Development of a Neural Radiance Field Technique for Volumetric Molecular Tagging

by

Sandra Holy Halder

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 8, 2026

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

Approved by

Brian Thurow, Chair, W. Allen and Martha Reed Professor of Aerospace Engineering
David Scarborough, Associate Professor of Aerospace Engineering
James Michael, Walt and Virginia Woltosz Associate Professor of Aerospace Engineering

Abstract

Tomographic reconstruction of three-dimensional flow fields from limited angular views is a fundamental challenge in experimental flow diagnostics. Conventional grid-based methods such as MART, SART, and RLD amplify noise and produce artifacts under the narrow angular sampling imposed by compact plenoptic imaging systems. In their discretized formulation, spatial resolution depends on the voxel grid spacing, the computational cost is high, and each time instant requires a separate reconstruction. Neural Radiance Fields provide a continuous, gridless alternative to these limitations, and this representation has been applied to flow diagnostics through FluidNeRF. FluidNeRF has not, however, been demonstrated for Molecular Tagging Velocimetry. Existing implementations rely on perspective camera models with wide angular coverage and on Fourier positional encoding, which requires a large network and long training times. Existing FluidNeRF implementations rely on perspective camera models with wide angular coverage and on Fourier positional encoding, which requires a large network and long training times. In this work, the FluidNeRF framework was adapted to the orthographic ray geometry of the Fourier Integral Microscope (FIMic) for the reconstruction of time-resolved scalar fields in MTV. Multiresolution hash encoding additionally replaced Fourier positional encoding, which reduced the required network size and thereby the computational cost of training.The framework was tested on two synthetic datasets generated using the calibrated FIMic imaging geometry: Poiseuille flow and turbulent channel flow. A systematic investigation of the encoding parameters and network size was performed to optimize the network and to enable time-resolved reconstruction of two time instants simultaneously. The optimized framework was then applied to experimental volumetric MTV measurements of a axisymmetric stagnation jet. The hash-encoded framework slightly exceeded the reconstruction fidelity of the Fourier-encoded implementation while reducing training time by more than an order of magnitude. Beamlet morphology and axial resolution were

preserved throughout the volume. Applied to experimental data, a single network captured the temporal evolution of the tagged beamlets, and the reconstructions showed good agreement with Richardson–Lucy deconvolution. The results establish time-resolved neural implicit representations as a computationally efficient alternative to conventional grid-based tomography for plenoptic MTV.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this thesis, the following Artificial Intelligence (AI) tools were used for assistance: Claude (Anthropic) and ChatGPT (OpenAI). These tools were primarily used to assist with grammar correction, sentence refinement, and improvement of language clarity. These tools were utilized solely as writing aids and did not contribute to the development of the research methodology, data analysis, interpretation of results, or scientific conclusions presented in this work. The author acknowledges full responsibility for the intellectual content, accuracy, and integrity of this thesis. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain the scholarly integrity of this research.

Digital Accessibility Disclosure Statement

In the preparation of this thesis, the following digital accessibility tools were used to ensure this document complies with federal requirements: Adobe Acrobat Pro Accessibility Checker. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgments

I would like to express my deepest gratitude to my advisor, **Dr. Brian Thurow**, whose mentorship has been the cornerstone of this work. His guidance extended far beyond this research, shaping me into a better researcher, writer, and person. Throughout my graduate studies, he continually challenged me to think critically, communicate clearly, and approach research with curiosity and integrity. I am especially grateful for his patience, his open-door policy, even for my many unannounced visits, and his unwavering willingness to help regardless of the problem. His encouragement, wisdom, and genuine investment in my growth have had a lasting impact on both my academic and personal development, and I will always be grateful for everything he has done for me.

I would also like to thank my committee members, **Dr. David Scarborough** and **Dr. James Michael**, for their valuable time, insightful feedback, and thoughtful contributions to this work.

I would like to thank my collaborators at **The George Washington University** for their valuable collaboration and support throughout this research. I am also grateful to my labmates in the **Advanced Flow Diagnostics Laboratory (AFDL)** for creating a collaborative, and enjoyable research environment. Their encouragement, discussions, and friendship made my graduate experience truly memorable. Finally, I gratefully acknowledge the financial support provided by the **Office of Naval Research (ONR)**, which made this research possible.

On a personal note, I owe everything to my parents, whose unconditional love, constant encouragement, prayers, and blessings carried me through every difficult moment. Thank you for always believing in me, for listening to every rant, and for reminding me to keep going even when I wanted to give up. None of this would have been possible without your endless sacrifices and unwavering support.

I am also deeply thankful to my dear friend, **Kaniz Fatema Bristy**, whose guidance,

friendship, and wisdom on how to survive graduate school kept me grounded when I needed it most. Your support has meant more to me than words can express.

Finally, I would like to thank my partner, **Connor McLaughlin**, for believing in me when I could not believe in myself, for bringing me peace during the most overwhelming moments, and for standing by me throughout this journey. Your patience, encouragement, and constant support gave me the strength to keep moving forward. This accomplishment would not have been possible without you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Accurate characterization of near-wall flow fields remains a fundamental challenge in fluid mechanics. Understanding fluid behavior in the immediate vicinity of a surface is fundamental to many engineering applications, including internal flows, boundary layer studies, and aerodynamic surface design [1]. Considerable effort has been directed toward improving near-wall flow measurements through advances in instrumentation, experimental techniques, and reconstruction methods. Despite these developments, near-wall measurements remain difficult due to the small spatial scales involved, steep velocity gradients, and limited optical access [2]. The viscous length scale, ℓ_ν, can range from 0.7 to 11 μm in air [3, 4], and can be as small as 1.7 μm in water [5, 6], making accurate measurements challenging.

Several particle-based and molecule-based techniques have been developed to address these challenges. Particle-based approaches, such as Particle Image Velocimetry (PIV) and Particle Tracking Velocimetry (PTV), rely on the motion of seeded tracer particles [7]. In contrast, molecule-based techniques use tagged molecules as flow tracers. These molecular tracers follow the flow with negligible inertia and are not affected by particle lag or surface reflections, making them well suited for near-wall measurements [7, 8]. Molecular Tagging Velocimetry (MTV) is one such technique and has been successfully applied to small measurement volumes on the order of 1–10 mm^3 and regions of high shear where particle-based methods become difficult to apply [2, 9].

Early implementations of MTV were primarily planar measurement techniques capable of providing two velocity components within a two-dimensional measurement plane (2D–2C). While useful for many applications, realistic flow fields are inherently three-dimensional and often contain significant out-of-plane motion. As a result, MTV has been extended to volumetric measurements using stereoscopic imaging systems, tomographic

approaches, and plenoptic imaging techniques [7]. Multi-camera stereoscopic systems were among the first approaches used to obtain three-dimensional MTV measurements [8]. However, their use in confined near-wall environments is often limited by restricted optical access, sensor placement constraints, and the complexity of camera calibration and alignment.

More recently, Fourier Integral Microscopy (FIMic) has enabled volumetric MTV measurements using a single plenoptic camera system [2, 9]. Although this approach overcomes many of the practical limitations associated with multi-camera configurations, the elemental images recorded by a FIMic system provide only a limited number of projection angles determined by the microlens array geometry. The angular separation between neighboring elemental views is small, and the available projections span only a narrow angular range around the measurement volume. As a result, the reconstruction problem is inherently a limited-angle tomography problem.

Despite this instrumentation advance, the quality of the three-dimensional velocity measurement remains dependent on the accuracy of the tomographic reconstruction. Existing reconstruction algorithms can produce reasonable results; however, they often require long reconstruction time and can suffer from noise amplification and reconstruction artifacts, which can become more pronounced under limited-angle conditions [10, 11]. These limitations have motivated the investigation of alternative reconstruction approaches for volumetric MTV measurements. Moreover, realistic flow fields are inherently unsteady, and capturing their temporal evolution requires the scalar field to be reconstructed at multiple time instants, further increasing the computational cost of reconstruction. Recently, implicit neural representations have emerged as an alternative to voxel-based tomographic reconstruction, representing volumetric information as continuous functions learned directly from image observations. In this work, a Neural Radiance Field (NeRF)-based reconstruction framework is developed to address these challenges and enable time-resolved volumetric reconstruction.

1.1 Research Gap

Despite major advances in plenoptic imaging for volumetric flow diagnostics, current single-camera MTV methodologies remain dependent on classical tomographic reconstruction algorithms to recover the 3D scalar field from a limited number of elemental images. Existing voxel-based reconstruction approaches are computationally expensive and sensitive to measurement noise. The FluidNeRF framework introduced by Kelly and Thurow [10, 12, 13] addressed these limitations by representing the scalar field as a continuous function, alleviating the discretization artifacts of voxel-based methods. However, the original FluidNeRF framework was formulated for wide-angle multi-camera configurations with perspective projection geometry. It is therefore not directly compatible with the narrow-angle orthographic geometry of the FIMic system. The framework also relies on Fourier positional encoding, which requires a large neural network and is computationally expensive [14]. Furthermore, previous applications to experimental plenoptic MTV data lacked synthetic datasets with known ground truth, preventing rigorous quantitative evaluation of reconstruction accuracy.

These limitations are particularly critical for time-resolved volumetric MTV, where the flow evolves continuously in space and time and multiple temporal states must be reconstructed efficiently. Therefore, there is a clear need for a quantitatively validated reconstruction framework capable of recovering time-resolved 3D scalar fields from single-camera plenoptic MTV measurements under limited-angle imaging conditions. Such a framework would enable practical volumetric velocimetry in confined near-wall environments while substantially reducing reconstruction cost.

1.2 Research Objectives

This work leverages recent advances in neural radiance field methods and plenoptic imaging to establish a new approach for three-dimensional, time-resolved volumetric reconstruction of scalar intensity fields in molecular tagging velocimetry. Classical tomo-

graphic reconstruction methods treat each temporal instant independently and therefore do not exploit the continuous spatiotemporal structure of the underlying flow field. Standard NeRF formulations present two limitations in this application. First, they are designed for perspective camera geometry, whereas the Fourier Integral Microscope uses orthographic projection. Second, they rely on Fourier positional encoding, which incurs substantial training cost for time-resolved reconstruction, where a separate scalar field must be recovered at each time step.

These limitations are addressed by extending FluidNeRF for MTV applications. Specifically, the proposed framework adapts the ray formulation to the FIMic orthographic geometry, incorporates time as a direct network input to enable time-resolved reconstruction of dynamic scalar fields, and replaces conventional Fourier positional encoding with multiresolution hash encoding to substantially reduce training time without sacrificing reconstruction accuracy. The framework is validated on synthetic datasets with known ground truth and evaluated on experimental plenoptic MTV data, establishing implicit neural representations as a viable and computationally efficient alternative to classical tomographic inversion for volumetric flow diagnostics. The primary objectives of this work are summarized as follows:

- Develop and implement a NeRF-based volumetric reconstruction framework for the Fourier Integral Microscope, for three-dimensional, time-resolved scalar field recovery from plenoptic MTV data.
- Establish the proposed framework as a computationally efficient alternative to conventional tomographic reconstruction methods for plenoptic-MTV.
- Establish multiresolution hash encoding as a viable alternative to Fourier positional encoding within the NeRF framework, enabling faster reconstruction without loss of accuracy.
- Quantitatively and qualitatively evaluate the reconstruction performance of the pro-

posed framework on synthetic and experimental plenoptic MTV datasets through comparison with ground-truth volumes.

1.3 Thesis Organization

The thesis is organized as follows: Chapter 2 presents the theoretical foundations and experimental background relevant to this work, including Molecular Tagging Velocimetry, Fourier Integral Microscope imaging systems, tomographic reconstruction methods, Neural Radiance Fields (NeRF), and encoding techniques for NeRF. Chapter 3 describes FluidNeRF formulation for MTV, synthetic data generation methodology, and multiresolution hash encoding implementation. Chapter 4 evaluates the proposed framework using synthetic datasets with known ground-truth volumes. Reconstruction accuracy, computational performance, convergence behavior, and the influence of key network and encoding hyperparameters are systematically investigated. Chapter 5 demonstrates the application of the proposed framework to experimental FIMic-MTV measurements, including both static and time-resolved volumetric reconstructions, and compares the experimental results with the synthetic benchmarks. Finally, Chapter 6 summarizes the major findings of this work, discusses its limitations, and outlines recommendations for future research.

Chapter 2

Background

This chapter establishes the technical foundation required to develop and justify the reconstruction method of this thesis. It first introduces molecular tagging velocimetry and its volumetric extension, then the Fourier Integral Microscopy architecture that enables single-camera volumetric acquisition. It then describes the limitations of classical tomographic reconstruction methods under the limited-angle, limited-view conditions imposed by FIMic, motivating the need for an alternative. The chapter closes by introducing neural radiance fields and the coordinate encoding techniques that together form the basis of the proposed reconstruction approach.

2.1 Molecular Tagging Velocimetry and Volumetric Flow Diagnostics

Molecular Tagging Velocimetry (MTV) is a non-intrusive flow visualization and velocimetry technique in which fluid velocity is determined from the displacement of optically tagged molecular tracers rather than seeded particles [7, 15]. In this technique, a pulsed laser excites molecular tracers within the working fluid to create tagged regions, and the displacement of these regions between successive images gives the local fluid velocity [8]. MTV has been demonstrated across turbulent boundary layers, reacting flows, and multi-parameter configurations recovering velocity [15], temperature [16, 17], and pressure [18].

One of the primary advantages of MTV arises from its use of molecular tracers [7]. Particle-based techniques, such as Particle Image Velocimetry (PIV), infer velocity from the displacement of seed particles [19]. This inference depends on the particles' ability to track the fluid motion, which degrades in flows with strong accelerations, density gradients, or complex internal geometries [7]. Furthermore, particle seeding introduces complications, including non-uniform seeding in stratified flows, unwanted nucleation during

phase change, and light contamination near solid boundaries [7]. These limitations are most severe in near-wall, high-shear regions, where uniform seeding is difficult to achieve and particle slip relative to the fluid introduces additional measurement error. MTV addresses these limitations by tagging the fluid directly rather than introducing a discrete particle phase. Tagged molecules carry negligible inertia and diffuse readily, allowing them to follow the local flow with high fidelity and to penetrate close to solid boundaries [8, 15]. This combination of negligible inertial lag, high diffusivity, and resistance to particle-related artifacts makes MTV well suited to near-wall, high-shear flow measurement [2].

The operating principle of MTV is illustrated in Figure 2.1. MTV is implemented through a two-pulse write-read sequence. During the write step, a pulsed laser first defines a tagged region within the flow through photobleaching, photodissociation, or direct excitation of a molecular species. This process renders the tagged region optically distinguishable from the surrounding untagged fluid. After a prescribed delay, a second pulse interrogates the same region which is known as the read step, recording the displaced or deformed tagged pattern. Because the write-read interval is known, the resulting displacement yields the local flow velocity directly [7].

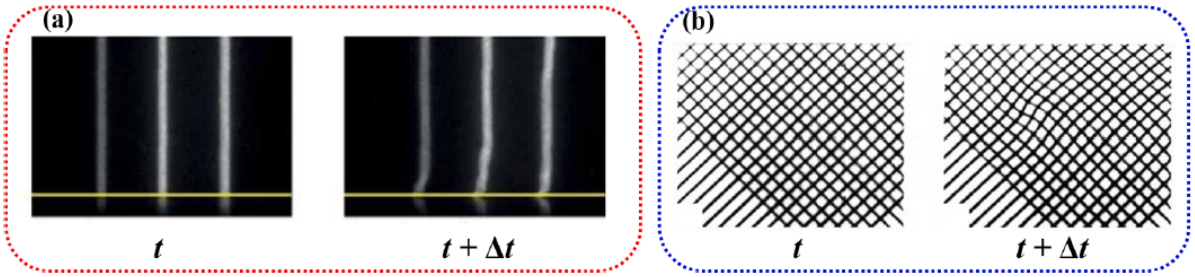


Figure 2.1: Examples of MTV tagging patterns used for velocity measurements, showing (a) the displacement of tagged lines and (b) the deformation of a tagged grid between times t and $t + \Delta t$. Adapted from [7].

The physical mechanism of tagging depends on the tracer species employed. Photobleaching irreversibly degrades a fluorescent dye, such as Rhodamine 6G, within the laser-illuminated region. The tagged region therefore appears dark against the bright

fluorescent background [2, 7]. Photodissociation instead generates a fluorescent or phosphorescent fragment from a parent molecule. Regardless of the tagging mechanism employed, the geometry of the tagging pattern determines how many velocity components can be recovered [20]. A single tagged line resolves only the velocity component normal to the line, giving a 1D-1C measurement. Intersecting beams, or a regular grid of beamlets, resolve two in-plane velocity components simultaneously, giving a planar 2D-2C measurement. Displacement between the undeformed and deformed images is determined via spatial cross-correlation. A source window centered on the tagged feature in the reference image is correlated against a larger interrogation window in the displaced image, and a polynomial fit to the correlation peak gives sub-pixel accuracy [7]. The choice of the write-read delay Δt governs a trade-off between dynamic range and accuracy. Increasing the delay enlarges the measurable displacement but allows greater molecular diffusion and signal decay, which widens the tagged feature and reduces the signal-to-noise ratio [8].

The implementation described above is inherently planar: the tagged pattern is confined to a single cross-sectional plane of the flow. Early MTV implementations were therefore limited to measuring velocity components within that plane [21]. Stereoscopic imaging was later introduced to recover the third, out-of-plane velocity component by viewing the tagged plane from two angles, but the measurement domain remained two-dimensional [7]. Many flows of practical interest, particularly near-wall turbulent flows, exhibit inherently 3D structure and require simultaneous measurement of all three velocity components throughout a finite volume.

To address this limitation, volumetric implementations of MTV were developed. Rather than measuring velocity within a single tagged plane, these approaches reconstruct a 3D scalar intensity field from multiple viewing angles, resolving the tagged structures throughout a volume [21]. Recently, the Plenoptic 3.0 architecture has enabled single-camera imaging for MTV, removing the optical-access and calibration difficulties associated with multi-camera arrangements [2, 9]. The resulting volumetric MTV measurement pro-

vides three-dimensional, three-component (3D-3C) velocity information by tracking the displacement of tagged beamlets between an undeformed reference volume and a deformed measurement volume [2]. The lateral displacement of each beamlet yields two velocity components directly, and the third component is recovered by applying the incompressible continuity equation [2]. Volumetric MTV therefore provides 3D instantaneous flow-structure information that planar implementations cannot capture.

2.2 Fourier Integral Microscopy

Fourier Integral Microscopy (FIMic), also referred to as Fourier light-field microscopy (FLFM), is a plenoptic 3.0 imaging architecture for single-camera volumetric microscopy and has demonstrated the highest spatial resolution among integral microscopy techniques [2]. It was recently applied to volumetric MTV by Huck et al. [2]. The motivation for FIMic arises from the limitations of conventional multi-camera volumetric imaging systems. 3D reconstruction requires multiple angular views of the measurement volume, which are traditionally acquired using several cameras distributed around the region of interest. Although such arrangements are well established in flow diagnostics, their implementation becomes difficult in near-wall and confined geometries, where optical access is limited and sensor placement around the measurement volume is restricted. In addition, multi-camera systems require precise alignment and volumetric calibration, both of which are susceptible to external perturbations such as facility vibrations [2]. FIMic provides a compact alternative by recording the angular information required for volumetric reconstruction on a single image sensor.

The optical architecture of the FIMic system is shown in Figure 2.2. The instrument consists of a microscope objective (MO), aperture stop (AS), tube lens (TL), field stop (FS), relay lens (RL), microlens array (MLA), and image sensor. The microscope objective and tube lens form the wide-field microscope portion of the system, while the relay lens and MLA introduce the plenoptic imaging architecture. The main optical parameters are

summarized in Table 2.1 [2].

Table 2.1: Parameters of the FIMic imaging system [2]

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

The defining feature of FIMic is the placement of the MLA at a plane conjugate to the Fourier plane, or aperture stop, of the infinity-corrected microscope objective. Since the Fourier plane is not always mechanically accessible, the relay system is used to conjugate the objective Fourier plane with the MLA. The image sensor is placed one microlens focal length downstream of the MLA. This arrangement causes the objective aperture to be subdivided into multiple angular views. The field stop limits the field of view of each elemental image and prevents overlap between elemental images on the sensor. As a result, the sensor records a grid of elemental images, where each elemental image is a complete view of the entire field of view from a distinct angular direction [2].

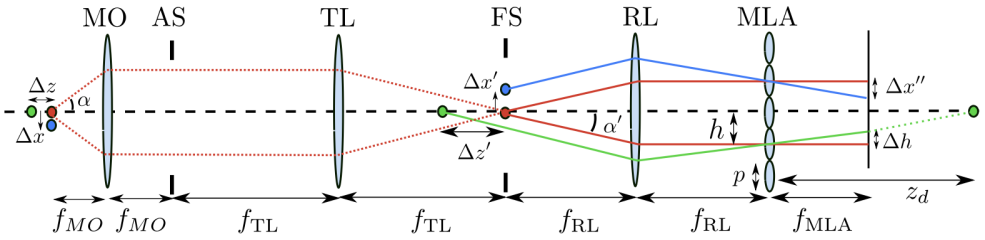


Figure 2.2: Schematic of the Fourier Integral Microscope (FIMic) optical layout [2]

The FIMic architecture differs from earlier plenoptic architectures in how the light field is sampled. In plenoptic 1.0, the MLA is positioned at the native image plane, and each elemental image corresponds to one spatial location observed from multiple angular directions. In contrast, FIMic, or plenoptic 3.0, places the MLA at the Fourier plane and forms elemental images of the entire scene. This sampling is reciprocal to plenoptic 1.0: instead of

emphasizing angular resolution at the cost of spatial resolution, FIMic provides high spatial resolution while preserving the angular diversity required for volumetric reconstruction [2].

The full sensor frame of Plenoptic 3.0 contains 23 full elemental images arranged in staggered rows. Each elemental image contains the full field of view, but each one corresponds to a different angular sample of the objective aperture. Therefore, the recorded sensor image contains both spatial information within each elemental image and angular information across elemental images. In the MTV implementation, these elemental images provide multiple angular views of the photobleached tagged beamlets, enabling reconstruction of the 3D intensity field from a single camera exposure.

2.2.1 Calibration

In the optical configuration of FIMic, the sensor frame contains 23 full elemental images arranged in staggered rows; however, only the central 19 elemental images are used by Huck et al. for volumetric reconstruction [2]. Therefore, calibration is performed for each of the 19 elemental views used in the reconstruction process. Accurate volumetric reconstruction requires a geometric calibration that maps object-space coordinates to pixel coordinates in each elemental image. The mapping is represented using a first-order affine camera model. For the n-th elemental image, a point in world coordinates (x_w, y_w, z_w) maps to pixel coordinates (x_c^n, y_c^n) according to

$$\begin{cases} x_c^n = x_w + z_w a_{n,x} + b_{n,x}, \\ y_c^n = y_w + z_w a_{n,y} + b_{n,y}. \end{cases} \tag{2.1}$$

The coefficients $a_{n,x}$ and $a_{n,y}$ represent the depth-dependent lateral displacement associated with the angular direction of the n-th elemental image, while $b_{n,x}$ and $b_{n,y}$ define the corresponding image offsets. This first-order affine mapping is consistent with the orthographic imaging geometry of FIMic and assumes negligible distortion [2]. The calibration therefore encodes the geometric configuration of the FIMic system by defining

how a 3D object-space point projects into each elemental image.

