GauSmoke: Hybrid Physics-Optical Gaussian Splatting for Sparse Smoke Reconstruction

WENRAN ZHANG, Nankai University, China
YUXIANG CAI, Nankai University, China
LETIAN HUANG, Nanjing University, China
DONGWEI YE, Nanjing University, China
JIE GUO, Nanjing University, China
BO REN*, Nankai University, China

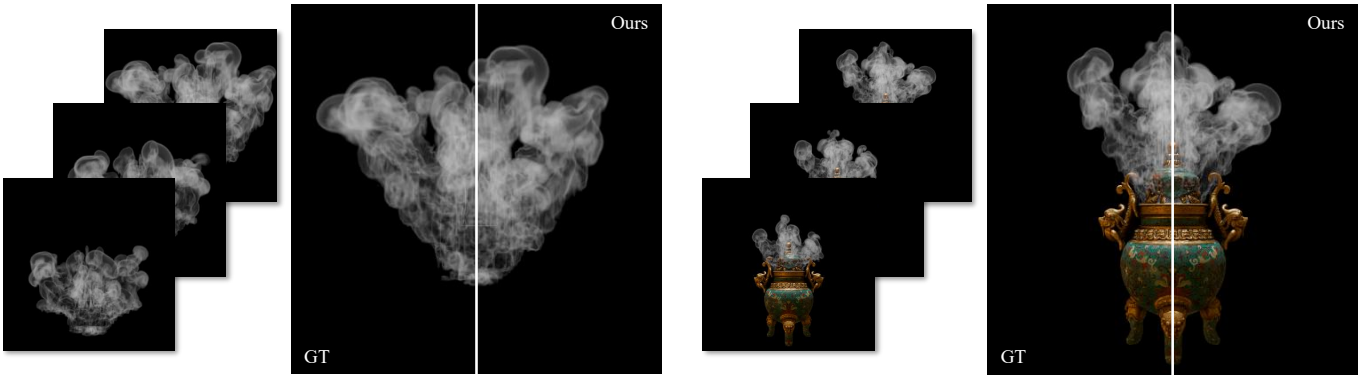


Fig. 1. **Smoke reconstruction from sparse views.** We propose a hybrid fluid representation that incorporates physical constraints into Gaussian Splatting, enabling joint reconstruction of velocity and density fields. Our method captures swirling structures and fine smoke details under sparse supervision, while supporting unified reconstruction of dynamic smoke and static solids.

Reconstructing dynamic fluids from sparse-view videos is a severely under-constrained problem due to complex volumetric visibility and turbulent, topology-changing motion. Current dynamic deformation-based 3D Gaussian Splatting (3DGS) methods use heuristic image-space warps, which often overfit limited observations via non-physical primitive scaling or drifting, producing floaters and geometric distortions. To address these challenges, we propose a hybrid, physics-aware fluid representation that injects physical constraints directly into the 3DGS pipeline. We develop a Beer–Lambert volumetric model mapping Gaussian primitives to a continuous, physically-grounded density field, bridging optical observations with fundamental physical laws. A physically-guided optimization evolves primitives under a global, volume-preserving velocity field, avoiding arbitrary per-particle deformations and effectively preserving volume while recovering fine-scale turbulent details. This unified formulation tightly couples fluid dynamics with differentiable rendering, enabling high-quality, joint predictions of velocity and density. Experiments on synthetic and real smoke datasets demonstrate that our approach surpasses state-of-the-art methods in both visual fidelity and physical consistency, achieving stable 3D reconstructions even from sparse views.

CCS Concepts: • **Computing methodologies** → **Image-based rendering**.

Additional Key Words and Phrases: 3D Gaussian Splatting, Smoke Reconstruction, Sparse-View

ACM Reference Format:
Wenran Zhang, Yuxiang Cai, Letian Huang, Dongwei Ye, Jie Guo, and Bo Ren. 2026. GauSmoke: Hybrid Physics-Optical Gaussian Splatting for Sparse Smoke Reconstruction. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3799902.3811148

*Corresponding author.

Authors' Contact Information: Wenran Zhang, zwr@mail.nankai.edu.cn, VCIP, College of Computer Science, Nankai University, Tianjin, China; Yuxiang Cai, caiyuxiang@mail.nankai.edu.cn, VCIP, College of Computer Science, Nankai University, Tianjin, China; Letian Huang, lthuang@smail.nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Dongwei Ye, dongweiye@smail.nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Jie Guo, guojie@nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Bo Ren, rb@nankai.edu.cn, VCIP, College of Computer Science, Nankai University, Tianjin, China.

This work is licensed under a Creative Commons Attribution 4.0 International License.
SIGGRAPH Conference Papers '26, Los Angeles, CA, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2554-8/2026/07
https://doi.org/10.1145/3799902.3811148

1 Introduction

Fluid phenomena such as smoke are ubiquitous in natural environments and visual effects, exhibiting far more complex shapes and motions than rigid or deformable solids. Classical graphics and CFD pipelines generate fluid motion using numerical forward solvers, typically based on Eulerian grids or Lagrangian particles [Koschier et al. 2020; Macklin et al. 2014; Park and Kim 2005]. These solvers evolve primary physical variables—velocity and pressure—while observable quantities like smoke density (soot concentration) are

treated as passive scalars advected by the flow [Bridson 2015]. Although forward solvers can faithfully simulate fluids under fully specified initial and boundary conditions, such information is rarely available in real-world scenarios. Consequently, recent work has increasingly focused on inverse reconstruction of fluids directly from RGB videos, casting fluid modeling as an optimization problem.

Despite advances in novel view synthesis, dominant 3D representations such as NeRF [Mildenhall et al. 2020] and 3DGS [Kerbl et al. 2023] are primarily optimized for static appearance reconstruction via differentiable rendering, lacking explicit motion. Recent dynamic 3DGS methods [Wu et al. 2024; Yang et al. 2024a] introduce deformation fields that warp primitives from a canonical space, but they assume topology-preserving, near-rigid motion, conflicting with the stochastic, topology-changing nature of turbulent smoke. Moreover, in-the-wild smoke capture is often limited by sparse views, further increasing the difficulty of reconstruction. Without physical constraints, deformation-based models easily overfit and produce blurry smoke structures in sparse-view scenarios.

To overcome the ill-posedness of sparse-view fluid reconstruction, a physically grounded representation governed by fluid laws is essential. Swirling and turbulent smoke structures critically rely on incompressibility, requiring the global velocity field to be divergence-free, i.e., $\nabla \cdot \mathbf{u} = 0$. Existing 3DGS-based frameworks, however, remain largely rendering-oriented, relying on heuristic image-space splatting without explicitly modeling or constraining the underlying dynamics. To address this, we propose a novel framework that imposes physical constraints on both motion estimation and rendering, yielding a hybrid representation that couples optical observations with temporally consistent fluid evolution.

Our approach achieves this coupling via two synergistic components. Optically, we develop a volumetric rendering model that represents smoke density with Gaussian primitives. By applying the Beer–Lambert law to enforce exponential extinction along 3D ray paths, we recover a continuous and globally consistent density field. This volumetric formulation further bridges the Eulerian and Lagrangian perspectives: we simultaneously reconstruct a reliable velocity field on an Eulerian grid while tracing Lagrangian Gaussian particles. By evolving Gaussian motion within this globally coordinated, volume-preserving field according to the transport equation, we avoid arbitrary per-particle deformations and steer the optimization toward physically admissible solutions. This dual representation allows Gaussian primitives to carry physically meaningful properties, such as density, while retaining the flexibility of the 3DGS optimization pipeline. Additionally, we introduce a refinement strategy to capture fine-scale turbulent structures. By explicitly coupling the dynamical evolution of Gaussian parameters with the underlying flow motions, our method achieves faithful smoke reconstructions even from sparse-view inputs.

