

Development of 3D Molecular Tagging Velocimetry with Fourier Integral Microscopy and Neural Radiance Fields

Brian Thurow^{1,*}, Sandra Halder¹, Bibek Sapkota¹, Mark Yamakaitis², Charles Fort², Philippe M Bardet²

1: Department of Aerospace Engineering, Auburn University, USA

2: Department of Mechanical and Aerospace Engineering, The George Washington University, USA

*Corresponding author: thurows@auburn.edu

Keywords: Neural Radiance Fields, Molecular Tagging Velocimetry, Tomographic Reconstruction.

ABSTRACT

Tomographic reconstruction of three-dimensional flow fields from limited angular views is a fundamental challenge in experimental fluid diagnostics. Conventional grid-based iterative methods amplify noise and produce reconstruction artifacts under the narrow angular sampling conditions imposed by compact imaging systems, and remain computationally expensive. Neural Radiance Fields provide a continuous, gridless reconstruction framework; however, conventional NeRF implementations based on Fourier encoding typically require large network architectures, resulting in a computationally expensive reconstruction process. In this work, the FluidNeRF framework was adopted and modified to support limited-angle orthographic ray formulation for MTV applications. A method was developed for synthetic data generation for the purpose of performance characterization. The results establish the modified FluidNeRF framework as a computationally efficient and viable alternative to classical tomographic approaches for time-resolved MTV reconstruction.

1. Introduction

Molecular Tagging Velocimetry (MTV) is a non-particle-based flow diagnostic in which molecular tracers are optically tagged and imaged at specified time intervals to capture the evolution of the flow field (Koochesfahani et al., 2004). MTV offers several advantages over particle-based methods, particularly for near-wall measurements. MTV is unaffected by particle dynamics or surface reflections, and its molecular tracers follow the flow with negligible inertia (Koochesfahani et al., 2004; Li et al., 2021; Huck et al., 2025). This makes it particularly well suited for near-wall measurements in small volumes ($\sim 1\text{--}10\text{ mm}^3$) and regions of high shear where particle-based techniques struggle. Although early MTV implementations were initially limited to planar, two-dimensional measurements, recent efforts have been made to capture full 3D-3C velocity information through multi-camera stereoscopic (Li et al., 2021). However, multi-camera configurations present significant practical challenges in near-wall environments. Confined spaces restrict optical access and

limit sensor placement around small measurement volumes. Furthermore, these systems require precise alignment and calibration, which is susceptible to disruption from facility vibrations and other external perturbations (Huck et al., 2025). Single-camera approaches therefore offer a compelling advantage for near-wall 3D velocimetry. The development of plenoptic (or light-field) camera based imaging presents an elegant solution whereby a microlens array can be used to generate multiple perspective images from a single image sensor and primary objective lens.

In a recent study, (Huck et al., 2025) demonstrated volumetric MTV using Fourier Integral Microscopy (FIMic), or Plenoptic 3.0, in which a microlens array is placed at the Fourier plane of an infinity-corrected microscope objective. In their work, Richardson-Lucy Deconvolution (RLD) was applied to the plenoptic camera generated focal stack to reconstruct the tagged MTV lines within the volume. The reconstructed intensity field was then interrogated using a peak detection scheme to identify the displacement of the tagged lines at successive time intervals, from which the two wall-parallel velocity components (u and v) were obtained. The wall-normal velocity component (w) was recovered via the continuity equation. The accuracy of such volumetric measurements depends on the accuracy of the tomographic reconstruction used to recover the 3D scalar field (Stull et al., 2014). Conventional approaches, such as MART (Jain et al., 2011), SART (Zhu et al., 2021), and deconvolution-based methods like Richardson-Lucy Deconvolution (RLD) (Ströhl and Kaminski, 2015), operate on discretized voxel grids. These methods are sensitive to noise amplification and discretization artifacts, and their performance degrades under the limited-angle, limited-view conditions inherent to plenoptic imaging systems (Kelly and Thurow, 2025; Barron et al., 2021).

To address these challenges, Neural Radiance Fields (NeRF) have recently been adapted for volumetric flow diagnostics as an alternative to voxel-based reconstruction (Kelly and Thurow, 2023). Rather than estimating intensity on a fixed grid, a neural network learns a continuous mapping from spatial coordinates to scalar intensity values by minimizing the discrepancy between rendered and measured projections (Mildenhall et al., 2021). This continuous representation avoids discretization artifacts and naturally regularizes the reconstruction through the network's inherent bias toward smooth solutions. Halder et al. (2026) extended the FluidNeRF framework to MTV applications by modeling the tagged molecular beamlets as a continuous spatial function instead of reconstructing individual frames on discrete voxel grids. This work demonstrated the adaptability of the original FluidNeRF framework to the experimental MTV data of Huck et al. (2025). However, the method remained computationally expensive and lacked ground truth data for a quantitative evaluation of reconstruction quality.

The present work addresses these limitations through several key developments. First, the scalar intensity field is modeled as a continuous function of both space and time (x, y, z, t) , which enables a single neural network to reconstruct the measurement volume across multiple time steps simultaneously. Second, multiresolution hash encoding (Müller et al., 2022) is adopted in place of frequency encoding, which allows the use of a significantly smaller network architecture and reduces

the time-resolved reconstruction time from several hours to approximately five minutes. Additionally, synthetic data is generated using the calibration information obtained from the FIMIC system (Huck et al., 2025). This provides the known ground truth necessary for a quantitative evaluation of reconstruction fidelity that was absent in prior work.

2. Background

Neural Radiance Fields (NeRF), first introduced by the computer vision community (Mildenhall et al., 2021), is a machine learning framework for novel view synthesis from a set of 2D projections. In this framework, a static scene is represented as a continuous function that maps a 5D input, three-dimensional spatial coordinate (x, y, z) and a two-dimensional viewing direction (θ, ϕ) , to predict volume density and a view-dependent radiance associated with that point. This gridless representation offers several advantages over conventional voxel-based methods: it avoids discretization artifacts (Kelly and Thurow, 2024), naturally regularizes the reconstruction through the network’s inductive bias toward smooth solutions (Mildenhall et al., 2021), and is flexible enough to incorporate advanced ray models (Barron et al., 2023; Engelmann et al., 2024), alternative rendering strategies (Warburg et al., 2023; Müller et al., 2022), and physics-based constraints (Qiu et al., 2024; Wu et al., 2025).

Building on this foundation, NeRF has recently been adopted for fluid mechanics applications by Kelly and Thurow (Kelly and Thurow, 2024) through the FluidNeRF framework, demonstrating continuous 3D scalar field reconstruction of a turbulent free jet from a distributed camera array. The framework further supports time-resolved reconstruction by incorporating the measurement instant t as a direct network input, and has been applied to experimental scalar field measurements of a smoke-filled free jet (Kelly and Thurow, 2025).

While Halder et al. (2026) demonstrated volumetric MTV reconstruction using the FIMic dataset, this implementation was limited to static reconstruction and did not incorporate time-resolved capability. The time-resolved FluidNeRF framework of Kelly and Thurow (2025), on the other hand, models each camera as a perspective projector, where rays originate from a single focal point and diverge through the image plane with large angular separation, an assumption that is fundamentally incompatible with the orthographic projection geometry of the FIMic system. Furthermore, both implementations rely on frequency-based positional encoding and a large network architecture, resulting in reconstruction times on the order of several hours per volume, a significant practical limitation for time-resolved measurements where multiple volumes must be reconstructed. Addressing the geometric incompatibility of the FIMic imaging system, the absence of time-resolved capability in the MTV implementation, and the computational cost of the original framework together constitute the primary motivation for the present work.

