopic experiments.

2 Theoretical Background

2.1 Tomography, Tomoscopy and Dynamic Imaging

X-rays are particularly well-suited for probing the three-dimensional composition and behavior of materials. Among the techniques leveraging this capability, time-resolved tomography stands out for its ability to capture the evolution of 3D structures dynamically. Conventional implementation acquires 4D datasets (three spatial dimensions plus time) by sequentially recording complete tomographic volumes at intervals faster than the timescale of the underlying phenomena. Each volume is reconstructed from a series of 2D projection images acquired over a sample rotation of at least $180°$ relative to the X-ray beam [2, 3].

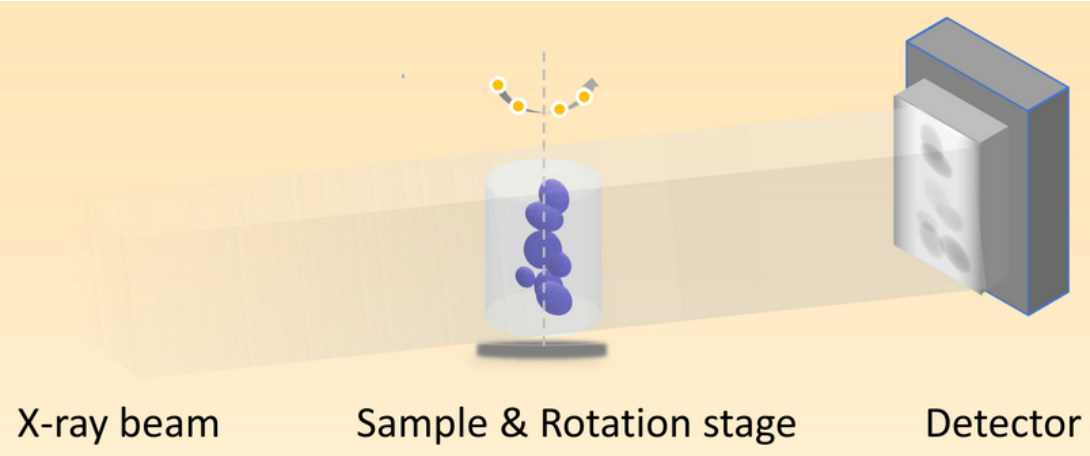


Figure 1: Computed tomography geometry showing sequential sampling [4, 5].

Despite its widespread adoption, this approach is constrained by physical and mechanical factors. First, the mechanical rotation of the sample can introduce forces that disturb the very processes under observation. Second, the sequential nature of data collection prevents true single-shot imaging. Third, both temporal and spatial resolution are inherently limited by the available photon flux (number of photons in the beam) from the source and the maximum available acquisition speed [2, 3].

However, this geometric shift introduces a new set of challenges. Although temporal resolution has been greatly improved, the number of unique angular views available for reconstruction at any given snapshot is significantly lower compared to conventional CT. The resulting sparse angular sampling, together with the phase-contrast nature of the measurements, makes the inversion problem ill-posed: reconstruction requires recovering both the 3D distribution of the complex refractive index and the phase of the wavefield from a limited number of intensity-only projections — a fundamentally more ambiguous task than classical absorption CT, as discussed in the following sections.

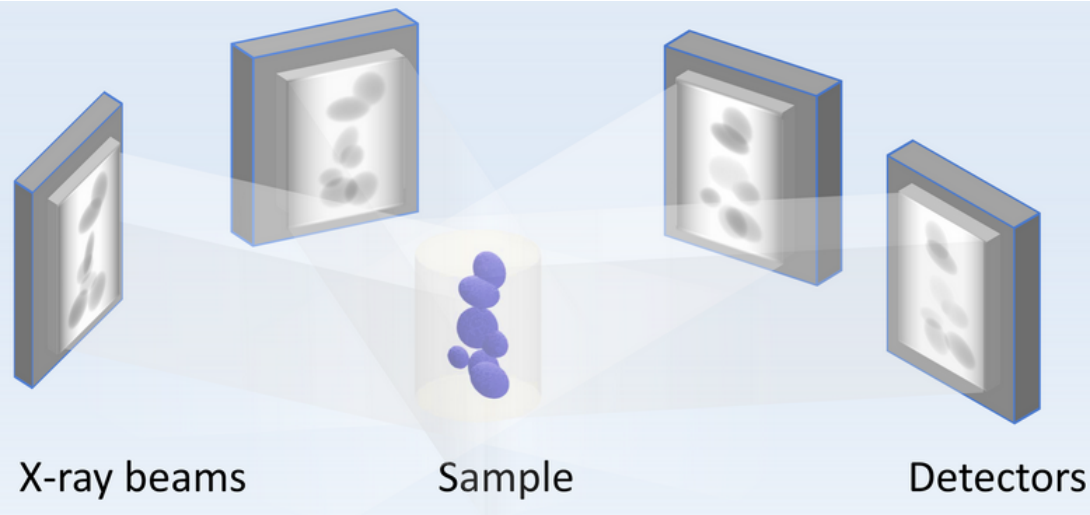


Figure 2: MHz-Tomoscopy / XMPI imaging setup [4, 5].

2.2 Phase-contrast imaging

Beyond sampling constraints, a fundamental limitation of conventional computed tomography (CT) arises from the physics of the X-ray interaction. Standard reconstruction algorithms rely on the Radon transform [6] applied to the attenuation coefficient derived from the intensity loss (Beer-Lambert law; see Eq. 2.8 and Section 2.3). This approach is often insufficient for phase objects—materials with low attenuation but significant phase-shift properties, such as soft tissues or polymers. In these cases, edges may be invisible or poorly defined in absorption-based reconstructions. To recover these structures, one must exploit the phase component of the X-ray wavefield [7].

The interaction of X-rays with matter is characterised by the complex refractive index n:

$$n = 1 - \delta + i\beta, \tag{2.1}$$

where β refers to absorption and δ represents the refractive index decrement responsible for phase shifts. For soft biological materials, δ is typically several orders of magnitude larger than β, meaning that phase effects dominate absorption effects [7, 8].

Under the projection approximation, the interaction of X-ray wavefield with the sample is described by the complex transmission function $T(x, y)$. Defining the magnitude of the wave-vector as $k = 2\pi/\lambda$, the transmission function is given by:

$$T(x, y) = \exp\left[-k \int \beta(x, y, z)\, dz\right] \cdot \exp\left[-ik \int \delta(x, y, z)\, dz\right]. \tag{2.2}$$

This can be decomposed into the amplitude $A(x, y)$ and phase $\phi(x, y)$ projections along the

optical axis z:

$$A(x,y) = \exp\left(-k \int \beta(x,y,z)\, dz\right) \tag{2.3}$$

$$\phi(x,y) = -k \int \delta(x,y,z)\, dz \tag{2.4}$$

Thus, $T(x,y) = A(x,y)\exp[i\phi(x,y)]$. [7]

The wavefield evolves over a propagation distance d via the Fresnel diffraction integral. Mathematically, this is represented as a convolution of the exit wavefield $U(x,y,0) = T(x,y)$ with the Fresnel propagator $P(x,y,z)$:

$$U(x,y,z) = U(x,y,0) * P(x,y,z) \tag{2.5}$$

where the free-space propagator is defined as:

$$P(x,y,z) = \frac{e^{ikz}}{i\lambda z} \exp\left[\frac{ik(x^2+y^2)}{2z}\right]. \tag{2.6}$$

The physical quantity captured by the detector is the intensity $I(x,y)$, which corresponds to the squared magnitude of the propagated wavefield:

$$I(x,y) = |U(x,y,z)|^2 = |U(x,y,0) * P(x,y,z)|^2. \tag{2.7}$$

The resulting intensity distribution encodes the phase information of the object through interference patterns [7]. Structurally, this implies that the detector records a signal sensitive to the curvature of the phase map. Structural boundaries are emphasised by bright-and-dark fringes—the well-known edge effect—where phase gradients manifest themselves as intensity variations [9].

Figure 3 illustrates how free-space propagation transforms a mixed absorption/phase object into these characteristic fringe patterns.

Computational regimes. From a computational perspective, the choice of propagation distance d defines the forward measurement model and introduces a trade-off between spatial resolution and the compactness of the signal in the frequency domain. Spatial resolution describes how well the imaging system can distinguish fine details or closely spaced features of the object, while frequency-domain compactness refers to how concentrated the signal representation is in the Fourier domain, which can simplify computational reconstruction.

In practice, this trade-off appears in different propagation regimes:

- **Far-field (Fraunhofer regime):** The object field undergoes a Fourier transform, and the detector measures only the intensity distribution $|\mathcal{F}\{U\}|^2$. Because phase information is not recorded, reconstruction requires phase-retrieval algorithms or additional

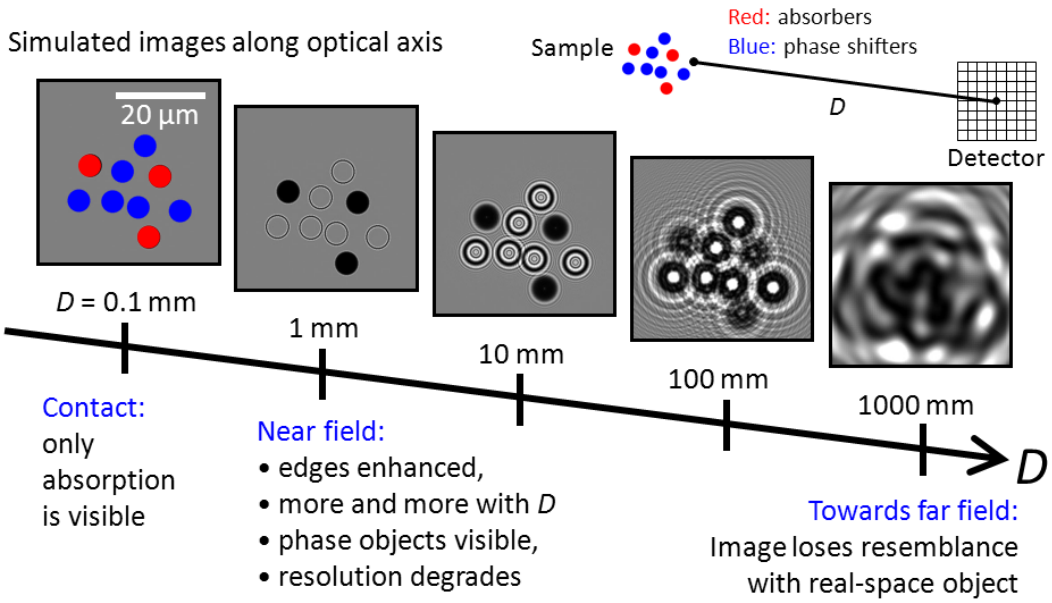


Figure 3: Free-space propagation behind an object composed of absorbing structures (red) and non-absorbing phase shifters (blue). Image adapted from [10].

constraints, such as support masks or reference waves.

- **Near-field (Fresnel regime):** The recorded intensity corresponds to the exit wave propagated by the Fresnel kernel, which preserves spatial structures in a phase-encoded form. The resulting pattern is a near-field diffraction pattern (hologram), characterised by interference fringes formed by the interaction between the scattered wave from the sample and the unscattered reference wave [11, 7, 3].

For small propagation distances z, the measured intensity remains close to the object plane, preserving a higher spatial resolution, although phase information may only be weakly encoded. As z increases, the diffraction becomes stronger and more phase information is encoded in the intensity pattern, which can facilitate computational reconstruction. However, this also spreads the signal spatially, leading to a more blurred intensity distribution. [11, 7, 3]

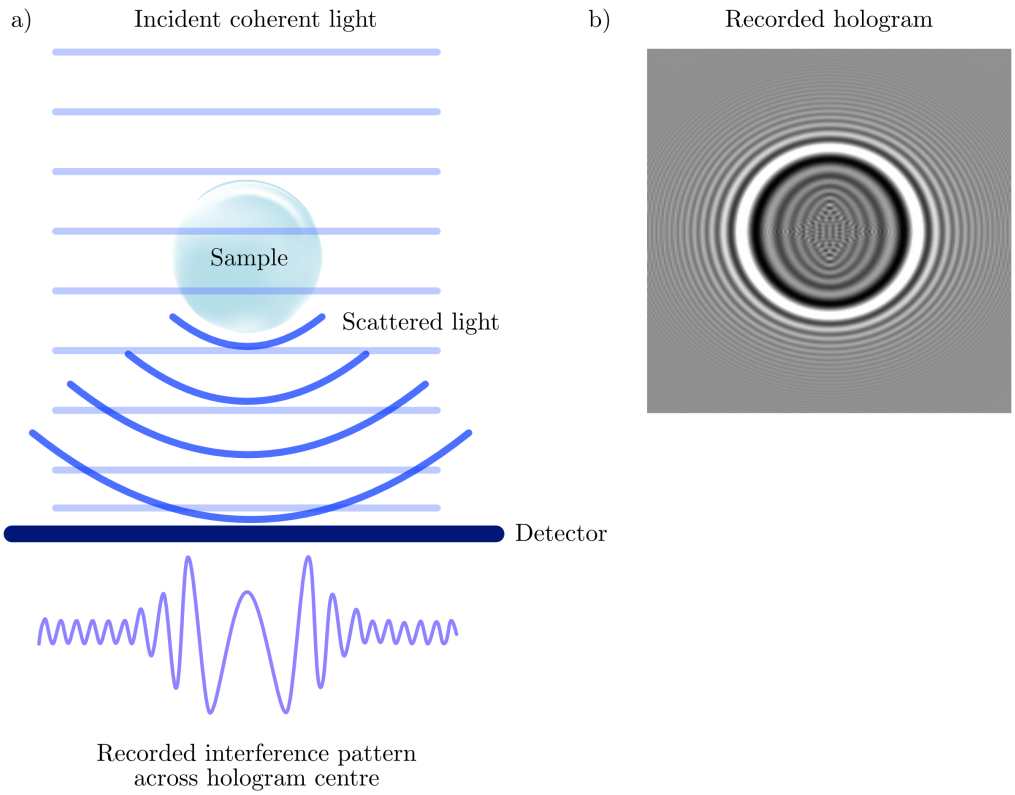


Figure 4: (a) Coherent X-ray wavefield interferes with light scattered from the sample, producing an interference pattern that encodes the object's phase and absorption. **(b)** The resulting hologram. Adapted from [3].

Phase retrieval Phase retrieval introduces a critical computational challenge. Standard absorption CT is a linear inverse problem — one simply recovers the absorption index β from measured intensities. Recovering the refractive-index decrement δ from intensity-only measurements, however, is non-linear: because phase information is discarded at the detector, many different objects can produce the same intensity pattern, making the problem inherently ill-posed. Without proper constraints, this non-linearity leads to reconstruction artefacts and ambiguities, which is particularly problematic when imaging dynamic, phase-dominant phenomena [7].

Uniqueness and stability can be recovered by imposing physically motivated constraints. The most common is a support constraint, which assumes the object is confined to a known bounded region of the field of view. Phase-retrieval algorithms belong to two broad families. **Deterministic methods** (e.g., transport-of-intensity approximations, analytic holographic reconstructions) yield closed-form estimates in a few analytical steps. **Iterative methods** (Gerchberg–Saxton, hybrid-input-output) treat the task as an optimisation: starting from an initial guess they repeatedly enforce data consistency and the constraints above until convergence. The choice between deterministic and iterative schemes is dictated by the available

prior information, noise level, and computational budget [12, 7].

By embedding these constraints into the reconstruction pipeline we mitigate the ambiguities intrinsic to the phase problem and obtain high-precision δ reconstructions from intensity-only measurements, thus fulfilling the promise of phase-contrast imaging for dynamic, soft biological specimens.

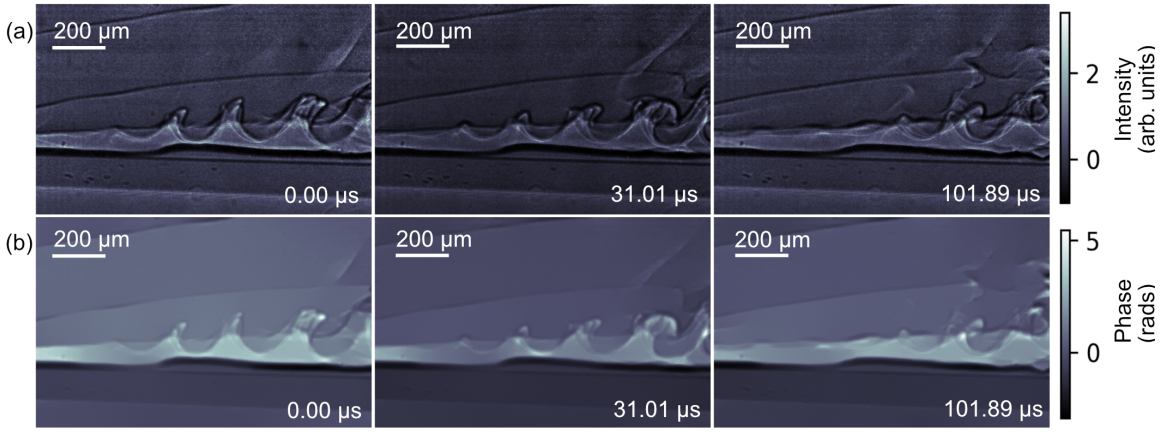


Figure 5: (a) actual experimental images from EuXFEL at three different time points and **(b)** the images after phase retrieval methods based on the alternating direction method of multipliers [13]. Image adapted from [14].

2.3 Tomographic Reconstruction

Classical tomographic reconstruction aims to recover the three-dimensional internal structure of an object from a series of two-dimensional projection images, typically acquired by rotating the sample through a defined angular range ($180°$ to $360°$). The fundamental unit of this data is the projection image, where each row of pixels corresponds to a specific "slice" of the object. By aggregating the same row of pixels from every angular projection and stacking them vertically, one constructs a sinogram [15]. In this transformed space, a single point in the object traces out a sine wave over the course of the rotation, and reconstruction amounts to "unwrapping" these sine waves back into their original spatial coordinates [3, 16]. This classical picture assumes dense, uniform angular sampling and purely absorptive contrast — both assumptions that are challenged in the sparse-view, phase-dominant regime of MHz-Tomoscopy addressed in this thesis.

The physical formation of these projections relies on the Projection Approximation within the framework of coherent X-ray optics. This model assumes that X-rays traverse the sample along straight ray paths parallel to the optical axis z. Under this assumption, the interaction of the X-ray wave with the sample is governed by the Beer-Lambert law, where the intensity loss is determined by the line integral of the absorption index $\beta(\mathbf{r})$ (accumulated absorption along the ray $\mathbf{r}$) [7]. Specifically, the measured intensity I is related to the incident

intensity I_0 by the following equation:

$$I = I_0 \exp\left(-\int \beta(\mathbf{r})\, ds\right). \tag{2.8}$$

The Beer-Lambert coefficient μ — the linear attenuation coefficient used in classical CT — captures the total reduction in beam intensity through two mechanisms: absorption (photon energy deposited in the material) and scattering (photons deflected out of the beam) [7]. In the wave-optics framework adopted here, this role is taken by the imaginary part β of the complex refractive index; the two are related by $\mu = 2k\beta$, which holds exactly for pure absorption and is an excellent approximation for soft biological tissue at hard X-ray energies where scattering contributes negligibly to total attenuation. β is therefore called the absorption index and is used consistently in place of μ throughout.

While phase contrast imaging exploits the refractive index decrement $\delta(\mathbf{r})$, standard CT focuses on recovering the distribution of $\beta(\mathbf{r})$. Mathematically, the mapping from the object to the projection data corresponds to the Radon transform. The forward transform converts the three-dimensional object into a set of two-dimensional projections (the sinogram), while the reconstruction process corresponds to inverting this transform to estimate the internal distribution of the absorption index β [7].

Because the projection model assumes straight ray paths (parallel geometry), the reconstruction is typically performed slice-by-slice. The inverse Radon transform is solved for a two-dimensional plane to yield a cross-section, which is then stacked vertically with other slices to generate the final three-dimensional volume. The quality of this reconstruction is highly dependent on the sampling density, as demonstrated in Figure 6, where insufficient angular sampling leads to significant degradation in the reconstructed volume [7, 3].

Standard reconstruction algorithms primarily fall into two categories: analytical methods and iterative methods. Filtered Back Projection (FBP) is the most widely used analytical technique due to its computational efficiency. FBP directly back-projects the filtered projections into the reconstruction space to "unsmear" the data. However, FBP assumes ideal sampling conditions and often produces artefacts when data are sparse or noisy. Alternatively, iterative reconstruction methods, such as the Algebraic Reconstruction Technique (ART) [17] or the Simultaneous Iterative Reconstruction Technique (SIRT) [18], explicitly model the projection process. These methods treat reconstruction as an optimisation problem, iteratively adjusting the volume estimate to minimise the difference between measured projections and simulated projections generated from the current volume. While these methods offer superior noise suppression and artefact reduction, they are computationally intensive and do not resolve the angular sparsity problem [3].