Our contributions are summarized as follows:

- We propose a hybrid, physics-aware fluid representation based on a Beer–Lambert volumetric Gaussian rendering model, mapping Gaussian primitive parameters to physically meaningful quantities and unifying physical constraints with differentiable rendering.
- We establish a physically-guided Gaussian optimization scheme that evolves primitives under a divergence-free velocity field, preserving volume and capturing fine-scale turbulent smoke details.
- We achieve high-quality joint estimation of velocity and density from sparse-view inputs, outperforming existing dynamic reconstruction methods in both visual fidelity and dynamic consistency.

2 Related Work

2.1 Neural Scene Representations

NeRF [Mildenhall et al. 2020] and 3D Gaussian Splatting [Kerbl et al. 2023] represent static scenes using implicit and explicit volumetric primitives, respectively. Subsequent 3DGS-based methods achieve faster rendering than implicit representations through a GPU-friendly splatting pipeline, while explicit primitives also provide convenient spatial control. Building on this success, a line of dynamic extensions introduces a temporal dimension to 3DGS [Liang et al. 2025; Wu et al. 2024; Yang et al. 2024a] by incorporating deformation networks that predict per-Gaussian displacements from a canonical space. This paradigm has demonstrated strong results for articulated objects and human bodies [Hu et al. 2024], especially under multi-view capture settings.

However, deformation-driven dynamic 3DGS is less suitable for turbulent smoke, especially under sparse-view supervision. Unlike rigid or articulated motion, smoke is not naturally represented as deformation from a canonical shape, but is better characterized as a time-varying density transported by a velocity field, with topology changes and approximate incompressibility.

2.2 Sparse-View Regularization for 3DGS

Sparse-view Gaussian reconstruction is ill-posed due to depth ambiguity and insufficient constraints due to the limited views, and recent 3DGS-based methods address this issue with additional priors. Depth-guided methods [Li et al. 2024; Wan et al. 2025; Xu et al. 2025] regularize Gaussian geometry using monocular or SfM-derived depth cues, thereby improving geometric consistency when observations are limited. In parallel, generative methods [Wang et al. 2025; Yang et al. 2024b; Yu et al. 2024b] effectively leverage learned image priors to compensate for missing views and reduce ambiguity beyond observed cameras.

These approaches substantially improve sparse-view reconstruction for static or general dynamic scenes, but they are not designed for time-varying participating media such as smoke. In particular, they do not explicitly account for volumetric image formation or transport-driven fluid motion, and therefore do not provide physically meaningful constraints for smoke reconstruction. Our method instead addresses sparse-view ambiguity through continuous volumetric rendering and physically grounded transport priors.

2.3 Neural Fluid Reconstruction

These challenges become particularly severe for smoke, where dynamic ambiguity and sparse-view incompleteness jointly make reconstruction highly ill-posed. As a result, prior fluid reconstruction methods have long relied on volumetric representations and

physics-aware priors. Neural Volumes [Lombardi et al. 2019] learns dynamic renderable volumes from multi-view images, while TomoFluid [Zang et al. 2020] and Global Transport [Franz et al. 2021] recover density and motion under sparse observations with stronger transport and physics-based constraints. More recent methods further move toward learned motion inference from limited views [Franz et al. 2023; Qiu et al. 2021].

NeRF-based approaches extend this line with implicit volumetric representations. Physics-informed methods [Chu et al. 2022] incorporate velocity fields into neural reconstruction, while later work enhances motion modeling with additional physical modules [Yu et al. 2023]. Voxel-assisted variants such as NeuSmoke [Qiu et al. 2024] reduce costly velocity-related derivative computations and deep MLP queries, improving efficiency to some extent, but remain largely implicit and limited in controllability.

Recently, 3DGS has also been applied to fluid modeling through explicit Gaussian primitives. In particular, FluidNexus [Gao et al. 2025] targets single-video 3D fluid reconstruction by synthesizing novel-view videos and then optimizing a physics-integrated visual-physical particle representation. Other concurrent Gaussian-based methods [Du et al. 2025; Xing et al. 2025] likewise provide a more efficient and spatially explicit alternative to implicit neural fields, but primarily model motion at the particle level, making it nontrivial to impose global transport and incompressibility constraints. FluidGS [Xie et al. 2025] introduces an Eulerian grid but models opacity rather than physical density. In contrast, our method couples Gaussian primitives with a continuous density field, Beer–Lambert rendering, and physically constrained transport for sparse-view smoke reconstruction.

2.4 Physics-Aware Dynamic Representations

Beyond reconstruction, recent work has explored learned representations for fluid dynamics through reduced-order models and differentiable simulation. Neural implicit reduced fluid simulation [Tao et al. 2024] models fluid evolution in a structured latent space with implicit neural geometry, while neural kinematic bases [Liu et al. 2025] learn physically constrained basis fields for real-time fluid animation; hierarchical differentiable fluid simulation [Kong et al. 2025] instead improves the scalability of gradient-based fluid control. These works highlight the broader value of combining compact dynamic representations with physical constraints, which also motivates our formulation.

3 Preliminary

Differentiable 3D Gaussian Splatting. Following the formulation of differentiable 3D Gaussian Splatting (3DGS) [Kerbl et al. 2023], a scene is represented as a collection of anisotropic 3D Gaussian primitives. Each Gaussian primitive is parameterized by a mean position $\mathbf{p}_i$ and a positive-definite covariance matrix $\mathbf{\Sigma}_i$, which together define a continuous volumetric kernel:

$$k_i(\mathbf{x}) = \exp\left(-\tfrac{1}{2}(\mathbf{p}_i - \mathbf{x})^\top \mathbf{\Sigma}_i^{-1}(\mathbf{p}_i - \mathbf{x})\right). \tag{1}$$

In addition, each 3D Gaussian is associated with a color c_i and an opacity o_i. Rendering is performed in image space, where each 3D Gaussian is projected onto the image plane via Elliptical Weighted Average (EWA) splatting [Zwicker et al. 2002] and composited to produce final pixel colors.

Specifically, the projection is approximated by the Jacobian of a projective transformation under an affine assumption. The resulting 2D covariance matrix is obtained by skipping the third row and column of the transformed 3D covariance, yielding an ellipse in the image plane. Differentiable rendering is then performed by alpha blending,

$$C = \sum_{i=1}^{N} c_i \hat{o}_i \prod_{j=1}^{i-1}(1 - \hat{o}_j), \tag{2}$$

where $\hat{o}_i$ denotes the 2D opacity obtained by multiplying the volumetric opacity o_i with the corresponding projected 2D Gaussian kernel $k_i^{2D}(\mathbf{x})$.