3. Methodology

3.1. Experimental Dataset

The experimental dataset used in this work was obtained from Huck et al. (2025), who demonstrated near-wall volumetric MTV in a turbulent stagnation jet using a Fourier Integral Microscope (FIMic), or Plenoptic 3.0. The measurement employs a two-step photochemical sequence: a 355 nm ultraviolet pulse photobleaches a structured grid of Rhodamine 6G molecules embedded in the flow, writing a pattern of dark beamlets into the fluorescent fluid. After a delay of $\Delta t = 100 \mu s$, a 527 nm read pulse of 150 ns duration illuminates a thin slab of the measurement volume, capturing the convected and deformed beamlet pattern whose displacement encodes the local velocity. A schematic of the optical layout is shown in Figure 1.

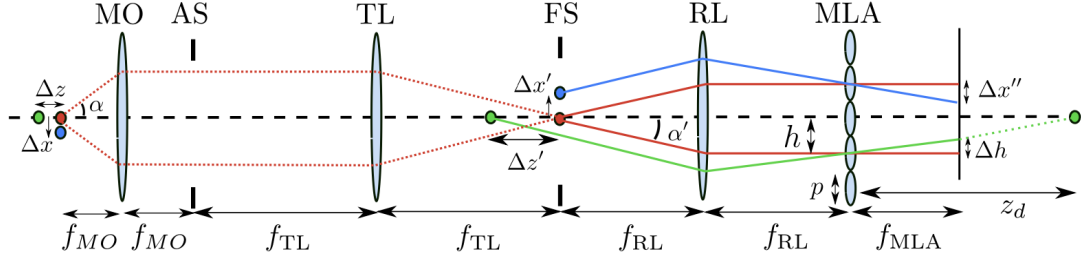


Figure 1. Optical layout of the Fourier Integral Microscope implemented by Huck et al. Huck et al. (2025).

The Fourier Integral Microscope (FIMic) described by (Huck et al., 2025) is an object-space telecentric imaging system, achieved by placing the aperture stop at the Fourier plane. Due to this telecentric configuration, the FIMic produces orthographic projections, meaning rays are parallel and there is minimal parallax when translating the view transversely. Each microlens in the array captures the measurement volume from a slightly different angle, producing a set of independent elemental images that together encode the three-dimensional angular content of the light field. The optical parameters governing this configuration are listed in Table 1, and play a direct role in the ray geometry used in the NeRF reconstruction developed in this work.

The FIMic system used in this work provided 23 angular perspectives of the MTV beamlet pattern, of which 19 elemental images contained clear projections of all six beamlets and were used in the reconstruction. Given the orthographic nature of the system, the projection of a 3D world-coordinate point (x_w, y_w, z_w) onto the camera plane of elemental image $n \in \{1, \dots, 19\}$ is accurately described by a first-order affine model:

$$x_c^{(n)} = x_w + z_w a_{n,x} + b_{n,x}, \quad (1)$$

$$y_c^{(n)} = y_w + z_w a_{n,y} + b_{n,y}, \quad (2)$$

Table 1. Parameters of the FIMic imaging system Huck et al. (2025)

Quantity	Symbol	Value	Unit
Microscope objective focal length	f_{MO}	20	mm
Microscope aperture stop diameter	AS	11.2	mm
Tube lens focal length	f_{TL}	200	mm
Relay lens focal length	f_{RL}	100	mm
Microlens array focal length	f_{MLA}	6.5	mm
Microlens array pitch	p	1	mm
Pixel pitch	δ	2.2	μm

where $a_{n,x}$ and $a_{n,y}$ are the angular sensitivities of the n -th elemental image in the horizontal and vertical directions, respectively, describing the lateral shift induced by a change in depth due to the differing view angle of each microlens. The terms $b_{n,x}$ and $b_{n,y}$ denote the lateral offsets of the optical axis for that elemental view.

In this work, the calibrated affine parameters from (Huck et al., 2025) are directly adopted and serve two purposes. First, they are used to generate orthographic camera rays for the NeRF reconstruction, where each elemental image is treated as an independent camera with its own affine projection model. For a given sample point in 3D space, the affine mapping defines the corresponding pixel location in each of the 19 elemental images, such that the volume rendering accurately models the true imaging geometry of the experimental system. Second, the calibration parameters are reused to generate synthetic MTV images that are geometrically consistent with the experimental system, providing reliable ground truth data for performance evaluation. Because the angular diversity between elemental images is set by the microlens geometry rather than by freely positioned cameras, the available projections span a narrow cone of directions. The angular limitation of this system represents the primary source of reconstruction difficulty - particularly near the wall where the tagged signal is most attenuated. This combination of characteristics makes the dataset a particularly well-suited experimental test case for the proposed framework.

3.2. FluidNeRF Framework

The overall reconstruction pipeline is illustrated in Figure 2. The framework follows the NeRF-based approach of (Kelly and Thurow, 2024), with modifications to accommodate the orthographic imaging geometry of the Plenoptic 3.0 FIMic system. The implementation reconstructs the scalar intensity field from the 19 elemental images acquired by the plenoptic camera.

The neural network receives the three-dimensional spatial coordinates (x, y, z) of each sample point along with the measurement time t . The spatial coordinates are encoded using the multiresolution hash encoder, producing a concatenated feature vector $\mathbf{y} \in \mathbb{R}^{LF}$, where L is the number of

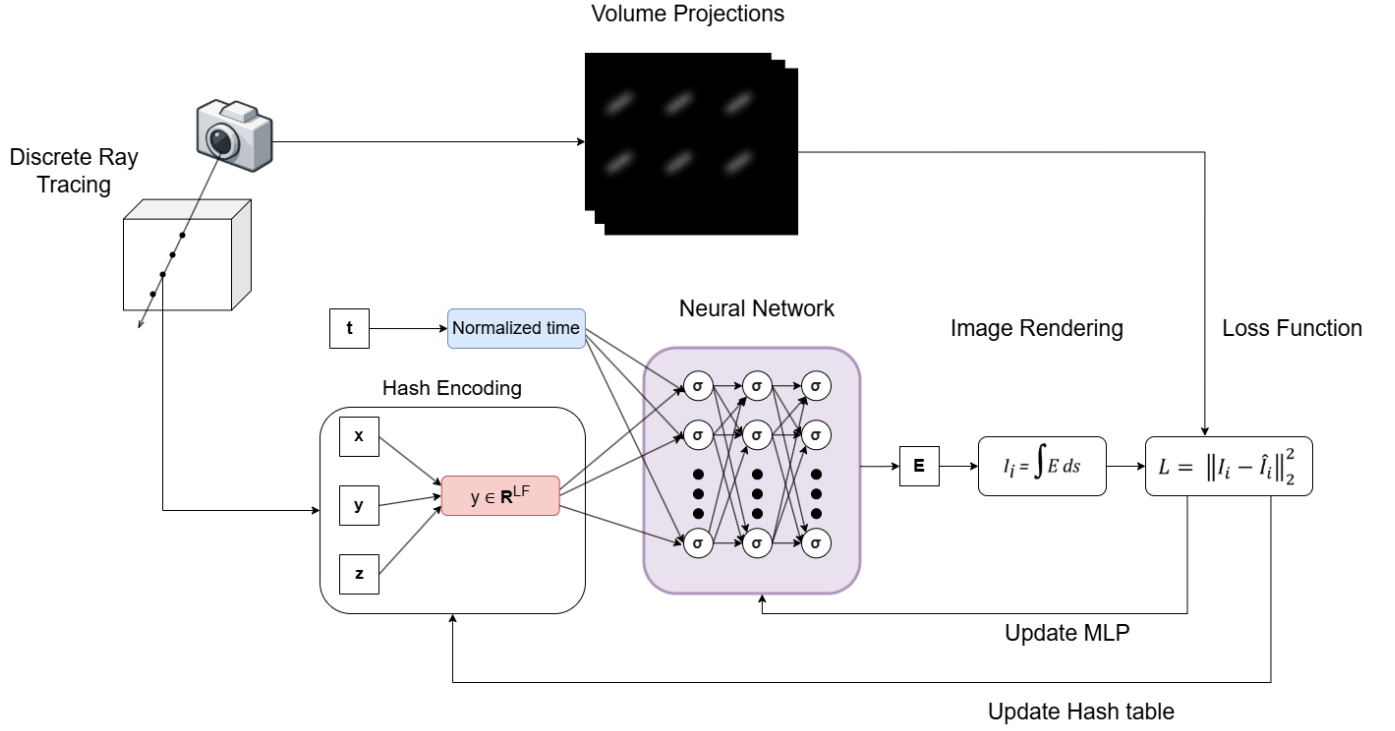


Figure 2. Schematic of the FluidNeRF algorithm for volumetric reconstruction. The neural network maps spatial and temporal coordinates using Hash encoding for intensity outputs.

hash levels and F is the number of features stored per level. The normalized time input t is then concatenated directly with the encoded spatial feature vector before being passed to the MLP. The combined feature vector is provided as input to the multilayer perceptron with ReLU activation functions to predict the scalar intensity field $E(x, y, z, t)$ corresponding to the tagged MTV beam-lets at the specified time instant. The network architecture and training hyperparameters used in this work are summarized in Table 2.