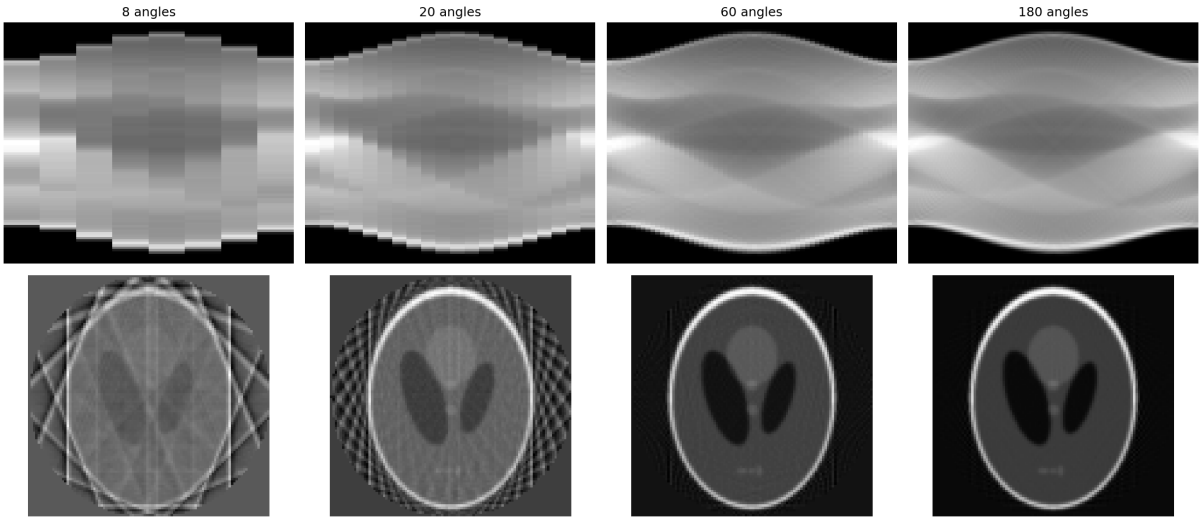


Figure 6: Effect of Angular Sampling on FBP Reconstruction. The top row displays sinograms with decreasing angular density; the bottom row shows the resulting degradation in reconstruction quality.

Principles of Raycasting

Since acquiring experimental data is costly and the current MHz Tomoscopy setup is limited to four simultaneous views, creating a reliable 3D model of the sample is extremely challenging. A digital twin of the experiment is therefore required so that reconstruction algorithms can be tested under controlled, reproducible conditions. The goal is to construct a 3D model of the sample under examination that can be rendered from arbitrary viewpoints.

Volumetric raycasting gives the Beer-Lambert law (Eq. 2.8) an explicit computational form by parameterising each X-ray as a ray with origin $\mathbf{o}$, unit direction $\mathbf{d}$, and path coordinate $t \in [0, D]$ [19, 20]:

$$I = I_0 \exp\left(-\int_0^D \beta(\mathbf{o} + t\mathbf{d})\, dt\right). \tag{2.9}$$

The quantity actually used in reconstruction is the projection data P, defined as the negative log-ratio of transmitted to incident intensity. This reduces the exponential to a line integral of the absorption index [20, 21]:

$$P = -\ln\left(\frac{I}{I_0}\right) = \int_0^D \beta(\mathbf{o} + t\mathbf{d})\, dt. \tag{2.10}$$

Because the volume is stored as a discrete grid of voxels, the continuous integral is approximated by a Riemann sum over N equal segments of length Δt [22]:

$$P \approx \sum_{i=0}^{N-1} \beta(\mathbf{o} + t_i\mathbf{d})\Delta t, \tag{2.11}$$

Anti-aliasing via multi-ray sampling (supersampling). A common issue in discrete raycasting is the appearance of aliasing artefacts, such as jagged edges at sharp boundaries. These arise when the sampling rate is insufficient to resolve fine detail in the volume. To mitigate these effects, rather than casting a single ray through the centre of each pixel, M rays are cast with randomised offsets within the pixel area. The final pixel intensity I_{pixel} is the average of these samples:

$$I_{\text{pixel}} = \frac{1}{M} \sum_{j=1}^{M} \sum_{i=0}^{N-1} \beta(\mathbf{o}_j + t_i \mathbf{d}_j) \Delta t, \tag{2.12}$$

where $\mathbf{o}_j$ and $\mathbf{d}_j$ are the origin and direction for the j-th ray. As shown in Figure 7, this multi-sampling strategy effectively suppresses high-frequency noise and yields smoother, more physically accurate projections [23].

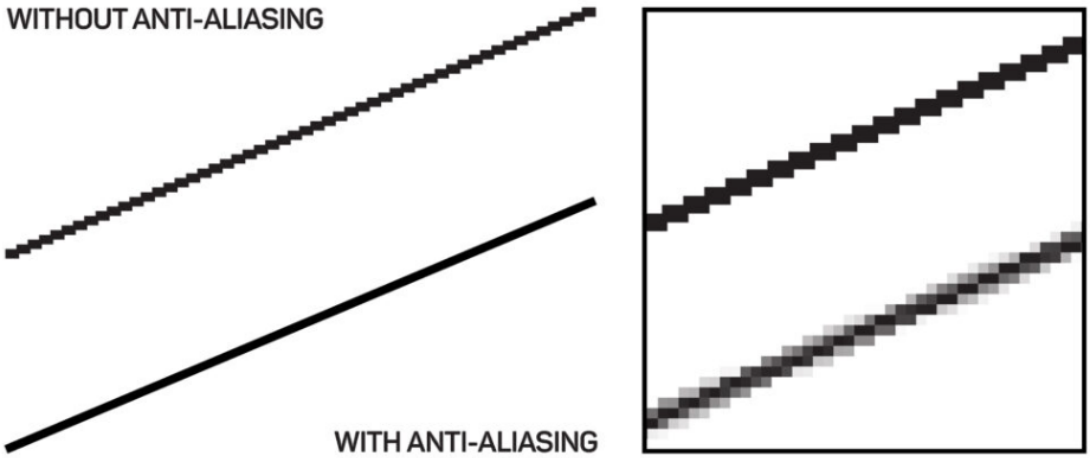


Figure 7: Comparison of rendered projections illustrating the reduction of moiré patterns and jagged boundary artefacts through supersampling anti-aliasing. Image source: [24].

Extension to phase-contrast raycasting. The geometric raycasting model described above assumes that image formation is governed primarily by X-ray absorption, following the scalar Beer-Lambert law (Eq. 2.8). However, as established in Section 2.2, X-ray phase-contrast imaging exploits the wave nature of radiation, where the refractive-index decrement (δ) often dominate over absorption (β). To accurately simulate this kind of imaging, the raycasting forward model must be extended beyond simple scalar absorption.

Rather than accumulating a single scalar value along the ray path, the raycasting process must integrate the complex refractive index to construct the transmission function $T(x, y)$, incorporating both the attenuation and phase-shift components. The final intensity recorded at the detector is not the direct result of this line integral but is derived after the wavefield propagates through free space via the Fresnel propagator (Eq. 2.5 in Section 2.2). This requires the rendering pipeline to handle two distinct volumetric fields (δ and β) and to simulate the resulting interference patterns. Consequently, the underlying data representation must be

capable of storing and optimizing these physical properties efficiently, supporting the high-fidelity reconstruction of low-density materials and abrupt structural changes without the memory constraints associated with traditional volumetric grids.

2.4 Common 3D Data Representations

In the realm of 3D reconstruction and processing, data representation plays a critical role in determining the fidelity and efficiency of the reconstruction pipeline. For X-ray MHz Tomoscopy (and tomography in general), the choice of representation is particularly constrained by the need to model internal material properties, such as optical density, attenuation (or complex index of refraction), rather than merely surface geometry. This section reviews four prominent methods: voxel grids, point clouds, particle-based splats, and neural fields, analyzing their suitability for volumetric data tasks.

Voxel Grids Voxel grids remain the standard representation in classical computed tomography reconstruction pipelines, where the 3D space is discretised into a regular grid of cubic elements. The number of scalar fields stored per voxel depends on the imaging modality: absorption CT requires a single value — the linear attenuation coefficient μ (equivalently the absorption index β, related by $\mu = 2k\beta$) [7] — whereas phase-contrast imaging requires two fields, the absorption index β and the refractive-index decrement δ, as employed in the simulation pipeline of this work. The primary advantage of the voxel representation is its direct, unambiguous mapping to physical space: forward-projection operators such as the Radon transform and standard reconstruction algorithms (FBP, SIRT, ART) operate on voxel grids natively, making the representation immediately compatible with the entire classical CT pipeline [25, 26, 7]. However, voxel grids are inherently memory-intensive; the storage requirement grows as $O(N^3)$ with linear resolution N, making high-resolution volumes quite large.

Point Clouds Point clouds represent 3D data as a set of discrete points in space, typically defined by coordinates (x, y, z) and attributes such as colour or intensity. Although widely generated by LiDAR and photogrammetry, point clouds are primarily surface-based representations [27]. In the context of tomography, this poses a significant limitation: point clouds capture the exterior geometry but lack the explicit internal connectivity or volumetric density information required to model X-ray attenuation through the object. While surface meshes can be derived from point clouds, they inherently do not represent the internal volume where material interactions occur [27]. Consequently, point clouds are often insufficient for tasks requiring the reconstruction of internal structural variations [26].

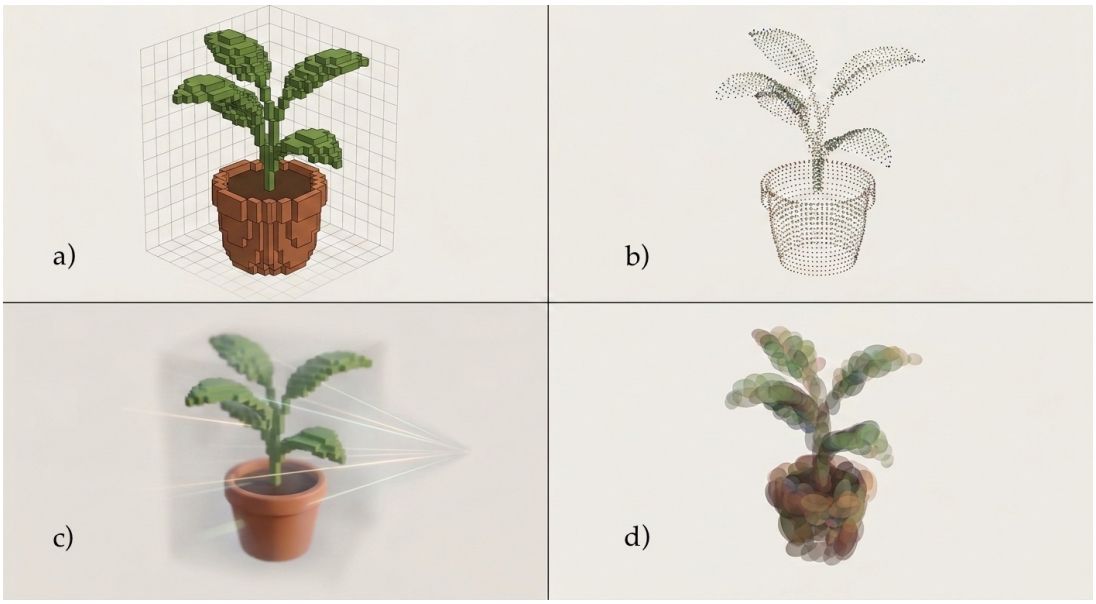


Figure 8: (a) Voxel Representation: Objects are constructed from fixed cubic volumes, creating a blocky, grid-based structure. (b) Point Clouds: The scene is defined by a sparse or dense collection of individual data points in 3D space. (c) Neural Radiance Fields (NeRF): A continuous volumetric representation that implicitly encodes scene geometry and appearance, enabling high-fidelity rendering with realistic view-dependent lighting and transparency. (d) Gaussian Splatting: Surfaces are reconstructed using overlapping primitives to create a continuous, rendered appearance. Image generated by [28].

Splats Gaussian splatting can be understood as a direct extension of the point cloud representation. Starting from an initial set of 3D points, each point is augmented with additional parameters—a 3D covariance matrix (encoding orientation and scale), opacity, and appearance attributes—transforming it from a bare coordinate into a volumetric Gaussian primitive [29]. These overlapping Gaussians collectively fill the interior of the object, giving rise to smooth, continuous density distributions that can encode material properties throughout the volume. This overcomes the core limitation of raw point clouds—the absence of volumetric interior information—while retaining the flexibility and compactness of an unstructured, particle-based layout. Unlike implicit neural representations like Neural Radiance Fields (NeRF) [30], 3D Gaussian Splatting (3DGS) rasterises all primitives in a single differentiable forward pass via a tile-based GPU rasteriser, avoiding the per-sample MLP queries that NeRF requires at every point along every ray. This yields substantially faster rendering and shorter optimisation times. From a memory and compression standpoint, each Gaussian primitive stores only a compact set of parameters — position, rotation quaternion, three scale values, opacity, and appearance attributes [29] — amounting to a few tens of floating-point values per primitive. For sparse scenes composed of distinct objects (such as the RBC in microfluidic channel studied here), a moderate primitive count suffices to represent the scene, making the total footprint competitive with or smaller than the $O(N^3)$ voxel grid needed to achieve equivalent spatial resolution. Against NeRF, both memory and speed advantages are more consis-

tent: NeRF must store network weights and evaluate the MLP at every sample point, whereas 3DGS accesses only the primitive parameters [29, 31]. It should be noted, however, that for highly dense or uniform volumes the primitive count can grow large and the per-primitive parameter overhead may erode the memory advantage over fine voxel grids. The parameters of all primitives are optimised end-to-end through the differentiable rasteriser, allowing the representation to adapt its spatial layout to the scene structure rather than allocating resources uniformly across empty space [29].

Neural Radiance Fields Neural Radiance Fields (NeRFs), or implicit neural representations, utilize neural networks to encode geometry and appearance as continuous functions mapping spatial coordinates to density and colour. This approach has gained traction for its ability to represent complex shapes with high fidelity and for its capability in reconstructing the objects in sparse-view scenarios. Neural fields enable applications such as high-quality view synthesis and shape interpolation. Nevertheless, they are computationally demanding; training these networks typically requires significant resources and time (tens of GB's of GPU memory and sometimes up to days of training times). For tomographic tasks involving rapid changes or large datasets, the training complexity and inference latency of neural fields can present a bottleneck compared to explicit representations [30, 3].

2.5 Implicit Neural Scene Representations (NeRF)

In computer graphics, a radiance field is a mathematical representation used to describe the spatial distribution of light within a scene. It is expressed as a five-dimensional function where the first three dimensions correspond to spatial position (x, y, z), and the remaining two specify direction (θ, ϕ) in spherical coordinates. This function outputs a colour vector $c = (r, g, b)$ and a volumetric density σ, which acts as a differential opacity regulating the accumulation of radiance along a ray. Formally, the radiance field is defined as:

$$L(x, y, z, \theta, \phi) = (c, \sigma) \tag{2.13}$$

By incorporating this radiance field into the rendering equation, realistic lighting effects such as shadows and specular reflections can be simulated. This formulation forms the basis of Neural Radiance Fields (NeRF), a method proposed to address the problem of novel view synthesis—generating new views of a scene from previously unseen camera positions.

NeRF models a scene as a continuous volumetric function learned by a deep neural network. This network takes 5D input coordinates and outputs a view-dependent RGB colour and density, learned through a multilayer perceptron (MLP) without any convolutional layers. The model is trained using only a sparse set of 2D input images [30].

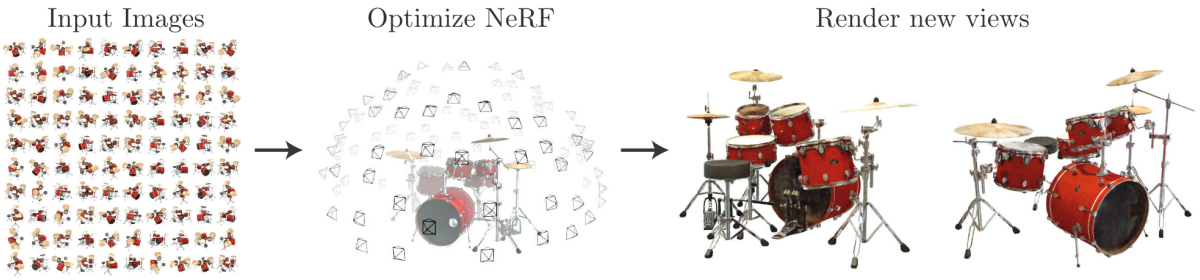


Figure 9: A continuous 5D neural radiance field representing a scene is optimised via volume rendering. Given 100 input views of the synthetic Drums scene sampled from a hemisphere, two novel views are generated using the trained NeRF model [30].

Scene-specific learning Unlike conventional deep learning approaches that generalise across large datasets, NeRF uses a scene-specific training paradigm. Each 3D scene is encoded in the neural network weights, optimised from a small number of views. This results in a compact and high-fidelity representation, outperforming voxel-based methods which are less efficient and more memory-intensive [30].

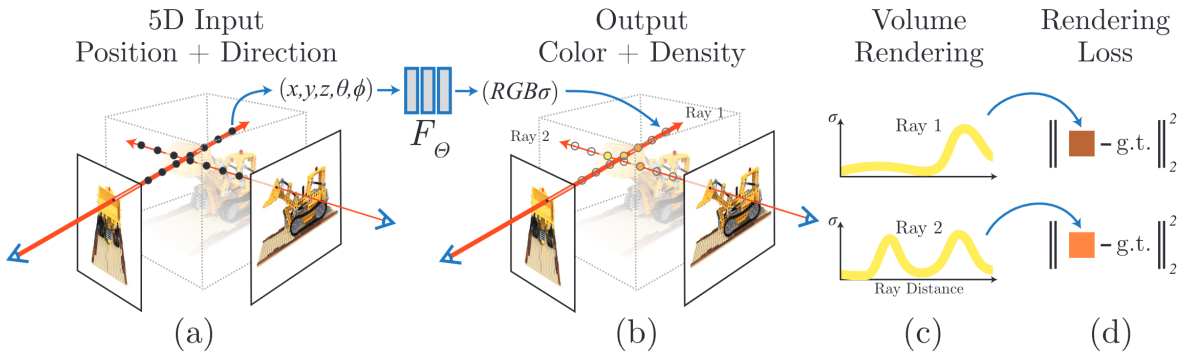


Figure 10: (a) Sample 5D coordinates along camera rays. (b) Use an MLP to map these to colour and density. (c) Apply volume rendering. (d) Compute the loss by comparing generated and ground-truth images [30].

Volume rendering and optimisation To render an image, NeRF computes the colour of each pixel by casting a ray from the camera into the scene and accumulating the colour and density contributions of all points along that ray. Formally, the expected colour $C(r)$ along a ray $r(t) = \mathbf{o} + t\mathbf{d}$, where $\mathbf{o}$ is the ray origin and $\mathbf{d}$ is the direction, is given by:

$$C(r) = \int_{t_n}^{t_f} \mathcal{T}(t)\, \sigma(r(t))\, c(r(t), \mathbf{d})\, dt, \tag{2.14}$$

where $\sigma(r(t))$ is the volume density and $c(r(t), \mathbf{d})$ is the view-dependent colour at each point. The term $\mathcal{T}(t)$ is the accumulated transmittance, representing the probability that the ray travels from the near bound t_n to a point t without being absorbed:

$$\mathcal{T}(t) = \exp\left(-\int_{t_n}^{t} \sigma(r(s))\, ds\right). \tag{2.15}$$

Note that $\mathcal{T}(t)$ here denotes the NeRF ray transmittance and is distinct from the complex transmission function $T(x, y)$ introduced in Section 2.2. This formulation naturally accounts for both occlusion and transparency, enabling the generation of photorealistic renderings.

The network is trained by minimising the error between the predicted pixel colour $\hat{C}(r)$ and the ground-truth pixel colour $C(r)$ across all rays $\mathcal{R}$ in the training images:

$$\mathcal{L} = \sum_{r \in \mathcal{R}} \left\| \hat{C}(r) - C(r) \right\|_2^2. \tag{2.16}$$

This squared ℓ_2 loss penalises large per-ray colour errors more heavily than small ones, encouraging the network to accurately reconstruct both fine detail and broad colour distributions. Through this optimisation, the network learns an implicit 3D representation of the scene entirely from 2D supervision, without any explicit geometric prior, and can subsequently be queried to synthesise novel viewpoints [30, 19].

2.6 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) [32] pursue a more ambitious goal than standard data-driven models. Rather than fitting a network purely to observations, the aim is to train a model that represents a complete spatio-temporal field—in the context of X-ray tomography, the full 4D evolution of a material's physical properties across space and time—while enforcing the governing partial differential equations (PDEs) as constraints throughout the domain. In this framework, the network does not merely interpolate between measurements; it learns a solution that is physically consistent everywhere, even where no data exist. This makes PINNs particularly compelling for MHz-Tomoscopy, where the 3D object evolves continuously in time yet only sparse, projection-based measurements are available at each instant.