Volumetric Rendering. Volumetric rendering [Kajiya and Von Herzen 1984] models a radiance field as a continuous participating medium defined over 3D space by a density field $\sigma(\mathbf{x})$ and a view-dependent radiance function $c(\mathbf{x}, \mathbf{d})$. Given a camera ray parameterized as $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$, the accumulated color is obtained by integrating the volumetric contributions of $\sigma(\mathbf{x})$ along the ray:

$$C(\mathbf{r}) = \int_0^\infty T(0,s)\,\sigma(\mathbf{r}(s))\,c(\mathbf{r}(s),\mathbf{d})\,ds, \tag{3}$$

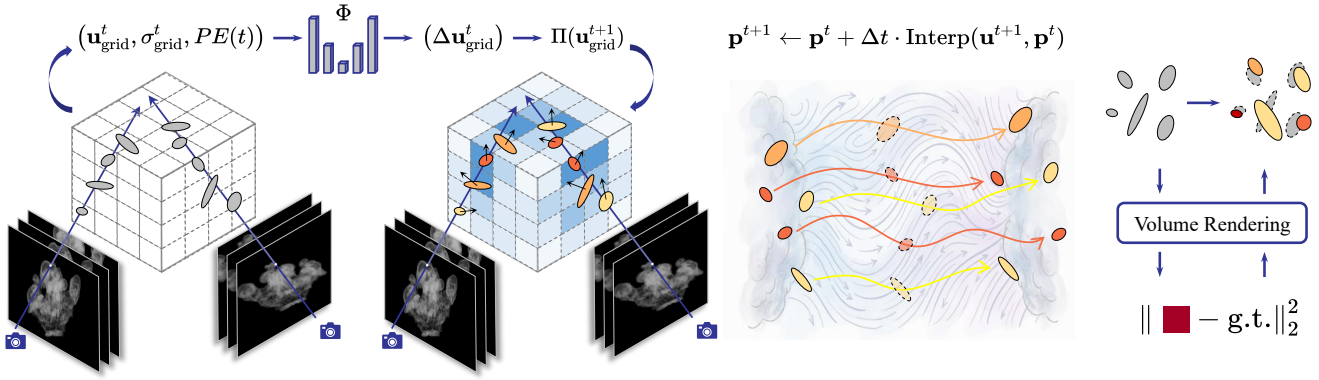
$$T(0,s) = \exp\left(-\int_0^s \sigma(\mathbf{r}(\tau))\,d\tau\right), \tag{4}$$

Following the Beer–Lambert law, where $T(0,s)$ denotes the accumulated transmittance along the ray segment. Besides, $\sigma(\mathbf{x})$ denotes the volumetric extinction coefficient (often informally referred to as "density") that quantifies light attenuation due to absorption and scattering, and thus directly reflects the concentration of participating media such as smoke or fog. This formulation highlights that the density field $\sigma(\mathbf{x})$ is the sole physical quantity governing volumetric visibility and occlusion.

4 Neural Scene Representation for Fluids

Recovering 3D fluid evolution from sparse multi-view videos remains highly challenging. Leveraging the fast rendering capabilities of 3D Gaussian Splatting (3DGS), we augment the discrete particle representation with a continuous physical field. This hybrid model integrates physical dynamics with optical observations, guiding 3DGS optimization with constraints to enable faithful smoke reconstruction from sparse observations.

As shown in Fig. 2, our framework uses a dual-constraint approach that enforces optical consistency via the Beer–Lambert law and physical plausibility through incompressible flow. We reinterpret 3DGS as a continuous density field, where Gaussian primitives act as volumetric basis functions, ensuring exponential extinction along rays (Sec. 4.1). A hybrid Eulerian-Lagrangian architecture governs motion: a neural Eulerian grid computes a divergence-free velocity field (Sec. 4.2), while Gaussian primitives advection volumetric attributes (Sec. 4.3). This coupling preserves fine-scale structures and anisotropic deformations without numerical diffusion. Together with a joint optimization strategy (Sec. 4.4), these components enable the reconstruction of physically consistent and visually faithful 3D smoke fields from sparse RGB inputs via differentiable rendering



(a) Volumetric Density Representation **(b) Neural Velocity Field** **(c) Gaussian Evolution** **(d) Joint Optimization**

Fig. 2. **Overview of the GauSmoke framework.** (a) Given sparse-view inputs, we use a Beer-Lambert volumetric formulation to reinterpret 3D Gaussian primitives into a continuous density field. (b) A neural network Φ estimates an incremental velocity $\Delta\mathbf{u}^t_{\text{grid}}$ update on an Eulerian grid by perceiving the current density distribution. The field is then projected via $\Pi(\cdot)$ to satisfy the divergence-free (incompressibility) constraint. (c) The velocity field advects Gaussian primitives, thereby integrating physical fluid dynamics into the 3DGS optimization. (d) Rendering and physical losses are jointly optimized for end-to-end density and velocity refinement.

4.1 Volumetric Gaussian Density Field

To address the severe reconstruction ambiguity and ill-posedness inherent in sparse-view smoke capture, and inspired by [Talegaonkar et al. 2025], we propose an enhanced volumetric Gaussian rendering model within the 3DGS framework (Fig. 2 (a)). Standard 3DGS relies on a 2D splatting operation for fast rasterization, which approximates transmittance linearly, neglecting self-occlusion of Gaussian primitives and losing depth information. This simplification compromises the physical accuracy of the reconstruction.

To overcome this, we incorporate the Beer–Lambert law into the 3DGS pipeline, explicitly reinterpreting Gaussian primitive parameters as a continuous density field. By performing an accurate integral of the extinction along 3D ray paths, our model ensures rigorous exponential extinction, regularizing the solution space and enabling the recovery of physically consistent volumetric structures that remain stable across unseen viewpoints. This volumetric design forms the first pillar of our hybrid approach, narrowing the search space to plausible 3D density distributions and enabling a tight coupling with fluid dynamics.

Continuous Density Representation. Our key insight is that fluid phenomena require consistency in continuous physical fields, whereas 3DGS offers an efficient Lagrangian particle representation. To bridge these two perspectives, we parameterize the continuous smoke density field $\sigma(\mathbf{x})$ by treating the 3D Gaussian primitives as volumetric basis functions:

$$\sigma(\mathbf{x}) = \sum_{i=1}^{N} \sigma_i \, k_i(\mathbf{x}), \tag{5}$$

where σ_i is a physical density coefficient assigned to each Gaussian kernel k_i. By expressing the density as a smooth superposition of kernels, we ensure that the field is everywhere differentiable. This enables the density to be transported by the learned velocity field in a mass-conserving manner, yielding a temporally consistent representation that supports both physically-grounded advection and gradient-based optimization.

Volumetric Rendering. To enable efficient and differentiable evaluation of the volumetric integral, we implement a discrete approximation that preserves the 3D structure of the Gaussian density field. For each camera ray $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$, we sort the intersecting Gaussian primitives in ascending depth order. The pixel color is then computed through a discrete volumetric integration scheme:

$$C(\mathbf{r}) = \sum_{i=1}^{N} T_{i-1} \, \alpha_i \, c_i, \qquad T_{i-1} = \prod_{j=1}^{i-1} (1 - \alpha_j), \tag{6}$$

Crucially, unlike the heuristic alpha-blending in standard 3DGS, the local opacity α_i is derived from the *segment transmittance* $\hat{T}_i$ based on the Beer–Lambert law, where the volumetric density acts as the extinction factor:

$$\alpha_i = 1 - \hat{T}_i, \qquad \hat{T}_i = \exp\big(-\sigma_i \mathcal{I}_i\big), \tag{7}$$

Here, $\mathcal{I}_i$ represents the analytical line integral of the i-th Gaussian kernel along the ray, capturing its exact 3D spatial contribution. By evaluating attenuation directly in 3D space, this formulation ensures physically-aware self-occlusion and view-consistency, effectively overcoming the geometric artifacts and "popping" effects inherent in 2D image-space EWA splatting. The derivation of the closed-form integral factor $\mathcal{I}_i$ is provided in the Appendix.

4.2 Neural Velocity Field

Under sparse-view supervision, smoke reconstruction becomes critically under-constrained. This often causes deformation-based assumptions to exploit non-physical scaling and drifting of Gaussian primitives to overfit limited observations, leading to floaters and geometric distortions. Enforcing incompressibility reshapes this ill-posed search space by embedding Gaussian motion in a globally coordinated, volume preserving velocity field, eliminating arbitrary per-particle deformations and guiding optimization toward physically admissible solutions. On the other hand, from a physical perspective, divergence-free flows naturally induce compensatory circulation and spontaneous vortex formation. Accordingly, one

would expect that if the optimization of Gaussian primitives can, to some extent, match a physically meaningful velocity field, more physically plausible and visually convincing smoke motion under unseen viewpoints could be achieved.