The network was trained by minimizing the squared difference between the rendered and measured pixel intensities. For each ray i , the loss function is defined as

$$\mathcal{L} = \sum_i \|I_i - \hat{I}_i\|_2^2, \quad (3)$$

where I_i is the measured pixel intensity and $\hat{I}_i$ is the corresponding intensity rendered through volume integration along the orthographic projection rays. Optimization was performed using the Adam optimizer, which jointly updated the MLP weights and hash table parameters at each iteration.

Table 2. Network and training hyperparameters.

Parameter	Value
Number of hidden layers	3
Hidden units per layer	128
Activation function	ReLU
Samples per ray	64
Rays per iteration	4096
Learning rate	1×10^{-4} , exponential decay

3.2.1. Input Encoding

The original FluidNeRF framework employed the sinusoidal positional encoding to map 3D spatial coordinates into a higher-dimensional feature space before passing them to the MLP. This encoding is non-parametric, as its basis functions are fixed sinusoidal functions with no trainable parameters (Müller et al., 2022). As a result, the MLP must be made correspondingly large to achieve sufficient representational capacity, which becomes computationally expensive (Deng et al., 2024).

To address this limitation, the multiresolution hash encoding proposed by (Müller et al., 2022) is adopted to encode the three-dimensional spatial coordinates (x, y, z) of each sample point. The hash encoding belongs to the class of parametric encodings (Müller et al., 2022), in which trainable feature vectors are organized within a spatial data structure and optimized jointly with the network weights during training. A key advantage of parametric encodings is their ability to efficiently capture multiscale spatial information, simultaneously representing both coarse global structure and fine local detail (Müller et al., 2022; Deng et al., 2024). This improved spatial representation enables the use of a significantly smaller MLP while maintaining reconstruction quality.

In the present framework, only the spatial coordinates are encoded using the hash encoding, while a normalized time input is provided directly to the MLP as a separate input. The encoding is implemented using the tiny-cuda-nn framework (Müller, 2021). A complete description of the hash encoding formulation and implementation details can be found in Müller et al. (2022).

3.2.2. Ray Sampling

The Plenoptic 3.0 system produces orthographic projections in which rays within each elemental image are mutually parallel. The ray direction differs from one elemental image to the next, producing the set of angular perspectives described in Section 3.1, while the ray origin varies pixel by pixel within each view due to the lateral offset of each microlens. These two properties - the per-view ray direction and the pixel-dependent ray origin - are encoded directly within the affine

calibration coefficients. For elemental image n , a ray passing through pixel i is parameterized as

$$\mathbf{r}(s) = \mathbf{o}_{n,i} + s \mathbf{d}_n, \quad (4)$$

where $\mathbf{o}_{n,i}$ is the pixel-dependent ray origin, $\mathbf{d}_n$ is the view-dependent ray direction extracted from the affine calibration coefficients, and $s \in [s_{\text{near}}, s_{\text{far}}]$ parameterizes depth along the ray. Since the tagged molecular beamlets occupy only a small, sparse fraction of the measurement volume, an occupancy grid (Müller et al., 2022) is maintained alongside the network to identify regions of non-negligible intensity within the volume. During ray marching, sample points falling within empty grid cells are skipped, which concentrates computation within the occupied regions and provides significant acceleration in training.

3.3. Synthetic Data Generation

Synthetic MTV images were generated using the calibrated affine projection model of the FIMic system described in Huck et al. (2025). The synthetic measurement volume consisted of six MTV beamlets arranged in a 3×2 configuration with a center-to-center spacing of $500 \mu\text{m}$, a diameter of $90 \mu\text{m}$, and an axial length of $1800 \mu\text{m}$, consistent with the experimental setup. To generate the synthetic elemental images for $t = 0$, sample points were randomly distributed within the prescribed beamlet structures in object space. The point distribution followed a Gaussian profile in the transverse (x, y) directions and a uniform distribution along the axial (z) direction, producing beamlets with realistic transverse intensity decay and approximately constant intensity along the beamlet length. Each sample point was then projected onto the sensor plane of each elemental image using the calibrated affine parameters, and the intensity contributions were accumulated across all points to produce the final synthetic elemental images. A representative example of the generated synthetic plenoptic images alongside the corresponding experimental images is shown in Figure 3.

The number of sample points per beamlet was determined through an iterative convergence study, in which points were added in batches of 10,000 and the normalized RMS change in the projected sensor image was monitored between successive batches. Convergence was declared when this change fell below 0.5%, yielding a total of 80,000 sample points per beamlet.

To evaluate the framework under time-resolved conditions, the sample points from $t = 0$ s were advected through a fully developed Poiseuille flow field between two parallel flat plates separated by a channel height of $H = 1800 \mu\text{m}$, with a peak centerline velocity of $u_{\text{max}} = 0.4 \text{ mm/s}$. The streamwise velocity profile is given by

$$u(y) = u_{\text{max}} \left(1 - \left(\frac{2y}{H} \right)^2 \right), \quad (5)$$

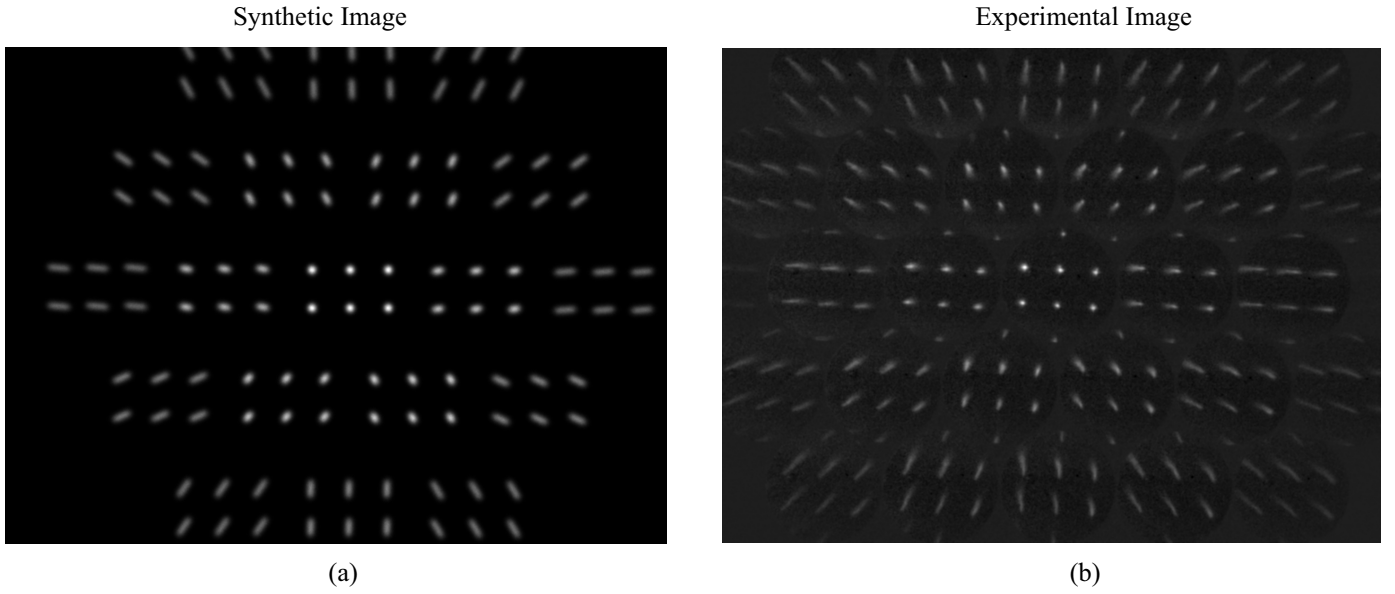


Figure 3. (a) Generated synthetic image and (b) experimental image of tagged MTV lines for the Fourier Integral Microscopy (FIMic) system.