The affine coefficients are obtained by traversing a resolution target through the measurement volume [2]. At each axial position, the target center is detected in each of the 19 elemental images using cross-correlation. The detected image coordinates are then used to determine the affine coefficients through least-squares regression. The resulting mapping gives an average residual error below 0.1 pixels for all 19 views [2]. Because the calibration target cannot be placed directly inside the stagnation jet facility, the camera model is further refined under experimental conditions using volume self-calibration, giving an average reprojection error of 0.35 pixels and a maximum reprojection error below 0.8 pixels [2].

2.2.2 Ray Geometry

The ray geometry of FIMic follows from the object-space telecentricity of the microscope objective. In a microscope, telecentricity is achieved by placing the aperture stop at the back focal plane of the objective, which is also the Fourier plane. Because the MLA is placed at a plane conjugate to this Fourier plane, the elemental images inherit the telecentric imaging geometry of the microscope objective. Consequently, the elemental images are orthographic angular images rather than perspective projections [2].

Orthographic imaging means that the rays within a given elemental image are parallel in object space. The ray direction changes from one elemental image to another because each microlens samples a different portion of the objective aperture, but the rays within an individual elemental image remain parallel. This geometry has several important consequences. First, there is no parallax when the view is translated transversely. Second, the object does not shift laterally as it comes into focus. Third, the lateral magnification remains constant across the field of view. Finally, the point spread function is transversely shift-invariant, which is required for deconvolution-based volumetric reconstruction [2]. These properties justify the use of the affine camera model in Eq. 2.1 and distinguish

the FIMic ray geometry from the perspective camera model used in conventional pinhole-camera formulations.

2.3 Classical Tomographic Reconstruction and Its Limitations

Tomographic reconstruction is the process of recovering a three-dimensional volumetric distribution from a series of 2D projection measurements acquired from different viewing directions. The underlying problem is an inverse problem in which the unknown volumetric scalar field is estimated such that its projections are consistent with the measured images. Tomographic reconstruction has become a standard tool in volumetric flow diagnostics such as Tomographic Particle Image Velocimetry (Tomo-PIV) [22] and Molecular Tagging Velocimetry (MTV) [2]. In volumetric flow measurements, the reconstructed scalar field serves as the basis for extracting velocity information; therefore, the accuracy of the measured velocity field depends directly on the fidelity of the tomographic reconstruction [23].

Most classical tomographic reconstruction methods represent the measurement volume using a discrete 3D voxel grid. The reconstruction is then formulated as an optimization problem in which the voxel intensities are iteratively updated until the projections generated from the reconstructed volume agree with the experimentally measured projections. Over the past several decades, numerous iterative reconstruction algorithms have been developed for this purpose. Among the most widely used algebraic methods are the Algebraic Reconstruction Technique (ART) and the Multiplicative Algebraic Reconstruction Technique (MART) [24, 25, 26], along with SART [27, 28] and deconvolution-based approaches such as Richardson–Lucy deconvolution [29, 2]. These methods repeatedly update the voxel intensities by minimizing the discrepancy between measured and calculated projections while enforcing consistency among all available viewing directions. Due to their robustness and relative simplicity, these iterative algorithms have become the standard reconstruction techniques for many volumetric flow diagnostics.

Although these approaches have demonstrated good performance in applications where a large number of projection views are available, they remain fundamentally ill-posed inverse problems. The recovered solution is generally not unique and is highly sensitive to measurement noise and incomplete projection data [10, 30]. Furthermore, because the reconstruction is performed on a discrete voxel grid, the achievable spatial resolution is limited by the voxel discretization. Grid-based representations can introduce discretization artifacts, staircase effects along object boundaries, and increased memory requirements as the desired spatial resolution increases. Since the computational complexity and memory consumption scale directly with the total number of voxels, high-resolution volumetric reconstructions become computationally expensive, often requiring hundreds of iterative updates before convergence [10].

These limitations become considerably more severe for Fourier Integral Microscopy (FIMic). Unlike conventional tomographic systems that acquire many projections over a wide angular range, the FIMic system generates only a limited number of elemental images with relatively small angular separation between neighboring views. Consequently, the available projections span only a narrow angular aperture around the measurement volume, producing a limited-angle, limited-view reconstruction problem. Under these conditions, neighboring elemental images contain highly correlated information, reducing the amount of independent information available for reconstruction. As a result, conventional iterative algorithms become increasingly susceptible to reconstruction artifacts, noise amplification, and reduced depth resolution, leading to degraded reconstruction fidelity [13, 11].

2.4 Neural Radiance Fields

Neural Radiance Fields (NeRF), first introduced by the computer vision community [30], is a machine-learning framework for novel view synthesis from a set of two-dimensional projections. In the original formulation, a static scene is represented as a continuous implicit function parameterized by a multilayer perceptron (MLP). The network maps a

five-dimensional input, consisting of a three-dimensional spatial coordinate (x, y, z) and a two-dimensional viewing direction (θ, ϕ), to a volume density and a view-dependent radiance associated with that point. In this representation, the volumetric scene is not stored as a discrete voxel array. Instead, the scene is encoded through a coordinate-based neural function, allowing the field to be evaluated at arbitrary spatial locations. Images are synthesized from this representation through volume rendering: for each pixel, a ray is cast through the scene, the network is evaluated at sampled points along the ray, and the resulting densities and radiances are numerically integrated to produce the pixel value [30]. Training proceeds by minimizing the difference between rendered and measured images, so that the continuous field is optimized directly against the two-dimensional observations.

The use of NeRF for volumetric reconstruction is motivated by the ill-posed nature of tomographic inverse problems described in Section 2.3. Under limited-angle conditions, the measured projections do not uniquely determine the underlying scalar field, and voxel-based reconstructions are consequently sensitive to noise amplification, reconstruction artifacts, and grid-dependent discretization errors. NeRF addresses this problem from a different direction: rather than estimating intensities on a fixed voxel grid, the scalar field is represented as a continuous function whose projections are optimized to match the measured images.

The continuous implicit representation of NeRF avoids discretization artifacts and naturally regularizes the reconstruction through the network's inductive bias toward smooth solutions [13]. The framework is also flexible, supporting advanced ray models [31, 32, 33], alternative rendering strategies [14, 34, 35], and physics-based constraints [36, 37]. Time-resolved formulations further incorporate the measurement instant as a direct network input, allowing the spatiotemporal scalar field to be represented within a single unified model in which temporal continuity provides additional regularization [13, 35]. NeRF has recently been applied to fluid mechanics by Kelly and Thurow [10, 12, 13] through the FluidNeRF framework, demonstrating scalar-field tomography from multi-view camera arrays and

time-resolved reconstruction on experimental data. However, existing implementations rely on distributed camera arrays modeled as perspective projectors [10, 13], where rays converge toward a single focal center.

2.5 Encoding Techniques

NeRF represents a scene as a continuous volumetric function parameterized by a coordinate-based MLP. The inputs to this network are three-dimensional spatial coordinates in object space [30] and, in the case of dynamic scenes, a temporal coordinate [38, 13]. These raw coordinate values are inherently low-dimensional scalars that carry limited information about local spatial structure [30]. Neural networks trained on such inputs are subject to a well-documented spectral bias: gradient-based optimization preferentially fits low-frequency components of the target function first, and convergence to high-frequency structure is slow or fails entirely without intervention [39, 40]. This bias is particularly problematic in NeRF reconstruction, where accurate representation of fine geometric and appearance detail is required. Input encoding addresses this by mapping raw coordinates to a higher-dimensional vector representation before they reach the MLP. This expanded representation embeds multi-scale spatial information and provides the network with a richer description of the surrounding spatial structure at each query point [30]. This substantially improves the conditioning of the optimization problem and accelerates convergence to high-quality reconstructions [14, 40].

Two broad categories of input encoding have been developed for neural field methods. The first comprises fixed encodings, which map coordinates to a higher-dimensional representation using a predefined set of basis functions, the most widely adopted of which is positional encoding based on sinusoidal functions [30, 41]. The second comprises parametric encodings, which augment the network with additional trainable parameters arranged in an auxiliary spatial data structure such as a grid, octree, or hash table [41, 14]. Parametric encodings offload a substantial portion of the learning task from the MLP to the

data structure, enabling the use of smaller and more computationally efficient networks without sacrificing reconstruction quality [14]. The following subsections describe positional encoding and multiresolution hash encoding in detail, as both are directly relevant to the method developed in this thesis.

2.5.1 Positional Encoding

Positional encoding maps input coordinates to a higher-dimensional representation by evaluating a sequence of sinusoidal basis functions at geometrically increasing frequencies. For a scalar input x, the encoding takes the form

$$\text{enc}(x) = \left[\sin(2^0\pi x), \cos(2^0\pi x), \sin(2^1\pi x), \cos(2^1\pi x), \ldots, \sin(2^{L-1}\pi x), \cos(2^{L-1}\pi x)\right], \tag{2.2}$$

producing a $2L$-dimensional output vector [30]. The inclusion of multiple frequency levels allows the network to distinguish fine-scale differences between neighboring coordinates that would otherwise appear nearly identical as raw scalar inputs [40]. This scheme was originally introduced for position awareness in transformer attention mechanisms and was subsequently adopted by Mildenhall et al. to encode the five-dimensional spatio-directional inputs of the NeRF formulation, where it substantially improved the reconstruction of high-frequency scene detail relative to unencoded inputs [30].

The principal advantage of positional encoding is its simplicity and analytical differentiability. The encoding used in NeRF introduces no trainable parameters and imposes no memory overhead beyond the expansion in input dimensionality. Because the encoding is a smooth, closed-form function of the input coordinate, derivatives of the network output with respect to the input are well-defined everywhere [30, 41]. The central limitation of positional encoding is that representing high-frequency structure requires either increasing L or enlarging the MLP, both of which increase computational cost [14, 41]. In the NeRF

setting, a standard positional encoding combined with an MLP of 8 hidden layers and 256 neurons per layer requires on the order of hours to converge to high-quality reconstructions [30].

2.5.2 Multiresolution Hash Encoding

Multiresolution hash encoding, introduced by Müller et al. [14], replaces the fixed analytic basis of positional encoding with a set of trainable feature vectors organized in a hash table. The encoding uses L independent resolution levels, where each level contains a hash table of up to T feature vectors of dimension F. The resolution at level l is defined by a geometric progression between a minimum resolution N_{min} and a maximum resolution N_{max}:

$$N_l = \left\lfloor N_{\text{min}}\, b^{\,l} \right\rfloor, \qquad b = \exp\!\left(\frac{\ln N_{\text{max}} - \ln N_{\text{min}}}{L-1}\right). \tag{2.3}$$

For a query coordinate $\mathbf{x} \in \mathbb{R}^d$, the encoding is performed independently at each resolution level. First, $\mathbf{x}$ is scaled by the grid resolution of the corresponding level, and the surrounding voxel vertices are identified. Each vertex is then mapped to a hash-table entry using a spatial hash function,

$$h(\mathbf{x}) = \left(\bigoplus_{i=1}^{d} x_i \pi_i\right) \bmod T, \tag{2.4}$$

where $\oplus$ denotes bitwise XOR and π_i are large prime constants. The F-dimensional feature vectors associated with the 2^d surrounding vertices are then d-linearly interpolated according to the relative position of $\mathbf{x}$ within the voxel. The interpolated features from all L levels are concatenated, together with any auxiliary input $\xi \in \mathbb{R}^E$, to produce the encoded input $\mathbf{y} \in \mathbb{R}^{LF+E}$ to the MLP [14]. The primary advantage of multiresolution hash encoding over positional encoding is training efficiency. Feature-vector gradients are sparse because only the 2^d vertices surrounding each query point are updated for each

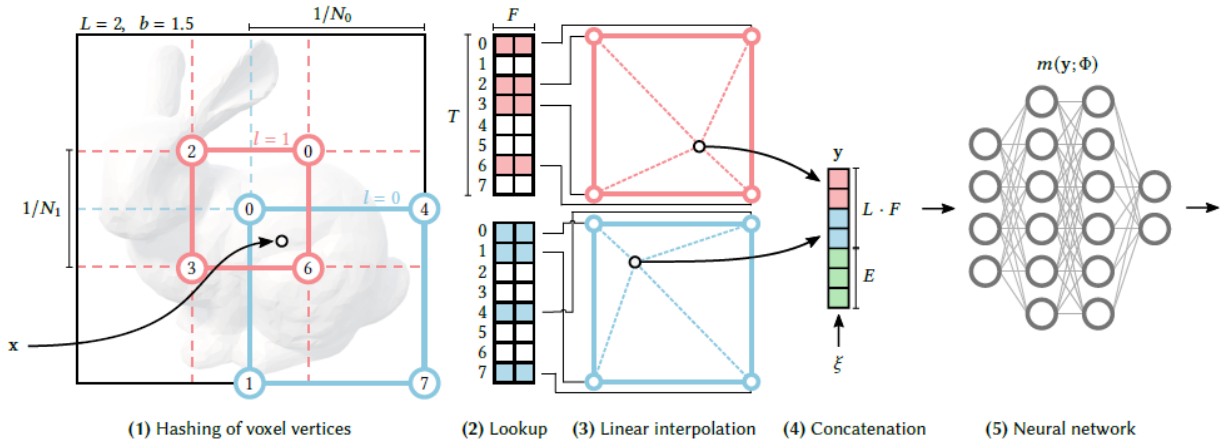


Figure 2.3: Schematic of the multiresolution hash encoding pipeline, showing the mapping of an input coordinate to trainable feature vectors across resolution levels. Adapted from [14].

sample. This allows the reconstruction to use a substantially smaller downstream MLP without sacrificing representation quality [41]. Müller et al. [14] demonstrated that using $L = 16$, $F = 2$, $T \in [2^{14}, 2^{24}]$, and an MLP with two hidden layers of 64 neurons can match or exceed the reconstruction accuracy of standard positional-encoding NeRF while requiring only a fraction of the training time. The multiresolution structure also provides implicit spatial adaptivity. Coarse levels capture low-frequency structure with fewer collisions, while fine levels represent higher-frequency detail, with hash collisions handled implicitly by the MLP through gradient averaging [41, 14]. The main limitation of hash encoding is that the piecewise linear interpolation between discrete grid vertices can produce derivatives that are discontinuous at voxel boundaries. This requires careful treatment in applications where derivatives of the network output with respect to input coordinates are used directly [41]. In addition, hash collisions at fine resolution levels can introduce low-amplitude spatial noise in the learned function, which may persist even with additional training [14].

Chapter 3

Methodology

This chapter describes the reconstruction framework developed to recover the volumetric MTV scalar field from FIMic elemental images. It first presents the overall pipeline and then details each component: the affine camera and ray model, the neural scalar-field representation, the multiresolution hash encoding, the volume rendering step, and the training procedure. It then describes the generation of the synthetic dataset used to provide known ground-truth volumes, including the flow fields used to produce time-resolved test cases. The chapter concludes with the metrics used to evaluate reconstruction fidelity.

3.1 Overview of the Reconstruction Framework

The reconstruction framework follows the NeRF-based FluidNeRF pipeline [13], with modifications to accommodate the orthographic projection geometry of the Plenoptic 3.0 FIMic system [2]. The framework uses the 19 elemental images acquired by the FIMic system at each measurement instant as projection constraints and reconstructs a continuous scalar intensity field, $E(x, y, z, t)$, corresponding to the tagged MTV beamlet pattern. As shown in Figure 3.1, the pipeline consists of four main components: ray generation from the affine calibration model, spatial encoding of sample-point coordinates, MLP-based scalar intensity prediction, and volume rendering. The three-dimensional sample coordinates (x, y, z) are encoded using multiresolution hash encoding, while the normalized measurement time t is concatenated with the encoded spatial features before being passed to the MLP. The network predicts the scalar intensity field $E(x, y, z, t)$, which is rendered along the calibrated orthographic rays to generate synthetic elemental-image projections. Training is performed by minimizing the difference between the rendered projections and the measured FIMic elemental images.

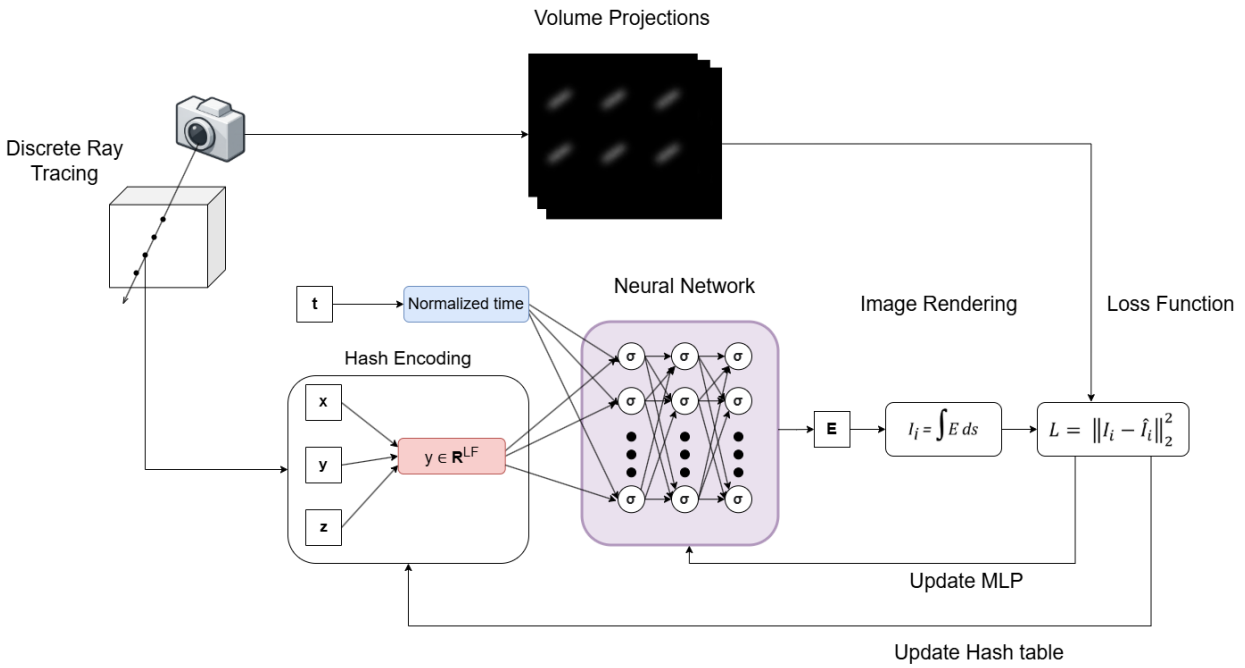


Figure 3.1: Schematic of the FluidNeRF reconstruction framework for FIMic-based MTV. The pipeline consists of ray generation from the affine calibration model, multiresolution hash encoding, MLP-based scalar field prediction, volume rendering, and projection-based loss minimization.

3.1.1 Camera and Ray Model

The dataset used in this work was obtained from Huck et al. [2], where near-wall volumetric MTV was performed using a Fourier Integral Microscope (FIMic), or Plenoptic 3.0 system. The FIMic dataset provided 19 elemental images that contained clear projections of all six beamlets, and these images were used in the reconstruction. These 19 elemental images form the projection data used to adapt the FluidNeRF reconstruction framework to the orthographic imaging geometry of the FIMic system.

The key modification is in the camera model. In the original FluidNeRF formulation, the imaging system is represented using perspective cameras. In the present framework, this projection model is replaced with the calibrated affine model of the FIMic system. Each elemental image is treated as an independent orthographic camera, with its projection geometry defined by the affine calibration coefficients from Huck et al. [2]. The imaging model

assumes geometric optics and therefore neglects wave-optical effects such as diffraction and interference. For a three-dimensional sample point, the affine model determines its corresponding pixel location in each of the 19 elemental images. This allows the volume rendering step to use the actual FIMic projection geometry rather than a perspective camera approximation.

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

$$\mathbf{r}(s) = \mathbf{o}_{n,i} + s\,\mathbf{d}_n, \tag{3.1}$$

where $\mathbf{o}_{n,i}$ is the pixel-dependent ray origin, $\mathbf{d}_n$ is the view-dependent ray direction obtained from the affine calibration coefficients, and $s \in [s_{\text{near}}, s_{\text{far}}]$ parameterizes depth along the ray. During reconstruction, sample points along these calibrated rays are evaluated by the neural field, rendered into synthetic elemental-image intensities, and compared with the measured FIMic elemental images. In this way, the experimental dataset and calibration from Huck et al. are used not only as image inputs, but also to define the orthographic ray model required for FIMic-based NeRF reconstruction.

3.1.2 Neural Representation

The tagged MTV beamlet pattern is represented as a continuous spatiotemporal scalar intensity field, $E(x, y, z, t)$, parameterized by a multilayer perceptron (MLP). This representation follows the FluidNeRF formulation, in which the unknown volumetric scalar distribution is modeled by a neural function rather than by a fixed voxel grid. The purpose of this representation is to reconstruct the MTV intensity field continuously within the measurement volume while preserving the dependence of the tagged pattern on measurement time.

For each sample point, the network receives spatial information associated with the coordinate (x, y, z) and the corresponding measurement instant t. The spatial coordinates are first converted into an encoded feature vector, as described in the following section.

The normalized time value t is then concatenated directly with the spatial feature vector before being passed to the MLP. The network output is the scalar intensity value $E(x, y, z, t)$ at the queried spatial location and time instant.

Including time as a direct network input allows the scalar field to be represented as a continuous function of both space and time. This is important for time-resolved MTV reconstruction because the tagged beamlet pattern evolves between measurement instants due to flow-induced displacement and deformation. Instead of reconstructing each time instance using a separate network, the same MLP represents the scalar field across multiple time steps. This shared spatiotemporal representation promotes consistency between reconstructed time instances and provides a unified model for the evolving MTV signal.

The MLP architecture used in this work consists of two hidden layers with 128 neurons per layer and ReLU activation functions. This architecture is substantially smaller than the Fourier-encoded FluidNeRF implementation used for comparison, where a larger network is required to obtain sufficient representational capacity. In the present framework, the use of an encoded spatial feature representation allows the MLP to remain compact while still representing the three-dimensional beamlet structure. The network architecture and training hyperparameters used in this work are summarized in Table 3.1. These values define the baseline configuration; the influence of individual network and training parameters on reconstruction quality is investigated in Chapter 4.

Table 3.1: Network and training hyperparameters.