In contrast to purely data-driven approaches, physical systems such as those studied in MHz-Tomoscopy are typically governed by well-established principles, including conservation laws and transport equations. This prior knowledge—often in the form of PDEs, empirically validated laws, or domain expertise—remains underutilised in conventional machine learning pipelines. When integrated into the training of neural networks, this structured information acts as a regularisation mechanism, significantly constraining the solution space and improving both convergence and generalisation, even in data-scarce scenarios [32]. For example, physically implausible solutions—such as negative densities or violations of mass conservation—can be automatically rejected by the model when these constraints are explicitly embedded.

In the context of materials science and X-ray imaging, this is particularly valuable. Consider, for example, the simulation of fluid penetration in a porous solid during an imaging experiment. Capturing the dynamic interaction between fluid flow, changing material boundaries, and X-ray absorption requires solving coupled PDEs under evolving boundary conditions. These systems are often governed by the Navier–Stokes equations, which describe the conservation of mass and momentum in fluid dynamics. Solving these equations accurately is critical for modelling the behaviour of incompressible or slightly compressible fluids under mechanical or thermal loading.

PINNs provide a mesh-free, flexible solution framework that can model these interactions by learning spatio-temporal fields directly—such as velocity, pressure, or density—while ensuring adherence to underlying physical laws. This approach enables the network to infer hidden states or dynamic field variables even when only partial or indirect measurements are available [32].

Earlier efforts to integrate prior physics into machine learning models used Gaussian processes for linear systems, enabling uncertainty quantification and analytic tractability. However, Gaussian processes face scalability issues and require linearisation or discretisation when extended to non-linear dynamic systems. PINNs overcome these limitations by leveraging the universal function approximation capabilities of deep neural networks and the computational efficiency of automatic differentiation to compute spatial and temporal derivatives without the need for finite difference schemes [32, 33].

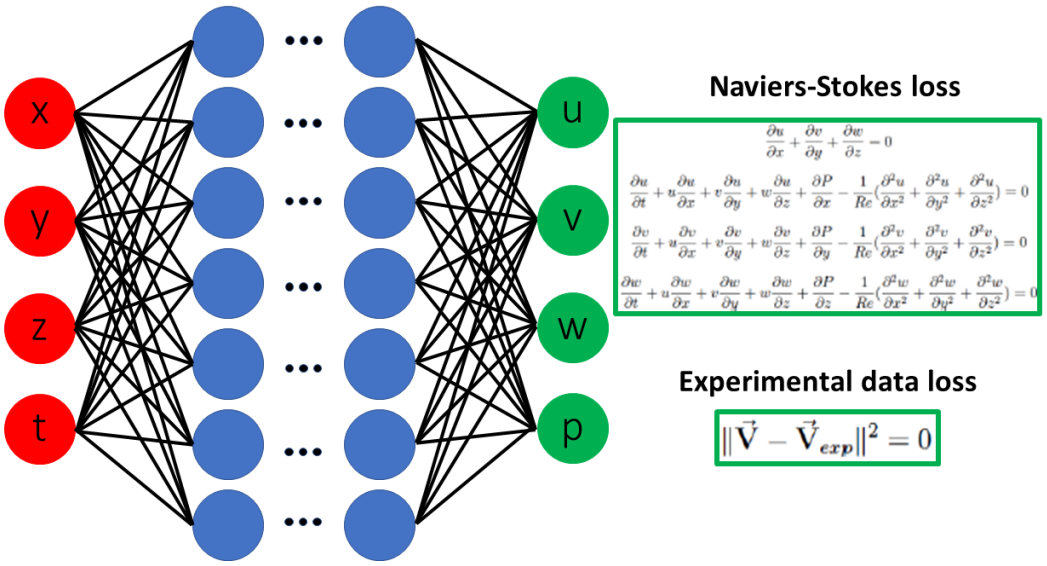


Figure 11: Physics-informed neural networks for solving Navier–Stokes equations. Adapted from [34].

2.7 3D Gaussian Splatting

3D Gaussian Splatting (3DGS), introduced by Kerbl et al. [29], represents a paradigm shift from implicit neural representations such as NeRF. Rather than encoding the scene geometry in the weights of an MLP, 3DGS represents a scene as a collection of explicit 3D Gaussian primitives. This point-based, explicit approach offers significant advantages in computational efficiency, physical interpretability, and the ability to model dynamic deformations, properties particularly relevant for MHz-Tomoscopy reconstructions.

Figure 12: Conceptual difference between NeRF and Gaussian Splatting. Image adapted from [35].

2.7.1 Scene Representation

A key limitation of implicit representations is their "black box" nature: extracting geometric information from a NeRF requires expensive post-processing, which can introduce errors and suppress fine detail. In 3DGS, geometry is explicit by construction - each primitive directly encodes a localized region of the scene, enabling straightforward analysis.

Each Gaussian $G(\mathbf{x})$ is defined by a center $\boldsymbol{\mu} \in \mathbb{R}^3$, a covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$ controlling the ellipsoid's shape and orientation, an opacity $\alpha \in [0, 1]$, and a view-dependent colour c encoded via spherical harmonics (SH) coefficients. The spatial contribution at point $\mathbf{x}$ is:

$$G(\mathbf{x}) = \alpha \cdot \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$

To ensure numerically stable optimization, $\boldsymbol{\Sigma}$ is not optimised directly. Instead, it is factored into a rotation matrix $\mathbf{R}$ and a diagonal scale matrix $\mathbf{S} = \mathrm{diag}(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^3$: $\boldsymbol{\Sigma} = \mathbf{R}\mathbf{S}\mathbf{S}^\top\mathbf{R}^\top$ [29].

2.7.2 Rendering Pipeline

Producing an image from the 3D Gaussian scene happens in three stages: projection, depth sorting, and alpha-blending compositing.

Projection to 2D (Splatting) The first stage gives the method its name. To render from a given viewpoint, every 3D Gaussian is forward-projected onto the image plane. The Gaussian's centre is transformed into camera (view) space via the world-to-view matrix $\mathbf{W}$, and its

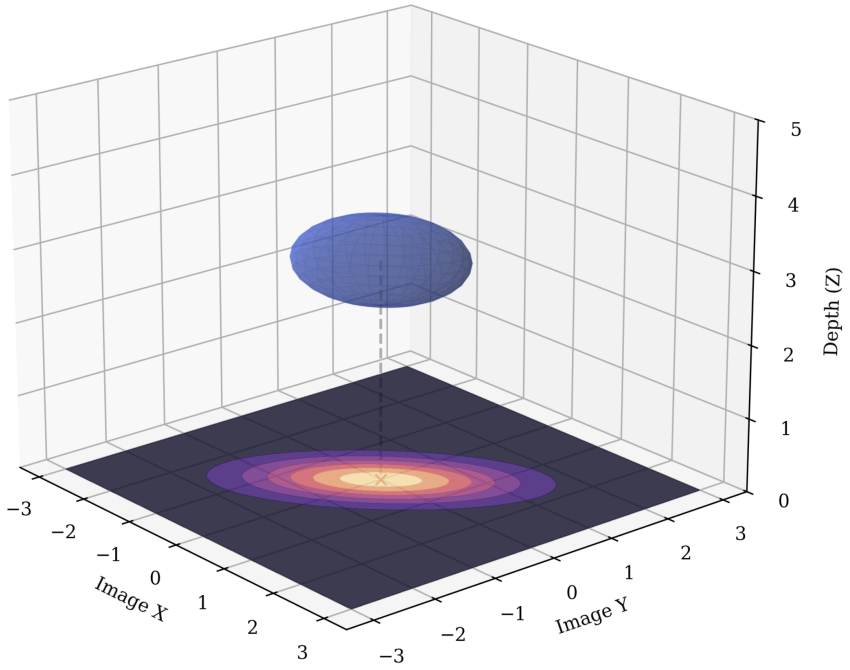


Figure 13: Projection of a 3D Gaussian onto the 2D image plane.

3D covariance $\boldsymbol{\Sigma}$ is projected to 2D using a local affine approximation via the Jacobian $\mathbf{J}$ of the projection mapping [29]:

$$\boldsymbol{\mu}_v = \mathbf{W}\boldsymbol{\mu}, \qquad \boldsymbol{\Sigma}' = \mathbf{J}\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^\top\mathbf{J}^\top.$$

The approximation is necessary for perspective cameras because a perspective projection technically distorts a Gaussian into a non-Gaussian shape [36]. For orthographic (parallel-beam) projection the mapping is already linear, so no approximation is needed. In both cases the outcome is a 2D Gaussian ellipse characterised by the projected mean $\boldsymbol{\mu}'$ and covariance $\boldsymbol{\Sigma}'$, ready to be "splatted" onto the image.

Depth Sorting Once all Gaussians have been projected, they must be ordered before compositing can take place. Gaussians that fall entirely outside the camera frustum are culled at this point. The surviving primitives are then sorted by their depth value—their distance from the camera along the viewing direction. In the official implementation this is performed efficiently using tiles: the screen is divided into small rectangular tiles, and a GPU radix sort orders the Gaussians within each tile. This tile-based organisation dramatically reduces the cost of the sort and enables massively parallel processing on the GPU. The sorted order establishes the front-to-back sequence required by the subsequent compositing step [29].

Rasterization and Alpha-Blending Compositing With the Gaussians sorted by depth, the renderer sweeps through them from the closest to the furthest and accumulates their contributions pixel by pixel. At each pixel location $\mathbf{x}'$, the colour contribution of the n-th Gaussian

depends on its learnt colour c_n, the footprint of its 2D ellipse at that pixel (encoded in α'_n, the Gaussian-weighted opacity), and the amount of light that has already passed through all nearer Gaussians. The latter quantity is the accumulated transmittance T_n (distinct from the complex transmission function $T(x, y)$ introduced in Section 2.2):

$$T_n = \prod_{j=1}^{n-1} (1 - \alpha'_j).$$

The final pixel colour is the weighted sum over all N Gaussians:

$$C = \sum_{n=1}^{N} c_n \, \alpha'_n \, T_n.$$

Intuitively, T_n starts at 1 (fully transparent path) and decreases with every Gaussian encountered; a Gaussian that lies behind a nearly opaque one contributes negligibly. This allows an early-exit optimisation: once T_n drops below a small threshold the traversal terminates, skipping all remaining occluded primitives. The result is the discrete counterpart of NeRF's continuous volumetric rendering integral (Eq. 2.14), but operating on explicit splats for substantially faster evaluation [29].

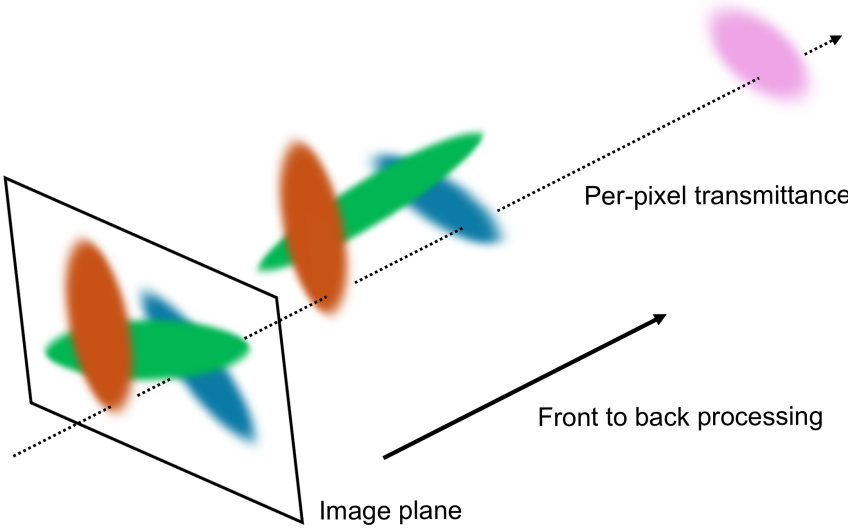


Figure 14: Compositing of Gaussian splats: sorted primitives are blended front-to-back, with each Gaussian's contribution attenuated by the transmittance accumulated from all nearer splats. In this example, each Gaussian is fully opaque ($\alpha = 1$), so no alpha blending effects are visible on the image plane. Image from [37].

2.7.3 Optimization

Optimisation starts from a sparse point cloud. Each point is initialised with small covariances and opacities based on local point density. The training loss combines an ℓ_1 intensity loss with a structural similarity term:

$$\mathcal{L} = (1 - \lambda)\,|\hat{I} - I|_1 + \lambda\left(1 - \mathrm{SSIM}(\hat{I}, I)\right), \tag{2.17}$$

where $\hat{I}$ is the rendered image, I is the ground truth image, and λ — a weighting factor balancing pixel-wise and perceptual accuracy. The ℓ_1 term measures the average absolute difference between pixels and is less sensitive to outliers than an ℓ_2 loss. This makes it more robust to effects introduced, for example, by glossy materials. The two loss terms complement each other: the ℓ_1 term encourages accurate pixel values, while SSIM [38] helps preserve overall structure, contrast, and sharpness. Together, they promote reconstructions that are both accurate and visually consistent, which is important for novel view synthesis.

Gradients are propagated through the differentiable rasteriser, and all Gaussian parameters are updated using the Adam optimiser. During training, the set of Gaussians is also adjusted to better fit the scene:

- **Densification**: Gaussians with large positional gradients (indicating poorly represented regions) are duplicated; large Gaussians are split into smaller ones.
- **Pruning**: Gaussians with very low opacity (below some defined threshold) or overly large Gaussians are removed.

Training usually converges in a few minutes, which is much faster than NeRF methods that can take hours or days [29, 31].

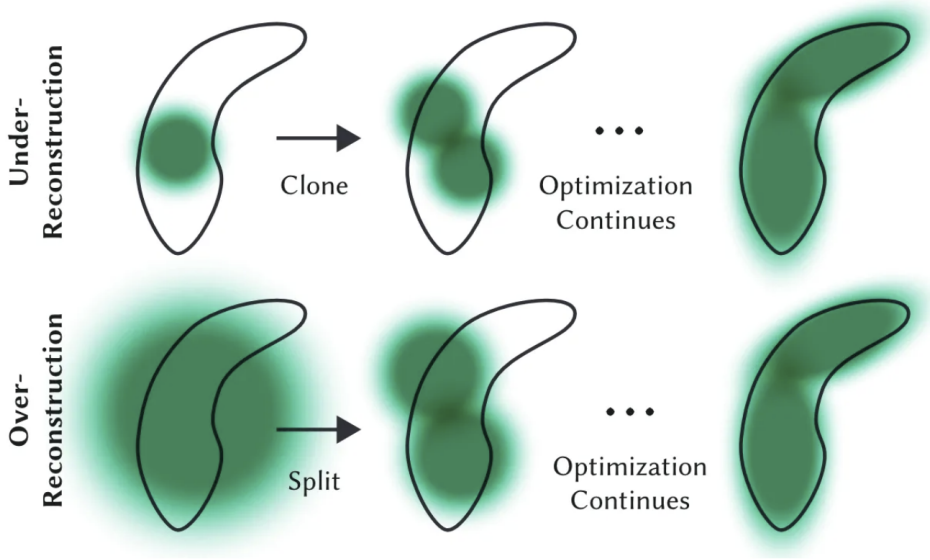


Figure 15: Adaptive densification and pruning. Adapted from [29].

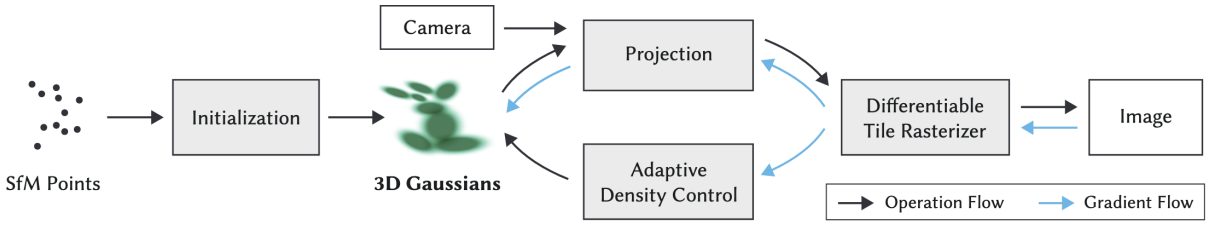


Figure 16: The 3DGS optimization pipeline, alternating gradient-based updates with adaptive densification and pruning. Adapted from [29].

2.8 ONIX

Standard NeRF requires extensive per-scene optimization from dozens or hundreds of viewpoints, making it impractical for sparse-view tomography. PixelNeRF [39] addresses this limitation through a fundamentally different training strategy: rather than fitting a separate network to each scene, a single network is trained across a dataset of multiple scenes, learning general scene priors. At inference, this shared network is conditioned on input images via a CNN encoder that extracts spatial feature volumes. For each query point along a camera ray, the corresponding image feature $W(\pi(x))$ is sampled from the feature volume at the projected pixel location $\pi(x)$. These features, together with spatial coordinates, are fed into the NeRF network to predict density and colour. This architecture enables reconstruction from as few as one or several views without requiring scene-specific optimization [39].

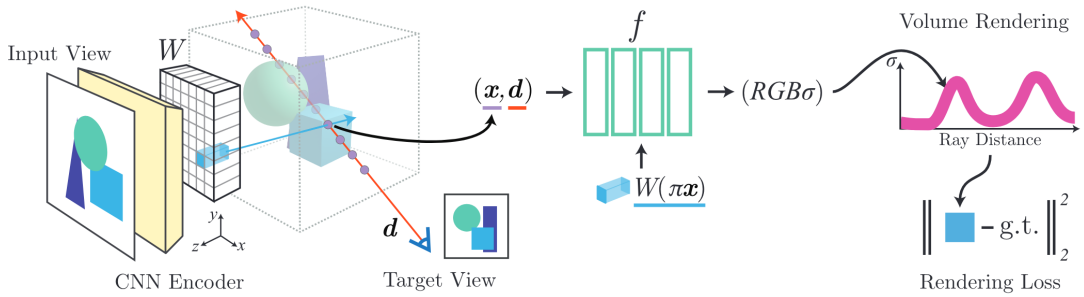


Figure 17: PixelNeRF architecture for single-view reconstruction. The network is conditioned on spatial image features extracted by a CNN encoder. Adapted from [39].

ONIX [40] adapts this few-shot framework to X-ray tomography by incorporating the physics of X-ray propagation. The key modification is that instead of predicting RGB colour and volumetric density, ONIX predicts the linear attenuation coefficient, modelling X-ray absorption via the Beer-Lambert law (Eq.2.8). The imaging geometry is fixed to parallel-beam, as shown in Figure 18.

Reconstruction proceeds by casting a parallel ray through each detector pixel. For each ray, the network predicts the attenuation coefficient at regular depth intervals, which are integrated via Beer-Lambert absorption to form the final projected intensity.

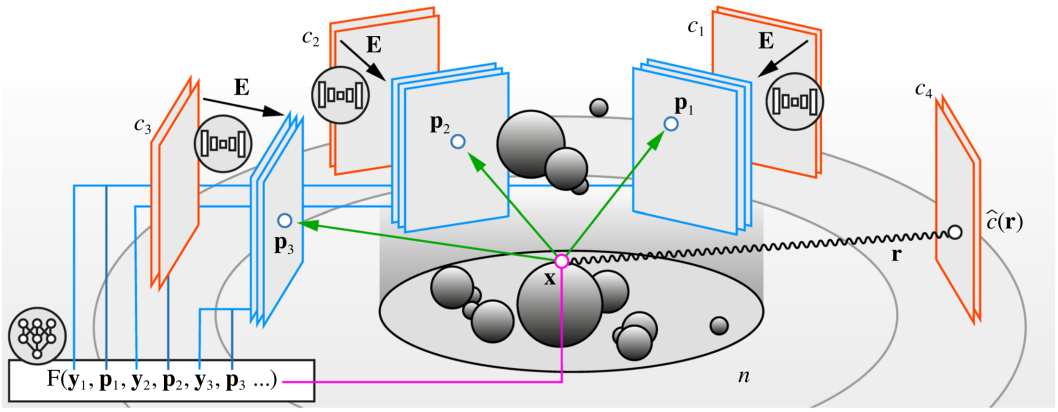


Figure 18: ONIX architecture. An encoder E maps input X-ray projections to latent feature volumes; the network F predicts the linear attenuation coefficient from 2D projected coordinates and corresponding latent vectors. Adapted from [40].