where y is the wall-normal coordinate measured from the channel centerline, H is the plate spacing, and $u_{\max}$ is the maximum velocity at the channel centerline. The displaced sample positions were then projected onto the sensor plane using the same calibrated affine model to generate the synthetic elemental images at $t = 0.1$ s. Each elemental view was modeled as an independent orthographic camera, consistent with the object-space telecentric configuration of the FIMic imaging system. The main parameters of the synthetic data generation procedure are summarized in Table 3.

Parameter	Value
Beamlet diameter	90 μm
Beamlet length	1800 μm
Beamlet spacing	500 μm
Total sample points per frame	480,000 (6 beamlets $\times$ 80,000)
Sensor resolution	1800 $\times$ 2400 px
Image bit depth	16 bit

Table 3. Synthetic plenoptic image generation parameters.

The use of two time instances allows assessment of the network’s ability to learn time-resolved data from a single spatiotemporal representation, while providing known ground truth at each time step for quantitative evaluation of reconstruction fidelity. It should be noted that, due to the orthographic projection geometry of the FIMic system, the camera model and reconstruction framework developed in this work are specific to this imaging configuration and are not directly

transferable to conventional perspective-based camera systems.

4. Results and Discussions

The reconstruction results and quantitative evaluation of the proposed framework using the synthetic dataset are presented in the following sections. To assess the impact of the encoding strategy, isosurface reconstructions obtained using both hash encoding and Fourier encoding are compared against the synthetic ground truth. The hash-encoded implementation follows the methodology described in Section 3. For comparison, the Fourier-encoded implementation follows the approach of Kelly and Thurow (2025), employing a multilayer perceptron with 12 hidden layers, 400 neurons per layer, and a positional encoding frequency of $L_x = 12$. The significantly larger network architecture required by the Fourier-encoded implementation contributes directly to the substantial increase in training time relative to the hash-encoded approach, as demonstrated in the following sections.

4.1. Synthetic Analysis

Figure 4 compares the training performance of the multiresolution hash encoding and conventional Fourier encoding on the synthetic MTV dataset. Both models were trained for 4000 iterations using identical ray-sampling points, training hyperparameters, and hardware (NVIDIA RTX 5040 GPU), ensuring that differences in performance arose primarily from the encoding strategy.

The training loss, validation NRMSE, and validation SSIM shown in Fig. 4 were computed from the rendered elemental image projections and their corresponding synthetic ground-truth projections. As shown in Fig. 4(a), both approaches exhibit a rapid decrease in training loss during the initial iterations; however, the hash-encoded implementation converges to a significantly lower loss value. The corresponding validation metrics in Figs. 4(b) and (c) show that the hash-encoded model achieves substantially lower NRMSE and higher SSIM throughout training. Although the validation metrics indicate that the hash-encoded model reaches convergence within the first few hundred iterations, training was continued to 4000 iterations to further refine fine-scale beamlet features that are not fully captured by global image-quality metrics. Figure 4(d) further highlights the computational advantage of the proposed approach. Under identical training conditions, the hash-encoded implementation completed 4000 training iterations in 789 s (13.1 min), compared with 13,481 s (224.7 min) for the Fourier-encoded implementation, corresponding to a reduction in training time of approximately 94%. These results demonstrate that multiresolution hash encoding not only improves reconstruction fidelity but also significantly accelerates convergence, making it well suited for time-resolved MTV reconstruction.

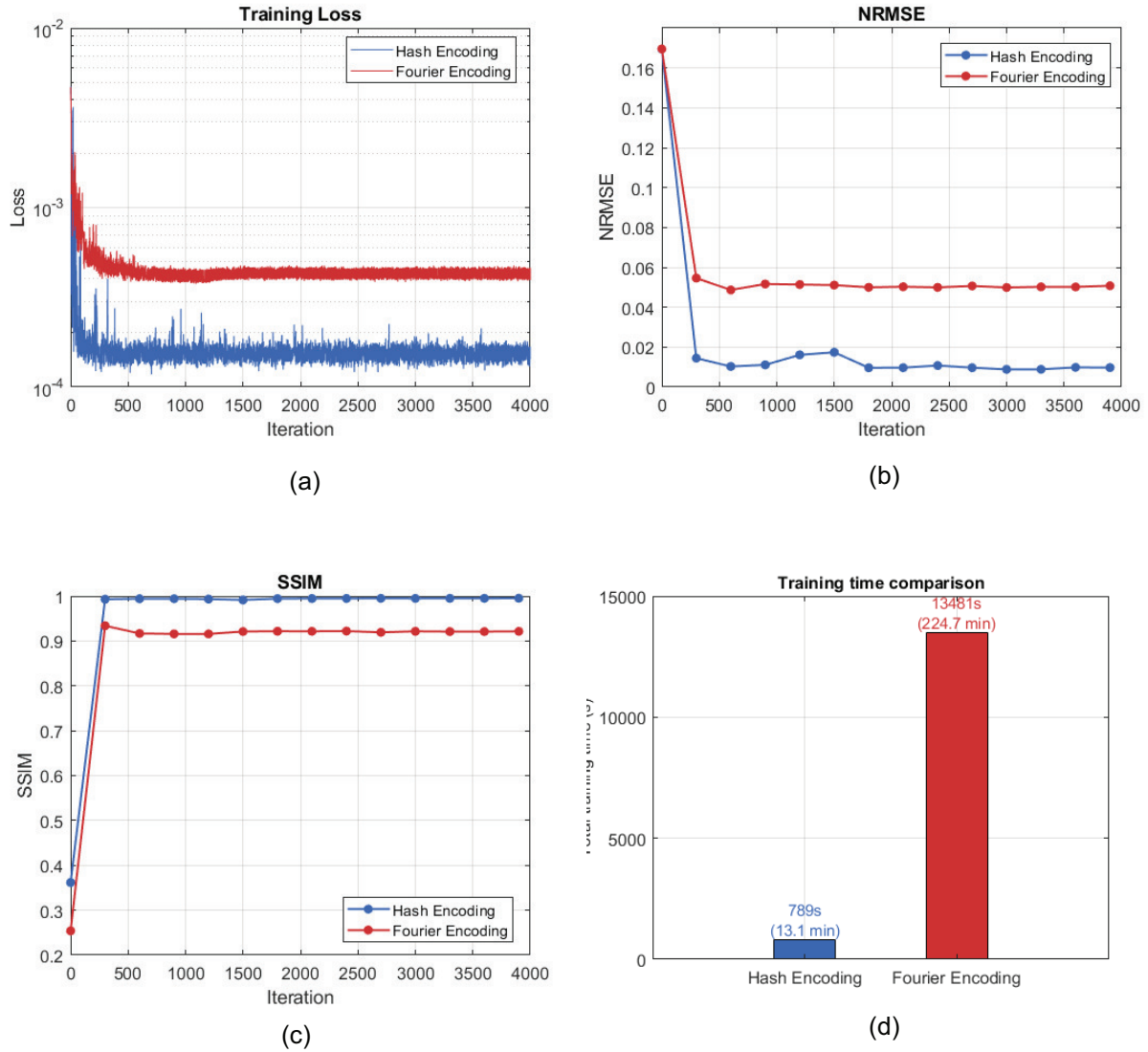


Figure 4. Comparison of multiresolution hash encoding and Fourier encoding during training on the synthetic MTV dataset. (a) Training loss as a function of iteration. (b) Normalized root mean square error (NRMSE). (c) Structural similarity index (SSIM). (d) Total training time.