Parameter	**Value**
Number of hidden layers	2
Hidden units per layer	128
Activation function	ReLU
Samples per ray	64
Rays per iteration	2048
Learning rate	1×10^{-4}, exponential decay

3.1.3 Hash Encoding

The multiresolution hash encoding used in this framework is implemented using the `tiny-cuda-nn` framework [42]. The original FluidNeRF framework employed sinusoidal positional encoding to map the three-dimensional spatial coordinates into a higher-dimensional feature space before passing them to the MLP. In the present framework, this frequency-based positional encoding is replaced with the multiresolution hash encoding proposed by Müller et al. [14].

The hash encoder is applied only to the spatial coordinates (x, y, z) of each sample point. The encoding maps the spatial coordinate to a concatenated feature vector, where L is the number of hash levels and F is the number of features stored per level. The trainable feature vectors are organized within a multiresolution spatial hash table and are optimized jointly with the MLP weights during training.

For each queried sample point, the spatial coordinate is evaluated at multiple resolution levels. At each level, the coordinate is associated with the surrounding grid vertices, and the corresponding trainable feature vectors are retrieved from the hash table. These features are interpolated according to the relative position of the sample point within the grid cell. The interpolated features from all hash levels are then concatenated to form the encoded spatial feature vector that is passed to the MLP.

The normalized time input t is not encoded using the hash encoder. Instead, time is provided directly to the MLP as a separate input by concatenating the normalized time value with the encoded spatial feature vector. Thus, the network input consists of the hash-encoded spatial representation and the measurement time, while the network output is the scalar intensity field $E(x, y, z, t)$.

3.1.4 Volume Rendering

Volume rendering is used to project the predicted scalar intensity field onto the elemental image plane. For each calibrated orthographic ray, the MLP is evaluated at sampled locations along the ray, and the predicted scalar intensities are accumulated through the measurement volume to obtain the rendered pixel intensity.

For a ray associated with pixel i in elemental image n, For a ray associated with pixel i in elemental image n, defined by Eq. 3.1, the rendered intensity is computed as

$$\hat{I}_i = \int_{s_{\text{near}}}^{s_{\text{far}}} E(\mathbf{r}(s), t)\, ds, \tag{3.2}$$

where $E(\mathbf{r}(s), t)$ is the scalar intensity predicted by the MLP at the sampled location along the ray at measurement time t.

In the implementation, this continuous integral is approximated numerically using $N_S = 128$ sample points along each ray. The scalar intensity values predicted at these sample points are accumulated to obtain the rendered elemental-image intensity $\hat{I}_i$. This quadrature makes two simplifying assumptions. First, each ray is treated as an infinitesimally thin line, so the cross-section of the ray is assumed small and its divergence through the volume is neglected. Second, the scalar intensity is assumed piecewise uniform between adjacent sample points, which allows the line integral to be evaluated by discrete summation and corresponds to a piecewise-constant quadrature rather than a higher-order scheme. These assumptions are consistent with the geometric optics ray model described

in Section 3.1.1. These rendered intensities provide the synthetic projections of the learned scalar field and are used in the subsequent network-training step.

3.1.5 Network Training

The network is trained by minimizing the discrepancy between the rendered and measured elemental-image intensities. For each sampled ray, the measured pixel intensity I_i is compared with the corresponding rendered intensity $\hat{I}_i$ obtained from volume integration along the calibrated orthographic ray. The loss function is defined as

$$\mathcal{L} = \sum_i \left(I_i - \hat{I}_i\right)^2, \tag{3.3}$$

where I_i is the measured pixel intensity and $\hat{I}_i$ is the corresponding rendered intensity.

Optimization is performed using the Adam optimizer, which jointly updates the MLP weights and the trainable hash-table parameters at each iteration. The network is trained for 4000 iterations on an NVIDIA RTX 5040 GPU. At each iteration, 2048 rays are sampled, and 64 sample points are evaluated along each ray during volume rendering. The learning rate is initialized as 1×10^{-4} and is reduced using exponential decay. Both time instants are included during training, allowing a single network to represent the scalar intensity field $E(x, y, z, t)$ across multiple measurement times.

3.2 Synthetic Data Generation

A synthetic dataset is generated using the calibrated affine projection model of the FIMic system. The purpose of the synthetic dataset is to provide known ground-truth volumes for quantitative evaluation of reconstruction fidelity. Since the experimental MTV measurements do not provide direct volumetric ground truth, the synthetic data are generated using the same projection geometry as the FIMic system. This allows the reconstructed NeRF volume to be compared against a known reference volume.

The synthetic measurement volume consists of six MTV beamlets arranged in a 3×2 configuration. The beamlets have a center-to-center spacing of 500 μm, a diameter of 90 μm, and an axial length of 1800 μm, consistent with the experimental configuration. These beamlet parameters define the initial scalar intensity distribution used to generate both the synthetic elemental images and the corresponding ground-truth volumes. The main parameters of the synthetic data generation procedure are summarized in Table 3.2.

Table 3.2: Synthetic plenoptic image generation parameters.

Parameter	**Value**
Beamlet diameter	90 μm
Beamlet length	1800 μm
Beamlet spacing	500 μm
Total sample points per frame	360,000 (6 beamlets $\times$ 60,000)
Sensor resolution	1800×2400 px
Image bit depth	16 bit

3.2.1 Forward Rendering for the Synthetic Dataset

Synthetic elemental images are generated by forward projecting the prescribed beamlet distribution through the calibrated affine projection model of the FIMic system. Sample points are first distributed within the six prescribed MTV beamlets in object space. The transverse distribution of each beamlet is modeled using a Gaussian intensity distribution function, while the axial distribution is taken to be uniform over the beamlet length.

Each sample point $\mathbf{x}_p = (x_p, y_p, z_p)$ is then projected onto the sensor plane of each elemental image using the calibrated affine model, and the intensity contributions from all projected sample points are accumulated to form the synthetic elemental image for each view. A representative example of the generated synthetic plenoptic images alongside the corresponding experimental images is shown in Figure 3.2.

The number of sample points per beamlet is determined through an iterative con-

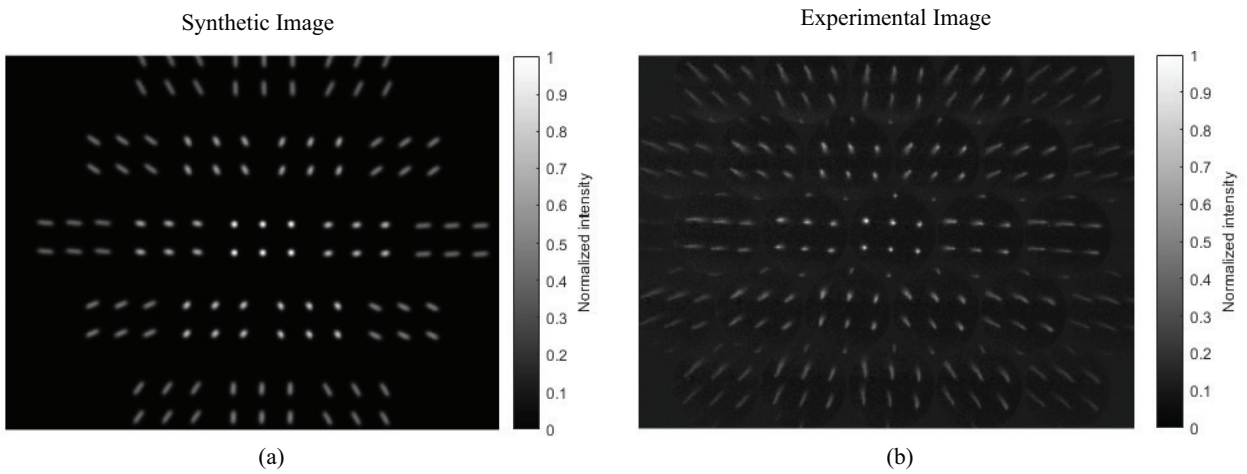


Figure 3.2: (a) Generated synthetic image and (b) experimental image of tagged MTV beamlets for the Fourier Integral Microscopy (FIMic) system.

vergence study. Points are added in batches of $10{,}000$, and the normalized RMS change in the projected sensor image is monitored between successive batches. Convergence is declared when this change falls below 0.5%, yielding $60{,}000$ sample points per beamlet and $360{,}000$ total sample points per time instant. Generating one set of 19 synthetic elemental images requires approximately 15 s on an NVIDIA GeForce RTX 5040 GPU. Further implementation details are provided in Appendix 1.

3.2.2 Flow Field Parameterization

A second time instant was required to evaluate the applicability of the reconstruction framework to time-resolved scalar-field recovery. For this purpose, two synthetic flow cases were generated in addition to the initial beamlet configuration at $t = 0\ s$: (1) an analytical fully developed Poiseuille flow between two parallel plates at $t = 0.1\ s$ and (2) a turbulent channel flow obtained from the Johns Hopkins Turbulence Database (JHTDB) at $Re_\tau = 5200$ [43]. The analytical case provides a simple, well-defined displacement field for controlled validation, whereas the JHTDB case represents a realistic turbulent flow with complex spatial variations.

For the Poiseuille flow case, the sample points from the initial beamlet configuration

at $t = 0$ s were advected to $t = 0.1$ s through a fully developed laminar channel flow. The channel height was set to $H = 1800$ μm, and the peak centerline velocity was set to $u_{\max} = 0.4$ mm/s. The streamwise velocity profile is given by

$$u(y) = u_{\max}\left[1 - \left(\frac{2y}{H}\right)^2\right], \tag{3.4}$$

where y is the wall-normal coordinate measured from the channel centerline.

The second synthetic case was generated using velocity data from the Johns Hopkins Turbulence Database (JHTDB) channel flow dataset. The `channel5200` dataset, which contains a direct numerical simulation (DNS) of fully developed turbulent channel flow at $Re_\tau = 5200$, was used [43]. The sample points from the initial beamlet configuration were advected using the turbulent velocity field obtained from the DNS data to generate the time-evolved dataset.

For velocity lookup, the beamlet volume was mapped to the lower ten percent of an 18 mm channel, corresponding to a channel half-height of $\delta = 9$ mm. The extracted DNS velocity field was converted to physical units and used to advect the particles using a fourth-order Runge–Kutta scheme with a timestep of $\Delta t = 0.00325$ convective units over a total advection time of $t = 0.026$ convective units.

Table 3.3 summarizes the turbulent length scales of the DNS dataset alongside the beamlet geometry, expressed both in viscous units and in the physical units of the present mapping. The measurement slab extends from the wall to 1.8 mm and therefore spans the viscous sublayer, the buffer layer, and the lower logarithmic region. The beamlet diameter of 90 μm exceeds the combined thickness of the viscous sublayer and buffer layer, and is more than an order of magnitude larger than the Kolmogorov scale in the same region. The tagged beamlets therefore sample the displacement produced by the energy-containing near-wall motions rather than the dissipative scales.

This synthetic turbulent case therefore evaluates the reconstruction framework

Table 3.3: Turbulent length scales of the JHTDB channel 5200 dataset and the corresponding beamlet geometry, expressed in viscous units and in physical units for the present mapping ($\delta = 9$ mm, $\delta_\nu = 1.74$ μm).

Quantity	Viscous units	Physical units
Viscous length scale, δ_ν	1	1.74 μm
Kolmogorov scale near the wall, η	~ 2	~ 3.5 μm
Viscous sublayer	$y^+ < 5$	< 8.7 μm
Buffer layer	$5 < y^+ < 30$	8.7–52 μm
Logarithmic region	$30 < y^+ < 1037$	52 μm–1.8 mm

under a realistic near-wall turbulent displacement field at the spatial scales accessible to the measurement system.

The displaced sample positions for both cases were then projected onto the sensor plane using the same calibrated affine projection model applied at the initial time instant. This produced synthetic elemental images of the displaced beamlet configuration for the Poiseuille and turbulent channel flow cases.

3.2.3 Ground Truth Volume

The ground-truth volume is constructed analytically from the prescribed Gaussian beamlet definitions. This volume provides a known reference scalar field against which the NeRF reconstruction can be quantitatively evaluated. Since the synthetic elemental images are generated from a prescribed beamlet distribution, the corresponding three-dimensional intensity field is known independently of the reconstruction process.

The ground truth is defined on a $256 \times 256 \times 256$ voxel grid spanning the physical extent of the synthetic measurement volume. At each voxel location, the scalar intensity is evaluated from the analytical beamlet definition. The transverse intensity distribution is defined by the Gaussian beamlet model, while the axial extent follows the prescribed beamlet length. This produces a volumetric reference field consistent with the synthetic beamlet geometry used during forward rendering.

Ground-truth volumes are generated at both $t = 0$ s and $t = 0.1$ s. The first volume corresponds to the undeformed beamlet configuration, while the second corresponds to

the Poiseuille-flow-displaced beamlet configuration. These two reference volumes are used to evaluate the reconstruction fidelity of the time-resolved NeRF framework through quantitative metrics and visual comparison of the recovered beamlet structure.

3.3 Evaluation Metrics

The performance of the proposed reconstruction framework was evaluated quantitatively using three metrics: the Structural Similarity Index (SSIM), the Peak Signal-to-Noise Ratio (PSNR), and the Normalized Root Mean Squared Error (NRMSE). These metrics are in standard use for assessing image and volumetric reconstruction quality [30, 14, 13]. The evaluation was performed during training, by comparing the rendered images with the corresponding views from the dataset, and after training, by comparing the reconstructed 3D scalar field with the ground-truth volume. These metrics were required because PSNR and NRMSE quantify the elementwise error while SSIM quantifies the preservation of local structure, and both are necessary to establish that the reconstruction is accurate.

The Structural Similarity Index compares two fields over local windows in terms of luminance, contrast, and structure. It evaluates the spatial arrangement of intensity rather than its absolute magnitude and therefore indicates directly whether the beamlet structure is preserved. The values of SSIM range from 0 to 1, with unity denoting perfect structural agreement. The Normalized Root Mean Squared Error quantifies the deviation between the reconstructed and reference intensities. Normalization by the intensity range of the reference renders the metric dimensionless and therefore comparable across datasets of differing intensity scale. The Peak Signal-to-Noise Ratio expresses the same deviation relative to the dynamic range of the signal on a logarithmic scale, and higher values denote lower reconstruction error. The three metrics above require a ground-truth field and were therefore applied only to the synthetic datasets.

Qualitative evaluation was performed through slice-wise intensity comparison, line profile comparison, and measurement of the Full Width at Half Maximum. The FWHM

was evaluated by extracting one-dimensional intensity profiles through the center of one reconstructed beamlet at multiple depth locations. It corresponds to the width of the profile at 50% of its peak intensity and provides a direct measure of the reconstructed beamlet diameter.

Chapter 4

Synthetic Data Results

This chapter evaluates the proposed reconstruction framework on synthetic data with known ground truth. It first presents a systematic hyperparameter investigation that identifies the encoding, network, and view-count settings giving the best balance between reconstruction accuracy and computational cost. The optimized configuration is then applied to the synthetic Poiseuille flow case, and the reconstructed volumes are assessed qualitatively and quantitatively against the analytical solution. The chapter then compares the multiresolution hash encoding with the original Fourier-encoded framework, and closes by recovering the velocity field for both the Poiseuille and turbulent channel flow cases.

4.1 Hyperparameter Investigation

Hyperparameter optimization plays an important role in neural radiance field (NeRF) reconstruction because it directly influences both the representation capacity of the network and the computational cost of training [14]. The performance of the multiresolution hash encoding depends on the selection of encoding hyperparameters, including the number of encoding levels, feature dimensions, spatial resolution hierarchy, and hash table size, which together determine the expressive capability of the encoding and its computational efficiency [14]. Similarly, the network architecture, and number of input observations influence the reconstruction quality and computational complexity of the optimization process.

This section presents a systematic investigation of the hyperparameters used in FluidNeRF for the reconstruction of the synthetic Poiseuille flow dataset. The performance of the reconstructed three-dimensional volumes is evaluated using the structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), normalized root mean square error (NRMSE), and total training time, with all metrics computed relative to the corresponding

ground truth volumes.

The investigation is divided into four parts. First, the influence of the multiresolution hash encoding parameters is examined. Next, the effects of the network architecture are evaluated. The influence of the number of elemental views is then investigated to determine the minimum number of observations required for accurate reconstruction under the limited-angle imaging configuration. Finally, the findings from the individual studies are combined to identify an optimized set of hyperparameters that provides the best balance between reconstruction accuracy and computational efficiency.

4.1.1 Encoding Parameters

An initial hyperparameter investigation of the encoding parameters was performed using the synthetic Poiseuille flow dataset at $t = 0.1$, where the beamlet displacement provides sufficient deformation to evaluate the reconstruction quality under nontrivial flow conditions. The baseline configuration consisted of a hash table size of $T = 2^{19}$, $L = 16$ encoding levels, $F = 2$ features per level, a coarse resolution of 16, a fine resolution of 512, 2048 rays per elemental image at each training iteration, 128 sample points evaluated along each ray, a network depth of 2 layers, and a network width of 128 neurons. To isolate the influence of each hyperparameter, only one parameter was varied at a time from the baseline configuration while all remaining hyperparameters were kept fixed. The resulting reconstruction quality and computational cost are summarized in Table 4.1.

Overall, varying the individual hyperparameters produced only minor changes in SSIM, PSNR, and NRMSE, indicating that the reconstruction quality is relatively insensitive to small adjustments of the multiresolution hash encoding parameters over the ranges investigated. This behavior is likely attributed to the characteristics of the synthetic Poiseuille flow dataset, which consists of sparse beamlet structures with relatively simple spatial features. Consequently, the baseline configuration already provides sufficient representational capacity to accurately reconstruct the three-dimensional scalar field, and further increasing

the encoding capacity yields only marginal improvements in reconstruction quality. In contrast, the total training time exhibited considerably larger variations, particularly for configurations with increased encoding capacity. These observations suggest that, for the sparse synthetic dataset considered in this work, the primary benefit of hyperparameter optimization lies in reducing computational cost while maintaining essentially unchanged reconstruction accuracy.

Table 4.1: Effect of selected hyperparameters on synthetic data reconstruction quality and training time.

Parameter	Value	SSIM	PSNR (dB)	NRMSE	Time (s)
Features	1	0.9612	25.05	0.0559	353.6
	2	0.9614	25.06	0.0559	387.4
	4	0.9611	25.05	0.0559	767.7
	8	0.9615	25.05	0.0559	1359.7
Levels	4	0.9603	25.01	0.0562	280.4
	8	0.9611	25.03	0.0560	291.7
	12	0.9618	25.09	0.0557	326.8
	16	0.9620	25.04	0.0559	443.9
Coarse Resolution	4	0.9612	25.04	0.0560	510.3
	8	0.9609	25.01	0.0562	400.6
	16	0.9620	25.04	0.0559	443.9
	32	0.9612	25.06	0.0558	387.1
	64	0.9606	25.13	0.0554	419.2
	128	0.9492	24.95	0.0566	428.0
Fine Resolution	128	0.9618	25.08	0.0557	383.2
	256	0.9621	25.06	0.0558	400.9
	512	0.9620	25.04	0.0559	443.9
	1024	0.9619	25.05	0.0559	415.7
	2048	0.9616	25.05	0.0559	425.4
Table Size	2^{18}	0.9614	25.05	0.0559	342.1
	2^{19}	0.9620	25.04	0.0559	443.9
	2^{20}	0.9614	25.05	0.0559	478.3
	2^{21}	0.9615	25.05	0.0559	641.9
	2^{22}	0.9617	25.09	0.0556	1080.7

While Table 4.1 showed that varying the encoding parameters individually had only a negligible effect on reconstruction quality, noticeable differences were observed in computational cost. One of the primary objectives of this work is to optimize the network

to reduce training time without sacrificing reconstruction accuracy. Since the number of encoding levels, L, and the number of features per level, F, together determine the capacity of the multiresolution hash encoding, these two hyperparameters are inherently coupled. Increasing either parameter increases the total number of trainable encoding parameters, resulting in greater representational capacity but also higher computational cost. For this reason, Müller *et al.* [14] analyzed L and F jointly while maintaining an approximately constant encoding size. They identified $F = 2$ and $L = 16$ as a favorable optimal configuration that balanced reconstruction quality and computational performance across a variety of complex scenes.

However, the synthetic MTV dataset used in this work is considerably simpler than the datasets considered in Instant-NGP. Instead of reconstructing complex natural scenes, the network is trained on sparse beamlet structures with relatively simple spatial features. Consequently, the default Instant-NGP configuration may provide more encoding capacity than necessary. Therefore, the coupled effects of L and F were investigated specifically for the present application. The number of features per level was varied as $F = \{1, 2, 4, 8\}$, while the number of encoding levels was varied as $L = \{4, 8, 12, 16, 20\}$. The resulting reconstruction quality and training time are presented in Figure 4.1.

Figures 4.1(a)–(c) illustrate the combined influence of the number of encoding levels and features per level on the reconstruction quality. For all feature dimensions, increasing the number of encoding levels from 4 to 12 generally improves the reconstruction quality. SSIM increases and reaches its maximum near 12 levels for all feature configurations, while PSNR exhibits a similar trend with only minor variations between configurations. Correspondingly, NRMSE decreases as the number of levels increases from 4 to 12, indicating improved agreement with the ground truth volume. Beyond 12 levels, however, the reconstruction metrics either plateau or exhibit slight degradation. For example, the SSIM obtained with $F = 4$ and $F = 8$ decreases slightly when the number of levels is increased from 12 to 16, while the $F = 1$ configuration continues to degrade at 20 levels.