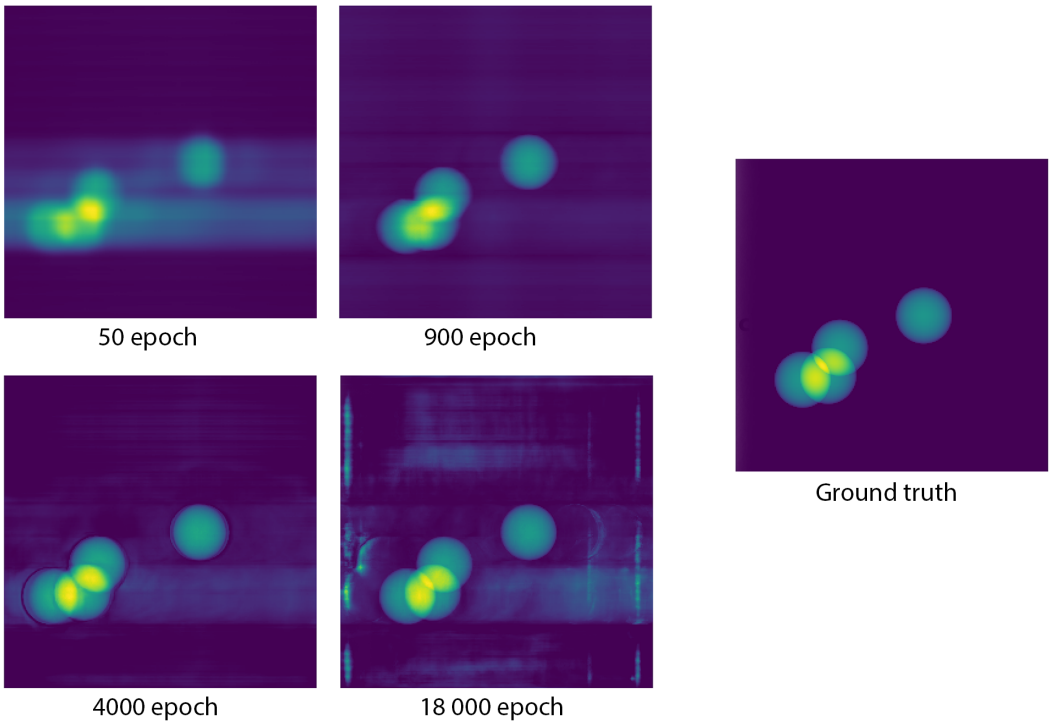


Figure 19: Novel viewpoints synthesised from five input projections (synthetic dataset) following extended training of 15 hours. Despite the sparse angular sampling, the reconstructed representation successfully preserves the overall geometry of the object. Adapted from [3].

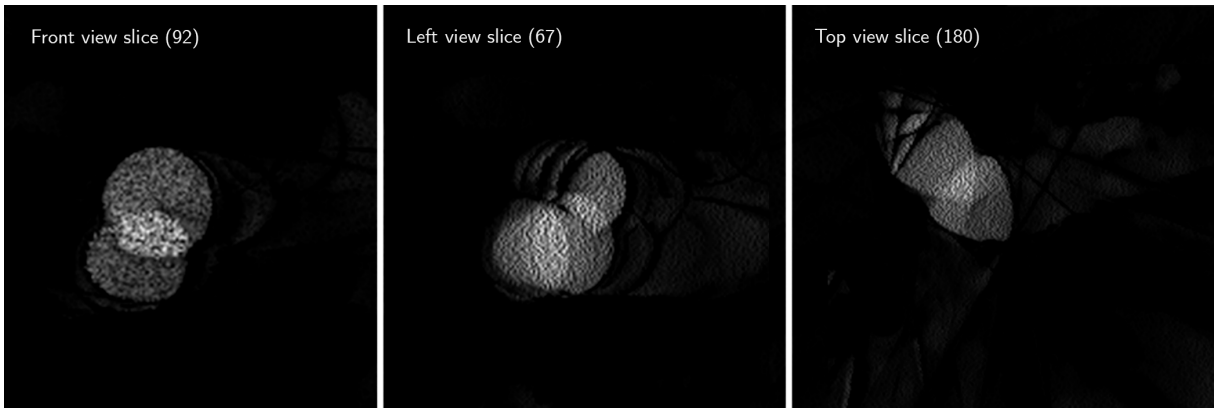


Figure 20: Cross-sectional slices of the reconstructed volume shown along three orthogonal planes, illustrating the internal structure of the recovered 3D density field. Adapted from [3].

ONIX has demonstrated state-of-the-art reconstruction quality for few-view X-ray tomography, particularly when compared to classical methods such as FBP and SIRT under highly sparse sampling conditions [3, 5]. However, a fundamental limitation is that it operates in the attenuation-only regime, modelling only absorption via the Beer-Lambert law. Furthermore, the implicit representation makes it difficult to apply physical constraints or to inspect individual scene parameters directly. This approach is valid for conventional absorption-contrast imaging but fails to capture phase information in holographic phase-contrast regimes, where X-rays undergo both absorption and phase shift.

In such settings, the detector no longer records a direct projection of the attenuation but instead captures interference fringes that encode the complex refractive index of the sample. Recovering the underlying structure from these holograms is inherently more ambiguous, as both the real and imaginary parts of the refractive index — δ and β respectively — must be jointly estimated. Extending ONIX to this regime would therefore require predicting both quantities simultaneously, alongside a physically accurate model of wavefront propagation from the sample to the detector. This represents a substantial increase in problem complexity, with direct implications for network design, training stability, and computational demand.

3 Materials and Methods

3.1 Simulation-Based Dataset Generation

A central challenge in applying novel-view synthesis to X-ray phase-contrast imaging is the absence of large, annotated training datasets. XFEL experiments are expensive and time-consuming, ground-truth three-dimensional structure is generally unknown, and the precise imaging conditions are difficult to reproduce. To overcome this, the training data used in this work were generated entirely in silico through a purpose-built forward simulation pipeline. The pipeline takes a single snapshot of a molecular dynamics simulation as its input and produces a set of synthetic phase-contrast projections that faithfully model the physics of Fresnel

near-field diffraction — the same mechanism as in MHz-Tomoscopy. Because every aspect of the scene is known exactly, the pipeline provides not only training images but also perfect geometric ground truth, giving full control over imaging parameters that are fixed once the experiment begins in reality.

3.1.1 Molecular Dynamics Simulation

The biological specimens of interest are Red Blood Cells (RBCs) flowing through a model microfluidic channel — a scenario directly relevant to medical applications of MHz-Tomoscopy. Their dynamics were simulated using the ESPResSo software package [41], specifically its PyOIF (Python Object-in-Fluid) extension [42]. ESPResSo provides a computational environment for solving Newton's equations of motion for a system of interacting particles, and PyOIF extends this to elastic objects immersed in fluid by representing each cell as a deformable triangulated mesh. The result is a physically realistic representation of the characteristic biconcave shape of RBCs under flow. The meshes used were provided by the Cell-in-Fluid Biomedical Modelling and Computation Group at the University of Žilina, Slovakia [43].

At selected points in time, the simulation state is exported as a set of VTK (Visualization Toolkit) Unstructured Grid files — one per cell — each encoding the current vertex positions and the triangular surface connectivity of the corresponding mesh. A single such snapshot, illustrated in Figure 21, is the starting point for all subsequent processing.

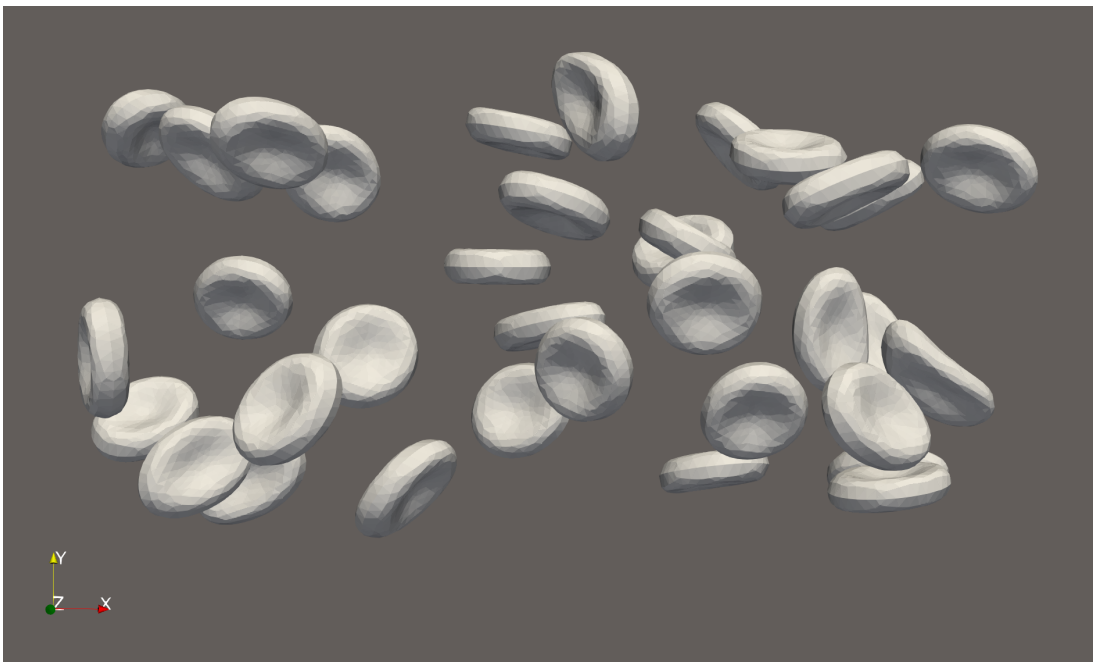


Figure 21: RBCs at one simulation snapshot, rendered from the corresponding VTK surface meshes in ParaView [44].

3.1.2 From Geometry to Phase-Contrast Projections

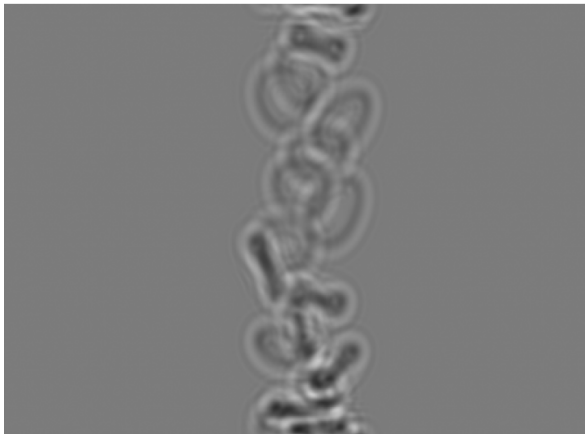
To turn the geometric snapshot into synthetic X-ray images, the pipeline must bridge two domains: the discrete triangulated surfaces produced by the MD simulation and the continuous wavefield physics of X-ray propagation.

The first step is to assign X-ray optical properties to each mesh - complex index of refraction (β - absorption, δ - phase shift) discussed in Section 2.2. Both quantities depend on the elemental composition and mass density of the material as well as on the photon energy, and are retrieved for each mesh from the `xraydb` library [45] based on its chemical formula and density. RBC and plasma meshes are assigned haemoglobin by default, with support for custom material configurations via an external file. At the operating energy of $20\,\mathrm{keV}$, the phase contribution dominates strongly over absorption. This large difference is precisely why conventional attenuation-based imaging produces almost no contrast from soft biological cells at hard X-ray energies, and why phase-contrast detection is essential.

With material properties assigned, the surface meshes are rasterised onto a shared three-dimensional Cartesian voxel grid, producing two scalar volumes: V_β and V_δ, which encode the spatial distribution of the absorption and phase indices throughout the scene. The grid resolution is set to match the coordinate units of the VTK files, ensuring that the physical scale of the simulation is preserved without rescaling.

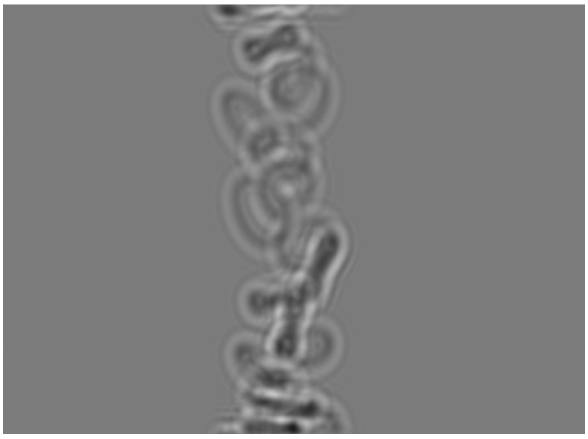
Parallel-beam projections are then computed by casting a grid of rays through these volumes for each viewing angle, accumulating the line integrals of β and δ along each ray path to obtain the projection maps B and D in physical units. The raycasting is accelerated on GPU via CUDA when available, with a CPU fallback for broader compatibility. A full set of views is acquired by rotating the virtual detector around the sample over $180°$, providing the all-around coverage that novel-view synthesis requires.

Finally, the projection maps are passed through the phase-contrast forward model described in Section 2.2. As detailed there, the model constructs the complex transmission function $T(x, y)$ (Section 2.2) at the sample exit plane, propagates the resulting wavefield to the detector using the Fresnel transfer function, and computes the squared modulus of the propagated field as the detected intensity. Then a Gaussian point spread function accounting for the finite resolution of the detector is applied. The output is a physically realistic phase-contrast hologram I_{pred}, exhibiting the characteristic bright–dark edge fringes produced by Fresnel near-field diffraction, as illustrated in Figure 22.



(a) $\theta = 0°$

(b) $\theta = 45°$



(c) $\theta = 168.8°$

Figure 22: Simulated phase-contrast projections of RBCs at three viewing angles (X-ray energy $E = 20\,\mathrm{keV}$, sample-to-detector distance $z = 10\,\mathrm{cm}$). The bright–dark fringes at cell boundaries are characteristics of Fresnel near-field diffraction and carry the phase information that the 3DGS model is trained to reproduce.

3.1.3 Dataset Packaging for 3DGS Training

The final stage converts the computed projections into the format expected by the 3DGS training framework. Each projection is paired with the known camera pose for that viewing angle, represented as a camera-to-world transformation matrix. Because the imaging geometry is fully determined analytically, there is no need for structure-from-motion reconstruction or any other pose estimation step — the poses are computed directly from the angles used during raycasting.

All projections in the set are normalised jointly to the range $[0, 1]$ using a single global min–max mapping. Using a common scale across all views is important: different viewing angles through an asymmetric sample genuinely produce different mean intensities, and collapsing each view to the same range independently would erase this physically meaningful variation, which the model must learn to predict.

The views are partitioned into a training set and a test set, where test views are dis-

tributed uniformly across the full angular range rather than clustered at one end. An initial point cloud, used to seed the Gaussian parameters before optimisation, is generated by sampling points uniformly at random within the scene volume. The complete dataset — images and camera poses (train/test) — is written in a layout compatible with the NeRF Synthetic format used by the 3DGS framework, allowing the standard training pipeline to be used without modification.

3.2 Modified GS Pipeline for Phase-Contrast Tomoscopy

The standard 3DGS framework [29] was designed for photorealistic novel-view synthesis: each 3D Gaussian carries spherical harmonic (SH) colour coefficients and an opacity value, and a differentiable tile-based rasterizer alpha-composes them front-to-back into an RGB image optimised against multi-view photographs. In the present work, this pipeline is restructured into a physically motivated forward model for X-ray phase-contrast imaging. The key modifications are:

- **Scene representation.** SH colour coefficients and opacity are replaced by two scalar material parameters per Gaussian—the absorption index β and the refractive index decrement δ. A density parameter is kept only for adaptive density control (pruning) and does not enter the rendering equation.
- **Projection geometry.** Perspective projection is replaced by orthographic (parallel-beam) projection, consistent with MHz-Tomoscopy.
- **Rasterization.** Modified CUDA rasterizer outputs two channels—the accumulated absorption map B and the accumulated phase-shift map D—instead of three RGB channels.
- **Phase-contrast forward model.** This part converts the B and D maps into a predicted detector intensity I_{pred} by applying the complex transmission function, Fresnel free-space propagation and detector point-spread function (describes how a focused optical imaging system responds to a point source of light, resulting in a blurred image rather than a perfect point).
- **Optimisation.** Training proceeds in two phases: a material-only warmup in which geometry is frozen, followed by joint optimisation of all parameters once the material field has stabilised. The training objective is a composite loss combining L1 reconstruction error, structural similarity (SSIM) [38], a Fourier Ring Correlation (FRC) term [46, 47], and a simplified Total Variation (TV) regulariser [48]. The whole loss function and its components are explained later in this section (see Eq. 3.1). Multi-view batching and anisotropy-aware regularisation further stabilises training.

Figure 23 illustrates the end-to-end architecture of the proposed pipeline.

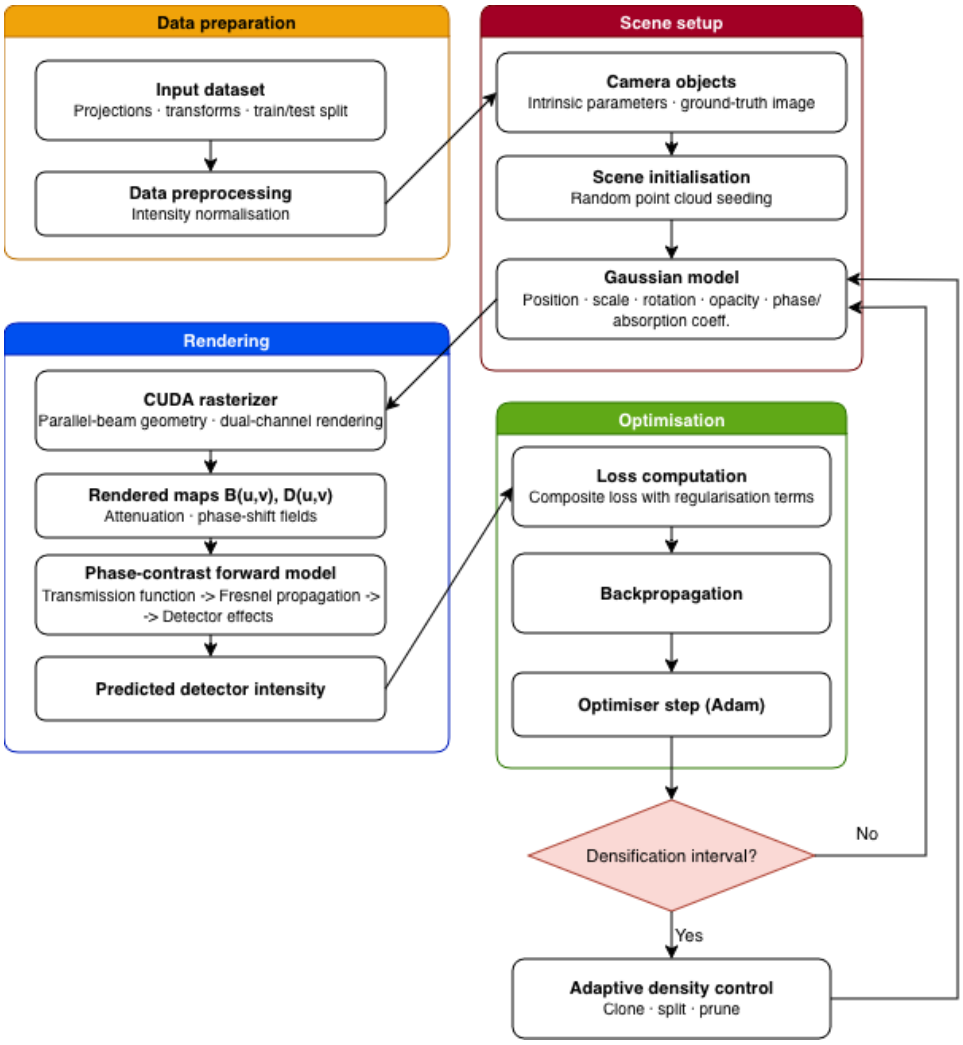


Figure 23: Proposed Gaussian Splatting pipeline for phase-contrast MHz-Tomoscopy.

Gaussian scene representation. The scene is modelled as a collection of N anisotropic 3D Gaussians. Rather than encoding colour and view-dependent opacity as in standard radiance fields, each Gaussian carries the physical properties of the sample—specifically the complex refractive index. The learnable parameters are summarised in Table 1.

Table 1: Learnable parameters of each Gaussian primitive.

Attribute	Symbol	Physical Interpretation
Position	$\mathbf{x}$	3D centroid in world coordinates
Scaling	$\mathbf{s}$	Semi-axes of the ellipsoid
Rotation	$\mathbf{q}$	Orientation via unit quaternions (compact rotation description)
Density	d	Geometric gating for pruning
Absorption index	β	Imaginary part of the refractive index
Refractive decrement	δ	Real-part deviation of the refractive index

Both β and δ are constrained to be non-negative, since X-ray materials can only attenuate or retard the wavefront. This is enforced transparently during optimisation via a smooth

activation function applied to the raw learnable values.

Each Gaussian has an ellipsoidal shape defined by its position, three scale values, and an orientation quaternion. Rather than emitting colour as in a standard radiance field, each primitive acts as a localised region of material: as the beam passes through the scene, it accumulates absorption and phase shift from all Gaussians along the ray path. Because X-ray imaging involves no occlusion — every Gaussian contributes additively regardless of depth — per-Gaussian opacity is fixed to one, and the density parameter is used only for pruning.