The convergence behavior observed in the validation metrics is also evident in the volumetric reconstructions shown in Figure 5. The hash-encoded implementation rapidly recovers the beamlet geometry and spatial arrangement within the first few hundred iterations, whereas the Fourier-encoded implementation exhibits slower convergence and increased reconstruction artifacts during the early stages of training. This qualitative observation is consistent with the lower NRMSE and higher SSIM achieved by the hash-encoded model throughout training.

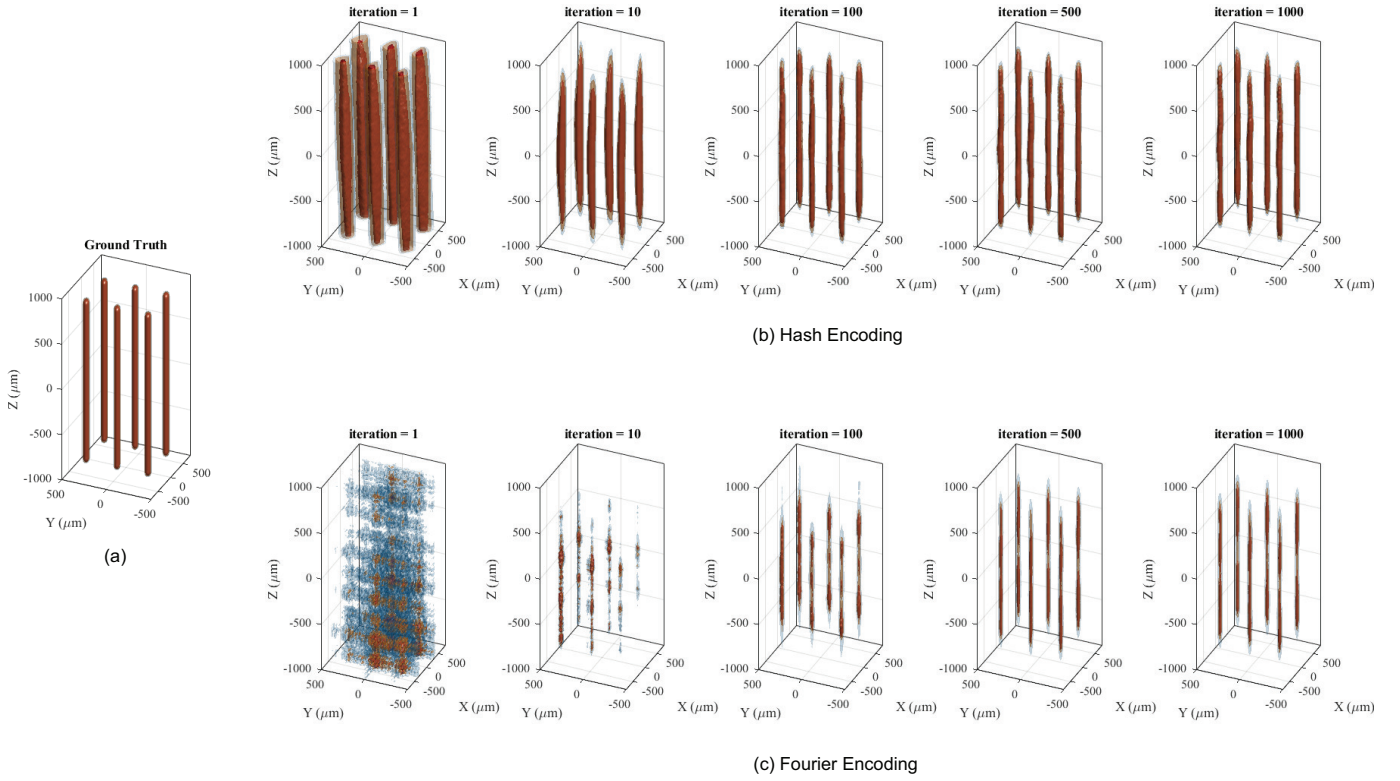


Figure 5. Evolution of the reconstructed synthetic MTV volume at $t = 0$ s during training. (a) Ground-truth volume generated from the synthetic dataset. (b) Reconstructions obtained using multiresolution hash encoding at iterations 1, 10, 100, 500, and 1000. (c) Reconstructions obtained using Fourier encoding at the same iterations.

4.2. Isosurface Comparison

Figure 6 presents 3D isosurface reconstructions of the synthetic MTV volume at $t = 0$ and $t = 0.1$ s. At both time instances, both the hash-encoded and Fourier-encoded reconstructions correctly reproduce the 3×2 beamlet arrangement, with each beamlet clearly resolved and spatially separated at the expected $500 \mu\text{m}$ center-to-center spacing, and the reconstructed isosurfaces appearing visually comparable in terms of beamlet geometry, spatial separation, and boundary sharpness. The reconstruction maintains good fidelity at $t = 0.1$ s, where the imposed Poiseuille flow introduces a spatially varying displacement across the beamlet array, visually evident in both reconstructed volumes. Minor reconstruction artifacts are visible near the boundaries of the region of interest, likely associated with limited angular sampling and edge effects.

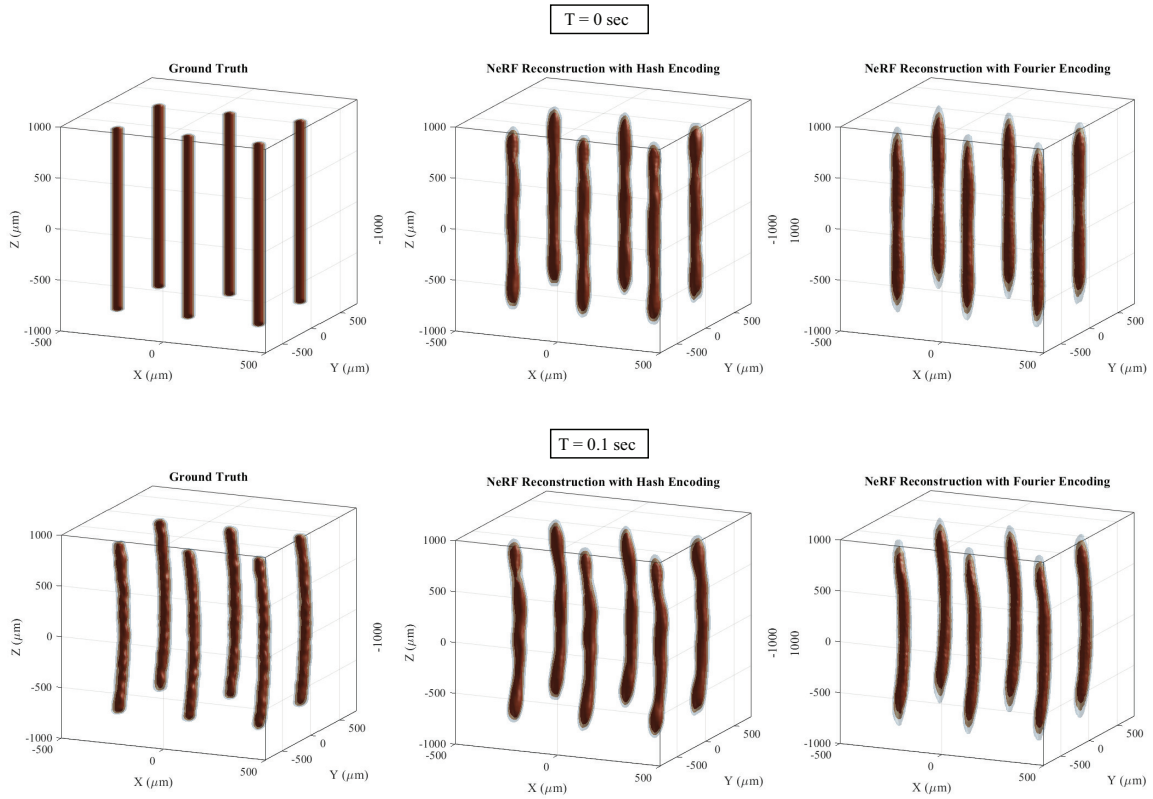


Figure 6. 3D isosurface reconstructions of the synthetic MTV volume at $t = 0$ and $t = 0.1$. Left column: ground truth volumes computed analytically from the Gaussian beamlet definitions. Middle column: FluidNeRF reconstructions obtained using multiresolution hash encoding. Right column: FluidNeRF reconstructions obtained using Fourier encoding.