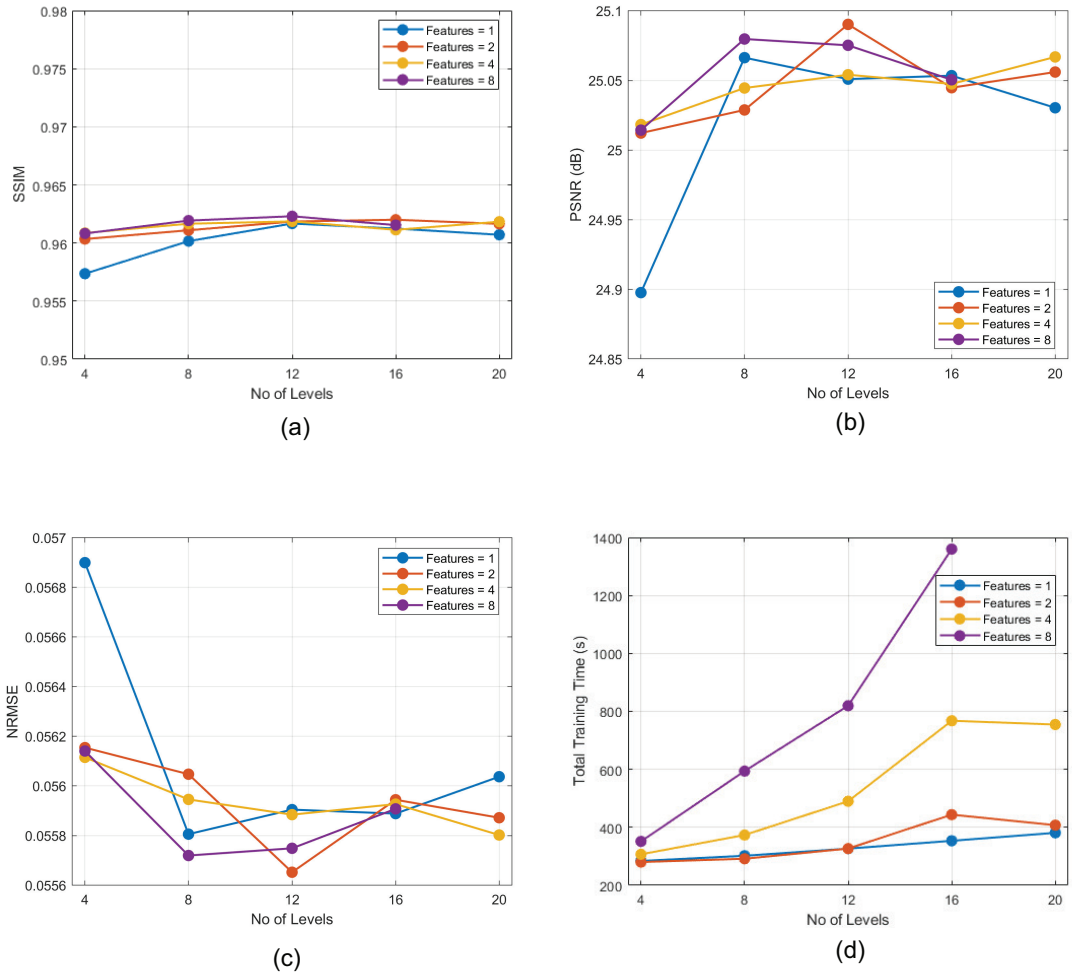


Figure 4.1: Effect of the number of encoding levels and features per level on synthetic data reconstruction. Reconstruction quality is evaluated using (a) SSIM, (b) PSNR, (c) NRMSE, and computational efficiency is assessed using (d) total training time.

Similar behavior is observed in the PSNR and NRMSE results, suggesting that additional encoding levels provide little benefit once sufficient multiresolution detail has been captured.

The influence of the number of features per level is also relatively limited. At lower encoding levels ($L = 4$ and $L = 8$), increasing the number of features from 1 to 2 or 4 noticeably improves the reconstruction quality, indicating that a single feature per level is insufficient to accurately represent the synthetic beamlets. However, once the number of levels reaches 12, the performance of the $F = 2$, $F = 4$, and $F = 8$ configurations becomes nearly identical across all three reconstruction metrics. Increasing the number of features

beyond two therefore provides negligible improvement in reconstruction quality, suggesting that the additional encoding capacity is largely redundant for the sparse synthetic MTV dataset.

In contrast, Figure 4.1(d) shows that the computational cost increases significantly with both hyperparameters. Training time remains relatively low for $F = 1$ and $F = 2$, but rises rapidly as either the number of encoding levels or features increases. The largest configuration ($F = 8$, $L = 16$) requires more than three times the training time of the smaller configurations while providing almost identical reconstruction quality. This indicates that the additional encoding capacity primarily increases computational cost rather than improving reconstruction accuracy.

Based on these results, $F = 2$ features per level and $L = 12$ encoding levels were selected for all subsequent experiments. This configuration achieves reconstruction quality comparable to the larger encoding configurations while significantly reducing training time. It therefore provides the best balance between reconstruction accuracy and computational efficiency for the sparse synthetic MTV dataset considered in this study.

Following the optimization of the number of encoding levels and features per level, the coarse and fine resolution parameters of the multiresolution hash encoding were investigated. These parameters define the range of spatial resolutions represented by the hash encoding. The base resolution determines the coarsest grid used to capture large-scale spatial variations, whereas the finest resolution specifies the highest-resolution grid responsible for representing fine geometric details. Müller et al. [14] recommended a base resolution of 16 together with a finest resolution between 512 and 524,288, depending on the scene complexity. These values were shown to provide good reconstruction performance across a wide range of natural scenes. However, the synthetic MTV dataset considered in this work contains sparse beamlet structures with significantly lower geometric complexity. Consequently, the optimal resolution hierarchy for the present application may differ from the default Instant-NGP configuration. Therefore, the base resolution was varied as

4, 8, 16, 32, 64, 128 while the finest resolution was varied as 128, 256, 512, 1024, 2048.

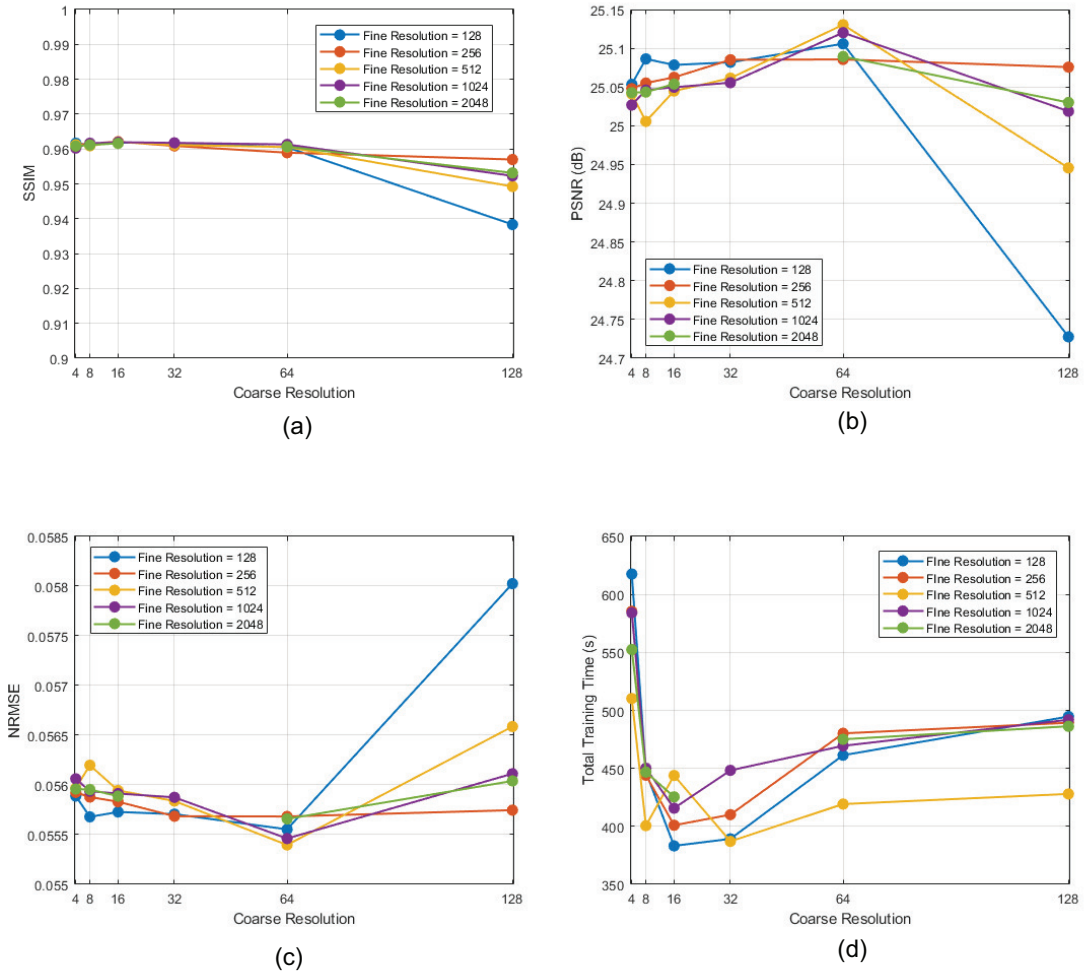


Figure 4.2: Effect of the coarse and fine resolution parameters of the multiresolution hash encoding on synthetic data reconstruction. The corresponding reconstruction quality and training efficiency are evaluated using (a) SSIM, (b) PSNR, (c) NRMSE, and (d) total training time.

Figures 4.2(a)–(c) show that for all fine-resolution configurations, SSIM and PSNR remain nearly constant as the coarse resolution increases from 4 to 64, while NRMSE exhibits only minor variations. This indicates that the multiresolution hash encoding is able to accurately represent the synthetic beamlet structures without requiring a very high base resolution. However, when the coarse resolution is increased to 128, a noticeable degradation in reconstruction quality is observed, particularly for the lower fine-resolution configurations (128 and 256). In these cases, SSIM decreases, PSNR drops, and NRMSE

increases, indicating that an excessively coarse hierarchy limits the network's ability to capture the underlying spatial information. Increasing the fine resolution beyond 512 provides only marginal improvements, with the performance for fine resolutions of 512, 1024, and 2048 remaining nearly identical across all reconstruction metrics.

The computational cost, shown in Figure 4.2(d), exhibits a different trend. Training time decreases substantially as the coarse resolution increases from 4 to approximately 16, after which it remains relatively stable for all fine-resolution configurations. Although the choice of fine resolution has only a modest influence on training time, larger fine resolutions generally require slightly longer training due to the increased encoding capacity. Since reconstruction quality changes very little beyond a fine resolution of 512, the additional computational cost associated with higher resolutions offers little practical benefit.

Based on these observations, a coarse resolution of 64 and a fine resolution of 512 were selected for all subsequent experiments. This configuration provides reconstruction quality comparable to the larger-resolution configurations while maintaining low computational cost, resulting in an effective balance between reconstruction accuracy and training efficiency for the synthetic MTV dataset.

4.1.2 Network Parameters

Following the optimization of the multiresolution hash encoding parameters, the influence of the network architecture on reconstruction performance was investigated. The architecture of the multilayer perceptron (MLP) determines the nonlinear mapping between the encoded spatial features and the volumetric density field, thereby influencing both the representational capacity of the network and the computational cost of training [30].

Unlike conventional NeRFs that rely on relatively large MLPs, Instant-NGP employs a compact fully connected network because the multiresolution hash encoding captures much of the spatial information required for reconstruction. Consequently, a relatively small MLP is sufficient to learn the remaining mapping, resulting in substantially faster

training without sacrificing reconstruction quality. Based on this observation, Müller *et al.* [14] adopted a network consisting of two hidden layers with 64 neurons per layer as the default architecture. However, since the synthetic MTV dataset considered in this work is significantly less complex than the natural scenes reconstructed by Instant-NGP, the optimal network architecture may differ. Therefore, the network width was varied as 32, 64, 128, and 256 neurons, while the network depth was varied from one to four hidden layers. The resulting reconstruction quality and computational cost are presented in Figure 4.3.

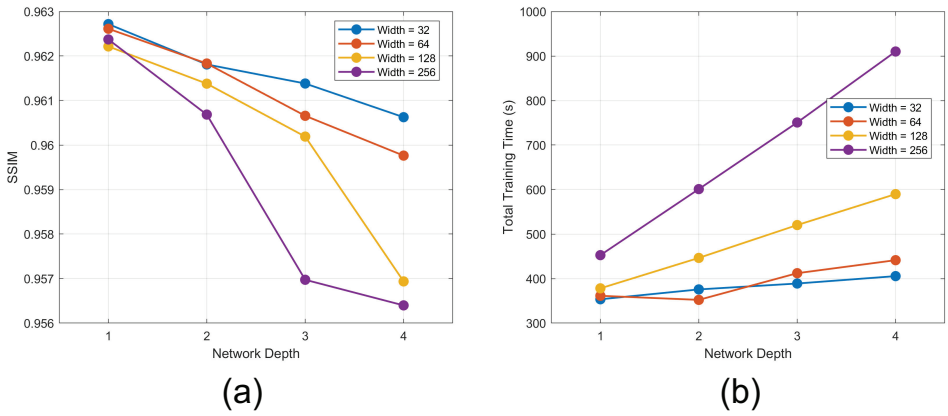


Figure 4.3: Effect of the network architecture on synthetic data reconstruction. (a) SSIM and (b) total training time as a function of network depth for different network widths.

As shown in Figure 4.3(a), increasing the network depth generally reduces the reconstruction quality for all investigated network widths. The decrease in SSIM is relatively small for the narrower networks (32 and 64 neurons) but becomes more pronounced as the network width increases to 128 and 256 neurons. This indicates that the compact MLP already provides sufficient representational capacity for the sparse synthetic Poiseuille flow dataset. Since the multiresolution hash encoding captures the dominant spatial information, increasing the network complexity provides little additional benefit to the reconstruction quality.

The slight degradation with added capacity is consistent with mild overfitting. Once

the hash encoding supplies the spatial information, the remaining mapping is simple, and the additional parameters of the larger networks are free to fit sensor noise and reconstruction ambiguity in the sparsely constrained regions rather than the underlying signal. The larger networks are also more difficult to optimize under the fixed training budget, which further limits any benefit from the increased capacity.

The effect of the network architecture on computational cost is shown in Figure 4.3(b). For all network widths, the total training time increases monotonically with increasing network depth. Likewise, increasing the network width substantially increases the training time at every network depth, with the 256-neuron configuration requiring the longest training time. Therefore, although larger network architectures increase the computational cost, they do not produce a corresponding improvement in reconstruction quality. Based on the overall hyperparameter investigation and subsequent experiments, a network consisting of three hidden layers with 64 neurons per layer was selected for all remaining reconstructions. This configuration provides a favorable balance between reconstruction quality and computational efficiency while maintaining a compact network architecture.

4.1.3 Effect of Number of Elemental Views

The number and angular distribution of projection views play a fundamental role in determining the quality of volumetric reconstruction in tomographic imaging, as each additional view provides complementary information about the three-dimensional object and reduces reconstruction ambiguity [44, 45]. Previous studies have shown that reducing the number of projection views or limiting the angular coverage generally degrades reconstruction quality due to insufficient sampling of the object, making limited-angle and sparse-view reconstruction inherently ill-posed [46, 47]. Unlike conventional tomographic systems, the Fourier Integral Microscopy (FIMic) system employed in this work operates under a limited-angle imaging geometry, where the elemental images are separated by approximately $5.7°$, resulting in a total angular coverage of only about $23°$ [2]. Consequently,

neighboring elemental views contain highly correlated information and do not fully span the measurement volume. This raises an important practical question regarding the minimum number of elemental views required to achieve accurate volumetric reconstruction. To investigate this trade-off, the number of elemental views was varied from 1 to 19 while keeping all network parameters fixed. This analysis was performed using the synthetic Poiseuille flow dataset at $t = 0.1$, where the beamlet displacement introduces sufficient deformation to evaluate reconstruction performance under nontrivial flow conditions.

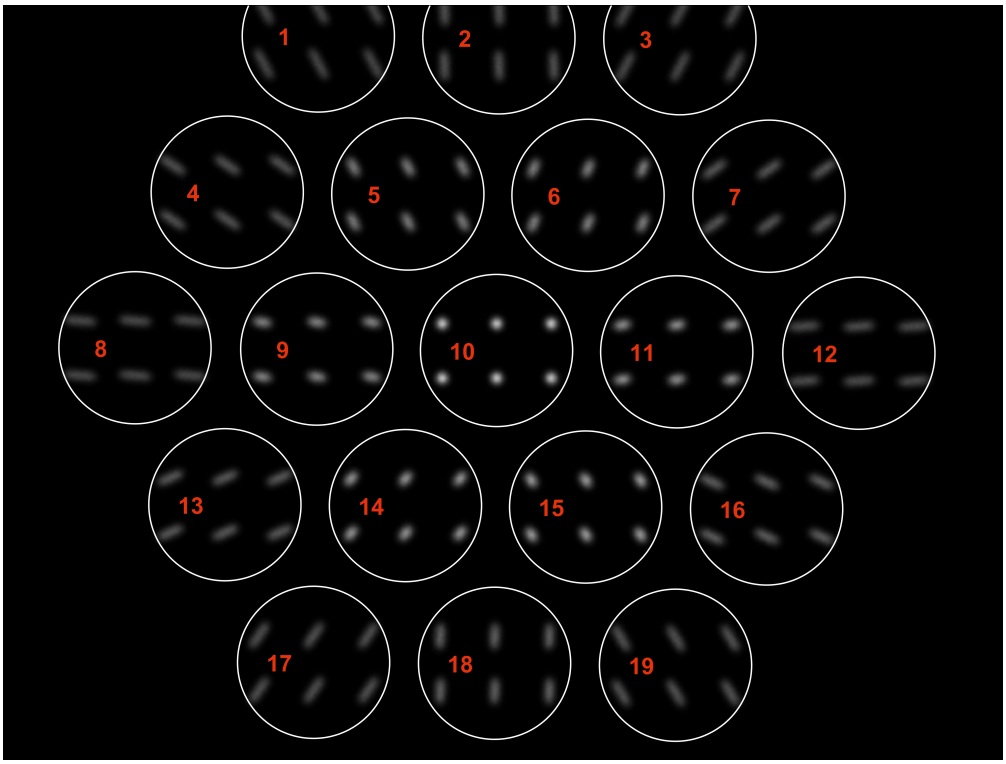


Figure 4.4: The 19 ground-truth elemental images produced by the FIMic microlens array, numbered by view.

Figure 4.4 shows the 19 ground-truth elemental images produced by the FIMic microlens array, each numbered by view. For each test case, a subset of views was removed from the elemental image set and FluidNeRF was trained on the remaining views. In each case the views were eliminated in such a way that the angular coverage available to the reconstruction was progressively limited. Table 4.2 lists the views used in each test case.

Table 4.2: View subsets used in each case of the view-count study.

Test case	Number of views	Views used
Case 1	1	10
Case 2	3	9, 10, 11
Case 3	5	9, 10, 11, 14, 15
Case 4	7	5, 6, 9, 10, 11, 14, 15
Case 5	9	4, 5, 6, 9, 10, 11, 14, 15, 16
Case 6	11	4, 5, 6, 9, 10, 11, 13, 14, 15, 16, 18
Case 7	16	1–6, 9, 10, 11, 13–19
Case 8	19	1–19 (all)

Figures 4.5(a)–(c) show that reconstruction quality is highly sensitive to the number of elemental views when only a small number of observations are available. Using a single elemental image results in substantially lower SSIM and PSNR together with a significantly higher NRMSE, demonstrating that one projection alone does not provide sufficient information to accurately recover the three-dimensional scalar field. This behavior is expected because depth information cannot be reliably inferred from a single viewpoint.

A significant improvement is observed when the number of elemental views is increased to three. Reconstruction quality continues to improve slightly for five and seven views, after which all three reconstruction metrics exhibit only minor changes. In particular, SSIM reaches a plateau beyond approximately seven elemental views, indicating that the global beamlet structures have been successfully recovered. Similar trends are observed in PSNR and NRMSE, where only negligible improvements are obtained by incorporating additional viewpoints. These results suggest that, for the sparse synthetic Poiseuille flow considered here, a relatively small number of elemental images is sufficient to reconstruct the dominant volumetric features.

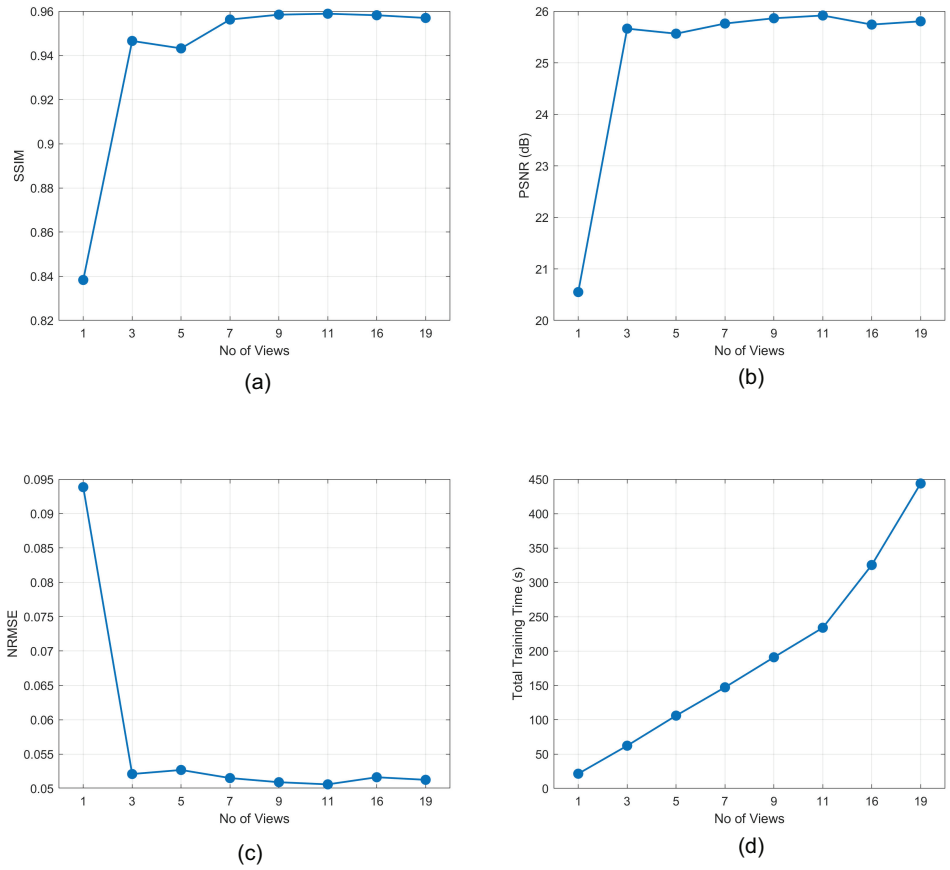


Figure 4.5: Effect of the number of elemental views on synthetic data reconstruction using multiresolution hash encoding. Reconstruction performance is evaluated using (a) SSIM, (b) PSNR, (c) NRMSE, and (d) total training time.

Figure 4.5(d) shows that the total training time increases monotonically with the number of elemental views because each additional image contributes more rays that must be processed during training. Although comparable reconstruction quality can be achieved using as few as five to seven elemental views, all 19 available views were retained for the remainder of this study. While the synthetic Poiseuille flow contains relatively simple beamlet deformations, more complex flow fields are expected to exhibit smaller-scale structures and stronger spatial gradients. In such cases, utilizing all available elemental views provides additional observational information that may improve the reconstruction of fine-scale flow features while maximizing the information captured by the FIMic system.

4.1.4 Final Network Configuration

Based on the results of the preceding hyperparameter investigations, an optimized network configuration was selected for all subsequent synthetic and experimental reconstructions. The selected parameters represent the best compromise between reconstruction accuracy and computational efficiency for the sparse MTV datasets considered in this work. While several hyperparameter combinations produced comparable SSIM, PSNR, and NRMSE values, the selected configuration consistently minimized the computational cost without sacrificing reconstruction quality. The final hyperparameter configuration is summarized in Table 4.3. All subsequent synthetic and experimental reconstructions presented in this thesis were performed using these parameters.

Table 4.3: Final optimized FluidNeRF hyperparameter configuration.