3.3 Training Stabilisation

Adapting 3DGS to sparse-view X-ray optimisation introduces several challenges: instability during early iterations, degenerate Gaussian shapes, and high-frequency artefacts in the predicted projections. A set of complementary enhancements addresses each of these in turn.

Composite training objective. Each predicted projection $\hat{I}$ is compared to the ground-truth measurement I using a composite loss:

$$\mathcal{L} = 0.75\,\ell_1(\hat{I}, I) + 0.20\,\big(1 - \mathrm{SSIM}(\hat{I}, I)\big) + \mathcal{L}_{\mathrm{TV}}(\hat{I}) + w_{\mathrm{FRC}}(t)\,\mathcal{L}_{\mathrm{FRC}}(\hat{I}, I). \tag{3.1}$$

The dominant terms are the pixel-wise L1 error and the structural-similarity penalty $(1 - \mathrm{SSIM})$, which together drive the reconstruction towards accurate intensity values and correct spatial structure. A lightweight Total Variation term $\mathcal{L}_{\mathrm{TV}}$ [48] suppresses high-frequency noise in the rendered projections without blurring the fringe patterns that carry the phase signal. A Fourier Ring Correlation term $\mathcal{L}_{\mathrm{FRC}}$ [46] further penalises mismatches at high spatial frequencies, where fringe detail is concentrated; its weight $w_{\mathrm{FRC}}(t)$ starts at zero and increases gradually over the first few thousand iterations, so that it only activates once the coarser L1 and SSIM fit is already established.

Simultaneous multi-view optimisation. Standard 3DGS optimises against a single randomly selected view per iteration, which tends to overfit individual viewpoints and produce conflicting gradient updates. Instead, V views are sampled simultaneously per gradient step. The loss in Eq. 3.1 is evaluated per view and averaged over the shared Gaussian parameters before backpropagation. Views whose loss exceeds the batch gradient median have their gradients detached, preventing any single poorly fitting view from destabilising the entire parameter set. The effect of varying V is examined in Section 4.2; in practice, moderate V offers the best quality–cost trade-off for the scene complexity used here.

Anisotropy-aware regularisation and pruning. Multi-view training tends to produce highly flattened Gaussians (needle or sheet-like shapes) that cause view-dependent artefacts.

A regularisation term penalises Gaussians whose smallest-to-largest scale ratio falls below a configurable minimum, weighted by a small factor. Complementarily, during periodic densification steps, Gaussians whose scale ratio falls below the same threshold are pruned outright alongside those with low opacity, removing persistent degenerate primitives that regularisation alone cannot suppress.

Taken together, these enhancements yield more stable training dynamics, cleaner geometry, and improved reconstruction quality with negligible additional computational overhead.

4 Results and Discussion

Holographic 3D Gaussian Splatting presents a reconstruction challenge that is fundamentally a joint compression and phase-retrieval task: the representation must simultaneously produce a compact, physically interpretable scene description and recover two distinct material parameters — the absorption index β and the refractive-index decrement δ — from intensity-only detector measurements in which phase information is encoded solely through the spatial structure of Fresnel fringe patterns.

The hologram intensity recorded at the detector is the squared magnitude of the propagated complex wavefield. This measurement is irreversible: many (β, δ) distributions can produce indistinguishable intensity patterns [7]. The reconstruction is therefore structurally equivalent to an embedded phase retrieval problem, entangled within a larger scene optimisation with no explicit phase-consistency constraint. The results that follow should be interpreted with this fundamental ill-posedness in mind, in particular, the gap between training and test performance reflects the inherent ambiguity of the inverse problem rather than straightforward overfitting.

4.1 Experimental Setup and Evaluation Metrics

All experiments were conducted on a single time-step snapshot of the RBC molecular dynamics simulation, using the simulation-based projection dataset described in Section 3.1. Sixteen views were distributed uniformly on 180° of rotation and divided into 12 training and 4 held-out test views using the interleaved schema described in Section 3.2. The imaging geometry was fixed throughout: pixel size $3.2\,\mu\mathrm{m}$, sample-to-detector propagation distance $z = 10\,\mathrm{cm}$, and a Gaussian detector point-spread function with $\sigma = 0.05$ px.

Each run was optimised for 30 000 iterations with the Adam optimiser. The material parameters were updated in isolation during the first 1 000 iterations (the warmup phase; Section 3.3), after which the geometry was unlocked and adaptive densification was active between iterations 3 000 and 15 000. All other optimisation hyperparameters were kept constant unless explicitly stated.

Four categories of design choices were systematically varied:

1. **Views per gradient step** At each optimisation step, the model can be updated using gradients accumulated from one or several projection views simultaneously. Using more views per step provides a richer training signal but increases memory and computation per iteration. We tested $V \in \{1, 2, 4, 8\}$ views, keeping the initial Gaussian count fixed at $N_0 = 8\text{k}$ and evaluating both 8 and 16 projections dataset.

2. **Initial Gaussian density** The reconstruction begins from a cloud of randomly placed Gaussians that are progressively refined. Starting with more Gaussians gives a denser initial coverage of the volume but increases memory usage and training time. We tested $N_0 \in \{8\text{k},\ 32\text{k},\ 98\text{k},\ 244\text{k}\}$ at fixed $V = 4$, again across both 8 and 16 projection directions.

3. **Learning-rate sensitivity** Two sets of learning rates govern how quickly Gaussian positions and shapes are updated: the position learning rate (`plr`) controls spatial movement, while the scaling+rotation learning rate (`slr`) controls how fast each Gaussian's shape and orientation adapt. Starting from the 8k, $V = 4$ baseline, we halved and doubled each rate independently to assess sensitivity and identify a stable operating range. The tested values were $\text{plr} \in \{8 \times 10^{-5},\ 1.6 \times 10^{-4},\ 3.2 \times 10^{-4}\}$ and $\text{slr} \in \{5 \times 10^{-4},\ 1 \times 10^{-3},\ 2 \times 10^{-3}\}$, where the middle value is the default in each case.

4. **Physics and densification settings** We tested four additional settings that control physical realism and how the Gaussian cloud grows: the FRC weight (`frc_max_weight`), how long geometry is frozen at the start (`material_warmup`), the anisotropy penalty (`lambda_scale_reg`), and the densification threshold (`densify_grad_threshold`). Each one was changed individually while keeping everything else the same. None of these changes performed better than the default settings described in Section 3.3, so these runs are not included in the eight main configurations reported below.

Following a systematic exploration of the optimisation space described above, eight representative configurations were selected; their settings and final quantitative metrics are summarised in Table 2.

Choice of evaluation metrics. A central finding of this work is that the choice of primary metric has a decisive effect on which configuration appears "best". Reconstruction quality is measured on the each training run using five complementary metrics:

- **PSNR** (dB) and **L1** — pixel-wise intensity error. These are the traditional primary metrics in computer-vision novel-view synthesis.

- **SSIM** (Structural Similarity Index) — measures perceptual structural similarity including luminance, contrast, and spatial structure.

- **FRCM** (Fourier Ring Correlation Metric) — a radially weighted mean of per-ring Fourier Ring Correlation values across all spatial frequencies, where rings at higher radii receive proportionally greater weight to emphasise fine fringe structure. FRCM lies in $[0, 1]$; higher values (towards 1.0) indicate greater spectral agreement between two images [46, 47].

Table 2: Training configurations and final test-split metrics for selected eight runs. N_0 — initial Gaussian count; V — views per gradient step; N_f — final Gaussian count at iteration 30 000; $\overline{\delta/\beta}$ — mean per-Gaussian refractive-to-absorption ratio at convergence. Best values in each metric column are in **bold**; for SSIM, FRCM, and PSNR higher is better, for L1 lower is better. Learning rate (LR)-tuned runs are marked with †.

Configuration	N_0	V	$\text{SSIM}_{\text{test}}$	$\text{FRCM}_{\text{test}}$	$\text{PSNR}_{\text{test}}$	L1_{test}	N_f	$\overline{\delta/\beta}$
8k / $V=1$	8k	1	0.692	0.072	15.1	0.126	86 300	115
8k / $V=4$	8k	4	0.756	0.068	16.0	0.117	18 900	165
8k / $V=8$	8k	8	0.803	0.072	17.2	0.103	2 200	45
32k / $V=4$	32k	4	0.539	0.025	18.0	0.097	17 200	15
98k / $V=4$	98k	4	0.447	0.022	18.5	0.091	31 400	12
244k / $V=4$	244k	4	0.449	0.028	**19.7**	0.080	36 600	5
† 8k / plr = 3.2e-4	8k	4	0.832	**0.096**	18.4	0.088	14 000	26
† 8k / slr = 2e-3	8k	4	**0.870**	0.094	19.5	**0.070**	10 300	20

PSNR as a misleading metric. Before examining the individual trained models, it is essential to establish why SSIM and FRCM, rather than PSNR, serve as the primary quality metrics in this work. In holographic reconstruction, PSNR is a misleading primary metric because it rewards matching the mean background intensity without requiring correct fringe structure. Throughout this chapter, SSIM and FRCM are therefore adopted as the primary quality metrics, with PSNR retained for reference and comparability with the broader 3DGS literature.

Figure 24 illustrates this problem directly, presenting the final test-split values of all three metrics for every run, sorted by PSNR rank in each panel so that rank inversions become immediately visible. The 244k configuration achieves the highest PSNR (19.7 dB) yet records the lowest SSIM (0.449) and near-lowest FRCM (0.028). Conversely, the learning-rate-tuned runs ($\text{slr} = 2 \times 10^{-3}$, $\text{plr} = 3.2 \times 10^{-4}$) reach only moderate PSNR (18.4–19.5 dB) while dominating on both structural metrics: SSIM 0.832–0.870 and FRCM 0.094–0.096.

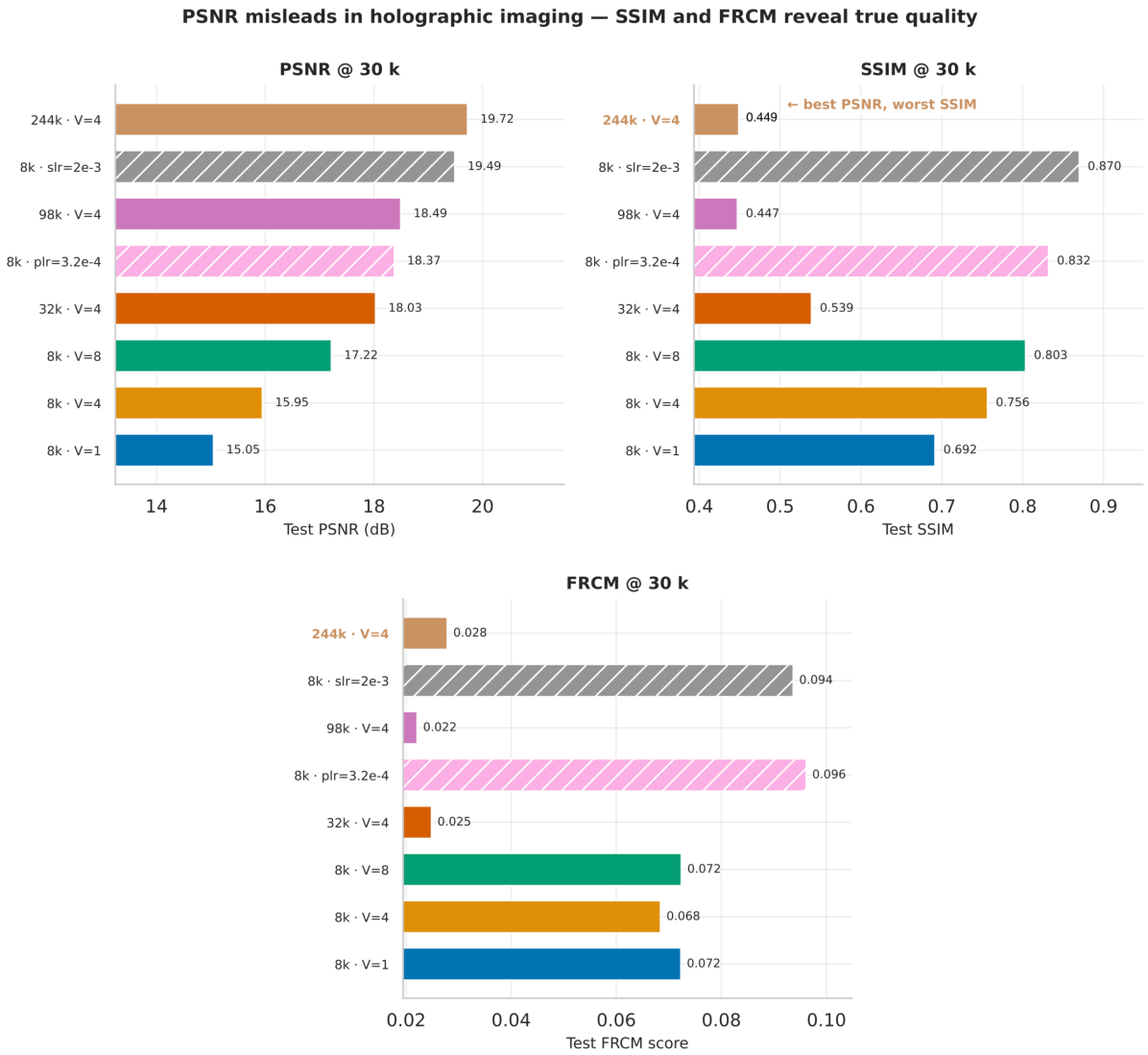


Figure 24: PSNR, SSIM, and FRCM at iteration 30 000, with all panels sorted by PSNR rank. The 244k run achieves the highest PSNR (19.7 dB) yet the lowest SSIM (0.449) and near-lowest FRCM (0.028), while the learning-rate-tuned runs dominate on both structural metrics despite only mid-range PSNR. This rank inversion demonstrates that PSNR rewards matching the mean background intensity without requiring correct Fresnel fringe structure.

This rank inversion has a clear physical explanation. PSNR is a pixel-wise intensity metric: a dense Gaussian cloud (244k primitives) can learn an accurate absorption fit that matches pixel intensities on average, converging to a degenerate absorption-only solution that never resolves the Fresnel fringes encoding the phase signal. SSIM and FRCM expose this failure because they measure spatial structure and frequency-domain agreement, respectively, penalising solutions that are photometrically close but structurally incorrect.

The implication extends beyond this experiment. Standard 3DGS benchmarks rank methods primarily by PSNR, but for any imaging modality where the signal of interest is encoded in spatial structure rather than absolute intensity, structural and frequency-domain metrics are indispensable. All subsequent analysis in this chapter therefore uses SSIM and

FRCM as the primary criteria.

Training dynamics. All eight runs converge steadily, though with quite different dynamics depending on the configuration. During the material-only warmup (iterations 1–1 000), the loss decreases smoothly as β and δ adjust while geometry remains frozen.

A distinctive feature of multi-view training is the aggressive pruning that follows the end of densification at iteration 15 000. The final Gaussian populations in Table 2 vary by more than an order of magnitude: the $V = 1$ run retains approximately 86 000 primitives, whereas the $V = 8$ run converges to only 2 200. This contrast — illustrated in Figure 30 (bottom image) — reflects a fundamental difference in how the two regimes constrain individual Gaussians. In single-view mode, a primitive that explains one training view well is never penalised for failing on others; it accumulates weak gradient signal and survives pruning. In multi-view mode, a Gaussian that cannot simultaneously account for multiple angularly separated views generates conflicting gradient updates, drives its effective density down, and is eventually removed. Multi-view optimisation therefore acts as an implicit regulariser that enforces cross-view structural consistency at the cost of a leaner final representation.

4.2 Model evaluation

Reconstruction quality overview. Figure 25 presents the primary quality overview: test SSIM trajectories over training for all eight runs, together with bar charts of the final SSIM and FRCM values. Two findings stand out immediately. First, the two LR-tuned runs (dashed curves) dominate all others by a substantial margin: slr achieves a final SSIM of 0.870 and plr reaches 0.832, compared to 0.803 for the third-best run ($V = 8$). Second, SSIM trajectories for the high-density runs (32k, 98k, 244k) degrade steadily after iteration 5 000, converging to values of 0.45–0.54 — far below even the baseline $V = 4$ run at 0.756. This pattern is the opposite of what PSNR suggests, underscoring the importance of the metric choice established in previous Section 4.1.

The training loss curves shown in Figure 26 confirm this picture at the optimisation level: LR-tuned runs maintain a consistently lower composite loss throughout training, whereas the high-density runs exhibit stagnation or divergence after the densification phase ends, mirroring the SSIM collapse observed above.

Figure 28 presents a side-by-side comparison of the rendered hologram outputs (test views) and the ground-truth projection across all eight runs. Runs with high SSIM and FRCM successfully reproduce the characteristic Fresnel fringe structure: bright halos and dark edge fringes at cell boundaries are visible in both the prediction and the ground truth. The overall fringe spacing, relative contrast, and the intensity distribution across the field of view are captured with good fidelity. The remaining runs either overfitted the fringe patterns or failed to generalise to test views.

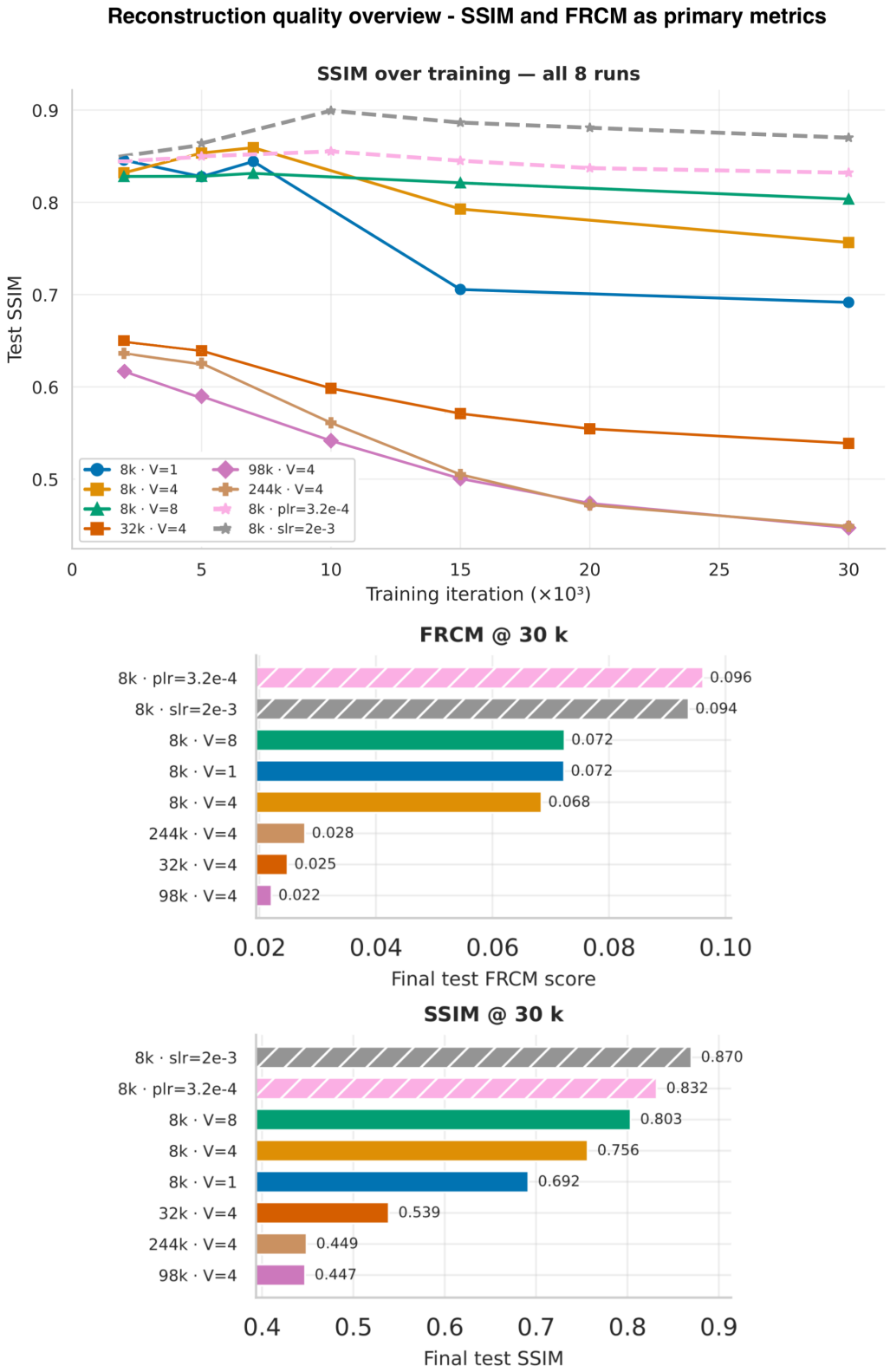


Figure 25: Reconstruction quality overview. **Top:** Test SSIM over training for all eight configurations. Dashed curves denote LR-tuned runs. **Middle:** Final test FRCM score. The LR-tuned runs (slr and plr) dominate both metrics by a wide margin. **Bottom:** Final test SSIM at iteration 30 000 (sorted; striped bars = LR-tuned).