However, the quantitative metrics presented in Table 4 reveal meaningful differences that are not readily apparent from visual inspection alone. The SSIM, MSE, and PSNR values were computed from the reconstructed 3D volumes corresponding to the isosurface reconstructions shown in Fig. 6. At both time instances, the hash-encoded implementation achieves slightly higher SSIM but substantially lower MSE and higher PSNR than the Fourier-encoded implementation. The quantitative metrics in Table 4 reveal that the hash encoding achieves MSE values on the order of 10^{-4} , compared to 10^{-3} for the Fourier encoding, representing an order of magnitude reduction in volumetric reconstruction error. Furthermore, the hash encoding yields a PSNR improvement over the Fourier-encoded approach for both time instances. These advantages are consistent across both time instances, confirming that visual inspection alone is insufficient to discriminate between reconstruction methods when both produce structurally plausible outputs. Notably, the hash encoding achieves these improved reconstruction metrics in a significantly reduced training time compared to the Fourier-encoded approach. Given these improvements, the remainder of the analysis is presented using the hash-encoded implementation.

Table 4. Quantitative comparison between ground truth volumes and FluidNeRF reconstructions using multiresolution hash encoding and Fourier encoding.

Metric	$t = 0$ s		$t = 0.1$ s	
	Hash	Fourier	Hash	Fourier
SSIM	0.9795	0.9603	0.9672	0.9571
MSE	0.000411	0.003127	0.000698	0.002742
PSNR (dB)	33.87	25.05	31.56	25.62

4.3. Depth-Resolved Cross-Sections

Figure 7 compares the 2D slices of ground-truth and FluidNeRF-reconstructed intensity volumes using hash encoding at two time instances, $T = 0$ s and $T = 0.1$ s. Figures 7(a,c) show transverse XY slices at $Z = 0$ μm , while Figures 7(b,d) show longitudinal XZ slices at $Y = 250$ μm . At both time points, the reconstruction accurately reproduces the spatial arrangement of the three lines and their transverse intensity profiles.

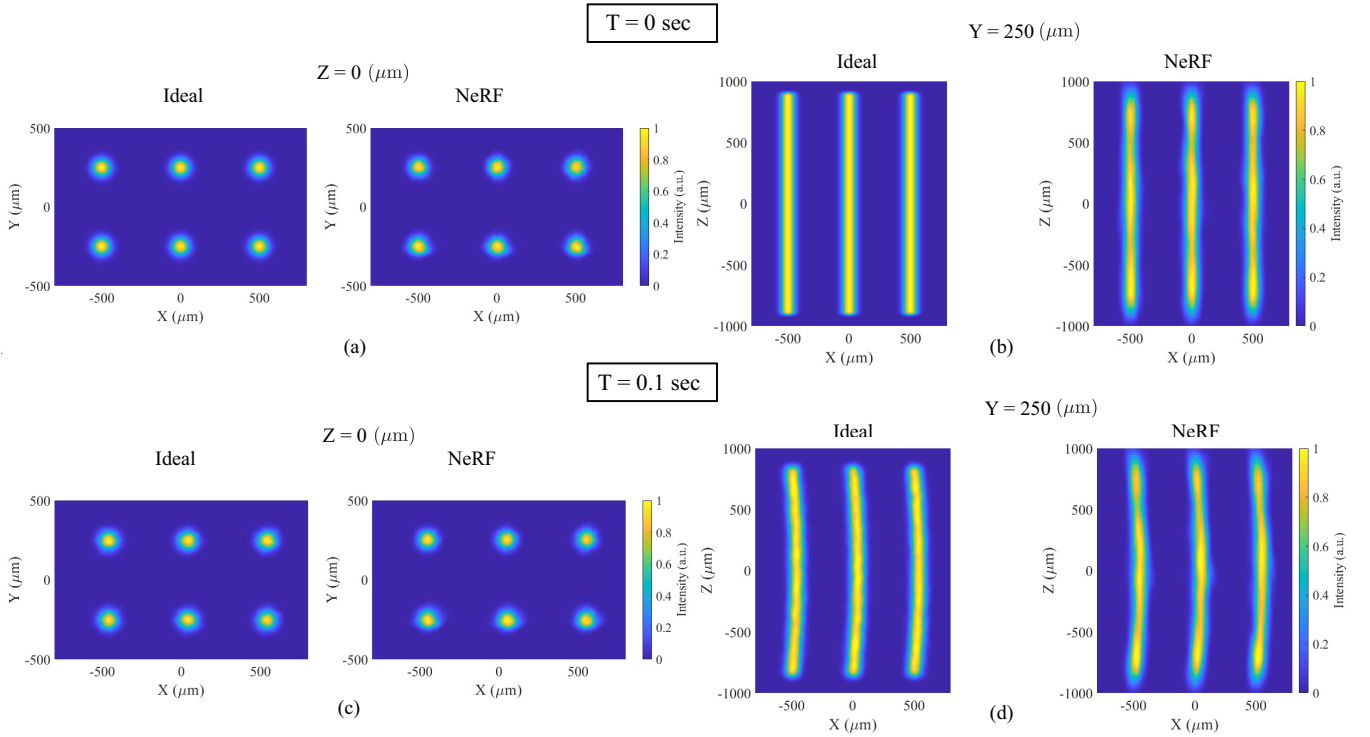


Figure 7. Comparison of ground-truth (*Ideal*) and FluidNeRF-reconstructed (*NeRF*) intensity volumes at $T = 0$ s and $T = 0.1$ s using hash encoding. (a) and (c) show transverse XY slices at $Z = 0$ μm , while (b) and (d) show longitudinal XZ slices at $Y = 250$ μm .

The XY slices show good agreement between the reconstructed and ground-truth volumes, with

the line features remaining well localized and clearly separated. In the XZ slices, FluidNeRF successfully captures the line trajectories and the time-dependent deformation observed at $T = 0.1$ s. Minor axial intensity variations are visible along the reconstructed lines due to limited angular information inherent to the plenoptic system; however, the overall structure and temporal evolution of the lines are well recovered.

4.4. Beamlet Width Characterization

Figure 8 shows a comparison between the normalized 1D intensity profiles extracted along the X -direction through the maximum intensity region of the reconstructed MTV line at several depth locations. The FluidNeRF reconstruction shows strong agreement with the synthetic ground-truth profiles, accurately capturing both the peak position and the overall intensity distribution throughout the measurement volume. The full width at half maximum (FWHM) measured at three representative depth planes was 91, 90, and 91 μm , respectively, compared with an ideal FWHM of 90 μm for all three planes at $T = 0$, as shown in Figure 8.