Category	**Parameter**	**Baseline**	**Selected**
Hash Encoding	Encoding levels (L)	16	12
	Features per level (F)	2	2
	Coarse resolution	16	64
	Fine resolution	512	512
	Hash table size (T)	2^{19}	2^{19}
Network	Depth	2	3
	Width	128	64
Input Data	Number of elemental views	19	19

4.2 Poiseuille Flow Reconstruction

The optimized network architecture identified in Section 4.1.4 was applied to the synthetic Poiseuille flow dataset described in Section 3.2. Since the synthetic beamlet geometry and imposed Poiseuille displacement are analytically prescribed, the corresponding volumetric intensity fields provide a known ground truth against which the reconstruction performance of the proposed framework can be rigorously evaluated. The reconstructed volumes are assessed through qualitative comparison with the analytical solution, followed by quantitative evaluation using volumetric image-quality metrics and beamlet profile analysis.

4.2.1 Training Convergence

The convergence behavior of FluidNeRF was first examined to verify that the network had reached a stable solution during training. Figure 4.6 shows the evolution of the training metrics over 4000 training iterations. It should be noted that the SSIM and NRMSE presented in this figure are computed using the 2D projection images during training rather than the reconstructed three-dimensional volume. These metrics therefore provide an indication of how well the network reproduces the measured projections throughout the optimization process.

As shown in Figure 4.6(a), the projection SSIM increases rapidly during the first few hundred training iterations and exceeds 0.96 early in the optimization. This indicates that the dominant global beamlet structures are recovered quickly. Such behavior is consistent with the multiresolution hash encoding proposed by Müller *et al.* [14], where the coarse-resolution levels capture the large-scale spatial structure during the early stages of training, while the finer-resolution levels progressively refine smaller-scale details as the optimization proceeds. Similarly, Figure 4.6(c) shows that the projection NRMSE decreases rapidly before reaching a nearly constant value, indicating stable convergence of the projected intensity fields.

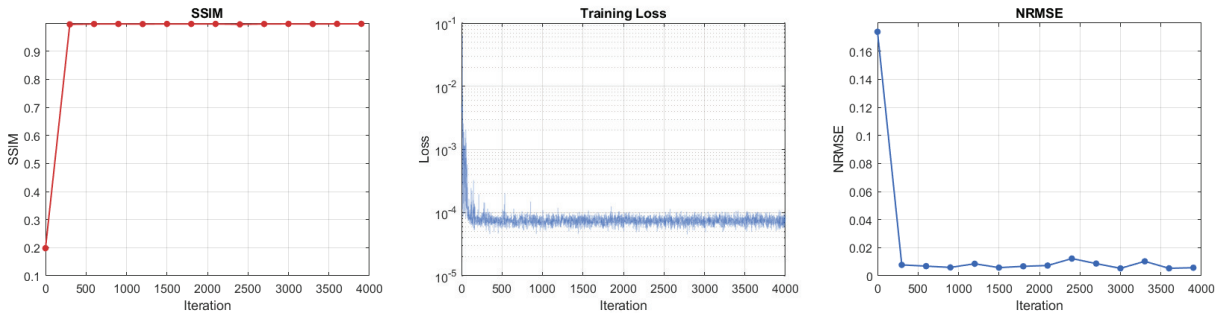


Figure 4.6: Training convergence for the synthetic Poiseuille flow reconstruction using FluidNeRF, showing (a) SSIM, (b) training loss, and (c) NRMSE over 4000 training iterations.

The corresponding training loss shown in Figure 4.6(b) exhibits a similar trend, decreasing rapidly before stabilizing with only minor fluctuations due to stochastic optimization. Although the projection metrics indicate that the network has largely converged within the first few hundred iterations, all reconstructions in this work were trained for 4000 iterations to ensure full convergence of the 3D volume and to maintain a consistent training procedure throughout the hyperparameter investigation and subsequent reconstruction studies.

4.2.2 Qualitative Evaluation

Figure 4.13 presents the 3D isosurface reconstructions of the synthetic MTV volume obtained using the optimized hash-encoded FluidNeRF framework at $t = 0$ and $t = 0.1$ s. Each reconstructed volume was normalized by percentile clipping between the 10th and 99.99th intensity percentiles and rescaled to the range zero to one. Isosurfaces were extracted at the normalized intensity levels of 0.5, 0.65, and 0.8. At both time instances, the reconstruction correctly reproduces the 3×2 beamlet arrangement, with each beamlet clearly resolved and spatially separated at the expected 500 μm center-to-center spacing. The reconstructed isosurfaces accurately preserve the beamlet geometry, spatial

separation, and boundary sharpness. The reconstruction also maintains high fidelity at $t = 0.1$ s, where the imposed Poiseuille flow introduces a spatially varying displacement across the beamlet array. Minor reconstruction artifacts are visible near the boundaries of the measurement volume. These appear as faint spurious intensity and slight blurring of the beamlet edges at the periphery of the reconstructed domain. They arise because the plenoptic imaging system provides limited angular coverage, so voxels near the volume boundary are constrained by fewer viewing rays than those at the center. The reduced ray coverage weakens the reconstruction where the beamlets approach the edge of the field of view, which produces the observed edge effects.

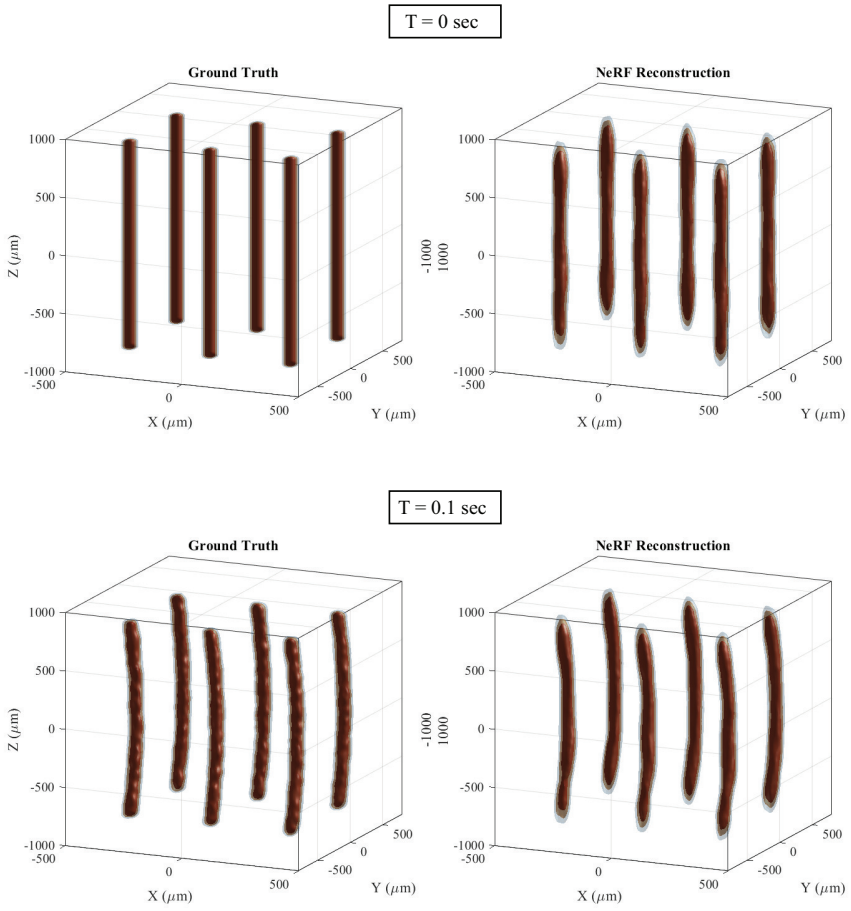


Figure 4.7: 3D isosurface reconstructions of the synthetic MTV volume at $t = 0$ and $t = 0.1$. Left column: ground truth volumes computed analytically from the Gaussian beamlet definitions. Right column: FluidNeRF reconstructions obtained using optimized Hash encoding.

Figure 4.8 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 4.8(a,c) show transverse XY slices at $Z = 0$ μm, while Figures 4.8(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.

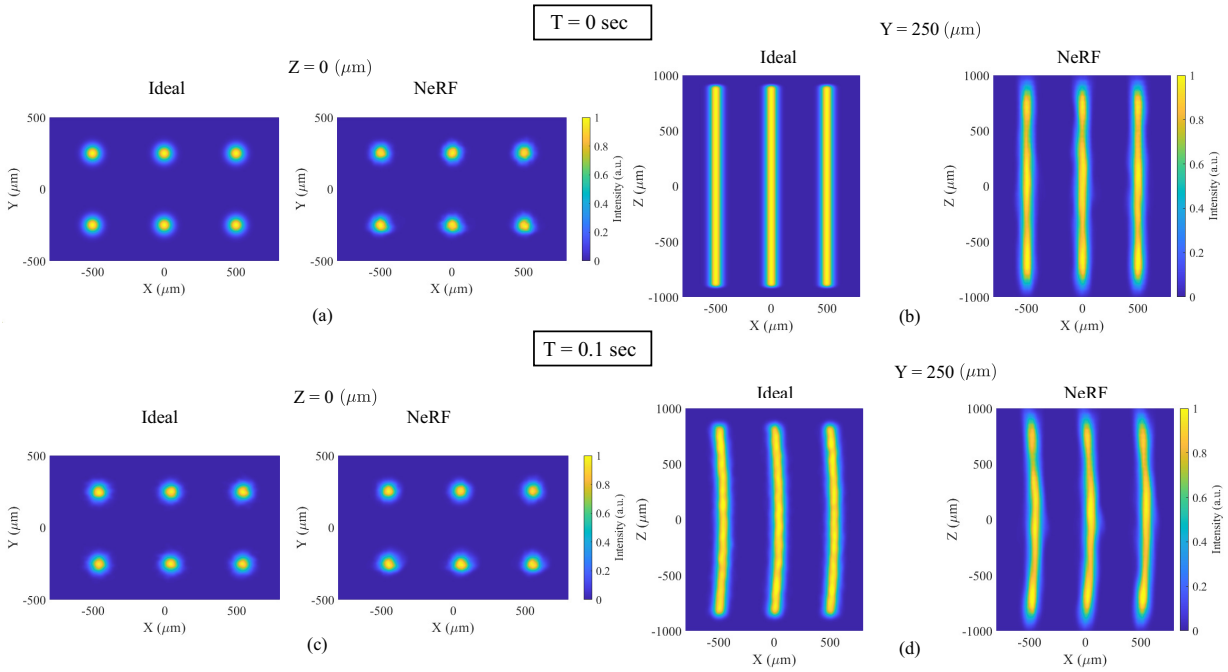


Figure 4.8: Comparison of ground-truth (*Ideal*) and FluidNeRF-reconstructed (*NeRF*) intensity volumes at $T = 0$ s and $T = 0.1$ s using hash encoding. (a) and (c) show transverse XY slices at $Z = 0\,\mu$m, while (b) and (d) show longitudinal XZ slices at $Y = 250\,\mu$m.

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.

4.2.3 Quantitative Reconstruction Results

While the qualitative comparisons demonstrate close visual agreement between the reconstructed and analytical volumes, quantitative evaluation provides a more objective assessment of the reconstruction fidelity. The reconstructed volumes were compared directly with the synthetic ground truth using the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Normalized Root Mean Square Error (NRMSE). In addition to these global volumetric metrics, one-dimensional beamlet intensity profiles and the corresponding full width at half maximum (FWHM) were evaluated to assess the accuracy of the reconstructed beamlet morphology.

Table 4.4: Quantitative comparison between the synthetic ground-truth volumes and the FluidNeRF reconstructions using the optimized hash-encoded implementation.

Metric	$t = 0$	$\Delta t = 0.026$ unit
SSIM	0.9795	0.9672
MSE	0.000411	0.000698
PSNR (dB)	33.87	31.56

Table 4.4 summarizes the quantitative comparison between the reconstructed and synthetic ground-truth volumes at the two time instances. The reconstructed volumes achieve SSIM values greater than 0.96, indicating a high degree of structural similarity with the analytical ground truth. Correspondingly, the low MSE values and high PSNR demonstrate that the reconstructed scalar intensity fields closely match the reference volumes with minimal reconstruction error. Although a slight decrease in SSIM and PSNR, accompanied by a small increase in MSE, is observed at $t = 0.1$ s, the quantitative metrics remain consistently high. This behavior is expected because the imposed Poiseuille flow produces greater spatial deformation of the beamlets, increasing the complexity of the underlying scalar field while preserving the overall beamlet morphology.

While these volumetric metrics provide a global assessment of the reconstruction

fidelity, a more detailed evaluation of the local beamlet structure is obtained through analysis of the reconstructed intensity profiles. Figure 4.9 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 4.9.

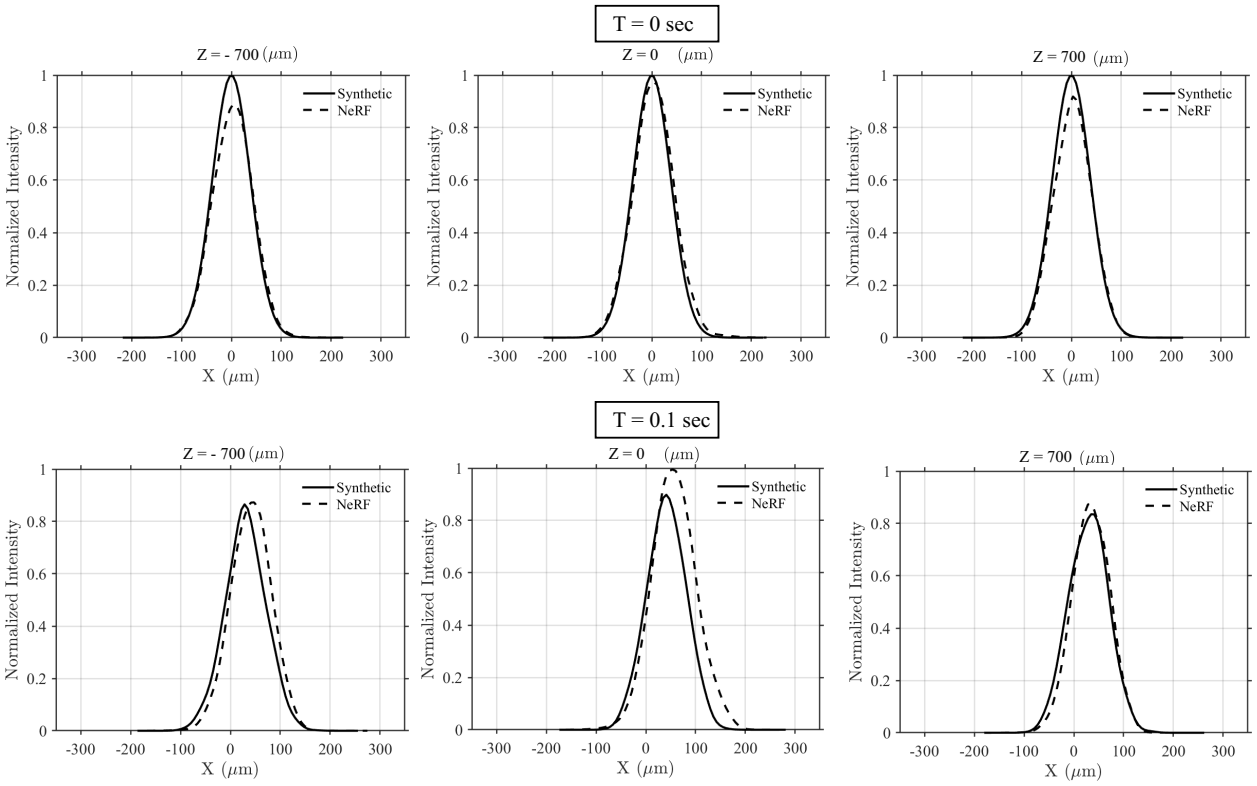


Figure 4.9: Comparison of reconstructed beamlet intensity profiles obtained using FluidNeRF with the corresponding synthetic ground-truth profiles at two time instances, $T = 0$ s and $T = 0.1$ s. Profiles are shown along the X-direction for three axial planes, $Z = -700$ μm, $Z = 0$ μm, and $Z = 700$ 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 synthetic reference at the same depth

planes. Unlike the reference at $T = 0$, which is constructed analytically and is therefore uniform across depth, the reference at $T = 0.1$ is obtained by advecting the sample points through the flow and binning them onto the voxel grid. This binning, combined with the depth-dependent Poiseuille displacement, introduces small plane-to-plane variations in the reference width. The reconstructed values track these variations and remain within a few micrometers of the reference at every plane. The reconstructed beam thickness is therefore well preserved across depth. Overall, FluidNeRF reconstructs the spatial morphology and intensity distribution of the MTV line while maintaining the characteristic beam width of the synthetic reference.

4.3 Turbulent Channel Flow Reconstruction

The Poiseuille case establishes that the reconstruction-to-velocity pipeline recovers a smooth, analytically prescribed displacement field. However, the parabolic profile varies only in the wall-normal direction and imposes a uniform, gradual deformation on the beamlets. Near-wall flows of practical interest are turbulent and produce displacement fields that vary in all three directions with fine spatial structure. The turbulent channel flow case was therefore introduced to evaluate the framework under a realistic near-wall displacement field. The JHTDB channel5200 dataset provides a direct numerical simulation at $Re_\tau = 5200$, giving a physically accurate turbulent velocity field at the spatial scales accessible to the measurement system. This case tests whether the reconstruction retains sufficient accuracy to recover velocity when the beamlets undergo the irregular, multi-directional deformation characteristic of near-wall turbulence.

4.3.1 Qualitative Evaluation

Figure 4.12 presents the 3D isosurface reconstructions of the synthetic turbulent MTV volume at the initial instant $t = 0$ and after a displacement time of $\Delta t = 0.026$ convective units. At $t = 0$, the reconstruction reproduces the undeformed 3×2 beamlet arrangement,

with each beamlet clearly resolved and separated at the expected 500 μm spacing. After advection, the beamlets exhibit the irregular spanwise waviness imposed by the turbulent velocity field. The reconstructed beamlets remain well separated at both instants and follow the ground-truth trajectories, including the non-uniform bending introduced by the turbulent motion. Small surface irregularities and boundary artifacts are visible near the edges of the measurement volume. These are caused by the limited angular coverage of the FIMic system, which spans only about 23° and provides weak depth information. The overall beamlet shape and the turbulent deformation are still recovered throughout the volume.

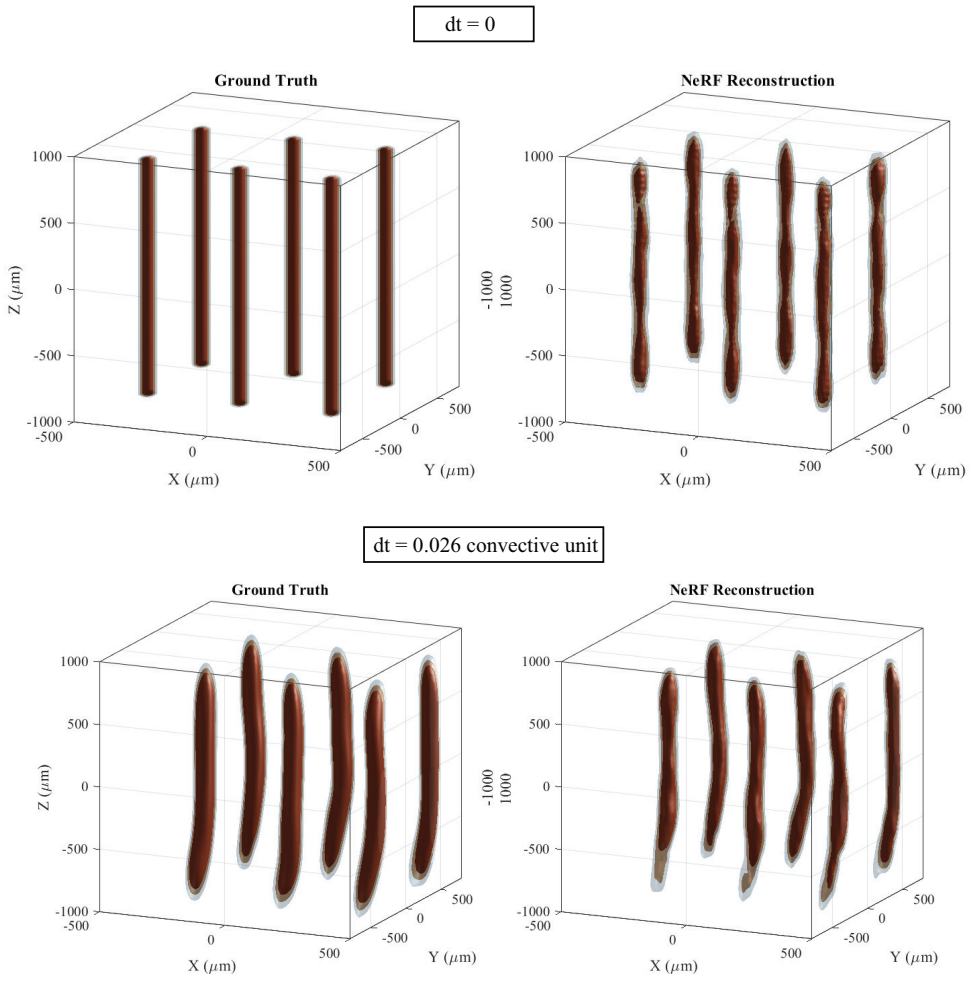


Figure 4.10: 3D isosurface reconstructions of the synthetic turbulent channel flow MTV volume at the initial instant ($\Delta t = 0$) and after advection ($\Delta t = 0.026$ convective units). Left column: ground-truth volumes generated from the JHTDB channel 5200 data. Right column: FluidNeRF reconstructions obtained using hash encoding.

Figure 4.11 shows transverse XY slices of the ground-truth and FluidNeRF-reconstructed intensity volumes for the synthetic turbulent channel flow at the advected instant $\Delta t = 0.026$ convective units. Both slices are taken at this instant but at two different depths: Figure 4.11(a) at $Z = 0$ μm and Figure 4.11(b) at $Z = 453$ μm. The undeformed case at $\Delta t = 0$ is not repeated here, as the initial beamlet arrangement was already presented for the Poiseuille flow in Section 4.2.2.