To make such merits realizable, in our approach, we model smoke motion using a neural velocity field $\mathbf{u}(\mathbf{x}, t)$ parameterized by a regular Eulerian grid. This grid-based formulation provides global spatial support and enables stable, numerically well-defined evaluation of PDE-based physical constraints, such as divergence and buoyancy-induced variations.

Eulerian-Lagrangian Coupling. A key aspect of our representation is the tight coupling between the Lagrangian Gaussian primitives and the Eulerian velocity grid. To bridge these two domains, at each time step t, we map the continuous Gaussian density field defined in Sec. 4.1 onto the regular grid to form a grid-aligned density volume σ^t_{grid}. This is achieved by sampling the superposition of Gaussian kernels at each voxel center: $\sigma^t_{\text{grid}}(\mathbf{x}_g) = \sum_i \sigma_i k_i(\mathbf{x}_g)$. This grid-based density serves as a crucial observation cue, allowing the neural update field to perceive the current smoke distribution and infer dynamics consistent with the observed mass transport.

Velocity Update and Physical Constraints. Following common practice in learned simulators, we model the temporal evolution of the velocity field using a residual form. Since the dataset lacks ground-truth velocity data, we initialize the Eulerian velocity field as $\mathbf{u}^0_{\text{grid}} = \mathbf{0}$ at time $t = 0$. At each subsequent time step t, a neural update field Φ predicts a per-voxel velocity increment:

$$\Delta \mathbf{u}^t_{\text{grid}} = \Phi(\mathbf{u}^t_{\text{grid}}, \sigma^t_{\text{grid}}, PE(t)), \tag{8}$$

where Φ is implemented as a 3D U-Net that takes the current grid-aligned velocity $\mathbf{u}^t_{\text{grid}}$, the mapped density field σ^t_{grid}, and a temporal encoding $PE(t)$ as inputs. The inclusion of σ^t_{grid} enables the network to sense the evolving smoke distribution, while $PE(t)$ enforces temporal consistency across frames.

The velocity field is then advanced through explicit integration:

$$\mathbf{u}^{t+1}_{\text{grid}} = \mathbf{u}^t_{\text{grid}} + \Delta \mathbf{u}^t_{\text{grid}}, \tag{9}$$

followed by a differentiable incompressibility projection:

$$\mathbf{u}^{t+1}_{\text{grid}} \leftarrow \Pi(\mathbf{u}^{t+1}_{\text{grid}}), \tag{10}$$

which enforces the physical constraint $\nabla \cdot \mathbf{u} = 0$ via an analytic divergence formulation implemented as a penalty-based differentiable operator. In practice, this parameterization mitigates drift over long horizons, while the per-step projection explicitly keeps the field within the divergence-free manifold, ensuring that the subsequent advection of Gaussian primitives (detailed in Sec. 4.3) remains physically grounded.

4.3 Physically-Guided Gaussian Evolution

Accurate fluid reconstruction requires a synergy between global physical consistency and the preservation of fine-scale appearance. While global physical constraints (e.g., incompressibility) are most stably evaluated on Eulerian grids, purely grid-based density advection often suffers from numerical diffusion that blurs turbulent details. To circumvent this, we adopt 3D Gaussian primitives as Lagrangian tracers. This hybrid approach leverages the grid to enforce physical laws while using particles to maintain sharp, high-frequency smoke structures.

Continuous Velocity Interpolation. To ensure the Lagrangian primitives move in accordance with the Eulerian velocity field defined in Sec. 4.2, we must define a continuous mapping from the grid to the particles. Inspired by kernel-based formulations in SPH [Koschier et al. 2020] and Gaussian Fluids [Yu et al. 2024a], we define the continuous velocity at any point $\mathbf{x}$ as a normalized interpolation from the surrounding grid nodes:

$$\mathbf{u}(\mathbf{x}) = \frac{\sum_{g \in \mathcal{N}(\mathbf{x})} \mathbf{u}_g \, k(\mathbf{x}_g - \mathbf{x})}{\sum_{g \in \mathcal{N}(\mathbf{x})} k(\mathbf{x}_g - \mathbf{x})}, \tag{11}$$

where $\mathbf{u}_g$ is the velocity at grid $\mathbf{x}_g$, and k is the Gaussian interpolation kernel. A significant merit of this formulation is that it admits an analytical closed-form expression for the spatial divergence $\nabla \cdot \mathbf{u}(\mathbf{x})$. Specifically, the divergence of the interpolated field is:

$$\nabla \cdot \mathbf{u}(\mathbf{x}) = \frac{1}{\sum_q k_q(\mathbf{x})} \sum_g k_g(\mathbf{x}) \big(\mathbf{u}_g - \mathbf{u}(\mathbf{x})\big)^\top \Sigma_g^{-1} (\mathbf{x}_g - \mathbf{x}), \tag{12}$$

where Σ_g is the spatial support of the kernel. This closed-form derivative enables numerically stable and fully differentiable enforcement of incompressibility constraints, ensuring that the velocity field experienced by the particles is strictly divergence-free.

Density Transport and Advection. The coupling between density and velocity is physically governed by the fluid transport equation $\frac{D\sigma}{Dt} = 0$. This principle dictates that the volumetric density σ conserved along Lagrangian trajectories. In our framework, this means that while the Gaussian centers $\mathbf{p}$ and their anisotropic geometric shapes evolve over time, the density coefficient σ_i of each primitive remains consistent with the recovered incompressible flow.

At each time step t, the Gaussian particles are advected by sampling the updated Eulerian velocity $\mathbf{u}^{t+1}_{\text{grid}}$ at their current locations:

$$\mathbf{p}^{t+1} = \mathbf{p}^t + \Delta t \cdot \mathbf{u}(\mathbf{p}^t), \tag{13}$$

where $\mathbf{u}(\mathbf{p}^t)$ is computed via the interpolation in Eq. (11). By evolving the Gaussian centers through this physically-constrained velocity field, we eliminate the need for heuristic deformation models.

This advection step, combined with the volumetric rendering described in Sec. 4.1, forms a closed, differentiable physical pathway: the Eulerian grid provides a divergence-free velocity field, which in turn dictates the Lagrangian evolution of Gaussian density carriers. This integration ensures that the reconstructed smoke dynamics are not only visually faithful but also adhere to the fundamental laws of fluid motion.

4.4 Physics-Optical Joint Optimization

To link optical observations with physical evolution, we cast smoke reconstruction as a physics-constrained inverse problem using our hybrid representation. The framework jointly recovers volumetric density and latent velocity fields from sparse RGB videos via a fully differentiable pipeline. Below, we describe the joint optimization procedure (Alg. 1) and the loss functions enforcing consistency between rendering and fluid dynamics.

Algorithm 1: Physics-Optical Joint Optimization

Input: Sparse-view RGB video $\{\mathbf{I}^t\}$ with calibrated cameras; initial Gaussians $\mathcal{G}^0$ and velocity grid $\mathbf{u}^0$; time step Δt, iterations K, K'.

Output: Time-varying Gaussians $\{\mathcal{G}^t\}$ and velocity fields $\{\mathbf{u}^t\}$.