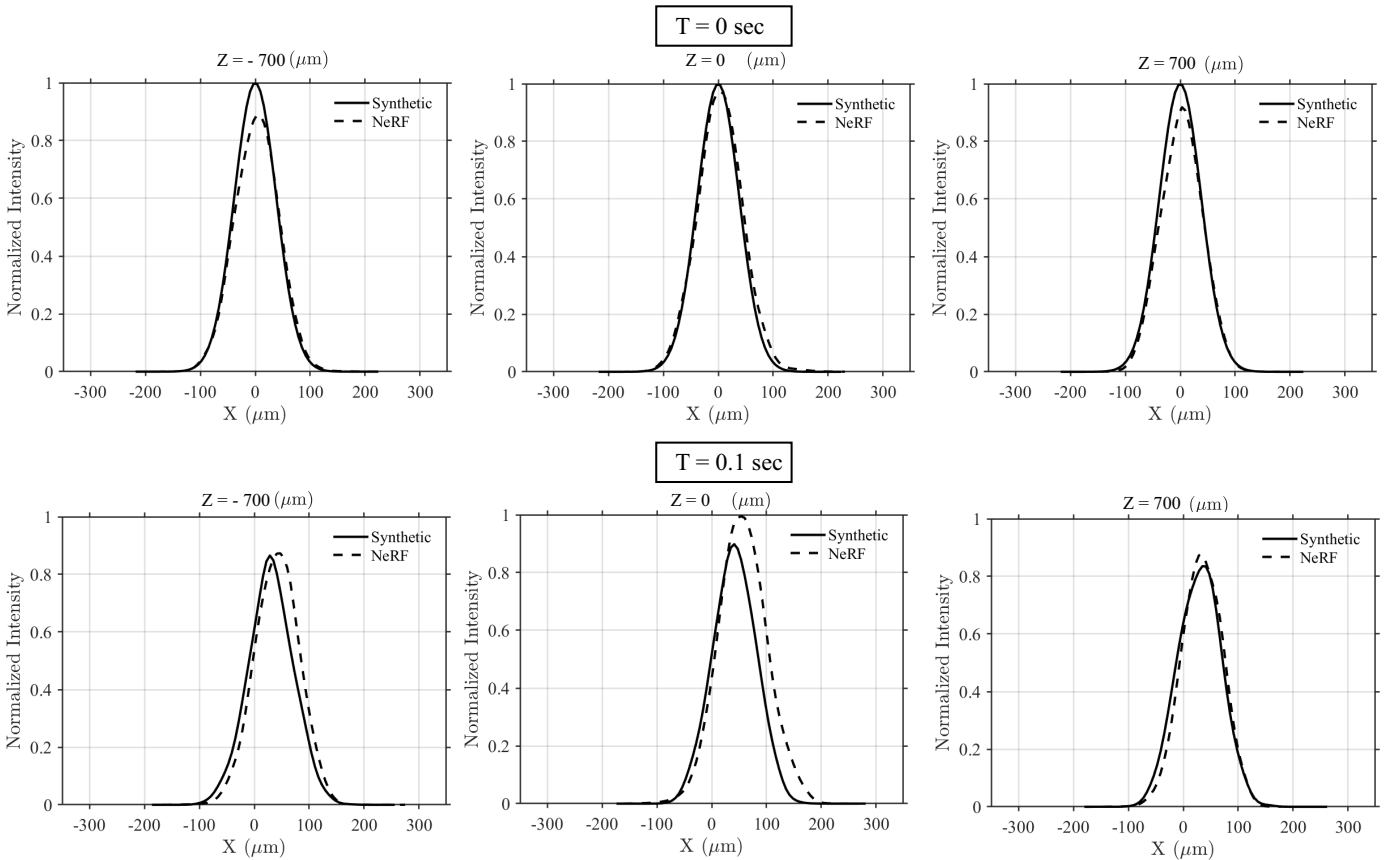


Figure 8. Comparison of reconstructed beamlet intensity profiles obtained using FluidNeRF with the corresponding synthetic ground-truth profiles at two time instances, $T = 0$ s and $T = 0.1$ s. Profiles are shown along the X -direction for three axial planes, $Z = -700$ μm , $Z = 0$ μm , and $Z = 700$ μm . Solid lines denote the synthetic ground truth, while dashed lines represent the reconstructed FluidNeRF intensity distributions.

At the later time step $T = 0.1$, the corresponding FWHM values were 96, 103, and 95 μm , compared with 97, 95, and 97 μm for the ideal synthetic reconstruction at the same depth planes. These results indicate that the reconstructed beam thickness is well preserved across depth, with only small variations relative to the synthetic reference. Overall, the comparison demonstrates that FluidNeRF reliably reconstructs the spatial morphology and intensity characteristics of the MTV line while maintaining the characteristic beam width of the synthetic reference solution.

4.5. Experimental Results

The experimental dataset was acquired at The George Washington University by Huck et al. (Huck et al., 2025), in which the 3D intensity field was reconstructed using Richardson-Lucy Deconvolution (RLD). A full description of the experimental setup, acquisition procedure, and RLD reconstruction methodology is provided in Huck et al. (2025). The same dataset was used to evaluate the FluidNeRF hash-encoded implementation, and Figure 9 presents a comparison of the RLD-reconstructed experimental volume and the corresponding FluidNeRF reconstruction at both time instances, $\Delta t = 0$ and $\Delta t = 200 \mu\text{s}$.

At both time instances, the FluidNeRF reconstruction successfully recovers the 3×2 beamlet arrangement, with each beamlet clearly resolved and spatially separated, consistent with the experimentally reconstructed volume. The overall beamlet geometry, axial extent, and spatial distribution are well preserved in the NeRF reconstruction. At $\Delta t = 200 \mu\text{s}$, the deformation and lateral displacement of the beamlet pattern introduced by the flow are clearly evident in both the experimental and FluidNeRF-reconstructed volumes. Both reconstructions exhibit consistent displacement patterns, with the FluidNeRF reconstruction closely reproducing the beamlet deformation observed in the RLD volume. Minor differences in boundary sharpness and intensity distribution between the two reconstructions are expected, given the limited angular diversity of the FIMic system and the fundamental differences between a continuous neural implicit representation and a voxel-based iterative reconstruction approach.