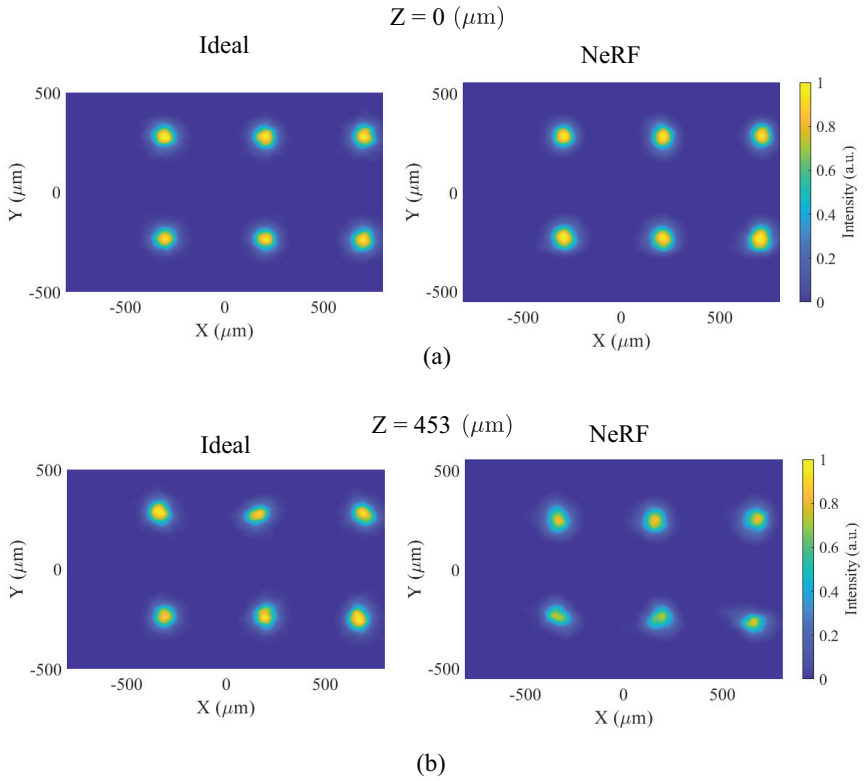


Figure 4.11: XY slices of the ground-truth and FluidNeRF-reconstructed intensity volumes for the synthetic turbulent channel flow. (a) Initial instant ($\Delta t = 0$) at $Z = 0$ μm. (b) Advected instant ($\Delta t = 0.026$ convective units) at $Z = 453$ μm.

At both depths, the reconstruction reproduces the position of all six beamlets, and each peak is well localized and clearly separated from its neighbors. Because the beamlets are deformed by the turbulent velocity field, they are no longer straight, so the two depth planes intersect the beamlets at different points along their length. This is why the beamlet positions shift between Figure 4.11(a) and Figure 4.11(b). The reconstruction follows these

shifts at both depths and matches the ground-truth arrangement.

The peak intensities of the reconstructed beamlets differ slightly from the ground truth. In the ground-truth slices the six peaks have nearly uniform intensity, whereas in the reconstructed slices the peak intensity varies from beamlet to beamlet. This difference is caused by the weak depth constraint of the limited-angle FIMic geometry, which recovers the in-plane beamlet positions more accurately than their intensity along the axial direction. The variation is small and does not affect the position or separation of the beamlets, so the reconstructed intensity distribution remains in good agreement with the ground truth at both depths.

4.3.2 Quantitative Reconstruction Results

Table 4.5 summarizes the quantitative comparison between the reconstructed and ground-truth volumes for the turbulent channel flow at the initial and advected instants. Both instants are reconstructed by a single time-resolved network. The reconstruction achieves an SSIM above 0.94 and a PSNR above 24 dB at both instants, indicating that the beamlet structure is well preserved throughout the reconstructed sequence.

Table 4.5: Quantitative comparison between the synthetic turbulent channel flow ground-truth volumes and the time-resolved FluidNeRF reconstruction using the optimized hash-encoded implementation.

Metric	$t = 0$	$\Delta t = 0.026$ unit
SSIM	0.9606	0.9416
MSE	0.003074	0.003373
PSNR (dB)	25.12	24.72

The metrics change only slightly between the two instants, with a small decrease in SSIM and PSNR and a small increase in MSE at the advected instant. This consistency is the key result, as it shows that a single network represents both the initial and the deformed beamlet fields without a meaningful loss of accuracy at the later instant. The

small reduction is expected, because the turbulent velocity field deforms the beamlets into irregular, non-planar shapes that are more complex to represent than the undeformed initial condition. The reconstruction preserves the position, separation, and overall morphology of the beamlets at both instants, confirming that the network configuration optimized on the Poiseuille flow generalizes to a realistic turbulent displacement field reconstructed in a time-resolved manner demonstrated in Section 4.2.3.

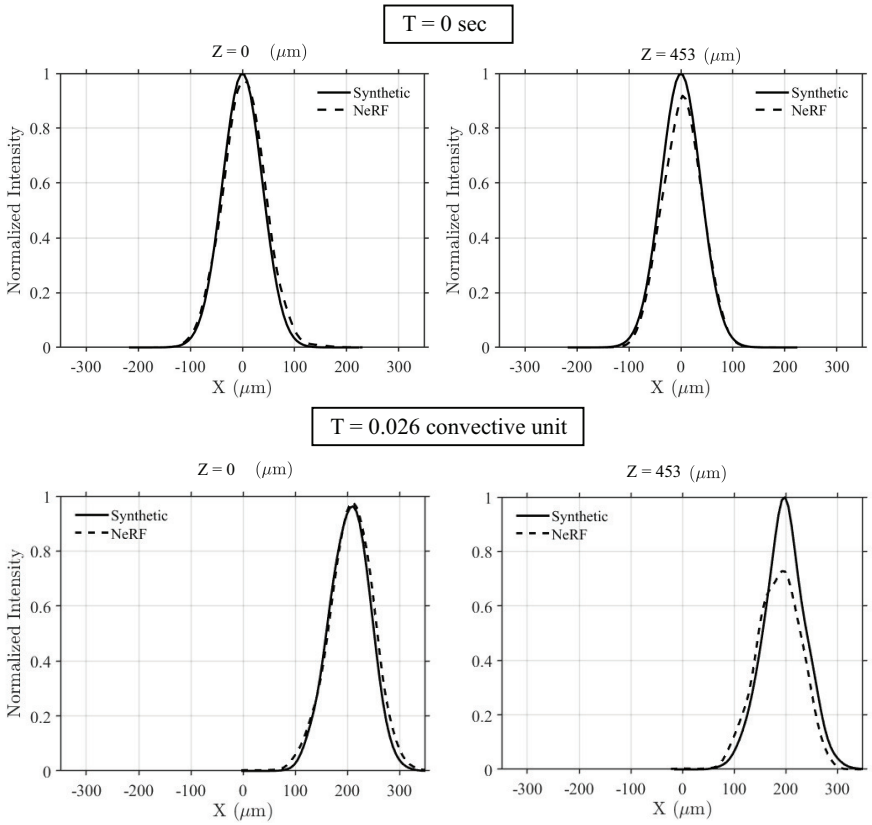


Figure 4.12: Comparison of reconstructed beamlet intensity profiles obtained using FluidNeRF with the corresponding synthetic ground-truth profiles for the turbulent channel flow at two instants, $t = 0$ and $\Delta t = 0.026$ convective units. Profiles are shown along the X-direction at two depth planes, $Z = 0$ μm and $Z = 453$ μm. Solid lines denote the synthetic ground truth, and dashed lines represent the reconstructed FluidNeRF profiles.

Figure 4.12 compares the normalized one-dimensional intensity profiles extracted along the X-direction through the beamlet cores for the synthetic turbulent channel flow. Profiles are shown at two depth planes, $Z = 0$ μm and $Z = 453$ μm, at the initial instant $t = 0$ and the advected instant $\Delta t = 0.026$ convective units. Solid lines denote the synthetic

ground truth and dashed lines the FluidNeRF reconstruction.

At the initial instant, the reconstructed profiles follow the ground truth at both depths. The peak position and the shape of the intensity distribution are recovered, and the reconstructed profile width matches the reference. At the advected instant, the profile peaks have shifted along the X-direction. This shift reflects the streamwise displacement imposed by the turbulent velocity field. The reconstruction recovers the shift at both depths. The time-resolved network therefore captures the beamlet displacement between the two instants.

At $t = 0$, the full width at half maximum was 91 and 91 μm for the reconstruction at $Z = 0$ μm and $Z = 453$ μm. The corresponding reference values were 91 and 89 μm. At the advected instant, the reconstruction gave 95 and 101 μm, compared with 96 and 93 μm for the reference. The reconstructed beamlet width is preserved across depth and between instants. The deviations from the synthetic reference are small.

A single shift value is not reported because the turbulent velocity field varies in all three directions. The streamwise displacement therefore differs between depths and between beamlets, unlike the uniform displacement of the Poiseuille case.

4.4 Comparison of Hash and Fourier Encoding

The original FluidNeRF framework proposed by Kelly and Thurow [12, 13] employs Fourier positional encoding for the reconstruction of three-dimensional scalar fields. In the present work, this encoding is replaced with multiresolution hash encoding to improve the computational efficiency of the reconstruction process. Although Instant-NGP demonstrated the computational advantages of multiresolution hash encoding for neural graphics applications [14], its performance for sparse, limited-angle volumetric reconstruction encountered in molecular tagging velocimetry has not been systematically evaluated. Therefore, a direct comparison between the original Fourier-encoded FluidNeRF implementation and the proposed hash-encoded framework was performed using the synthetic Poiseuille flow

dataset.

The Fourier implementation followed the methodology of Kelly and Thurow [13], employing a fully connected multilayer perceptron with 12 hidden layers, 400 neurons per layer, and a Fourier positional encoding with a positional encoding frequency of $L_x = 12$. The hash-encoded implementation used the optimized hyperparameters identified in Section 4.1.4. To ensure a fair comparison, both implementations were trained for 4000 iterations using 64 sample points per ray, 2048 sampling points along the beamlet axial (z) direction, identical optimization settings, and the same NVIDIA RTX 5040 GPU. Consequently, the primary differences between the two implementations are the encoding strategy and the corresponding network architecture required by each encoding. Reconstruction performance was compared using training convergence, qualitative reconstruction, quantitative volumetric metrics, and total training time.

4.4.1 Qualitative Comparison

Figure 4.13 compares the 3D isosurface reconstructions obtained using multiresolution hash encoding and Fourier encoding for the synthetic Poiseuille flow dataset at $t = 0$ and $t = 0.1$ s. The analytical ground-truth volumes are included for reference. Both encoding strategies successfully reconstruct the six-beamlet arrangement and accurately reproduce the imposed Poiseuille deformation at the later time step, indicating that each method is capable of learning the underlying three-dimensional scalar field from the limited-angle elemental images.

At both time instances, the reconstructed beamlets remain well separated and preserve the expected spatial arrangement and beamlet trajectories. The overall reconstruction quality of the two encoding strategies is visually comparable, with both methods accurately reproducing the global beamlet geometry and the curvature introduced by the Poiseuille flow. The Fourier-encoded reconstruction produces slightly smoother isosurfaces, whereas the hash-encoded reconstruction exhibits small surface irregularities. Despite these visual

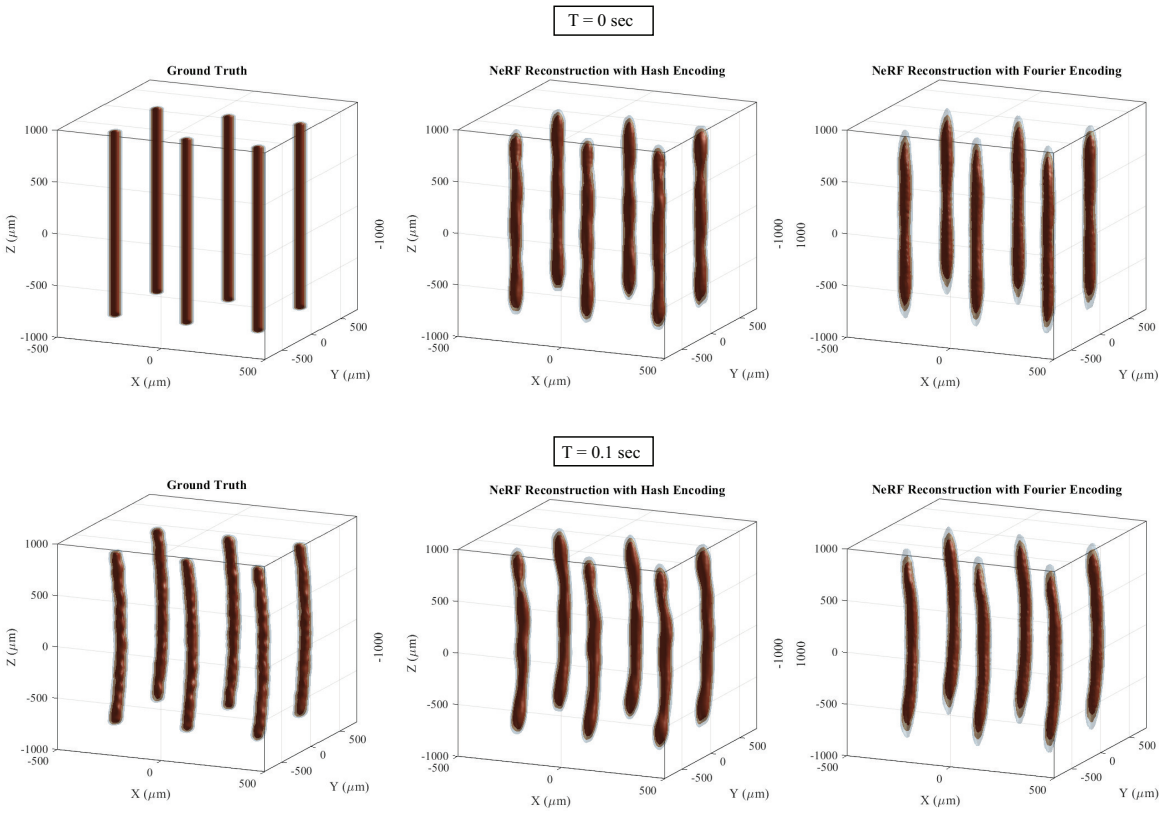


Figure 4.13: 3D isosurface reconstructions of the synthetic MTV volume at $t = 0$ and $t = 0.1$. Left column: ground truth volumes computed analytically from the Gaussian beamlet definitions. Middle column: FluidNeRF reconstructions obtained using multiresolution hash encoding. Right column: FluidNeRF reconstructions obtained using Fourier encoding.

differences, both methods closely match the analytical ground truth and preserve the overall beamlet morphology.

The qualitative comparison indicates that both encoding strategies provide accurate reconstructions of the sparse synthetic MTV volume. Therefore, a quantitative evaluation is required to determine whether the observed visual differences correspond to meaningful differences in reconstruction accuracy and computational efficiency.

4.4.2 Quantitative Comparison

While the qualitative comparison indicates that both encoding strategies successfully reconstruct the synthetic MTV volume, the volumetric image-quality metrics provide a more objective assessment of their reconstruction performance. Table 4.6 summarizes the SSIM,

MSE, and PSNR obtained by comparing the reconstructed volumes with the analytical ground truth at both time instances.

At $t = 0$, the multiresolution hash encoding achieves an SSIM of 0.9795, compared with 0.9603 for Fourier encoding. Similarly, the PSNR increases from 25.05 dB for Fourier encoding to 33.87 dB using hash encoding, while the MSE is reduced from 3.127×10^{-3} to 4.11×10^{-4}. Comparable improvements are also observed at $t = 0.1$ s, where the hash-encoded reconstruction achieves an SSIM of 0.9672 and a PSNR of 31.56 dB, compared with 0.9571 and 25.62 dB, respectively, for Fourier encoding. The corresponding MSE is reduced by approximately a factor of four, indicating a substantially closer agreement between the reconstructed and analytical scalar fields.

Table 4.6: 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

Although both methods accurately recover the overall beamlet geometry, the quantitative metrics consistently demonstrate the superior reconstruction accuracy of the proposed multiresolution hash encoding. The higher SSIM values indicate improved preservation of structural information, while the lower NRMSE and higher PSNR confirm that the reconstructed intensity field more closely matches the analytical solution. The improvement is observed at both time instances, indicating that the proposed framework maintains its reconstruction accuracy even after the beamlets undergo the spatial deformation introduced by the Poiseuille flow.

These results demonstrate that replacing Fourier positional encoding with multiresolution hash encoding not only substantially improves the computational efficiency of

FluidNeRF, but also produces more accurate volumetric reconstructions of the synthetic MTV dataset. The computational advantage of the proposed framework is examined in the following section.

4.4.3 Computational Performance

The computational performance of the proposed multiresolution hash encoding was compared directly with the original Fourier-encoded FluidNeRF framework under identical training conditions. Both implementations were trained using the synthetic Poiseuille flow dataset with identical ray sampling, optimization parameters, and hardware (NVIDIA RTX 5040 GPU). Consequently, the observed differences in convergence behavior and computational cost arise primarily from the choice of spatial encoding and the corresponding network architecture.

Figure 4.14(a)–(c) compares the evolution of the training loss, validation NRMSE, and validation SSIM throughout the optimization process. The validation metrics shown in Figures 4.14(b) and (c) were computed from the rendered elemental image projections and their corresponding synthetic ground-truth projections during training. Both implementations exhibit a rapid reduction in training loss during the initial stages of optimization. However, the multiresolution hash encoding consistently converges to a lower training loss while achieving lower NRMSE and higher SSIM than the Fourier-encoded implementation, indicating both faster convergence and improved reconstruction accuracy.

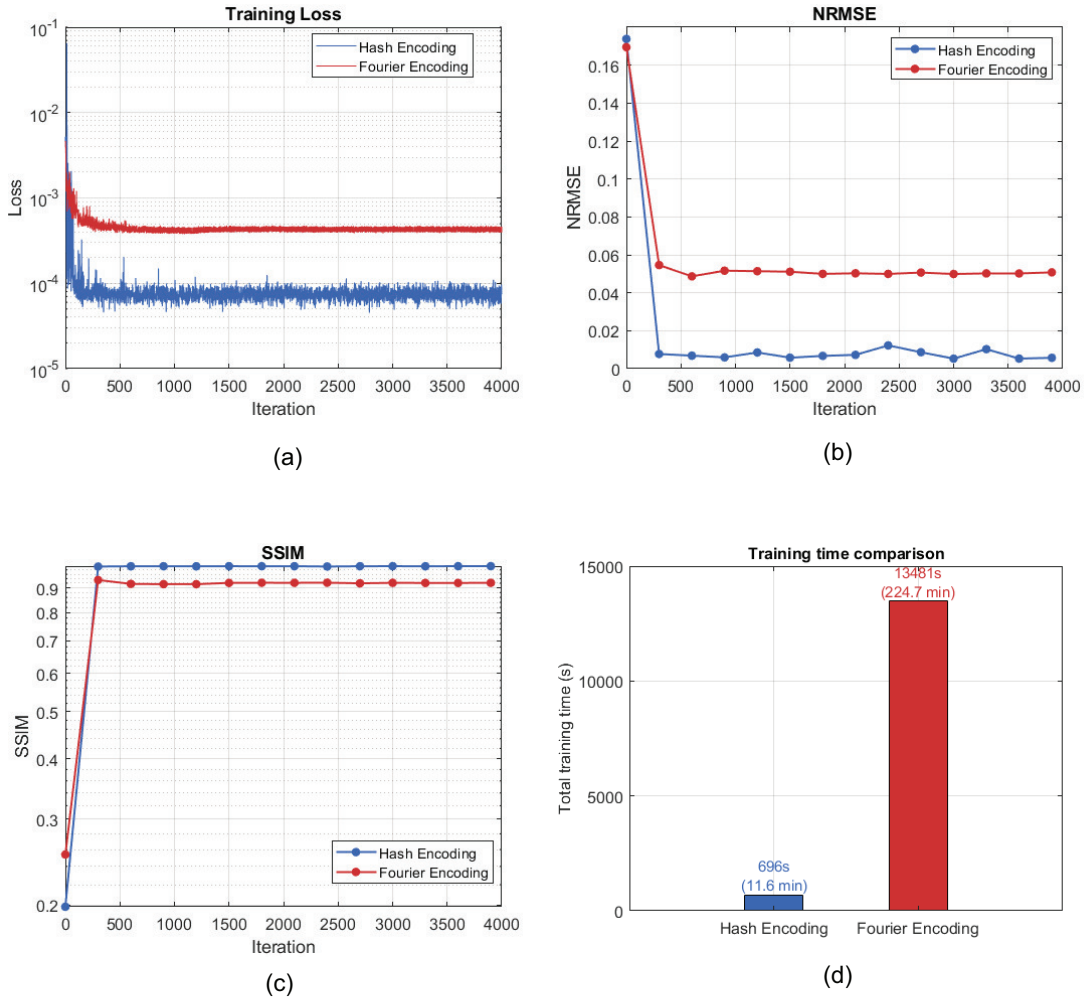


Figure 4.14: Comparison of multiresolution hash encoding and Fourier encoding during training on the synthetic MTV dataset. (a) Training loss as a function of iteration. (b) Normalized root mean square error (NRMSE). (c) Structural similarity index (SSIM). (d) Total training time.

The differences in convergence behavior are further illustrated in Figure 4.15, which shows the evolution of the reconstructed three-dimensional MTV volume during training. At the first few training iterations, neither encoding strategy has recovered the underlying beamlet structure. By 100 iterations, however, the multiresolution hash encoding has already reconstructed the six beamlets with well-defined geometry and correct spatial separation, whereas the Fourier-encoded implementation still exhibits fragmented beamlets and pronounced reconstruction artifacts. At 500 iterations, the hash-encoded reconstruction is visually close to the analytical solution, while the Fourier reconstruction continues to refine

the beamlet geometry. Even after 1000 iterations, the Fourier-encoded reconstruction still contains noticeable artifacts compared with the hash-encoded result. The rapid formation of the beamlet structure observed in Figure 4.15 explains the convergence trends shown in Figure 4.14, where the hash-encoded implementation reaches lower training loss, lower NRMSE, and higher SSIM within the first few hundred iterations. This behavior is consistent with the multiresolution hash encoding proposed by Müller *et al.* [14], where the coarse-resolution feature grids capture the dominant global structure early in training, while the finer-resolution levels progressively refine smaller spatial details as the optimization proceeds.