$\mathbf{u}^0 \leftarrow \mathbf{0}$;
for *each frame* t **do**
  repeat
    $\sigma^t_{(i)} = \mathcal{D}(\mathcal{G}^t_{(i)})$;
    $\Delta\mathbf{u}_{(i)} \leftarrow \Phi(\mathbf{u}^t, \sigma^t_{(i)}, PE(t))$;
    $\mathbf{u}^{t+1}_{(i)} \leftarrow \Pi(\mathbf{u}^t + \Delta\mathbf{u}_{(i)})$;
    $\mathbf{p}^{t+1}_{(i)} = \mathbf{p}^t + \Delta t \cdot \mathrm{Interp}(\mathbf{u}^{t+1}_{(i)}, \mathbf{p}^t)$;
    $\mathcal{G}^{t+1}_{(i)} = \mathrm{Optimize}(\mathcal{G}^t_{(i)}, \nabla(\mathcal{L}))$;
  until $i \geq K$;
  $\mathbf{u}^{t+1} \leftarrow \mathbf{u}^{t+1}_{(K)}$; $\mathbf{p}^{t+1} \leftarrow \mathbf{p}^{t+1}_{(K)}$;
  repeat
    $\mathcal{G}^{t+1}_{(i+1)} = \mathrm{Optimize}(\mathcal{G}^{t+1}_{(i)}, \nabla(\mathcal{L}_{\text{render}} + \mathcal{L}_{\text{refine}}))$;
  until $i \geq K'$;
  $\mathcal{G}^{t+1} \leftarrow \mathcal{G}^{t+1}_{(K')}$;

Optimization Objective. We jointly optimize the Gaussian parameters $\mathcal{G} = \{\mathbf{p}_i, \Sigma_i, \sigma_i, c_i\}$ and the network weights of the neural velocity update field Φ by minimizing a multi-objective loss:

$$\mathcal{L} = \mathcal{L}_{\text{render}} + \lambda_{\text{div}}\mathcal{L}_{\text{div}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{refine}}\mathcal{L}_{\text{refine}}, \tag{14}$$

where the terms enforce photometric consistency, incompressibility, advection transport, and trajectory regularization, respectively, and λ_{div}, λ_{adv}, and λ_{refine} denote the corresponding loss weights.

Photometric Consistency. The primary observation-driven supervision is the rendering loss, which minimizes the L_1 error between the observed RGB frames I^t and the images $\hat{I}^t$ produced by our volumetric Gaussian renderer:

$$\mathcal{L}_{\text{render}} = \sum_t \left\| I^t - \hat{I}^t(\mathcal{G}^t) \right\|_1. \tag{15}$$

By propagating gradients through the Beer–Lambert extinction model, this loss provides the necessary signal to recover the 3D density field from sparse views.

Eulerian Incompressibility Constraint. To ensure the inferred flow is physically admissible, we penalize the divergence of the Eulerian velocity field. Utilizing the closed-form formulation in Eq. (12), the loss is evaluated as:

$$\mathcal{L}_{\text{div}} = \left\| \nabla \cdot \mathbf{u}(\mathbf{x}, t) \right\|_2^2. \tag{16}$$

This term constrains the Eulerian grid to favor divergence-free solutions, which is fundamental for preserving fluid volume and inducing coherent vortical motion.

Lagrangian Advection Consistency. This loss serves as the core coupling mechanism between the density and velocity fields. According to the transport equation, the density $\sigma(\mathbf{x}, t)$ should be conserved along the flow. We therefore penalize the material derivative of the reconstructed density field:

$$\mathcal{L}_{\text{adv}} = \left\| \frac{\partial \sigma}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla_{\mathbf{x}} \sigma(\mathbf{x}, t) \right\|_2^2. \tag{17}$$

This term ensures that the recovered velocity field $\mathbf{u}$ is the true driver of the observed density transport, effectively bridging the Lagrangian Gaussian evolution with Eulerian grid dynamics.

Sub-grid Level Refinement Regularization. While the Eulerian grid captures macroscopic flow, fine-scale turbulent details often exceed the grid's resolution. We therefore allow the Lagrangian Gaussians to undergo small corrective displacements $\mathbf{p}^{\text{refine}}$ beyond the grid-advected positions $\mathbf{p}^{t+1}$ to better match the observation. To keep these updates physically plausible, we enforce:

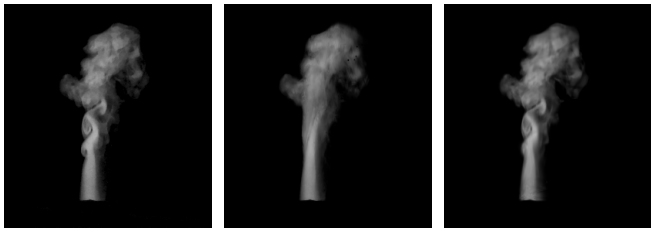
$$\mathcal{L}_{\text{refine}} = \left\| \mathbf{p}^{\text{refine}} - \mathbf{p}^{t+1} \right\|_2^2. \tag{18}$$

This loss treats the physically-advected trajectory as a strong motion prior, decoupling global flow consistency from high-frequency appearance refinement.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate our method on the publicly available **real-captured** dataset *ScalarFlow* [Eckert et al. 2019] and a **synthetic** smoke dataset created in Blender. *ScalarFlow* has a resolution of 270×480, with 5 views and 120 frames per sequence. The synthetic dataset includes scenes with static solids, such as *Censer*, *Coffee*, and *Steam Train*, following a sparse-view setting with fewer than 15 viewpoints and more than 120 frames at a resolution of 800×800. We further test a more challenging obstacle-interaction scene on the *FluidNexus-Ball* dataset, where upward-rising smoke passes through a small spherical obstacle. Together, these datasets provide both real-world and synthetic evaluations of our method.



Reference **4D-GS** **Ours**

Fig. 3. **Comparisons on ScalarFlow (training view).** Even with seen-view supervision, 4DGS fails to capture fine-scale vortex patterns.

Baselines and Metrics. We compare our method with four representative baselines: two dynamic 3DGS methods, **4D-GS** [Wu et al. 2024] and Deformable GS [Yang et al. 2024a] (denoted as **Deform. GS**), the fastest NeRF-based smoke reconstruction method **NeuSmoke** [Qiu et al. 2024], and **FluidNexus** [Gao et al. 2025], which uses a visual-physical Gaussian particle representation.

To evaluate appearance quality and novel-view synthesis, we report PSNR, SSIM, and LPIPS against novel-view references. Physical consistency is assessed using velocity visualizations and the

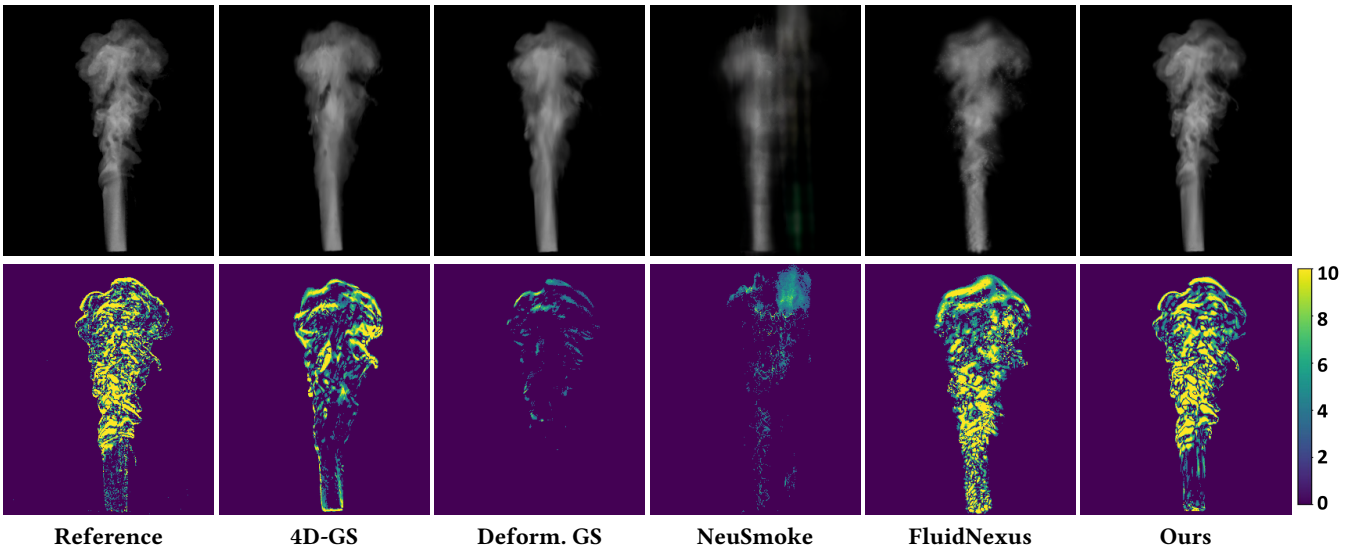


Fig. 4. **Qualitative comparison on ScalarFlow.** Top: rendered novel-view images. Bottom: temporal difference maps between consecutive frames, where darker colors indicate smaller temporal changes and brighter colors indicate larger changes.

time-averaged divergence of the velocity field, $\nabla \cdot \mathbf{u}$, over the reconstructed volume. Computational efficiency is measured by runtime and memory usage under identical hardware and resolution settings. Together, these metrics provide a comprehensive evaluation of both visual fidelity and physical accuracy.