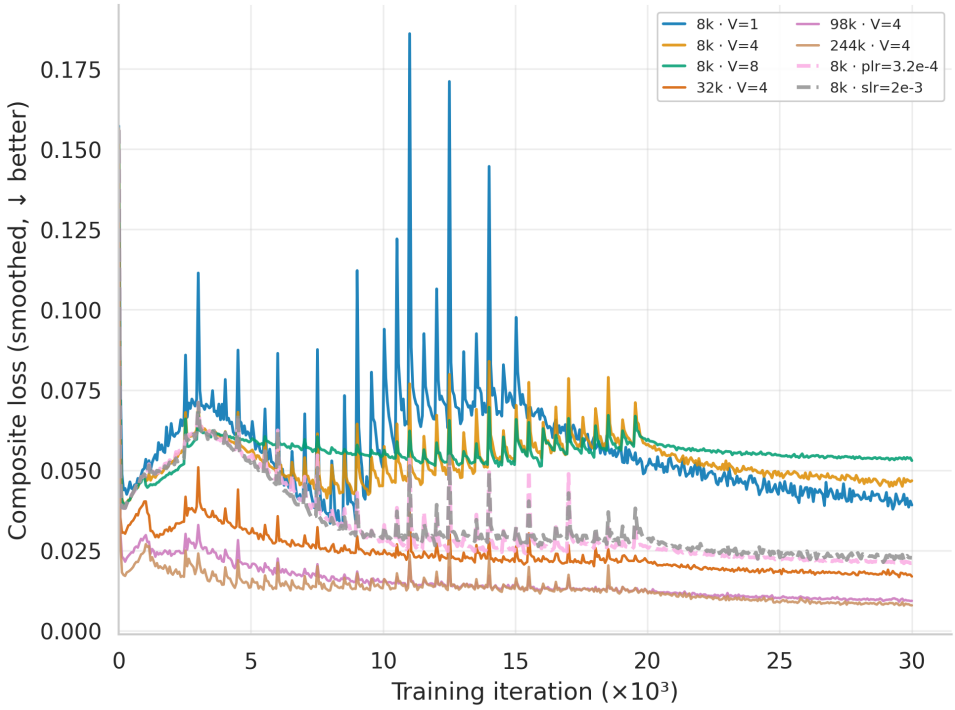


Figure 26: Composite training loss over 30 000 iterations across all eight configurations. Although the LR-tuned runs do not achieve the lowest training loss, they generalise more effectively, as reflected in the test metrics—an instance of the bias–variance trade-off.

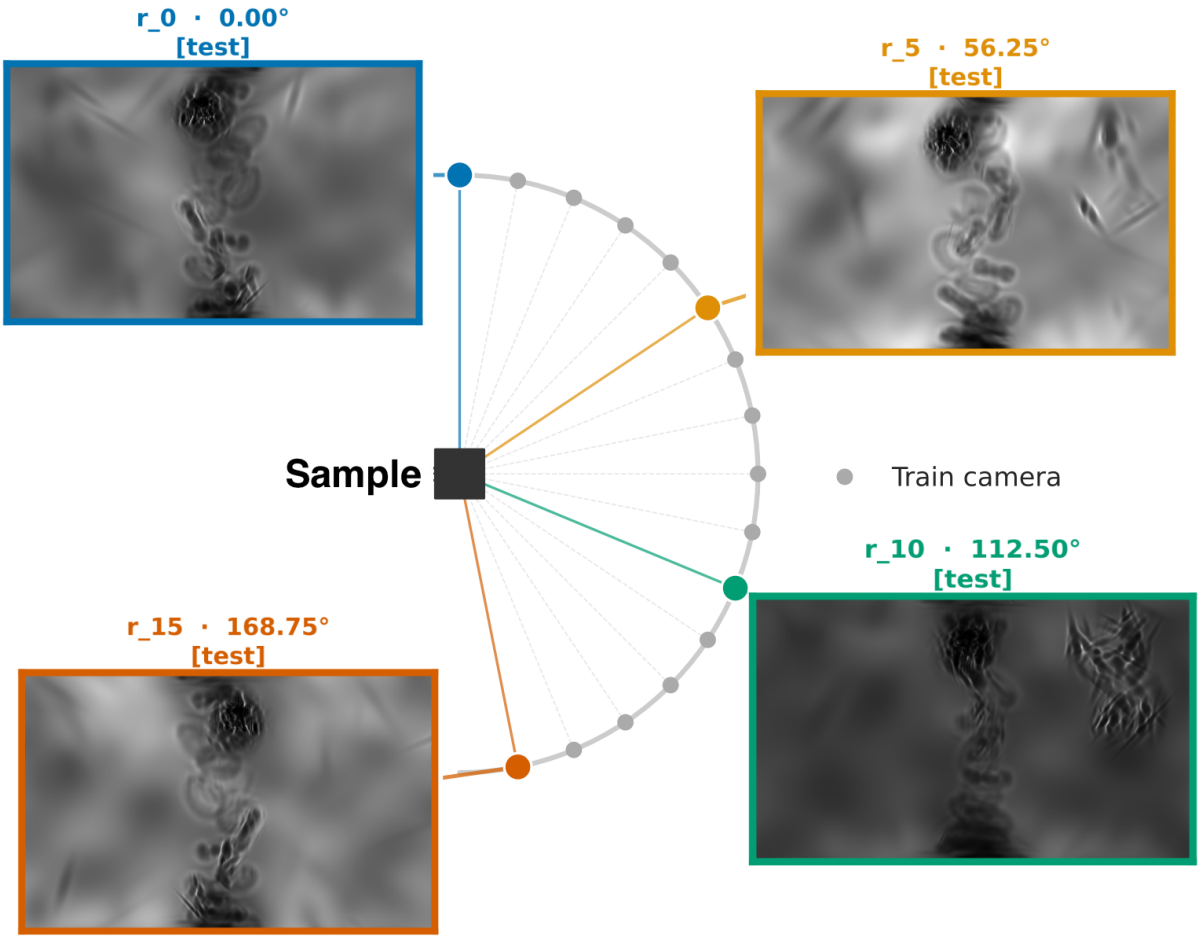


Figure 27: Camera geometry for the 180°, 16-projection setting ($N_0 = 8\text{k}$, $\text{plr} = 3.2 \times 10^{-4}$), with rendered test projections at the final iteration (30 000). Some overfitting is visible at this checkpoint; peak SSIM/FRCM performance is shown in Figure 28.

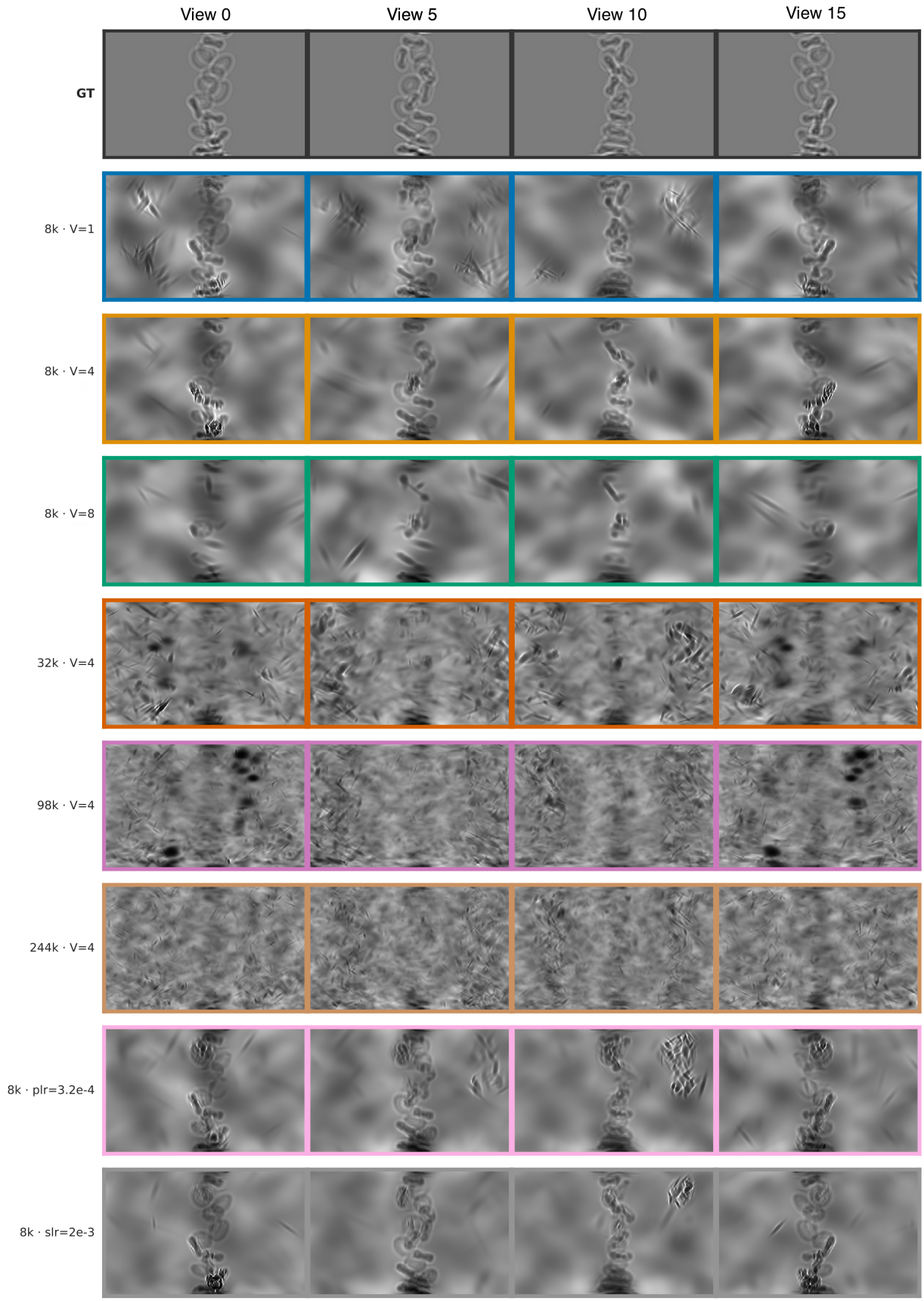


Figure 28: Reconstruction quality across eight representative runs for all four test views rendered at approximately iteration 10 000 (near the FRCM/SSIM peak for each run). Of the top-performing configurations, the $8\text{k}/V = 8$ run produces visually coherent holograms; however, the underlying geometry is not fully recovered, and certain RBCs are absent from some viewpoints.

The reconstruction illustrated in Figure 29 shows the predicted intensity and the learned B (absorption) and D (phase) projection maps at training iteration 30 000. At iteration 1 000, the rendered output is nearly featureless: the B and D maps carry little spatial structure and the predicted hologram matches only the broad background level. The D map already shows the earliest hint of the RBC position at its peak, yet as training progresses it begins to overfit individual views, and phase artefacts become visible in the final renders (Figure 28). By iteration 30 000, the fringe pattern is clearly reproduced and the D map has sharpened considerably, resolving cell boundaries with greater fidelity; however, some phase overfitting persists at this stage. The B map remains close to zero throughout, consistent with the near-transparent character of soft biological tissue at the X-ray energies used.

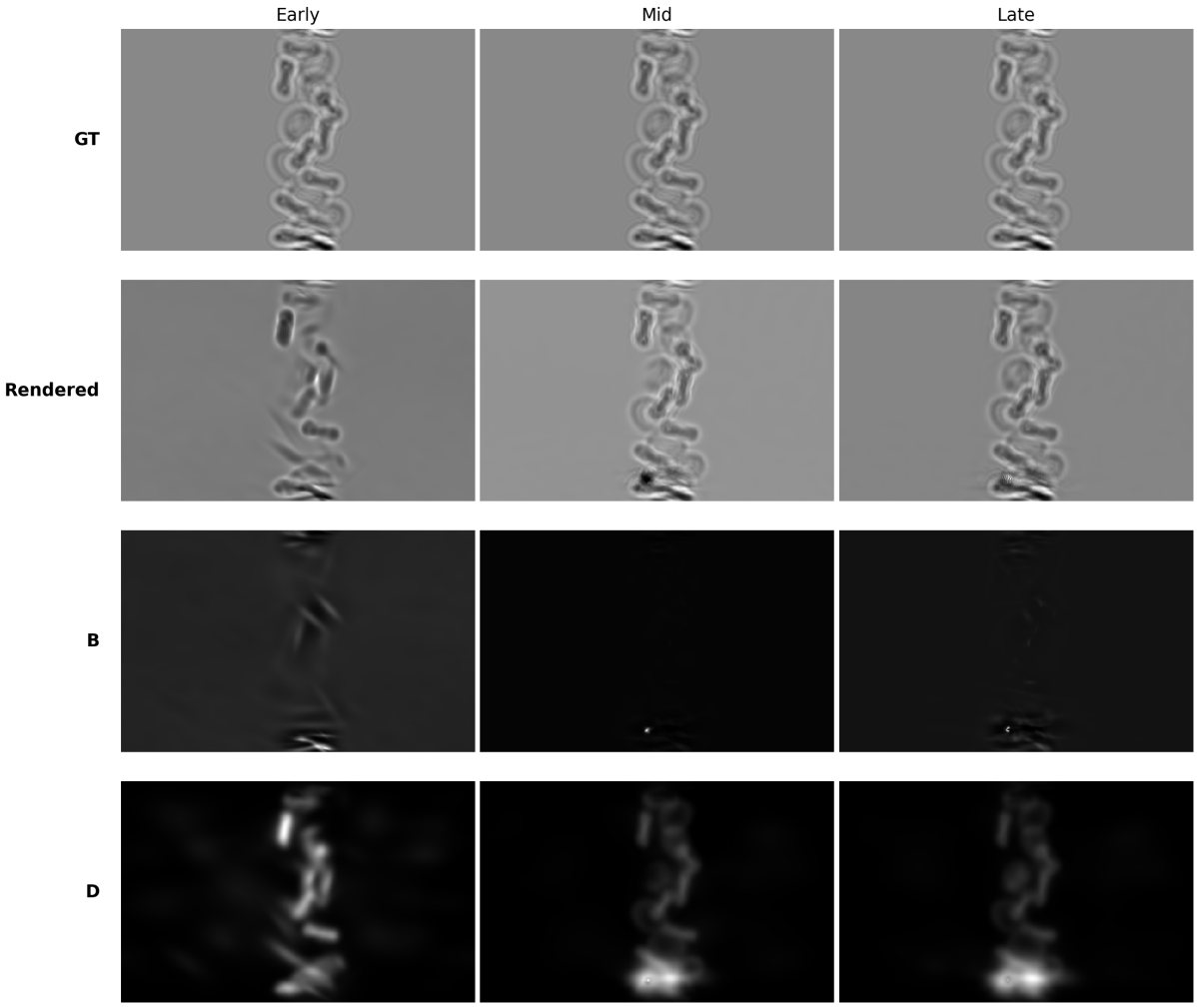


Figure 29: Progression of reconstruction quality for the $\mathrm{slr} = 2 \times 10^{-3}$ run (training views) at training steps 5 000, 15 000, and 30 000 (columns). Rows from top: ground truth hologram, rendered hologram, absorption projection map B, and phase projection map D. The D map evolves from a point cloud into a spatially structured representation of the phase-shifting cells; B remains low throughout, confirming the phase-dominant character of the scene.

Views per iteration. Figure 30 examines the effect of the multi-view batch size V on reconstruction quality and Gaussian population dynamics. The top panel plots FRCM over training

for $V \in \{1, 4, 8\}$ (all with $N_0 = 8\text{k}$), with the best overall run (slr) shown as a dashed reference. The bottom panel shows the corresponding Gaussian count evolution.

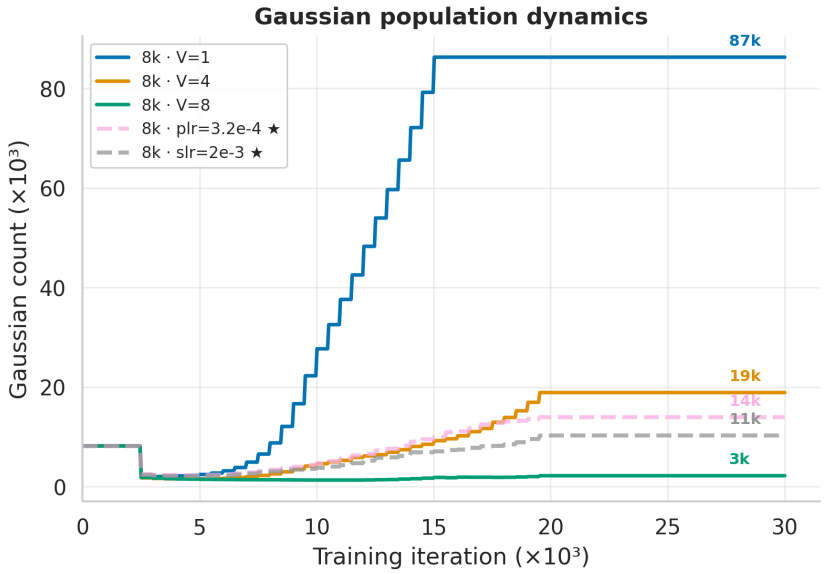


Figure 30: Overview of FRCM score and Gaussian population dynamics. **Top:** Test FRCM over training. Stars mark the peak FRCM for each V; the slr reference (dashed) is included for context. $V = 1$ and $V = 4$ both peak at iteration 7 000, while $V = 8$ peaks later at 15 000. All runs degrade slowly in FRCM after their peak as well as in SSIM (see Figure 25). **Bottom:** Gaussian population dynamics. The $V = 1$ run explodes to 86k Gaussians, while $V = 8$ prunes aggressively to only 2k.

Two patterns are evident. First, FRCM peaks mid-training (at iterations 7 000–15 000) rather than at the final checkpoint for all three configurations — an observation with important implications for early stopping (Section 4.2). Second, the $V = 1$ run grows to 86 000 Gaussians in the absence of cross-view pruning pressure, yet its peak FRCM (0.089 at iteration 7 000) is reached and then rapidly lost as over-densification buries the fringe signal under redundant primitives. Figure 25 further shows that SSIM is substantially lower for $V = 1$ (0.692 versus 0.870 for the best run) than for the other configurations. Without cross-view gradient averaging, the optimiser cannot exploit the geometric consistency information latent in the 12-view training set (4 test views), and many Gaussians settle at locally correct but globally inconsistent positions. The result is a dense, loosely constrained population that reproduces individual views with moderate fidelity but generalises poorly.

The $V = 4$ and $V = 8$ configurations impose stricter cross-view constraints. $V = 8$ prunes most aggressively, reducing the final Gaussian count to 2 200; however, this extreme pruning discards primitives that contribute to reconstruction fidelity, resulting in a final FRCM of 0.072 — comparable to the $V = 1$ baseline — despite achieving the highest SSIM of 0.803 among the V-ablation runs. Although the geometry appeared more consistent under this setting, certain RBCs were absent in some views. $V = 4$ retains a moderate final population of 18 900 Gaussians and achieves the highest baseline FRCM peak among the compared config-

urations, making it the natural candidate for further parameter tuning.

Initial Gaussian density. A natural expectation is that denser initialisation improves reconstruction quality by putting more Gaussians close to the true sample shape. However, neither SSIM nor FRCM support this expectation, as shown in Figure 31 and 32.

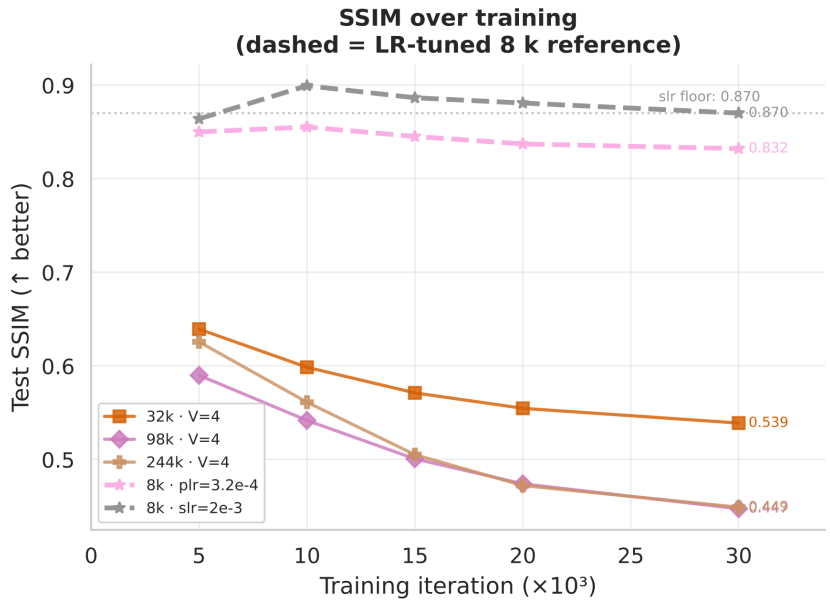


Figure 31: Density ablation ($V = 4$). Test SSIM over training for 32k, 98k, and 244k initial Gaussians, with the LR-tuned 8k runs (dashed) as reference.