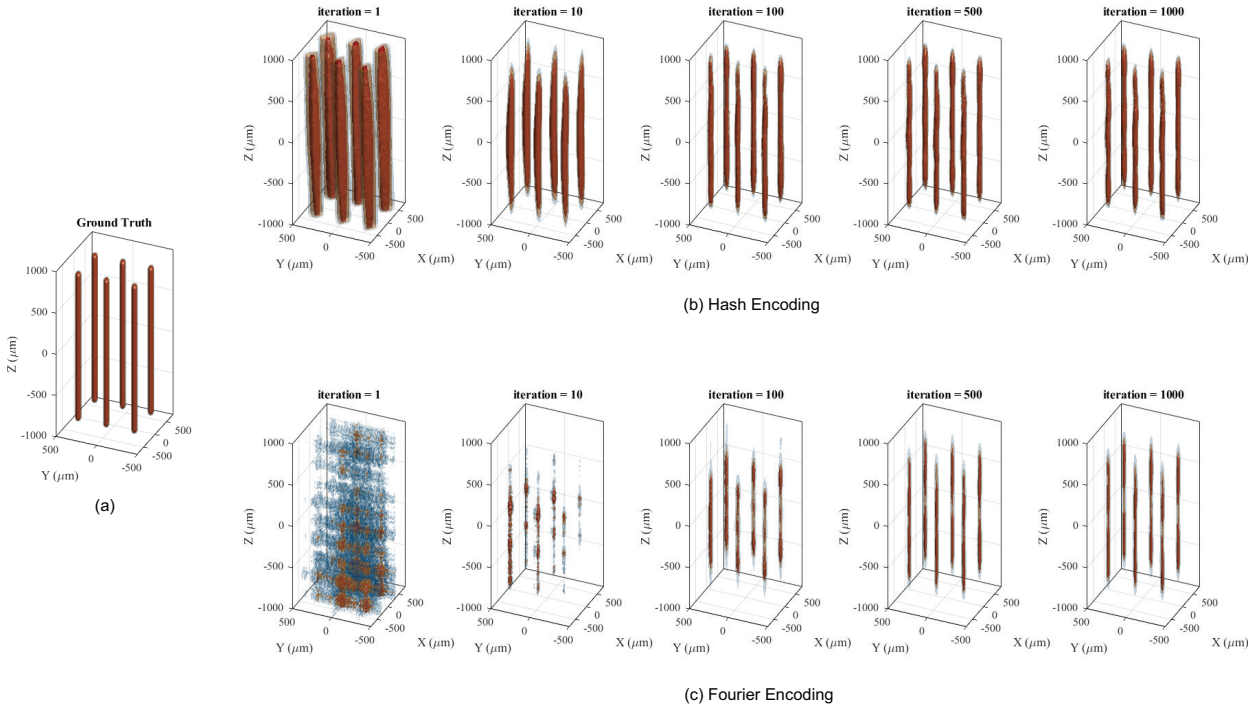


Figure 4.15: Evolution of the reconstructed synthetic MTV volume at $t = 0$ s during training. (a) Ground-truth volume generated from the synthetic dataset. (b) Reconstructions obtained using multiresolution hash encoding at iterations 1, 10, 100, 500, and 1000. (c) Reconstructions obtained using Fourier encoding at the same iterations.

The computational advantage of the proposed framework is summarized in Figure 4.14(d). Under identical training conditions, the multiresolution hash encoding completed 4000 training iterations in 696 s (11.6 min), whereas the Fourier-encoded implementation required 13 481 s (224.7 min). This corresponds to approximately a 95% reduction in

training time, or nearly a 19-fold improvement in computational efficiency. The substantial reduction in computational cost is primarily attributed to the trainable multiresolution hash encoding, which efficiently represents multiscale spatial information and therefore enables the use of a compact multilayer perceptron consisting of 3 hidden layers with 64 neurons per layer. In contrast, the Fourier-encoded implementation requires a significantly larger network consisting of 12 hidden layers with 400 neurons per layer to achieve sufficient representational capacity.

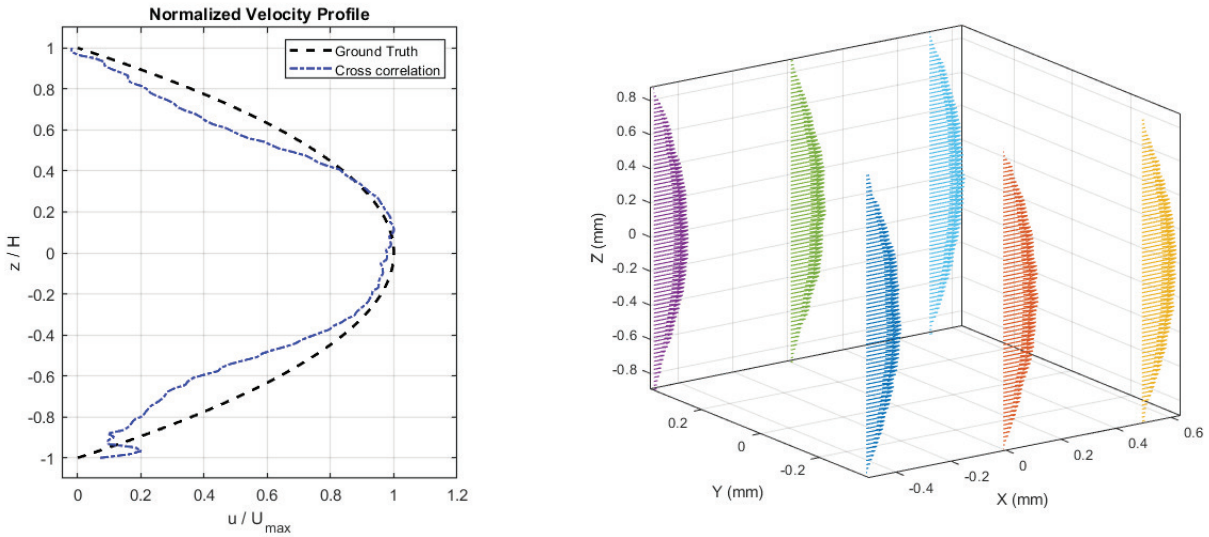
Overall, the comparison demonstrates that replacing Fourier positional encoding with multiresolution hash encoding not only improves the reconstruction accuracy but also dramatically accelerates convergence and reduces computational cost. These improvements make the proposed framework particularly well suited for time-resolved molecular tagging velocimetry, where a large number of three-dimensional volumes must be reconstructed within practical computational times.

4.5 Velocity Estimation

The purpose of the molecular tagging measurement, however, is to recover the velocity field, which is obtained from the displacement of the tagged beamlets between the two measurement instants. To demonstrate that the reconstructed volumes support this final step, the velocity field was estimated from the two reconstructed instants and compared with the analytical Poiseuille profile.

The beamlet displacement was measured by two-dimensional cross-correlation applied slice by slice along the beamlet axis, using a 64×64 voxel interrogation window centered on each beamlet. For every slice, the correlation peak between the two instants gave the local streamwise displacement, which was converted to velocity by dividing by the time separation Δt and normalized by its maximum value. Figure 4.16(a) compares the normalized velocity profile recovered from the reconstruction with the analytical ground-truth profile as a function of the wall-normal coordinate z/H. Figure 4.16(b) shows the

corresponding three-dimensional view of the displaced beamlets, in which the streamwise displacement varies parabolically with wall-normal position, reaching its maximum at the channel center and approaching zero at the walls.



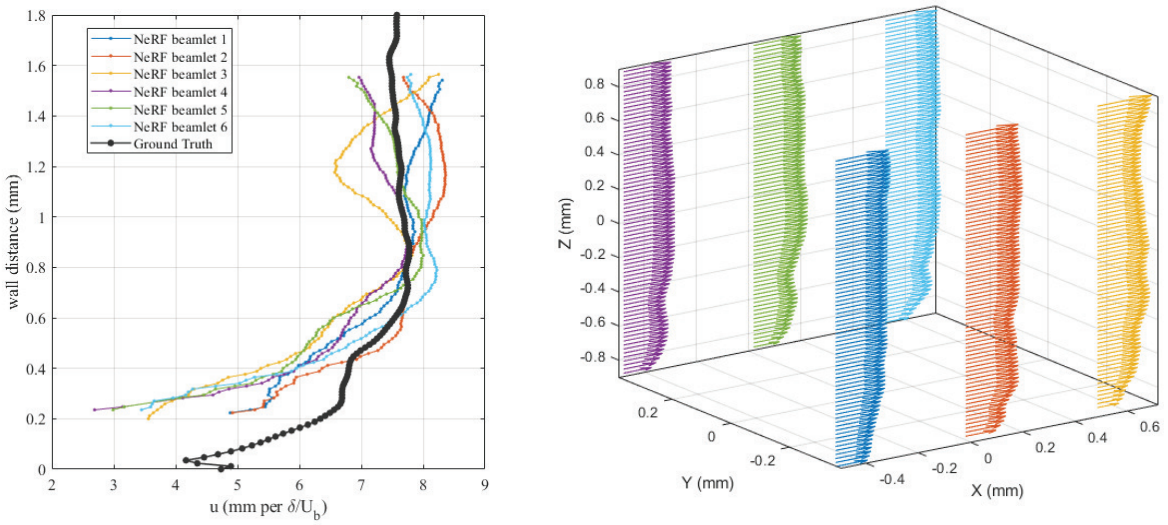
(a) (b)

Figure 4.16: Velocity estimation for the synthetic Poiseuille flow. (a) Normalized velocity profile recovered from the FluidNeRF reconstruction, compared with the analytical ground-truth profile. (b) 3D view of the displaced beamlets, showing the parabolic variation of the streamwise displacement with wall-normal position.

The recovered velocity profile follows the analytical parabolic distribution across the channel. The peak velocity near the channel center and the symmetry of the profile about $z/H = 0$ are both reproduced. Small deviations from the analytical profile are observed near the walls, where the beamlet displacement is smallest and the cross-correlation is therefore most sensitive to reconstruction noise. The agreement across the remainder of the channel confirms that the reconstructed scalar field is accurate enough to recover the underlying velocity field, establishing the full reconstruction-to-velocity pipeline for the synthetic Poiseuille case.

The velocity field for the turbulent channel flow was estimated using the same procedure applied to the Poiseuille case. Figure 4.17(a) compares the streamwise velocity profile recovered from each of the six reconstructed beamlets with the ground-truth profile

as a function of wall distance. Figure 4.17(b) shows the corresponding three-dimensional view of the displaced beamlets.



(a) (b)

Figure 4.17: Velocity estimation for the synthetic turbulent channel flow. (a) Streamwise velocity profiles recovered from the six reconstructed beamlets, compared with the ground-truth profile. (b) 3D view of the displaced beamlets, showing the irregular deformation produced by the turbulent velocity field.

Unlike the parabolic Poiseuille profile, the turbulent channel flow produces a mean velocity profile that rises steeply near the wall and flattens toward the outer edge of the slab, consistent with the near-wall structure of a turbulent boundary layer. The velocity profiles recovered from the reconstructed beamlets follow this ground-truth distribution, reproducing the steep near-wall gradient and the outer plateau. The 3D view shows that the beamlets are deformed irregularly by the turbulent motion rather than with the smooth curvature of the Poiseuille case.

The recovered profiles show greater scatter between beamlets than in the Poiseuille case. This is expected, because the turbulent displacement field varies in all three directions and imposes finer, non-uniform deformation on the beamlets, which is more demanding to recover under the limited-angle imaging geometry. The scatter is largest near the wall, where the beamlet displacement is smallest and the cross-correlation is most sensitive to

reconstruction noise. Despite this scatter, the reconstructed profiles capture the shape of the turbulent mean velocity profile across the slab, demonstrating that the reconstruction-to-velocity pipeline extends to a realistic turbulent displacement field.

Chapter 5

Experimental Data Results

This chapter applies the reconstruction framework to experimental FIMic measurements of near-wall MTV. Because no volumetric ground truth is available for the experimental data, the reconstructions are compared against the Richardson–Lucy deconvolution result used as a reference. A single network is trained across both measurement instants to demonstrate time-resolved reconstruction, and the recovered volumes are assessed through isosurface comparison, two-dimensional slices, and one-dimensional beamlet intensity profiles.

5.1 Experimental Setup

The experimental dataset used in this work was obtained from Huck et al. [2], who demonstrated near-wall volumetric molecular tagging velocimetry (MTV) 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 $\Delta t = 200\ \mu$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 flow motion.

The Fourier Integral Microscope is an object-space telecentric imaging system that produces orthographic projections by placing a microlens array at the Fourier plane of an infinity-corrected microscope objective. Owing to this telecentric configuration, the rays within each elemental image are mutually parallel, while each elemental image provides a different viewing direction of the measurement volume.

The FIMic system records a total of 23 elemental images of the tagged beamlet pattern. However, only 19 elemental images containing clear projections of all six beamlets

were used for reconstruction in the present work. The same experimental dataset previously reconstructed by Huck et al. [2] using Richardson–Lucy deconvolution (RLD) was used here to evaluate the performance of the proposed hash-encoded FluidNeRF framework. The RLD reconstruction serves as the reference for comparison in the following sections.

5.2 Isosurface Reconstruction

To demonstrate the time-resolved capability of the proposed framework, a single FluidNeRF network was trained simultaneously using the elemental images acquired at both measurement instants. This formulation enables the reconstruction of multiple temporal states within a single neural representation while maintaining a compact network architecture.

Figure 5.1 compares the three-dimensional isosurface reconstructions obtained using the proposed FluidNeRF framework with the experimental Richardson–Lucy Deconvolution (RLD) reconstruction at $\Delta t = 0$ and $\Delta t = 200$ μs. At the initial time instant, the reconstructed beamlet arrangement closely matches the experimental reference, with all six beamlets clearly resolved and maintaining their spatial distribution throughout the measurement volume.

At $\Delta t = 200$ μs, the deformation and lateral displacement of the beamlet pattern produced by the flow are clearly evident in both reconstructions. The FluidNeRF reconstruction successfully reproduces the overall beamlet trajectories, displacement patterns, and spatial deformation observed in the RLD reconstruction. The agreement between the two reconstructions demonstrates that the proposed framework accurately captures the temporal evolution of the experimental scalar field despite the limited-angle imaging geometry of the FIMic system.

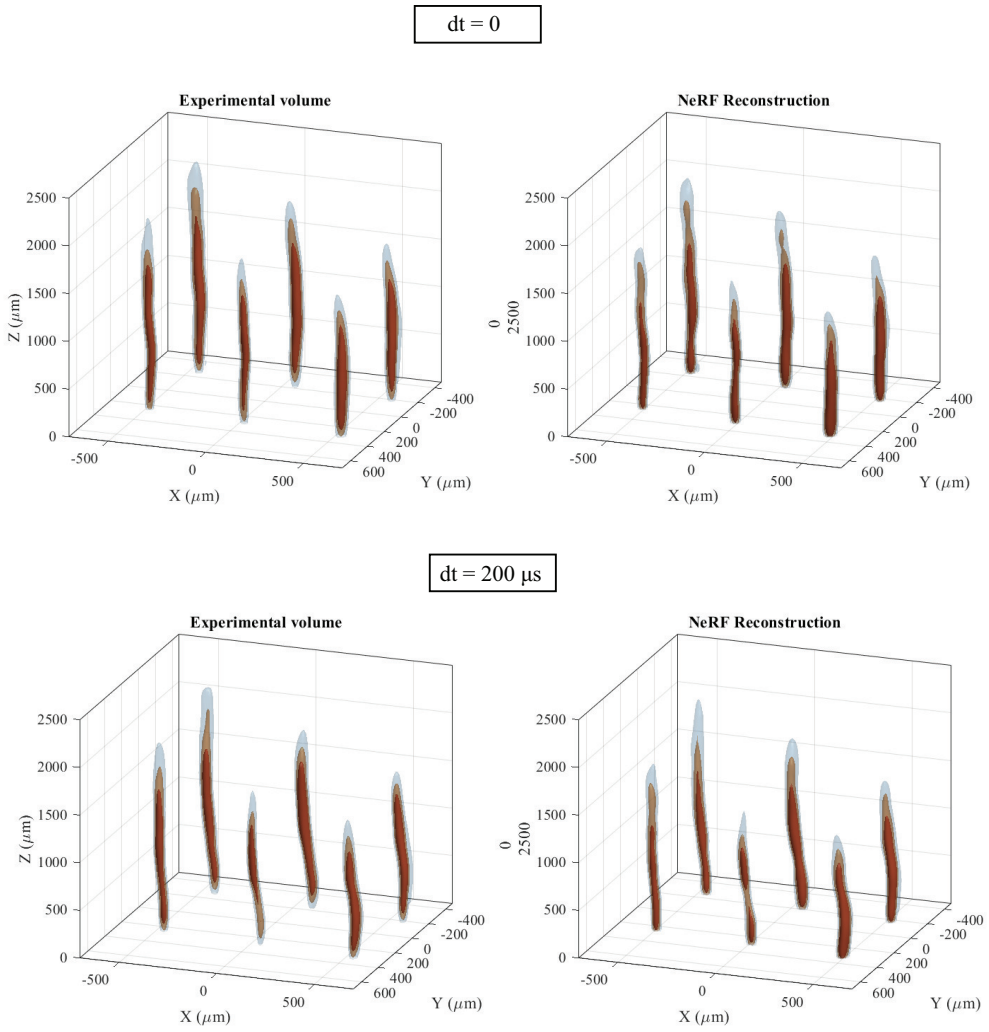


Figure 5.1: 3D isosurface reconstructions of the experimental MTV volume at $\Delta t = 0$ and $\Delta t = 200\ \mu$s. Left column: experimental volume reconstructed using Richardson–Lucy Deconvolution (RLD). Right column: corresponding FluidNeRF reconstructions using multiresolution hash encoding.

Although minor differences in beamlet boundary sharpness and local intensity distribution are observed between the two reconstructions, these differences are expected due to the limited angular diversity of the FIMic system and the fundamentally different reconstruction methodologies. Richardson–Lucy Deconvolution reconstructs the volume on a discrete voxel grid, whereas FluidNeRF represents the scalar field as a continuous neural implicit function. Overall, the results demonstrate that the proposed hash-encoded FluidNeRF framework successfully performs time-resolved volumetric reconstruction of ex-

perimental MTV measurements while producing reconstructions that are in close agreement with the established RLD reference.

5.3 Qualitative Comparison to RLD Reconstruction

Figure 5.2 compares two-dimensional slices of the experimental volumes reconstructed using Richardson–Lucy deconvolution (RLD) and the proposed FluidNeRF framework at both measurement instants. Figures 5.2(a) and 5.2(c) show transverse XY slices at $Z = 947$ μm, and Figures 5.2(b) and 5.2(d) show longitudinal XZ slices at $Y = 250$ μm. The top row corresponds to the initial instant $\Delta t = 0$ and the bottom row to the advected instant $\Delta t = 200$ μs. In the transverse slices at the initial instant, both reconstructions resolve the six beamlets as well-localized intensity peaks in the 3×2 arrangement. The peak positions in the FluidNeRF reconstruction match those of the RLD reference. At the advected instant, the beamlet peaks have shifted in response to the flow, and this displacement is reproduced by both reconstructions.

The longitudinal slices show the beamlets extending through the depth of the measurement volume. At the initial instant, the beamlets appear as near-vertical intensity bands in both reconstructions. At the advected instant, the bands are tilted and deformed by the flow, and the FluidNeRF reconstruction reproduces the same trajectories observed in the RLD reference. The background in the FluidNeRF slices is smoother than in the RLD slices, which reflects the continuous neural representation of the scalar field compared with the voxel-based RLD reconstruction. The position, separation, and deformation of the beamlets are in agreement between the two methods at both instants and across both slice orientations.

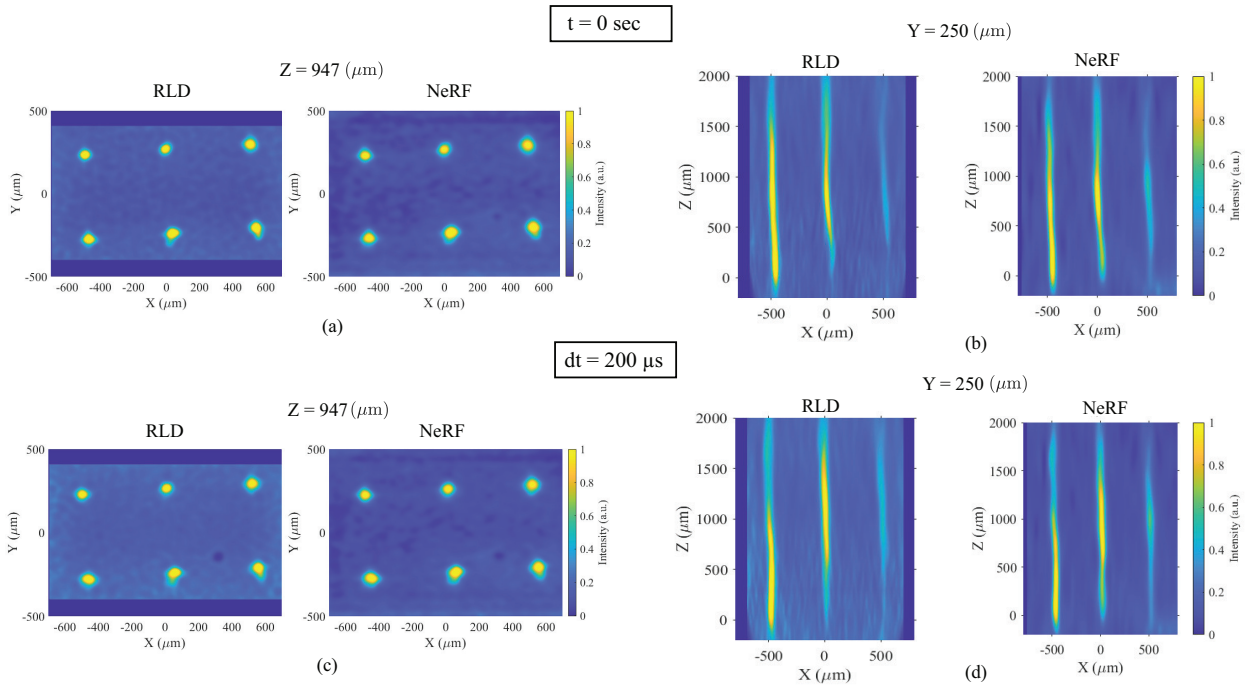


Figure 5.2: Two-dimensional slices of the experimental MTV volumes reconstructed using Richardson–Lucy deconvolution (RLD) and FluidNeRF at $\Delta t = 0$ and $\Delta t = 200$ μs. (a), (c) Transverse XY slices at $Z = 947$ μm. (b), (d) Longitudinal XZ slices at $Y = 250$ μm.

Figure 5.3 compares the normalized one-dimensional intensity profiles extracted along the X-direction through the beamlet cores for the experimental reconstructions. Profiles are shown at three depth planes, $Z = 188$ μm, $Z = 986$ μm, and $Z = 1624$ μm, at the initial instant $\Delta t = 0$ (top row) and the advected instant $\Delta t = 200$ μs (bottom row). Solid lines denote the experimental RLD reconstruction and dashed lines the FluidNeRF reconstruction.

At both instants and all three depths, the main peak of the FluidNeRF profile aligns with that of the RLD reference in both position and width. The peak location is recovered at each depth, and the shift in peak position between the two instants reflects the beamlet displacement produced by the flow. The agreement in the core of the profile is consistent across the depth of the measurement volume.