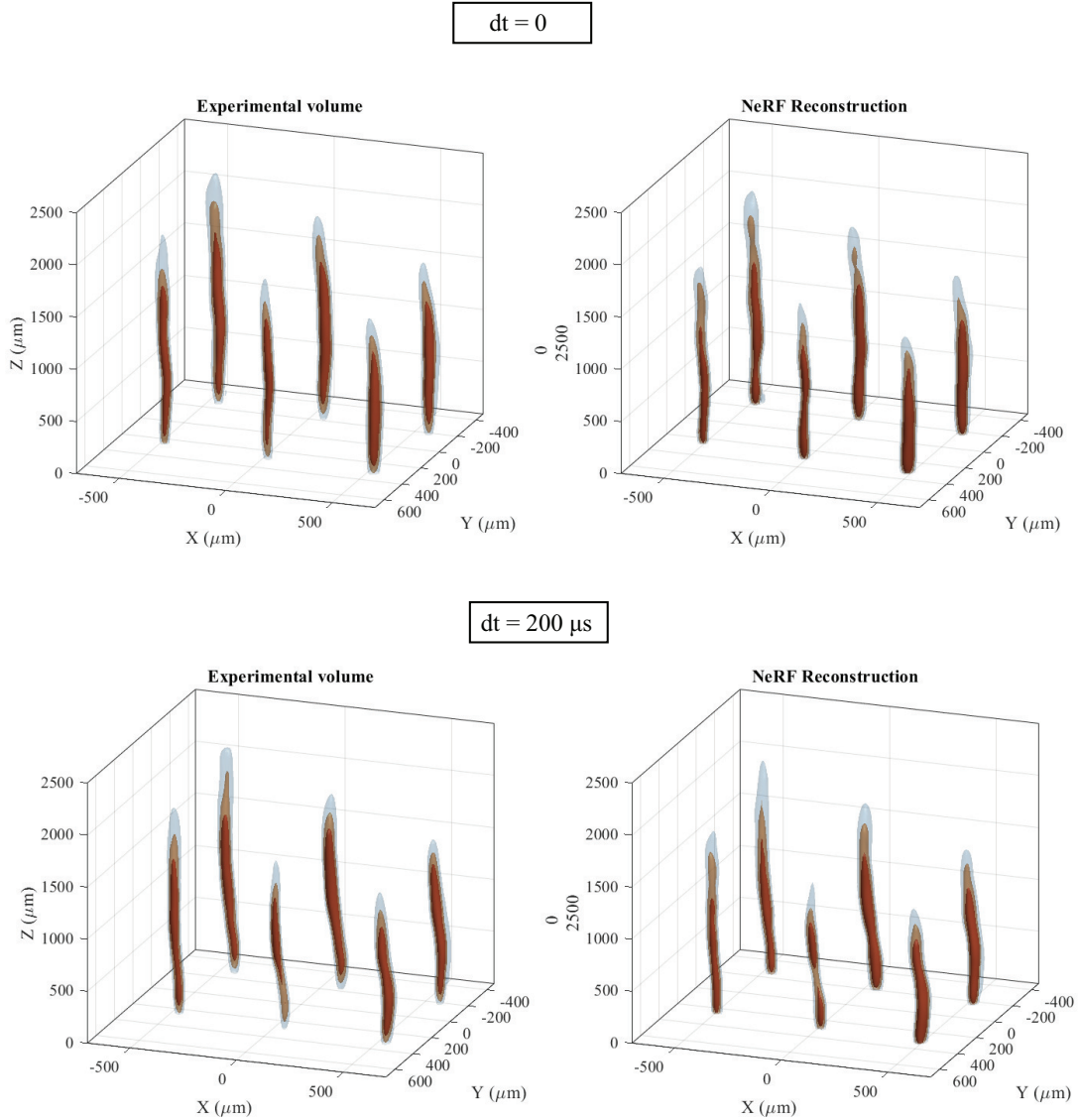


Figure 9. 3D isosurface reconstructions of the experimental MTV volume at $\Delta t = 0$ and $\Delta t = 200 \mu s$. Left column: experimental volume reconstructed using Richardson-Lucy Deconvolution (RLD) implemented by Huck et al. (2025). Right column: corresponding FluidNeRF reconstructions obtained using multiresolution hash encoding.

5. Conclusions

This work demonstrates the feasibility of a NeRF-based reconstruction framework for 3D scalar field recovery from FIMic-based molecular tagging velocimetry. By adapting the camera model to the orthographic projection geometry of the Plenoptic 3.0 system, incorporating a multiresolution hash encoding, and representing the scalar field as a continuous function of both space and time, the framework achieves accurate and computationally efficient time-resolved reconstruction of the tagged MTV beamlet pattern.

A central finding of this work is the substantial computational advantage of multiresolution hash encoding over conventional Fourier encoding. Under identical training conditions, the hash-encoded implementation completed reconstruction in 789 s (13.1 min), compared with 13,481 s (224.7 min) for the Fourier-encoded implementation. Fourier encoding contains no trainable parameters and therefore requires a larger MLP to capture fine spatial details. In contrast, multiresolution hash encoding introduces learnable features within the encoding itself, enabling a significantly smaller network to represent both coarse and fine-scale structures efficiently. This reduces the computational cost of training without compromising reconstruction as evident in the synthetic results. Comparison of reconstruction quality shows that the modified FluidNeRF reconstruction with hash encoding closely approximates the ideal volumetric reconstruction of the MTV lines, while demonstrating slightly improved performance over traditional Fourier encoding in both qualitative structural preservation and quantitative accuracy. Overall, the synthetic analysis demonstrates the capability of the proposed framework to accurately reconstruct three-dimensional MTV line structures and highlights the effectiveness of the modified FluidNeRF approach.

Future work will focus on extending the present framework from volumetric reconstruction to fully time-resolved three-dimensional flow field measurements. A key objective is to use optical flow techniques on consecutive reconstructed volumetric intensity fields to estimate the underlying three-dimensional velocity field. Furthermore, incorporating governing physical laws directly into the optimization process through physics-informed neural field formulations may improve reconstruction accuracy and ensure physically consistent solutions. Future studies will also investigate the influence of network architecture, encoding strategy, training hyperparameter, image noise, and experimental imaging conditions on reconstruction quality to better quantify the strengths, limitations, and uncertainty characteristics of the proposed approach.

Acknowledgements

Development of the tomographic reconstruction method for MTV using Neural Radiance Fields was supported by the Office of Naval Research under Award No. N00014-25-12398, managed by Dr. Yin-Lu (Julie) Young. Selected sections of the manuscript were refined for clarity, grammar, and language using ChatGPT (OpenAI, GPT-5 series). The authors reviewed and edited all AI-assisted text and take full responsibility for the final content.

References

- Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. (2021). Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of*

the IEEE/CVF international conference on computer vision, pages 5855–5864.

- Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2023). Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19697–19705.
- Deng, Z., Xiao, H., Lang, Y., Feng, H., and Zhang, J. (2024). Multi-scale hash encoding based neural geometry representation. *Computational Visual Media*, 10(3):453–470.
- Engelmann, F., Manhardt, F., Niemeyer, M., Tateno, K., Pollefeys, M., and Tombari, F. (2024). Open-nerf: open set 3d neural scene segmentation with pixel-wise features and rendered novel views. *arXiv preprint arXiv:2404.03650*.
- Halder, S. H., Yamakaitis, M. J., Sapkota, B., Fort, C., Bardet, P., and Thurow, B. S. (2026). Development of tomographic reconstruction method for mtv using neural radiance fields. In *AIAA SCITECH 2026 Forum*, page 2100.
- Huck, P. D., Yamakaitis, M. J., Fort, C., and Bardet, P. M. (2025). Near-wall volumetric molecular tagging velocimetry with a fourier integral microscope. *Experiments in Fluids*, 66(8):152.
- Jain, N., Raj, A., Kalra, M., Munshi, P., and Ravindran, V. (2011). Mart algorithms for circular and helical cone-beam tomography. *Research in Nondestructive Evaluation*, 22(3):147–168.
- Kelly, D. and Thurow, B. (2024). Investigation of a neural implicit representation tomography method for flow diagnostics. *Measurement Science and Technology*, 35(5):056007.
- Kelly, D. L. and Thurow, B. S. (2023). Fluidnerf: a scalar-field reconstruction technique for flow diagnostics using neural radiance fields. In *AIAA SciTech 2023 Forum*, page 0412.
- Kelly, D. L. and Thurow, B. S. (2025). Experimental proof-of-concept of a time-resolved gridless tomography method for 3d scalar fields. In *AIAA SCITECH 2025 Forum*, page 1059.
- Koochesfahani, M., Goh, A., and Schock, H. (2004). Molecular tagging velocimetry (mtv) and its automotive applications. In *The Aerodynamics of Heavy Vehicles: Trucks, Buses, and Trains*, pages 143–155. Springer.
- Li, F., Zhang, H., and Bai, B. (2021). A review of molecular tagging measurement technique. *Measurement*, 171:108790.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106.

- Müller, T. (2021). tiny-cuda-nn.
- Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15.
- Qiu, R.-Z., Yang, G., Zeng, W., and Wang, X. (2024). Feature splatting: Language-driven physics-based scene synthesis and editing. *arXiv preprint arXiv:2404.01223*.
- Ströhl, F. and Kaminski, C. F. (2015). A joint richardson—lucy deconvolution algorithm for the reconstruction of multifocal structured illumination microscopy data. *Methods and Applications in Fluorescence*, 3(1):014002.
- Stull, K. E., Tise, M. L., Ali, Z., and Fowler, D. R. (2014). Accuracy and reliability of measurements obtained from computed tomography 3d volume rendered images. *Forensic science international*, 238:133–140.
- Warburg, F., Weber, E., Tancik, M., Holynski, A., and Kanazawa, A. (2023). Nerfbusters: Removing ghostly artifacts from casually captured nerfs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18120–18130.
- Wu, S., Basu, S., Broedermann, T., Van Gool, L., and Sakaridis, C. (2025). Pbr-nerf: Inverse rendering with physics-based neural fields. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 10974–10984.
- Zhu, X., Wu, Z., Li, J., Zhang, B., and Xu, C. (2021). A pre-recognition sart algorithm for the volumetric reconstruction of the light field piv. *Optics and Lasers in Engineering*, 143:106625.