Implementation Details. Implemented in PyTorch, we define the velocity field on a 64^3 grid and initialize 30k Gaussian primitives. We use the Adam optimizer with initial learning rates of 1×10^{-3} for velocity and 5×10^{-4} for Gaussians, applying exponential decay (0.95) every 10 frames. Optimization parameters are set to $K = 250$ (advection), $K' = 50$ (refinement), with loss weights $\lambda_{\text{div}} = 0.02$, $\lambda_{\text{adv}} = 0.02$, and $\lambda_{\text{refine}} = 0.01$. Training a 120-frame sequence takes approximately 0.8 hours on a single NVIDIA RTX 3090 GPU, notably faster than NeuSmoke's 1.5–3 hours.

5.2 Comparison with State-of-the-Arts

To provide a comprehensive comparison of different methods, we conduct experiments and analysis on the real-captured *ScalarFlow* smoke dataset, as well as synthetic smoke and static geometry scenes. The overall results are summarized in Tab. 1. As shown, our method outperforms existing approaches across all scenarios, demonstrating the effectiveness of the smoke reconstruction technique we propose.

Comparisons on ScalarFlow. As shown in Fig. 3, a key observation is that existing dynamic baselines struggle to recover turbulent details, even under training views. Despite direct supervision, deformation-based methods fail to reconstruct fine-scale vortices, revealing a fundamental mismatch between rigid-like warping assumptions and the topology-changing nature of smoke dynamics. This limitation is further exacerbated in novel-view synthesis (Fig. 4, top row), where 4D-GS produces amorphous blobs and Deform. GS results in over-smoothed, blurred structures.

The corresponding temporal difference maps (Fig. 4, bottom row) further expose the poor motion quality of the baselines: while 4D-GS exhibits severe temporal jitter and NeuSmoke suffers from chaotic artifacts, Deform. GS fails to capture high-frequency motion, resulting in a significantly dulled difference signal. FluidNexus is less detailed, whereas our method recovers sharp vortical patterns in the difference maps that closely match the ground truth. This demonstrates that, by coupling the density field with a physically-constrained velocity field, our approach preserves the intrinsic swirling dynamics of smoke that are fundamentally lost in heuristic deformation models.

Consistent with this observation, Tab. 1 shows that on *ScalarFlow*, our method improves PSNR by 2.33 dB over the best baseline, while also achieving higher SSIM and reducing LPIPS by 58.0%.

Physical Fidelity. We further evaluate our framework on challenging scenarios. Since deformation-based baselines lack an explicit velocity representation, we estimate their pseudo-velocity from inter-frame displacements (Fig. 5). These visualizations show that their learned motions are chaotic and physically inconsistent. In contrast, our method reconstructs a structured upward flow, with higher velocities near the emission source and within the rising core. Along with captured vortical patterns, this faithfully reflects smoke dynamics, including acceleration and buoyancy. Velocity divergence statistics are presented in the ablation analysis (Sec. 5.3).

Our method demonstrates significant advantages in handling scenes with varying background and smoke colors. As shown in Fig. 6, traditional 3DGS methods, such as 4D-GS and Deform. GS, struggle to capture the interaction between vibrant yellow smoke and light, particularly in light absorption and scattering, leading to blurry results and color bleeding, especially when the smoke color is similar to the background. While NeuSmoke also struggles with background and smoke separation, our hybrid representation, leveraging a density-field model, robustly handles changes in smoke and

Table 1. **Quantitative comparisons of rendering results under novel views. Bold** represents the best results, underlined represents the second-best, and dashes indicate results that couldn't be reproduced.

Method	ScalarFlow			Smoke			Yellow			Censer			Coffee			Steam Train		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
4D-GS	29.22	0.933	0.0400	32.77	0.944	0.0628	30.24	0.958	0.1183	30.69	0.936	0.0554	32.44	0.952	0.0740	27.46	0.907	0.1587
Deform. GS	29.50	0.936	0.0416	28.05	0.922	0.0755	40.98	0.984	0.0885	29.54	0.933	0.0552	30.75	0.946	0.0792	34.25	0.969	0.1054
NeuSmoke	22.74	0.853	0.1060	31.54	0.928	0.0894	34.15	0.965	0.2173	27.81	0.888	0.1364	27.60	0.908	0.1481	28.62	0.839	0.1782
FluidNexus	32.50	0.954	0.0810	22.91	0.865	0.1200	–	–	–	–	–	–	–	–	–	–	–	–
Ours	**34.83**	**0.955**	**0.0168**	**37.98**	**0.975**	**0.0336**	**42.77**	**0.990**	**0.0596**	**36.98**	**0.976**	**0.0307**	**36.30**	**0.976**	**0.0411**	**38.21**	**0.961**	**0.0897**

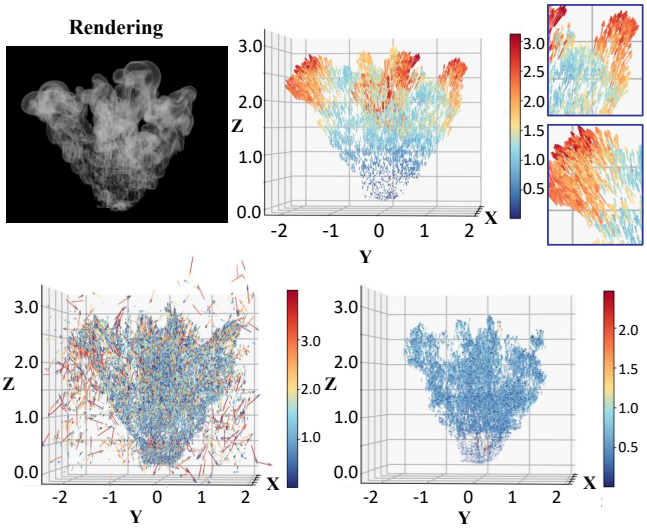


Fig. 5. **Velocity comparison.** The top row shows the rendering at frame 130 (left), followed by the velocity field for our method (middle), and a zoomed-in view (right) highlighting fine details. The bottom row presents the pseudo-velocity field from the deformation-based baselines, with 4D-GS (left) and Deform. GS (right). The color bar represents speed values. Note that NeuSmoke does not construct a velocity field, and FluidNexus fails to reconstruct the scene.