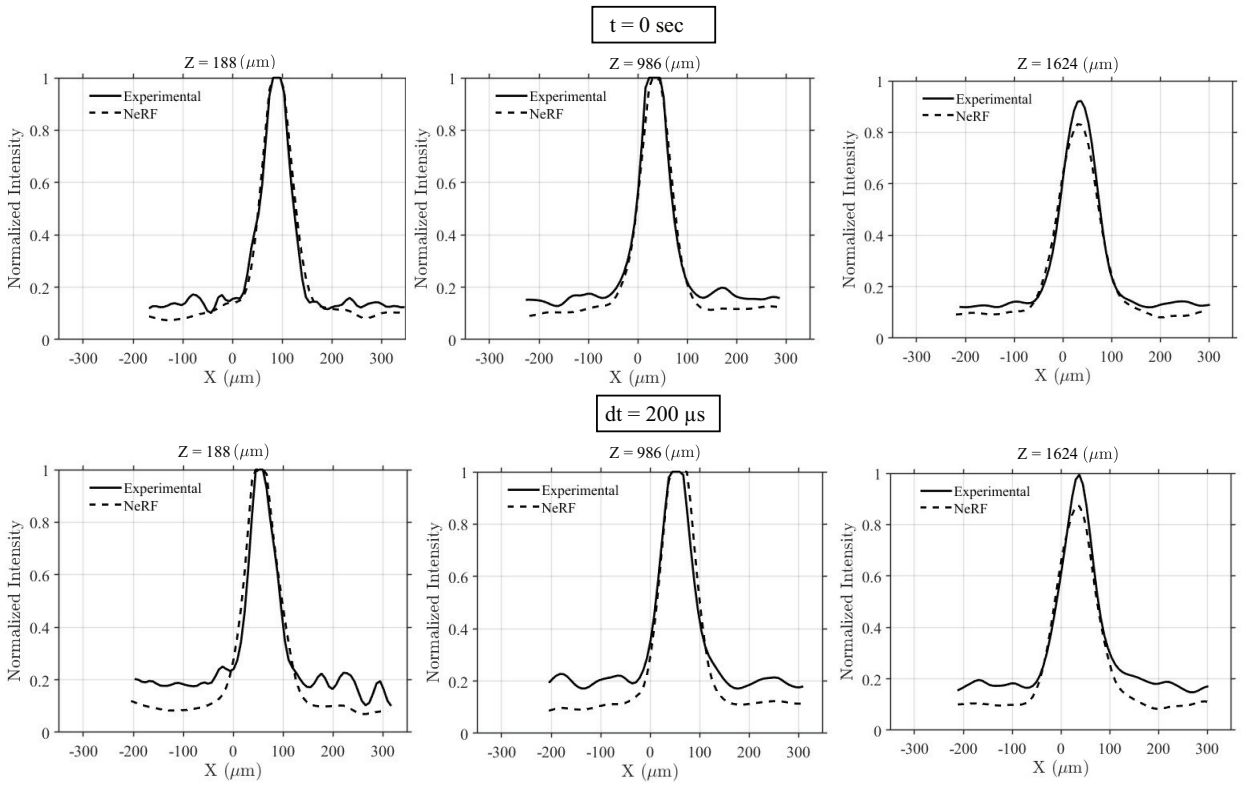


Figure 5.3: Comparison of reconstructed beamlet intensity profiles obtained using Fluid-NeRF with the experimental RLD reconstruction along the X-direction at three depth planes, $Z = 188$ μm, $Z = 986$ μm, and $Z = 1624$ μm. Top row: initial instant $\Delta t = 0$. Bottom row: advected instant $\Delta t = 200$ μs. Solid lines denote the experimental RLD reconstruction, and dashed lines represent the FluidNeRF reconstruction.

The two reconstructions differ in the background regions away from the beamlet cores. The RLD profiles show a higher and more irregular background level, whereas the FluidNeRF profiles have a lower and smoother background. This difference is more pronounced at the advected instant, where the RLD background exhibits stronger fluctuations. The lower background of the FluidNeRF reconstruction follows from the continuous neural representation of the scalar field, which does not reconstruct the volume on a discrete voxel grid and is therefore less sensitive to the voxel-level noise present in the RLD reconstruction. The agreement in the beamlet cores, combined with the reduced background, indicates that the reconstructed beamlet profiles are consistent with the reference while producing a cleaner intensity field.

Chapter 6

Conclusions

This thesis presented a Neural Radiance Field (NeRF)-based framework for the volumetric reconstruction of molecular tagging velocimetry (MTV) measurements acquired using a Fourier Integral Microscope (FIMic). The proposed framework extends the original FluidNeRF methodology by replacing conventional Fourier positional encoding with multiresolution hash encoding and adapting the reconstruction framework for efficient time-resolved scalar field reconstruction. The performance of the proposed method was evaluated using both synthetic and experimental datasets and compared with conventional reconstruction approaches.

6.1 Summary of Contributions

The primary objective of this research was to develop a computationally efficient neural reconstruction framework capable of accurately reconstructing three-dimensional scalar fields from the limited-angle projections produced by a Fourier Integral Microscope. The proposed framework was successfully demonstrated on both synthetic and experimental molecular tagging velocimetry datasets.

A synthetic plenoptic MTV dataset with known ground truth was first developed to provide a controlled environment for quantitatively evaluating reconstruction performance. Unlike previous experimental studies, the synthetic dataset enabled direct comparison between the reconstructed scalar field and the analytical solution using volumetric image-quality metrics. This dataset also provided a systematic platform for investigating the influence of network architecture, hash encoding parameters, ray-sampling parameters, and the number of projection views on the reconstruction accuracy and computational performance.

A comprehensive hyperparameter study was then performed to optimize the pro-

posed framework for sparse plenoptic MTV data. The investigation showed that many of the individual hash encoding parameters produced only negligible changes in reconstruction quality while having a significant influence on computational cost. A coupled investigation of the encoding levels and feature dimensions identified an optimized configuration consisting of 12 encoding levels and two features per level. Additional studies of the spatial resolution hierarchy, network architecture, sampling strategy, and projection views established a compact network configuration that achieved high reconstruction fidelity while substantially reducing training time.

Using the optimized configuration, the proposed framework successfully reconstructed synthetic Poiseuille flow volumes with high structural similarity to the analytical ground truth. The reconstructed beamlet geometry, intensity distributions, and beamlet widths closely matched the reference solution, demonstrating the ability of the proposed framework to accurately recover sparse three-dimensional scalar fields.

A direct comparison between multiresolution hash encoding and the original Fourier-encoded FluidNeRF implementation demonstrated that the proposed encoding strategy improved both reconstruction accuracy and computational efficiency. The hash-encoded implementation consistently achieved higher SSIM and PSNR values together with lower reconstruction error than the Fourier implementation. Furthermore, the optimized framework reduced the training time from approximately 225 minutes to less than 12 minutes, corresponding to nearly a 19-fold improvement in computational efficiency while maintaining excellent reconstruction fidelity.

Finally, the proposed framework was applied to experimental plenoptic MTV measurements obtained using a Fourier Integral Microscope. The reconstructed scalar fields showed good qualitative agreement with Richardson–Lucy deconvolution (RLD) reconstructions at both measurement instants. By incorporating normalized time directly as a network input, the proposed framework successfully reconstructed the temporal evolution of the experimental beamlet pattern within a single neural representation, demonstrating the

feasibility of time-resolved volumetric reconstruction using implicit neural representations.

6.2 Limitations

Although the proposed framework demonstrated accurate and computationally efficient reconstruction of sparse molecular tagging velocimetry data, several limitations remain.

First, the hyperparameter optimization presented in this work was performed using a synthetic Poiseuille flow dataset consisting of relatively sparse beamlet structures. While this simplified dataset enabled systematic evaluation of the reconstruction framework, more complex turbulent flow fields containing finer spatial structures may require different encoding and network configurations.

Second, the experimental validation was limited to a single stagnation jet dataset acquired using a Fourier Integral Microscope. Although good agreement with Richardson–Lucy deconvolution was observed, a quantitative assessment of the experimental reconstruction accuracy remains challenging because no ground-truth experimental scalar field is available.

Third, all reconstructions were performed using two measurement instants separated by a fixed time delay. Although the proposed framework incorporates time as a continuous network input, the capability of reconstructing longer time sequences containing many temporal states has not yet been investigated.

Finally, the present study focused on scalar field reconstruction rather than direct velocity estimation. The reconstructed beamlet volumes provide the foundation for subsequent velocity calculations, but the integration of velocity extraction within the neural reconstruction framework remains outside the scope of the present work.

A further limitation is that the current framework does not model diffraction. The forward model maps each sample point to the sensor through the calibrated projection geometry alone, without accounting for the diffraction-limited response of the microscope

objective. This is justified in the present work because the beamlet diameter of $90\ \mu$m is much larger than the diffraction-limited spot size of the objective, so diffraction has a negligible effect on the recorded beamlet pattern. For finer tagged structures, however, the diffraction-limited spot size sets a lower bound on the spatial resolution that the FIMic and FluidNeRF combination can achieve, and neglecting it would cause the reconstruction to underestimate the true blur of the imaging system.

6.3 Future Work

Several opportunities exist to further improve and extend the proposed framework.

The most immediate extension is the application of the framework to more challenging synthetic datasets generated from direct numerical simulation (DNS), such as the Johns Hopkins Turbulence Database. These datasets provide significantly more complex flow structures than the Poiseuille flow considered in this work and will enable evaluation of the proposed framework under realistic turbulent flow conditions.

Future work should also investigate reconstruction from longer time sequences by incorporating additional temporal measurements during training. Such studies would further exploit the continuous spatiotemporal representation provided by Neural Radiance Fields and may improve reconstruction accuracy through temporal regularization.

Although multiresolution hash encoding substantially reduced the computational cost, additional improvements may be achieved through alternative neural field architectures, adaptive ray sampling strategies, occupancy-grid optimization, or more advanced neural encodings. Investigation of these approaches could further reduce reconstruction time while maintaining or improving reconstruction fidelity.

The proposed framework may also be integrated directly with velocity estimation algorithms such as optical flow or beamlet tracking, enabling an end-to-end neural pipeline that reconstructs both the scalar field and the corresponding three-dimensional velocity field.

Another important extension is the incorporation of a diffraction-aware forward model into the FluidNeRF framework. The current implementation maps sample points to the sensor through the calibrated projection geometry alone and does not account for the diffraction-limited response of the microscope objective. A ray-tracing forward model that includes the point spread function of the FIMic system, either measured experimentally or computed from the objective and microlens parameters, would allow the network to reconstruct the underlying scalar field rather than the diffraction-blurred field. Because the FluidNeRF forward model is differentiable and represents the field continuously, this optical response can be applied to the rendered projections within the existing training loop. Such a model would become increasingly important at higher magnifications, where the diffraction-limited spot size approaches the size of the tagged structures and sets the limit on the achievable spatial resolution.

Finally, future experimental studies should apply the proposed framework to a wider range of flow configurations, including highly turbulent flows, near-wall boundary layers, and complex engineering flow fields. Demonstrating robust performance under these conditions would further establish multiresolution hash encoding as a practical reconstruction framework for time-resolved volumetric molecular tagging velocimetry.

References

[1] S. B. Pope. “Turbulent flows.” In: *Measurement Science and Technology* 12.11 (2001), pp. 2020–2021.

[2] P. D. Huck, M. J. Yamakaitis, C. Fort, and P. M. Bardet. “Near-wall volumetric molecular tagging velocimetry with a Fourier integral microscope.” In: *Experiments in Fluids* 66.8 (2025), p. 152.

[3] M. Hultmark, M. Vallikivi, S. C. C. Bailey, and A. Smits. “Turbulent pipe flow at extreme Reynolds numbers.” In: *Physical review letters* 108.9 (2012), p. 094501.

[4] C. E. Willert, J. Soria, M. Stanislas, J. Klinner, O. Amili, M. Eisfelder, C. Cuvier, G. Bellani, T. Fiorini, and A. Talamelli. “Near-wall statistics of a turbulent pipe flow at shear Reynolds numbers up to 40 000.” In: *Journal of Fluid Mechanics* 826 (2017), R5.

[5] G. F. Oweis, E. S. Winkel, J. M. Cutbrith, S. L. Ceccio, M. Perlin, and D. R. Dowling. “The mean velocity profile of a smooth-flat-plate turbulent boundary layer at high Reynolds number.” In: *Journal of fluid mechanics* 665 (2010), pp. 357–381.

[6] E. S. Winkel, J. M. Cutbirth, S. L. Ceccio, M. Perlin, and D. R. Dowling. “Turbulence profiles from a smooth flat-plate turbulent boundary layer at high Reynolds number.” In: *Experimental thermal and fluid science* 40 (2012), pp. 140–149.

[7] F. Li, H. Zhang, and B. Bai. “A review of molecular tagging measurement technique.” In: *Measurement* 171 (2021), p. 108790.

[8] M. Koochesfahani, A. Goh, and H. Schock. “Molecular Tagging Velocimetry (MTV) and its automotive applications.” In: *The Aerodynamics of Heavy Vehicles: Trucks, Buses, and Trains*. Springer, 2004, pp. 143–155.

[9] P. D. Huck, C. Fort, and P. M. Bardet. “Wall shear stress measurements using a MTV-based plenoptic microscope.” In: *AIAA SCITECH 2023 Forum*. 2023, p. 2491.

[10] D. L. Kelly and B. S. Thurow. “FluidNeRF: a scalar-field reconstruction technique for flow diagnostics using neural radiance fields.” In: *AIAA SciTech 2023 Forum*. 2023, p. 0412.

[11] F. Scarano. “Tomographic PIV: principles and practice.” In: *Measurement Science and Technology* 24.1 (2013), p. 012001.

[12] D. Kelly and B. Thurow. “Investigation of a neural implicit representation tomography method for flow diagnostics.” In: *Measurement Science and Technology* 35.5 (2024), p. 056007.

[13] D. L. Kelly and B. S. Thurow. “Experimental Proof-of-Concept of a Time-Resolved Gridless Tomography Method for 3D Scalar Fields.” In: *AIAA SCITECH 2025 Forum*. 2025, p. 1059.

[14] T. Müller, A. Evans, C. Schied, and A. Keller. “Instant neural graphics primitives with a multiresolution hash encoding.” In: *ACM transactions on graphics (TOG)* 41.4 (2022), pp. 1–15.

[15] M. Koochesfahani, R. Cohn, C. Gendrich, and D. Nocera. “Molecular tagging diagnostics for the study of kinematics and mixing in liquid phase flows.” In: *Developments in Laser Techniques and Fluid Mechanics: Selected Papers from the 8th International Symposium, Lisbon, Portugal 8–11 July, 1996*. Springer. 1997, pp. 125–144.

[16] M. R. Edwards, A. Dogariu, and R. B. Miles. “Simultaneous temperature and velocity measurements in air with femtosecond laser tagging.” In: *AIAA journal* 53.8 (2015), pp. 2280–2288.

[17] J. Ke and D. Bohl. “Effect of experimental parameters and image noise on the error levels in simultaneous velocity and temperature measurements using molecular tagging velocimetry/thermometry (MTV/T).” In: *Experiments in fluids* 50.2 (2011), pp. 465–478.

[18] R. Basu, A. M. Naguib, and M. M. Koochesfahani. “Feasibility study of whole-field pressure measurements in gas flows: molecular tagging manometry.” In: *Experiments in fluids* 49.1 (2010), pp. 67–75.

[19] R. J. Adrian and J. Westerweel. *Particle image velocimetry*. 30. Cambridge university press, 2011.

[20] J. R. Elsnab, J. P. Monty, C. M. White, M. M. Koochesfahani, and J. C. Klewicki. “Efficacy of single-component MTV to measure turbulent wall-flow velocity derivative profiles at high resolution.” In: *Experiments in Fluids* 58.9 (2017), p. 128.

[21] M. Koochesfahani. “Molecular Tagging Velocimetry (MTV)-progress and applications.” In: *30th Fluid Dynamics Conference*. 1999, p. 3786.

[22] T. Fahringer and B. Thurow. “Tomographic reconstruction of a 3-D flow field using a plenoptic camera.” In: *42nd AIAA fluid dynamics conference and exhibit*. 2012, p. 2826.

[23] K. E. Stull, M. L. Tise, Z. Ali, and D. R. Fowler. “Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images.” In: *Forensic science international* 238 (2014), pp. 133–140.

[24] D. Mishra, K. Muralidhar, and P. Munshi. “A robust MART algorithm for tomographic applications.” In: *Numerical Heat Transfer: Part B: Fundamentals* 35.4 (1999), pp. 485–506.

[25] N. Jain, A. Raj, M. Kalra, P. Munshi, and V. Ravindran. “MART algorithms for circular and helical cone-beam tomography.” In: *Research in Nondestructive Evaluation* 22.3 (2011), pp. 147–168.

[26] K. Lynch and F. Scarano. “An efficient and accurate approach to MTE-MART for time-resolved tomographic PIV.” In: *Experiments in Fluids* 56.3 (2015), p. 66.

[27] B. S. Kumar and G. Nagarajan. “Tomographic image reconstruction using SART algorithm.” In: *International Journal of Medical Engineering and Informatics* 8.3 (2016), pp. 239–248.

[28] X. Zhu, Z. Wu, J. Li, B. Zhang, and C. Xu. “A pre-recognition SART algorithm for the volumetric reconstruction of the light field PIV.” In: *Optics and Lasers in Engineering* 143 (2021), p. 106625.

[29] F. Ströhl and C. F. Kaminski. “A joint Richardson—Lucy deconvolution algorithm for the reconstruction of multifocal structured illumination microscopy data.” In: *Methods and Applications in Fluorescence* 3.1 (2015), p. 014002.

[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.” In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[31] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan. “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields.” In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 5855–5864.

[32] F. Engelmann, F. Manhardt, M. Niemeyer, K. Tateno, M. Pollefeys, and F. Tombari. “OpenNeRF: open set 3D neural scene segmentation with pixel-wise features and rendered novel views.” In: *arXiv preprint arXiv:2404.03650* (2024).

[33] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman. “Zip-nerf: Anti-aliased grid-based neural radiance fields.” In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 19697–19705.

[34] F. Warburg, E. Weber, M. Tancik, A. Holynski, and A. Kanazawa. “Nerfbusters: Removing ghostly artifacts from casually captured nerfs.” In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 18120–18130.

[35] S. Fridovich-Keil, G. Meanti, F. R. Warburg, B. Recht, and A. Kanazawa. “K-planes: Explicit radiance fields in space, time, and appearance.” In: *Proceedings*

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023, pp. 12479–12488.

[36] S. Wu, S. Basu, T. Broedermann, L. Van Gool, and C. Sakaridis. “PBR-NeRF: Inverse Rendering with Physics-Based Neural Fields.” In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, pp. 10974–10984.

[37] R.-Z. Qiu, G. Yang, W. Zeng, and X. Wang. “Feature splatting: Language-driven physics-based scene synthesis and editing.” In: *arXiv preprint arXiv:2404.01223* (2024).

[38] L. Song, A. Chen, Z. Li, Z. Chen, L. Chen, J. Yuan, Y. Xu, and A. Geiger. “Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields.” In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2732–2742.

[39] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville. “On the spectral bias of neural networks.” In: *International conference on machine learning*. PMLR. 2019, pp. 5301–5310.

[40] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng. “Fourier features let networks learn high frequency functions in low dimensional domains.” In: *Advances in neural information processing systems* 33 (2020), pp. 7537–7547.

[41] Z. Deng, H. Xiao, Y. Lang, H. Feng, and J. Zhang. “Multi-scale hash encoding based neural geometry representation.” In: *Computational Visual Media* 10.3 (2024), pp. 453–470.

[42] T. Müller. *tiny-cuda-nn*. Version 2.0. Apr. 2021. URL: `https://github.com/NVlabs/tiny-cuda-nn`.

[43] M. Lee and R. D. Moser. “Direct numerical simulation of turbulent channel flow up to.” In: *Journal of fluid mechanics* 774 (2015), pp. 395–415.

[44] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[45] G. T. Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[46] S. J. Kisner, E. Haneda, C. A. Bouman, S. Skatter, M. Kourinny, and S. Bedford. “Limited view angle iterative CT reconstruction.” In: *Computational Imaging X*. Vol. 8296. SPIE. 2012, pp. 66–74.

[47] J. Frikel and E. T. Quinto. “Characterization and reduction of artifacts in limited angle tomography.” In: *Inverse Problems* 29.12 (2013), p. 125007.

Appendix 1

Implementation Details

This appendix describes the implementation used to generate the synthetic FIMic dataset.

The generation code is implemented in Python. Numerical arrays and random sampling are handled with `NumPy`, the affine calibration coefficients are read from a CSV file using `pandas`, and image smoothing is performed with `scipy.ndimage.gaussian_filter`. Particle advection uses the `solve_ivp` routine from `scipy.integrate`, and the ground-truth volumes are written to MATLAB `.mat` files using `scipy.io.savemat`. Synthetic sensor images are written as 16-bit TIFF files. A fixed random seed is used so that the generated dataset is reproducible.

The number of sample points per beamlet is determined by an iterative convergence check rather than fixed in advance. After each batch is projected onto the sensor, the accumulated sensor image is normalized by its peak value and compared with the normalized image from the previous batch. The normalized root-mean-square change between successive images is computed and monitored. Sampling continues until this change falls below the convergence threshold or until a maximum number of points per beamlet is reached. The threshold and batch size are set so that convergence is reached at the point-count reported in Chapter 3. The convergence history is recorded and written to a diagnostic plot. Figure A1.1 shows the normalized root-mean-square change as a function of the number of sample points per beamlet. The change decreases monotonically as points are added and falls below the convergence threshold at the reported point-count, confirming that the projected sensor image has stabilized and that further sampling produces no appreciable change.

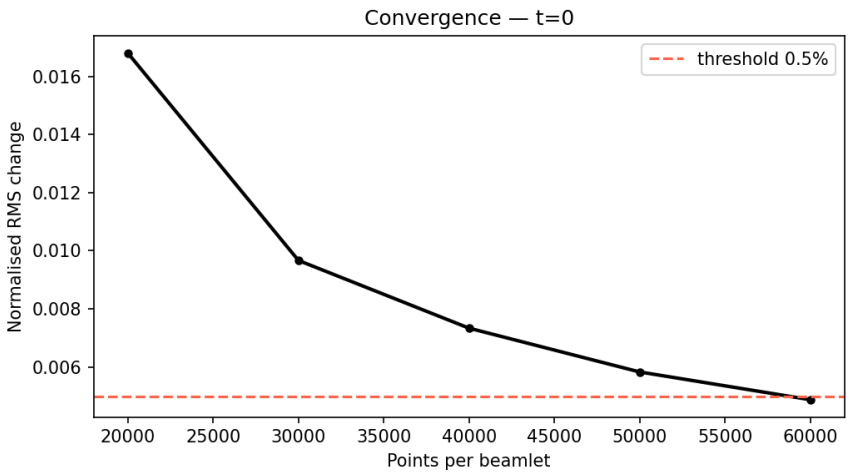


Figure A1.1: Monte Carlo convergence of the synthetic sensor image at $t = 0$. The normalized root-mean-square change between successive batches decreases as sample points are added and falls below the convergence threshold at the selected point-count.

The computed pixel coordinates are rounded to the nearest integer, and sample points falling outside the sensor bounds are discarded. The weights of the remaining points are accumulated into the sensor image using scatter-add, which correctly sums the contributions of multiple sample points that project to the same pixel. Both a combined full-sensor image and the individual per-view images are accumulated in the same pass.