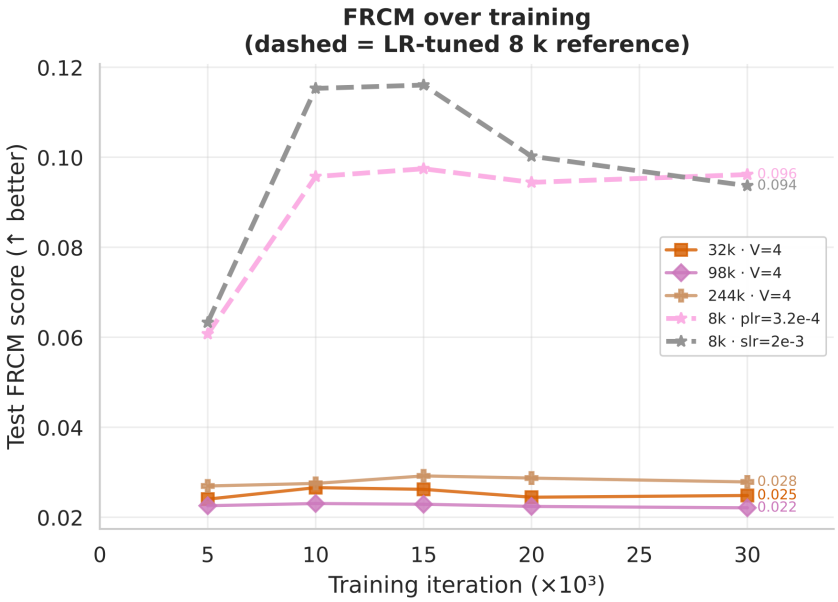


Figure 32: Density ablation ($V = 4$). Test FRCM over training for the same runs. All high-density runs converge to SSIM < 0.55 and FRCM < 0.03, far below the LR-tuned 8k references that achieve SSIM > 0.83 and FRCM > 0.09.

All three high-density runs — 32k, 98k, and 244k, each with $V = 4$ — converge to SSIM

values of 0.45–0.54 and FRCM values of 0.02–0.03, far below the 8k baseline ($V = 4$, SSIM = 0.756) and dramatically worse than the LR-tuned 8k runs (SSIM 0.83–0.87, FRCM 0.09–0.10). The degradation is progressive: the SSIM trajectories decline steadily from their initial values throughout training.

This counterintuitive result reflects a failure of the default learning-rate schedule at high Gaussian counts. At the default position learning rate, these Gaussians move too aggressively during densification, creating view-dependent absorption artefacts that improve PSNR (by matching the per-pixel mean intensity) but destroy structural fidelity. The result is a degenerate solution in which the model fits the background component of the hologram rather than the fringe pattern — precisely the failure mode detected by the SSIM/FRCM metric analysis in Section 4.1.

The key insight is that learning-rate tuning on a sparse point cloud initialisation ($N_0 = 8\text{k}$) outperforms a 30× increase in Gaussian count at the default learning rates on the metrics that matter for phase-contrast imaging. This implies that the bottleneck is not representational capacity but rather the optimisation dynamics — a finding with substantial practical importance, since lean initialisations are faster to train and require less GPU memory.

Early stopping and FRCM peaking. An unexpected finding from the model training is that FRCM peaks mid-training and then degrades, even as the total loss continues to decrease. Figures 33,34 examine this pattern across all eight runs.

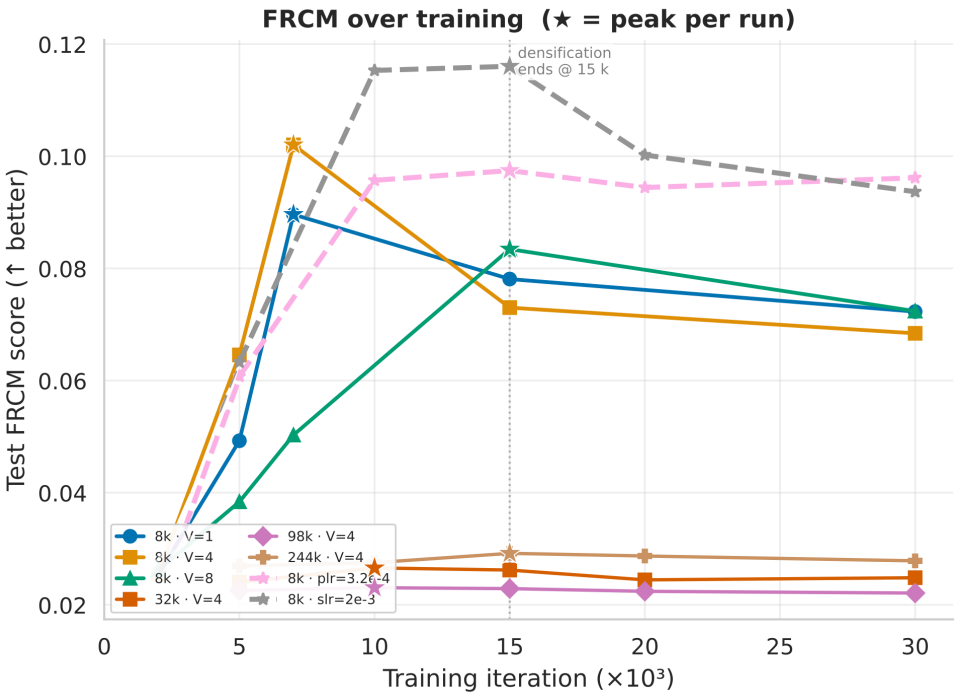


Figure 33: Test FRCM over training for all eight runs. Stars mark the peak FRCM for each run. The vertical dashed line marks the end of the densification window (iteration 15 000). Most runs peak before or near this boundary.

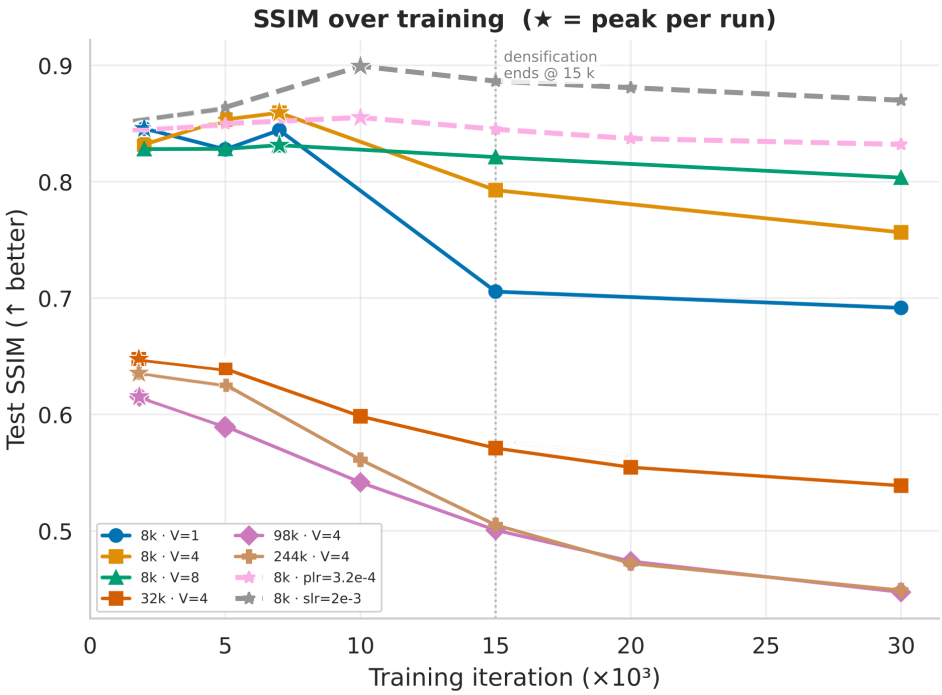


Figure 34: Test SSIM over training. SSIM either peaks early (high-density runs, $V = 1$) or grows steadily (LR-tuned runs) and peaks around 10k and decline afterwards.

For the $V = 1$ and $V = 4$ baselines, FRCM peaks near iteration 7 000; for $V = 8$, the peak occurs around 15 000; and for the LR-tuned runs, FRCM peaks around 10 000–15 000 before slowly declining. The pattern is consistent: continued optimisation after the densification window improves the pixel-level fit (reducing L1 loss) but smears high-frequency fringe detail. This seems to occur because late-stage training mostly adjusts material parameters (β, δ) on relatively settled geometry.

4.3 Discussion

Physical interpretability of learned parameters. Beyond image-quality metrics, an important question is whether the learned Gaussian parameters are physically meaningful — that is, whether the model has converged to values of β and δ that are consistent with the known material properties of the sample. For haemoglobin at the X-ray energies relevant to this experiment, the refractive-to-absorption ratio δ/β is expected to lie in the range of approximately $500–2000$ [45], reflecting the strongly phase-dominant regime of soft biological tissue at hard X-ray energies. Figure 35 presents the evolution of two key physical diagnostics across all runs: the mean per-Gaussian δ/β ratio.

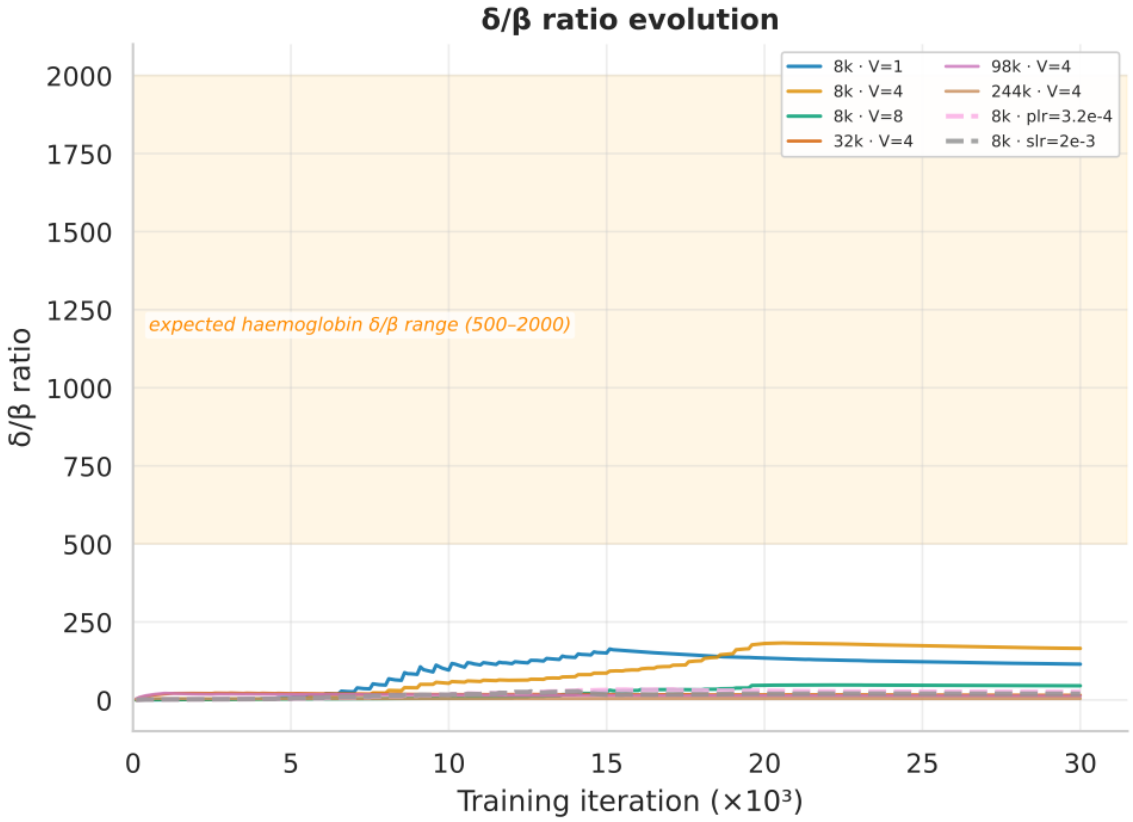


Figure 35: Material parameter evolution over training. Mean per-Gaussian δ/β ratio. The shaded band marks the physically expected range for haemoglobin (500–2000). All runs settle at much lower ratios ($\approx 5 - 165$).

All runs converge to much lower ratios ($\approx 5 - 165$) than expected, substantially below the haemoglobin reference value. This pattern admits a clear interpretation in terms of the ill-posedness of phase retrieval. Multi-view optimisation introduces constraints that restrict which parameter combinations can simultaneously explain multiple angular projections, but these constraints do not uniquely find out the right δ/β ratio — they just select one from the many feasible solutions.

The resulting local minimum happens to have smaller δ values relative to β, corresponding to a less phase-dominant material representation, even though the predicted holograms are reconstructed with higher structural fidelity. Nevertheless, this lower ratio could be the result of splats representing the background (zero δ), pulling the overall ratio down.

Crucially, the values in Figure 35 are consistent with the picture shown in Figure 29: B maps remain close to zero while D maps carry most of the reconstructed signal. This qualitative correctness — the model knows the sample is phase-dominant, even if the absolute scale of δ is not precisely matched — is a meaningful result in its own right.

This directly motivates the incorporation of a single-material prior (constraining the δ/β ratio to the known value of hemoglobin at given X-ray energy), which would collapse the search for parameters to a one-dimensional subspace, potentially improving the overall quality of reconstruction. However, such a prior would only be applicable to single-material

samples; for multi-material specimens the δ/β ratio can differ substantially.

The experimental results establish several key findings for phase-contrast Gaussian splatting:

1. Metric choice is decisive. PSNR misleads by rewarding background absorption fits. For any imaging modality where the signal of interest is encoded in spatial structure — as in holographic phase-contrast imaging — SSIM and frequency-domain metrics like FRCM should serve as the primary evaluation criteria. This finding has implications beyond this specific pipeline: it applies to any neural scene representation trained on holographic or interferometric data.

2. Learning-rate tuning outperforms brute-force scaling. The two best configurations ($\mathrm{slr} = 2 \times 10^{-3}$, SSIM = 0.870; $\mathrm{plr} = 3.2 \times 10^{-4}$, FRCM = 0.096) use only $N_0 = 8\mathrm{k}$ initial Gaussians — 30 times fewer than the 244k configuration that leads on PSNR. The bottleneck is not representational capacity but optimisation dynamics: the learning-rate schedule determines whether the model converges to a physically meaningful phase-contrast solution or a degenerate absorption-only fit.

3. Moderate Gaussian populations are optimal. Final Gaussian counts in the 8 000–15 000 range correlate with the best structural quality. Too few Gaussians (2 200 at $V = 8$) lack the spatial resolution to represent fringes; too many (86 300 at $V = 1$ or 36 600 at 244k) enable view-dependent overfitting.

4. Physical correctness. As established in Section 4.3, the learned δ/β ratios are substantially lower than the physically expected values for haemoglobin, confirming that structural fidelity and physical accuracy correspond to different optima within the phase-retrieval landscape. Incorporating a δ/β prior would constrain the search to the physically correct subspace and represents the most direct way of improving the accuracy of the reconstruction.

Sparse-view regime. The use of 12 training views over $180°$ represents a severely undersampled angular regime relative to standard computed tomography. The multi-view training strategy substantially mitigates this by jointly constraining the scene from multiple angles in each gradient step, but cannot recover information that is simply absent from the 16-view dataset. Increasing the number of available views would be expected to improve test generalisation considerably. However, when considered in the context of MHz-Tomoscopy, a set of 16 views is relatively generous. Moreover, recent extensions of 3D Gaussian Splatting specifically target the sparse-view setting and could be leveraged here. In particular, the availability of large-scale experimental datasets suggests the possibility of incorporating a dedicated neural component to enable 3DGS to better infer structure from a limited number of viewpoints.

Computational considerations. All runs completed in a relatively short time on a single GPU (NVIDIA A100, 40 GB), with a modest memory footprint. Even at the maximum setting of $V = 8$, memory usage remained around 10 GB throughout training. The learning-rate-tuned configuration with ($V = 4$) achieved an iteration time of approximately 30 ms, completing 30,000 iterations in about 15 minutes. Stopping at the FRCM/SSIM peak (iterations 7 000–10 000) would reduce this to approximately 5 minutes. This is well within the practical timescales of beamtime experiments, suggesting that near-online reconstruction is feasible with the proposed method.

Implications and outlook. The results demonstrate that the adapted 3DGS framework provides a viable, differentiable forward model for phase-contrast tomography as well as for MHz-Tomoscopy, capable of recovering a spatially structured scene representation from as few as 16 synthetic projections. The pipeline correctly models Fresnel propagation physics, reproduces qualitative fringe structure, and learns a representation in which the phase signal dominates over absorption. Preliminary observations on experimental XFEL data are consistent with these findings, suggesting the pipeline is viable beyond the synthetic setting.

The primary limitation seems to be the ill-posedness of the inverse problem, which manifests as physically inaccurate δ/β ratios and the ultra-sparse regime of the MHz-Tomoscopy. Besides the priors discussed in Paragraph 4.3 above, several more techniques could improve the reconstruction quality:

- Combined `plr` and `slr` tuning, potentially capturing both the structural (SSIM) gains of `slr` and the frequency-domain (FRCM) gains of `plr` simultaneously;
- SSIM/FRCM-based early stopping or checkpoint selection.

A further priority is extending the framework to multiple molecular dynamics timesteps for an initial assessment of 4D reconstruction capability — the ultimate objective of applying Gaussian Splatting to MHz Tomoscopy.

5 Conclusion

This thesis addressed the challenge of reconstructing volumetric structure from the severely limited number of projections available in MHz-Tomoscopy — a technique capable of imaging dynamic processes at sub-microsecond timescales but constrained, in practice, to only a handful of simultaneous angular views [2, 3]. A parallel challenge is the volume of data produced at MHz rates and the imperfections of the instrumentation, which together demand a scene representation that is compact enough for efficient storage and fast enough to support near-online post-processing. The central question was whether 3D Gaussian Splatting, adapted to the physics of X-ray phase-contrast imaging, could serve as an effective, computa-

tionally tractable, and compact representation that addresses both the reconstruction and the data-management challenges simultaneously.

To enable controlled evaluation in the absence of large annotated experimental datasets, a forward simulation pipeline was developed that generates physically realistic phase-contrast holograms from molecular dynamics snapshots of RBCs flowing through a model microfluidic channel. The pipeline models the complete X-ray imaging chain — from material assignment and voxelisation, through raycasting, to Fresnel free-space propagation — and outputs synthetic projections with the characteristic bright-dark edge fringes of near-field diffraction. The standard 3DGS framework was then significantly modified to accommodate this imaging physics: spherical harmonic colour coefficients were replaced by per-Gaussian absorption (β) and refractive-index decrement (δ) parameters, perspective projection was replaced by parallel-beam geometry, and a fully differentiable phase-contrast forward model was embedded in the rendering pipeline. Training was stabilised through a composite loss and simultaneous multi-view gradient accumulation.

The experiments demonstrated that the adapted framework can recover spatially structured phase-contrast reconstructions from 16 synthetic projections (12 training and 4 held-out test views) in approximately 15 minutes on a single GPU. A central finding concerns metric selection: PSNR proved to be a misleading indicator for holographic reconstruction, as it rewards matching the background intensity without penalising incorrect fringe structure. SSIM and the Fourier Ring Correlation Metric (FRCM) were found to be more appropriate primary metrics, and this observation likely generalises to any scene representation, whether neural or not, trained on holographic or interferometric data. Contrary to intuition, increasing the initial Gaussian count from 8k to 244k degraded structural quality, revealing that the bottleneck is optimisation dynamics rather than representational capacity. Targeted learning-rate adjustments on a lean 8k initialisation produced the best results (SSIM $= 0.870$, FRCM $= 0.096$), and multi-view training acted as an implicit regulariser that enforced cross-view geometric consistency. The learned material fields were qualitatively correct — phase-projection maps dominated, while absorption remained near zero — but the per-Gaussian δ/β ratios were substantially lower than the expected physical values, a consequence of the inherent ill-posedness of the embedded phase retrieval problem.