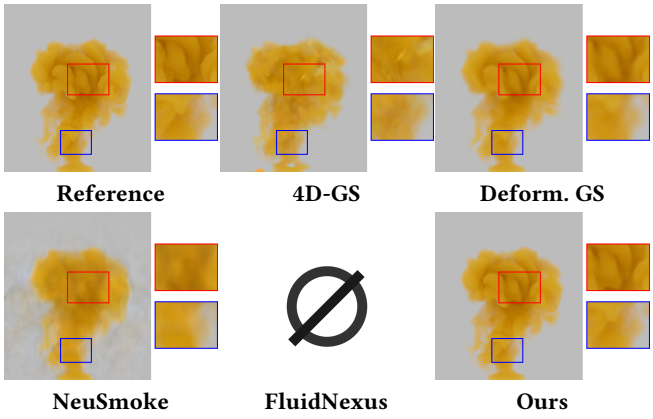


Fig. 6. **Generalization to diverse appearances.** Our hybrid representation demonstrates strong generalization across arbitrary smoke colors and background intensities. Note that *FluidNexus* fails to reconstruct the scene using the parameters and code provided by authors.

background colors. This approach accurately simulates absorption and scattering of light in different colors, resulting in more natural and realistic reconstructions.

Regarding FluidNexus, we successfully replicated its performance on the *ScalarFlow* dataset, as it provided the necessary parameters for reconstruction. However, for a new dataset, setting complex hyperparameters proved challenging, and some scene reconstructions failed. This highlights the limited generalization of FluidNexus to more complex scenes, where fine-tuned parameters are crucial for successful reconstruction.

Generalization to Scenes with Static Geometry. We further evaluate generalization to scenes containing static geometry (*Censer*, *Coffee*, and *Steam Train*). As shown in Fig. 11, our method cleanly separates smoke from static components while capturing swirling fluid motion. In comparison, baselines often corrupt static structures: NeuSmoke produces vertical streaks near boundaries (e.g., above the cup), 4D-GS exhibits strong temporal jitter in difference maps, and Deform. GS tends to over-smooth smoke details and solid–fluid interfaces. Specifically, our velocity network utilizes the density field as input to differentiate high-density static solids from smoke, effectively refining the separation at solid-fluid interfaces. Leveraging volumetric rendering (Sec. 4.1) and explicit static/dynamic primitives, our method preserves sharp object boundaries and recovers high-frequency vortical structures consistent with the reference.

In scenarios involving interactions with static obstacles and backgrounds, especially in the **FluidNexus-Ball dataset**, as shown in Fig. 9, our method demonstrates some blurring near the boundary between the smoke and the ball. Additionally, large areas of the background exhibit a similar blurring effect, and the separation between the smoke and background is not as clear, causing a loss of background detail. The interaction between the smoke and the background lacks precision, which impacts the overall visual quality. In contrast, 4D-GS and FluidNexus perform better in reconstructing the geometry of obstacles and backgrounds, capturing more accurate shapes and details. However, the smoke in these methods tends to appear as blurred particles, lacking finer details. In practice, FluidNexus even includes a separate stage for reconstructing the background, which improves background details but still struggles with capturing intricate smoke features.

While our method shows some limitations in handling these interactions, we provide a more detailed analysis of this issue in

Sec. 5.4, where we discuss the challenges and outline potential directions for future improvements.

Efficiency Analysis. We compare runtime and memory usage under the same hardware and training protocol. As shown in Tab. 2, on a *ScalarFlow* smoke scene trained for 30k iterations on a single RTX 3090 24GB, our method incurs moderate overhead compared to deformation-based Gaussian baselines due to Beer–Lambert rendering and physically constrained flow modeling. However, it remains significantly faster than NeuSmoke while offering better reconstruction quality.

Table 2. **Efficiency comparison with baselines on ScalarFlow.Bold** represents the best results, underlined represents the second-best.

Method	Train Time (h)↓	FPS↑	Storage(MB)↓	PSNR↑
4D-GS	**0.25**	**152**	**17**	29.22
Deformable GS	0.30	135	20	29.50
NeuSmoke	1.80	28	53	22.74
FluidNexus	0.60	102	123	32.50
Ours	0.80	114	42	**34.83**

Regarding memory usage, unlike NeuSmoke, which stores global MLP weights and positional encodings, our method uses a lightweight U-Net to learn velocity field changes. This reduces the need to store complex per-frame physical attributes, as seen in FluidNexus. By adopting a more compact architecture, we mitigate the storage burden, achieving high-quality reconstructions with lower memory consumption, effectively balancing performance and efficiency.

5.3 Ablation Study

To verify the effectiveness of the key components in our framework, we ablate three crucial components: physical constraints ($\mathcal{L}_{\text{div}}$, $\mathcal{L}_{\text{adv}}$), volumetric rendering, and sub-grid refinement ($\mathcal{L}_{\text{refine}}$). The quantitative results are summarized in Tab. 3.

Table 3. **Ablation Study of our proposed components.**

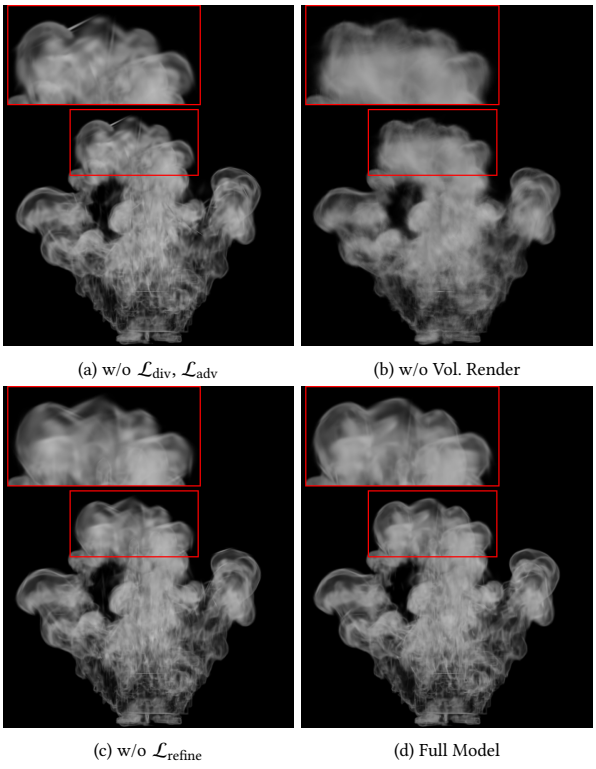
Evaluation	PSNR↑	SSIM↑	LPIPS↓	$\nabla \cdot \mathbf{u}$↓
4D-GS	32.77	0.944	0.0628	–
w/o $\mathcal{L}_{\text{div}}$, $\mathcal{L}_{\text{adv}}$	30.12	0.925	0.0810	0.402
w/o Vol. Render	30.45	0.932	0.0825	0.127
w/o $\mathcal{L}_{\text{refine}}$	33.87	0.964	0.0418	0.098
Full Model	**37.98**	**0.975**	**0.0336**	**0.085**

Effect of Physical Constraints. As shown in Fig. 7 (a), removing the physical constraints ($\mathcal{L}_{\text{div}}$, $\mathcal{L}_{\text{adv}}$) causes a significant drop in performance. Without these constraints, image loss dominates the optimization process, leading to non-physical velocity fields. This results in Gaussians drifting and overfitting sparse views, causing unrealistic fluid dynamics. As reflected in the PSNR and SSIM values (Tab. 3), these metrics drop significantly, while LPIPS increases, indicating a degradation in visual quality. Furthermore, the velocity field's divergence ($\nabla \cdot \mathbf{u}$) becomes abnormal, showing that the absence of physical constraints leads to inconsistent and unphysical flow dynamics. This highlights the critical role of physical constraints in ensuring realistic fluid motion and regularizing the optimization.

Impact of Volumetric Rendering. Fig. 7 (b) demonstrates the impact of removing our Beer–Lambert volumetric rendering. Without it, the smoke appears as a dense, opaque collection of surface-like clusters rather than a continuous, translucent medium. The PSNR drops to 30.45, showing the loss of fine details and transparency in the smoke. This "point-splat" appearance fails to capture the subtle transparency and light-scattering effects of fluids, underscoring the importance of our explicit density-field formulation for maintaining volumetric consistency and realistic fluid simulation.