Several improvements follow directly from the experimental analysis and are discussed in Section 4.3: incorporating a single-material δ/β or phase-retrieved prior, further adjusting and combining learning-rate modifications, adopting FRCM-based early stopping, and extending the framework to multiple time-steps for an initial evaluation of 4D reconstruction. Beyond these, two broader directions appear particularly promising. Because the Gaussian primitives are explicit and spatially localised, physical constraints can be embedded directly in the training objective — for instance, penalising changes in primitive volume to encourage incompressibility, or coupling neighbouring primitives to preserve local structural coherence. Such constraints could provide additional regularisation in the sparse-view regime where the

projection data alone cannot fully constrain the solution. The particle-based nature of the representation makes it natural to couple the reconstructed scene to a differentiable physical simulator, so that the imaging model and the dynamical model can be optimised jointly. This would open a path towards inferring both 3D structure and material dynamics from the experiments — the long-term objective of MHz-Tomoscopy.

Overall, 3D Gaussian Splatting proves to be a well-suited representation for MHz-Tomoscopy experiments: the sparse, particle-based scene description provides the compact storage and efficient forward-pass evaluation that the data-volume demands of these experiments require, the differentiable physics-based forward model enables quantitative recovery of the complex refractive index without ground-truth volume supervision, and the optimisation converges within timescales compatible with beamtime constraints. The framework is expected to find broad application across phase-contrast imaging modalities wherever sparse angular sampling and near-online reconstruction are requirements. The most immediate priority is validation against real XFEL projections, where the number of available views is fixed at four and the imaging conditions introduce noise and geometric imperfections not present in the synthetic setting.

6 Resumé

Táto diplomová práca sa zamerala na rekonštrukciu trojrozmernej štruktúry z obmedzeného počtu röntgenových projekcií, aký je k dispozícii v MHz-Tomoskopii. MHz-Tomoskopia je zobrazovacia technika, ktorá namiesto rotácie vzorky zachytáva viacero projekcií naraz pomocou jedného pulzu röntgenového lasera, čím umožňuje sledovať dynamické procesy — šírenie šokových vĺn, kavitáciu alebo vznik trhlín — pri opakovacích frekvenciách v rádoch MHz. Cenou za toto vysoké časové rozlíšenie je len niekoľko simultánnych uhlov pohľadu, čo robí inverznú úlohu rekonštrukcie vážne poddeterminovanou. Praktickú výzvu ešte prehlbuje objem dát produkovaných pri MHz frekvenciách a nedokonalosti zobrazovacieho inštrumentu: tieto faktory spoločne vyžadujú reprezentáciu scény, ktorá je dostatočne kompaktná na efektívne ukladanie a dostatočne rýchla na spracovanie v takmer reálnom čase. Problém sa ďalej komplikuje tým, že mäkké biologické tkanivá röntgenové žiarenie predovšetkým fázovo posúvajú a takmer neabsorbujú, takže klasické absorpčné rekonštrukčné metódy — FBP aj SIRT — pri takom malom počte projekcií produkujú závažné artefakty a nedokážu exploitovať fázový kontrast.

V práci sme skúmali, či metóda 3D Gaussian Splatting (3DGS), pôvodne vyvinutá pre počítačovú grafiku, dokáže po úprave na fyziku röntgenového zobrazovania efektívne rekonštruovať fázovo-kontrastné obrazy z obmedzeného počtu projekcií a zároveň poskytnúť kompaktnú reprezentáciu vhodnú pre tomoskopické experimenty. V teoretickej časti sme zhrnuli princípy MHz-Tomoskopie, fázovo-kontrastného zobrazovania a tomografickej rekonštrukcie, a porovnali dostupné prístupy k reprezentácii 3D scény — voxelové mriežky, point clouds, neurónové radiačné polia (NeRF) a Gaussian Splatting. Z hľadiska pamäte a škálovateľnosti má každý z týchto prístupov odlišné vlastnosti: voxelové mriežky rastú kubicky s rozlíšením ($O(N^3)$), NeRF vyžaduje vyhodnocovanie neurónovej siete pre každý vzorkovací bod, zatiaľ čo 3DGS uchováva iba kompaktnú množinu parametrov primitívov a rasterizuje ich v jedinom diferencovateľnom prechode bez volaní neurónovej siete. Pre scény s riedkym počtom objektov, akými sú napr. červené krvinky v mikrofluidickom kanáli, to znamená výrazne nižšiu pamäťovú stopu a kratší čas výpočtu oproti obom alternatívam. Ďalej sme ukázali, že NeRF-založený model ONIX, hoci sľubný pre rekonštrukciu z malého počtu pohľadov, pracuje výhradne v absorpčnom režime a nedokáže modelovať fázové efekty dominantné v MHz-Tomoskopii.

Keďže reálne experimentálne dáta zo zariadenia s viacerými simultánnymi projekciami zatiaľ nie sú dostupné v dostatočnej kvalite, vyvinuli sme simulačnú pipeline, ktorá slúži ako generátor referenčných dát. Červené krvinky prúdiace mikrofluidickým kanálom boli simulované softvérom ESPResSo s využitím coarse-grain geometrických modelov buniek; každej krvinke boli priradené röntgeno-optické vlastnosti — absorpčný index β a dekrement indexu lomu δ — vypočítané z Henkeho tabuliek atómových rozptylových faktorov pre hemoglobín. Následne sa vypočítali projekcie na detektore pomocou raycastingu, zostrojila sa komplexná transmisná funkcia a Fresnelovým propagátorom sa propagovalo vlnové pole na detektor, čím

vznikli hologramy s charakteristickými okrajovými prúžkami fázového kontrastu. Použitie syntetického datasetu nám umožnilo presne kontrolovať zobrazovacie podmienky a porovnávať výsledky rekonštrukcií voči ground truth.

Samotný 3DGS framework sme podstatne upravili: namiesto farby nesie každý Gaussián materiálové parametre absorpčného indexu β a dekrementu indexu lomu δ, perspektívna projekcia bola nahradená rovnobežnou geometriou zodpovedajúcou MHz-Tomoskopii a do pipeline sme zabudovali plne diferencovateľný fyzikálny model Fresnelovho šírenia vlnového poľa vrátane funkcie rozptylu bodu detektora. Vypustením sférických harmonických koeficientov farebného renderovania a ich nahradením dvojicou skalárov (β, δ) sa počet parametrov samotných primitívov výrazne znížil, čo priamo prispieva ku kompaktnosti reprezentácie. Tréning sme stabilizovali kompozitnou funkciou straty kombinujúcou L1 chybu a štrukturálnu podobnosť SSIM, pričom sme trénovali súčasne na viacerých pohľadoch naraz, čo slúžilo ako implicitný regularizátor vynucujúci konzistenciu naprieč rôznymi pohľadmi.

Experimenty sme realizovali na jednej časovej stope simulácie červených krviniek s 16 projekciami, z ktorých 12 slúžilo na trénovanie a 4 na testovanie. Rekonštrukcia zo 16 projekcií trvala približne 15 minút na jedinom GPU, čo je v rámci praktických časových možností beamtimových experimentov na EuXFEL. Kľúčovým zistením bolo, že bežne používaná metrika PSNR je pre hodnotenie holografickej rekonštrukcie nevhodná — odmeňuje zhodu priemernej intenzity pozadia, nie správnosť štruktúry Fresnelovych prúžkov. Potvrdili sme to experimentálne: konfigurácia s najvyšším počtom Gaussiánov (244k) dosiahla najvyšší PSNR, ale zároveň najnižší SSIM (0,449) a takmer najnižší FRCM (0,028), zatiaľ čo konfigurácie s doladenými pomermi učenia na riedkej inicializácii (8k Gaussiánov) dominovali v oboch štrukturálnych metrikách (SSIM = 0,870, FRCM = 0,096). Ukázalo sa tiež, že správne nastavenie pomeru učenia prekonáva tridsaťnásobné zvýšenie počtu primitívov, čo poukazuje na to, že hlavnou prekážkou kvalitnej rekonštrukcie je optimalizačná dynamika, nie kapacita reprezentácie.

3D Gaussian Splatting sa ukazuje ako vhodný prístup pre rekonštrukciu fázovo-kontrastných obrazov z obmedzeného počtu projekcií v tomoskopických experimentoch. Riedka reprezentácia scény pomocou Gaussových primitívov priamo rieši požiadavky na kompaktné ukladanie a rýchle spracovanie dát charakteristické pre MHz-Tomoskopiu: namiesto kubicky rastúcej voxelovej mriežky postačuje uložiť iba parametre aktívnych primitívov, čo pri riedkych scénach tohto typu predstavuje výraznú úsporu pamäte a umožňuje škálovanie na väčšie objemy bez zodpovedajúceho rastu pamäťových nárokov. Naučené materiálové polia boli kvalitatívne správne — fázové mapy dominovali a absorpcia zostala blízko nule — čo odráža fyzikálnu realitu fázovo dominantných biologických tkanív. Absolútne hodnoty pomeru δ/β však nezodpovedali fyzikálne očakávaným hodnotám pre hemoglobín, čo je dôsledkom inherentnej nejednoznačnosti fázovej rekonštrukcie; tento problém by mohol byť čiastočne vyriešený zavedením fyzikálneho prioru fixujúceho pomer δ/β. Predbežné pozorovania na experimentálnych dátach z EuXFEL sú konzistentné s týmito zisteniami, čo naznačuje, že pipeline je použiteľná aj mimo syntetického prostredia. Ďalšími smermi budúcej práce sú rozšírenie na

viacero časových krokov pre hodnotenie 4D rekonštrukcie a validácia pipeline na reálnych experimentálnych dátach so štyrmi simultánnymi projekciami, pre ktoré je MHz-Tomoskopia primárne určená. Očakávame, že metóda nájde široké uplatnenie vo fázovo-kontrastnom zobrazovaní.

References

[1] "MHz-Tomoscopy," Apr. 2025. [Online]. Available: https://tomoscopy.science.upjs.sk/ (visited on 2025-06-16).

[2] P. Villanueva-Perez, V. Bellucci, Y. Zhang, S. Birnsteinova, R. Graceffa, L. Adriano, E. M. Asimakopoulou, I. Petrov, Z. Yao, M. Romagnoni, A. Mazzolari, R. Letrun, C. Kim, J. C. P. Koliyadu, C. Deiter, R. Bean, G. Giovanetti, L. Gelisio, T. Ritschel, A. Mancuso, H. N. Chapman, A. Meents, T. Sato, and P. Vagovic, "Megahertz X-ray Multi-projection imaging," May 2023. [Online]. Available: http://arxiv.org/abs/2305.11920 (visited on 2024-03-20).

[3] D. Moško, "Object reconstruction from X-Ray multiprojection tomoscopy experiments," Bachelor's Thesis, University of Pavol Jozef Safarik, Kosice, May 2024, (visited on 2025-11-05).

[4] pvilla, "Pvilla/ONIX," Jan. 2026. [Online]. Available: https://github.com/pvilla/ONIX (visited on 2026-03-04).

[5] Y. Zhang, Z. Yao, R. Klöfkorn, T. Ritschel, and P. Villanueva-Perez, "4D-ONIX: A deep learning approach for reconstructing 3D movies from sparse X-ray projections," Jan. 2024. [Online]. Available: http://arxiv.org/abs/2401.09508 (visited on 2024-01-29).

[6] F. Natterer, *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, Jan. 2001. [Online]. Available: http://epubs.siam.org/doi/book/10.1137/1.9780898719284 (visited on 2026-04-24).

[7] D. Paganin, *Coherent X-Ray Optics*, ser. Oxford Series on Synchrotron Radiation. Oxford; New York: Oxford University Press, 2006, no. 6.

[8] A. Momose, T. Takeda, Y. Itai, and K. Hirano, "Phase–contrast X–ray computed tomography for observing biological soft tissues," *Nature Medicine*, vol. 2, no. 4, pp. 473–475, Apr. 1996. [Online]. Available: https://www.nature.com/articles/nm0496-473 (visited on 2026-04-29).

[9] A. Burvall, U. Lundström, P. A. C. Takman, D. H. Larsson, and H. M. Hertz, "Phase retrieval in X-ray phase-contrast imaging suitable for tomography," *Optics Express*, vol. 19, no. 11, pp. 10 359–10 376, May 2011. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?uri=oe-19-11-10359 (visited on 2026-03-06).

[10] T. Weitkamp, "Image formation behind an object composed of absorbing structures and non-absorbing phase shifters, as a result of X-ray free space propagation." Sep. 2006. [Online]. Available: https://commons.wikimedia.org/wiki/File:X-ray_free-space_propagation_-_from_near_field_to_far_field.png#/media/File:

X-ray_free-space_propagation_-_from_near_field_to_far_field.png (visited on 2026-02-06).

[11] "Fresnel and Fraunhofer Approximations," Jan. 2022. [Online]. Available: https://phys.libretexts.org/Bookshelves/Optics/BSc_Optics_(Konijnenberg_Adam_and_Urbach)/06%3A_Scalar_diffraction_optics/6.07%3A_Fresnel_and_Fraunhofer_Approximations (visited on 2026-02-17).

[12] C. Zuo, J. Li, J. Sun, Y. Fan, J. Zhang, L. Lu, R. Zhang, B. Wang, L. Huang, and Q. Chen, "Transport of intensity equation: A tutorial," *Optics and Lasers in Engineering*, vol. 135, p. 106187, Dec. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143816619320858 (visited on 2026-04-29).

[13] P. Villanueva-Perez, F. Arcadu, P. Cloetens, and M. Stampanoni, "Contrast-transfer-function phase retrieval based on compressed sensing," *Optics Letters*, vol. 42, no. 6, pp. 1133–1136, Mar. 2017. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?uri=ol-42-6-1133 (visited on 2026-02-16).

[14] J. C. P. Koliyadu, D. Moško, E. M. Asimakopoulou, V. Bellucci, Š. Birnšteinová, R. Bean, R. Letrun, C. Kim, H. Kirkwood, G. Giovanetti, N. Jardon, J. Szuba, T. Guest, A. Koch, J. Grünert, P. Szeles, P. Villanueva-Perez, F. Reuter, C.-D. Ohl, M. A. Noack, F. Garcia-Moreno, Z. Kuglerová-Valdová, L. Juha, M. Nikl, W. Yashiro, H. Soyama, D. Eakins, A. M. Korsunsky, J. Uličný, A. Meents, H. N. Chapman, A. P. Mancuso, T. Sato, and P. Vagovič, "Development of MHz X-ray phase contrast imaging at the European XFEL," *Journal of Synchrotron Radiation*, vol. 32, no. 1, pp. 17–28, Jan. 2025. [Online]. Available: https://journals.iucr.org/s/issues/2025/01/00/yi5159/ (visited on 2026-04-26).

[15] M. E. V. Álvarez, "Beyond traditional tomography: X-ray multi-projection imaging for additive manufacturing," Bachelor's Thesis, Lund University, 2023.

[16] B. Nett, "Filtered BackProjection (FBP) Illustrated Guide For Radiologic Technologists," Nov. 2020. [Online]. Available: https://howradiologyworks.com/filtered-backprojection-fbp-illustrated-guide-for-radiologic-technologists/ (visited on 2026-03-06).

[17] R. Gordon, R. Bender, and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, Dec. 1970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022519370901098 (visited on 2026-04-08).

[18] J. Trampert and J.-J. Leveque, "Simultaneous iterative reconstruction technique: Physical interpretation based on the generalized least squares solution," *Journal of*

Geophysical Research: Solid Earth, vol. 95, no. B8, pp. 12 553–12 559, 1990. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1029/JB095iB08p12553 (visited on 2026-04-08).

[19] J. T. Kajiya and B. P. Von Herzen, "Ray tracing volume densities," *ACM SIGGRAPH Computer Graphics*, vol. 18, no. 3, pp. 165–174, Jan. 1984. [Online]. Available: https://dl.acm.org/doi/10.1145/964965.808594 (visited on 2024-03-15).

[20] J. S. Jørgensen, "X-ray Computed Tomography: Forward problem and FBP reconstruction," Technical University of Denmark, Jan. 2023.

[21] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988.

[22] N. Max, "Optical models for direct volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 2, pp. 99–108, Jun. 1995. [Online]. Available: https://ieeexplore.ieee.org/document/468400 (visited on 2026-04-26).

[23] S. Rotenberg, "CSE 168: Rendering Algorithms," University of California, San Diego, 2017. [Online]. Available: https://cseweb.ucsd.edu/classes/sp17/cse168-a/CSE168_06_Antialiasing.pdf

[24] L. Anna, "Anti-Aliasing in Gaming: The Battle for Perfect Graphics," Aug. 2022. [Online]. Available: https://vokigames.com/anti-aliasing-in-gaming-the-battle-for-perfect-graphics/ (visited on 2025-06-11).

[25] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[26] Z. Wang, "3D Representation Methods: A Survey," Oct. 2024. [Online]. Available: http://arxiv.org/abs/2410.06475 (visited on 2026-04-08).

[27] "Point cloud," *Wikipedia*, Mar. 2026. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Point_cloud&oldid=1341468063 (visited on 2026-03-05).

[28] "Nano Banana 2 - Gemini AI image generator & photo editor." [Online]. Available: https://gemini.google/us/overview/image-generation/?hl=en-US (visited on 2026-04-24).

[29] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, Aug. 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3592433 (visited on 2025-01-01).

[30] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," Aug. 2020. [Online]. Available: http://arxiv.org/abs/2003.08934 (visited on 2024-01-19).

[31] G. Chen and W. Wang, "A Survey on 3D Gaussian Splatting," 2024. [Online]. Available: https://arxiv.org/abs/2401.03890 (visited on 2026-02-13).

[32] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations," Nov. 2017. [Online]. Available: http://arxiv.org/abs/1711.10561 (visited on 2025-06-09).

[33] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next," *Journal of Scientific Computing*, vol. 92, no. 3, p. 88, Jul. 2022. [Online]. Available: https://doi.org/10.1007/s10915-022-01939-z (visited on 2025-06-09).

[34] R. Munafò, "Physics-informed neural networks for solving Navier-Stokes equations." Jun. 2021. [Online]. Available: https://upload.wikimedia.org/wikipedia/commons/9/90/Physics-informed_nerural_networks.png (visited on 2026-02-06).

[35] Kate Yurkova, "A Comprehensive Overview of Gaussian Splatting," Dec. 2023. [Online]. Available: https://quantinsightsnetwork.com/a-comprehensive-overview-of-gaussian-splatting-by-kate-yurkova-dec-2023/ (visited on 2024-06-17).

[36] P. S. Heckbert, "Fundamentals of Texture Mapping and Image Warping," Master's thesis, University of California, Berkeley, CA 94720, Jun. 1989. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/1989/5504.html

[37] S. Choi, B. Chao, J. Yang, M. Gopakumar, and G. Wetzstein, "Gaussian Wave Splatting for Computer-Generated Holography," May 2025. [Online]. Available: http://arxiv.org/abs/2505.06582 (visited on 2026-04-15).

[38] Zhou Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. [Online]. Available: https://ieeexplore.ieee.org/document/1284395/ (visited on 2026-03-23).

[39] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelNeRF: Neural Radiance Fields from One or Few Images," May 2021. [Online]. Available: http://arxiv.org/abs/2012.02190 (visited on 2024-03-12).

[40] Y. Zhang, Z. Yao, T. Ritschel, and P. Villanueva-Perez, "ONIX: An X-ray deep-learning tool for 3D reconstructions from sparse views," *Applied Research*, Apr. 2023. [Online].

Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/appl.202300016 (visited on 2023-09-15).

[41] "ESPResSo Extensible Simulation Package for Research on Soft Matter," Jan. 2012. [Online]. Available: https://espressomd.org/wordpress/ (visited on 2026-02-05).

[42] I. Jančigová, K. Kovalčíková, R. Weeber, and I. Cimrák, "PyOIF: Computational tool for modelling of multi-cell flows in complex geometries," *PLOS Computational Biology*, vol. 16, no. 10, p. e1008249, Oct. 2020. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008249 (visited on 2026-02-05).

[43] "Compute with PyOIF – Cell in Fluid." [Online]. Available: https://cellinfluid.fri.uniza.sk/resources-espresso/ (visited on 2026-02-05).

[44] "ParaView - Open-source, multi-platform data analysis and visualization application." [Online]. Available: https://www.paraview.org/ (visited on 2026-04-18).

[45] "X-ray DB: X-ray Reference Data in SQLite — X-ray DB: X-ray Reference Data in SQLite." [Online]. Available: https://xraypy.github.io/XrayDB/ (visited on 2026-02-11).

[46] S. Koho, G. Tortarolo, M. Castello, T. Deguchi, A. Diaspro, and G. Vicidomini, "Fourier ring correlation simplifies image restoration in fluorescence microscopy," *Nature Communications*, vol. 10, no. 1, p. 3103, Jul. 2019. [Online]. Available: https://www.nature.com/articles/s41467-019-11024-z (visited on 2026-03-25).

[47] Y. Zhang, M. A. Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, and P. Villanueva-Perez, "PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets," *Optics Express*, vol. 29, no. 13, pp. 19 593–19 604, Jun. 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?uri=oe-29-13-19593 (visited on 2026-03-23).

[48] "TV-L1 Image Denoising Algorithm," Mar. 2026. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/57604-tv-l1-image-denoising-algorithm (visited on 2026-03-25).