Necessity of Sub-grid Refinement. Fig. 7 (c) shows the effect of omitting the sub-grid refinement term ($\mathcal{L}_{\text{refine}}$). Without refinement, the model loses fine-scale turbulent details, resulting in an over-smoothed reconstruction. PSNR improves when refinement is applied, restoring sharp vortical patterns typical of turbulent smoke. This demonstrates that sub-grid refinement is essential for capturing small-scale details while maintaining global physical consistency.

In summary, each component contributes significantly to the final performance. The physical constraints ensure realistic velocity fields, volumetric rendering maintains the true appearance of smoke, and sub-grid refinement recovers small-scale turbulent structures critical for high-quality fluid simulation.



(a) w/o $\mathcal{L}_{\text{div}}$, $\mathcal{L}_{\text{adv}}$ (b) w/o Vol. Render

(c) w/o $\mathcal{L}_{\text{refine}}$ (d) Full Model

Fig. 7. **Qualitative ablation of framework components.** Each component is vital for high-fidelity results: omitting physical constraints leads to motion artifacts, while removing volumetric rendering significantly degrades the sharpness of fine-grained vortical structures.

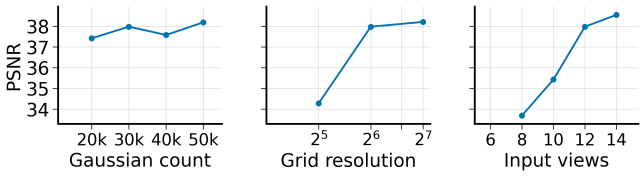


Fig. 8. **PSNR performance vs. different settings.** The plots illustrate the effect of various hyperparameters on reconstruction quality. We vary the Gaussian count, grid resolution, and input views to determine the optimal configuration for the best performance.

Performance–Efficiency Tradeoff. Fig. 8 shows the tradeoff between performance and efficiency in our framework. The initial Gaussian count has only a marginal effect on reconstruction quality, with PSNR remaining nearly constant when the count varies from 20k to 50k. This suggests that increasing the Gaussian count within this range yields limited performance gains. In contrast, grid resolution directly influences the level of detail captured in the physical velocity field, and higher resolutions consistently improve PSNR. However, this improvement comes at a computational cost, increasing memory consumption by 34.6% and training time by 12.5%. The number of training views also plays a significant role, as more views provide richer observational constraints, leading to better reconstruction quality. After considering all factors, we select the final hyperparameters to balance reconstruction accuracy and computational efficiency.

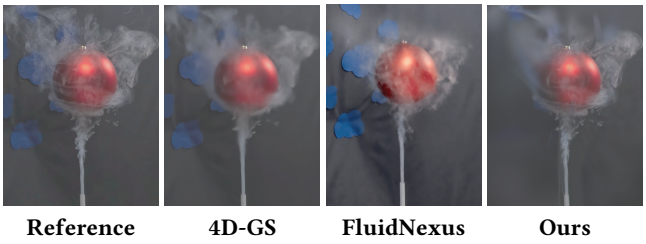


Fig. 9. **Reconstruction results on the FluidNexus-Ball dataset.** Comparing the interaction with static objects and background.

5.4 Conclusion and Limitations

Despite the superior performance in fluid reconstruction, our method still faces several limitations and challenges.

Background Assumption. In line with common sparse-view smoke reconstruction settings, our experiments primarily use relatively simple backgrounds to isolate the core fluid reconstruction problem. In more complex real-world scenes with textured backgrounds, ambiguous colors, or occlusions (as shown in Fig. 9), background details consume Gaussian capacity and computation, potentially introducing artifacts such as ghost density, which make foreground-background separation more challenging. Incorporating explicit foreground-background decomposition or segmentation-aware modeling is a natural direction for improving robustness in unconstrained environments.

Fluid–Solid Interaction. Our method is primarily designed for fluid dynamics and does not explicitly model fluid–solid interactions. Specifically, we do not enforce strict boundary conditions (e.g., no-penetration constraints), leading to density leakage in challenging obstacle configurations, such as in Fig. 9, where our method shows blurry boundaries between the smoke and the ball. Furthermore, when solids exhibit complex or dynamic motion, the learned velocity field—tailored for incompressible flow—cannot faithfully capture rigid body dynamics. As shown in Fig. 10, although our method preserves the global smoke flow better than baseline methods, the moving teapot lid (red inset) exhibits local geometric distortion and loss of rigidity during reconstruction. This suggests that the fluid-oriented motion prior is insufficient for jointly modeling smoke transport and rigid object motion. Future work could explore geometry-aware boundary constraints (e.g., SDF- or mesh-based representations) and hybrid motion models to better bridge fluid dynamics with rigid object motion.

Acknowledgments

This work was supported by the Natural Science Foundation of China (No. 62595772, No. 62132012) and the Fundamental Research Funds for the Central Universities (Nankai University, No. 63263248).

References

Robert Bridson. 2015. *Fluid simulation for computer graphics.* AK Peters/CRC Press.

Mengyu Chu, Lingjie Liu, Quan Zheng, Aleksandra Franz, Hans-Peter Seidel, Christian Theobalt, and Rhaleb Zayer. 2022. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Transactions on Graphics (ToG)* 41, 4 (2022), 1–14.

F Du, Y Zhang, Y Ji, X Wang, C Yao, J Kosinka, S Frey, A Telea, and X Ban. 2025. GaussFluids: Reconstructing Lagrangian Fluid Particles from Videos via Gaussian Splatting. In *Pacific Graphics 2025.*

Marie-Lena Eckert, Kiwon Um, and Nils Thuerey. 2019. ScalarFlow: a large-scale volumetric data set of real-world scalar transport flows for computer animation and machine learning. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–16.

Aleksandra Franz, Barbara Solenthaler, and Nils Thuerey. 2021. Global Transport for Fluid Reconstruction with Learned Self-Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.*

Aleksandra Franz, Barbara Solenthaler, and Nils Thuerey. 2023. Learning to Estimate Single-View Volumetric Flow Motions without 3D Supervision. In *International Conference on Learning Representations.*

Yue Gao, Hong-Xing Yu, Bo Zhu, and Jiajun Wu. 2025. FluidNexus: 3D fluid reconstruction and prediction from a single video. In *Proceedings of the Computer Vision and Pattern Recognition Conference.* 26091–26101.

Shoukang Hu, Tao Hu, and Ziwei Liu. 2024. Gauhuman: Articulated gaussian splatting from monocular human videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 20418–20431.

James T Kajiya and Brian P Von Herzen. 1984. Ray tracing volume densities. *ACM SIGGRAPH computer graphics* 18, 3 (1984), 165–174.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

Xiangyu Kong, Arnaud Schoentgen, Damien Rioux-Lavoie, Paul G. Kry, and Derek Nowrouzezahrai. 2025. Hierarchical Differentiable Fluid Simulation. *Computer Graphics Forum* 0, 0 (2025), e70226. doi:10.1111/cgf.70226

Dan Koschier, Jan Bender, Barbara Solenthaler, and Matthias Teschner. 2020. Smoothed particle hydrodynamics techniques for the physics based simulation of fluids and solids. *arXiv preprint arXiv:2009.06944* (2020).

Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. 2024. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 20775–20785.

Yiqing Liang, Numair Khan, Zhengqin Li, Thu Nguyen-Phuoc, Douglas Lanman, James Tompkin, and Lei Xiao. 2025. Gaufre: Gaussian deformation fields for real-time dynamic novel view synthesis. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).* IEEE, 2642–2652.

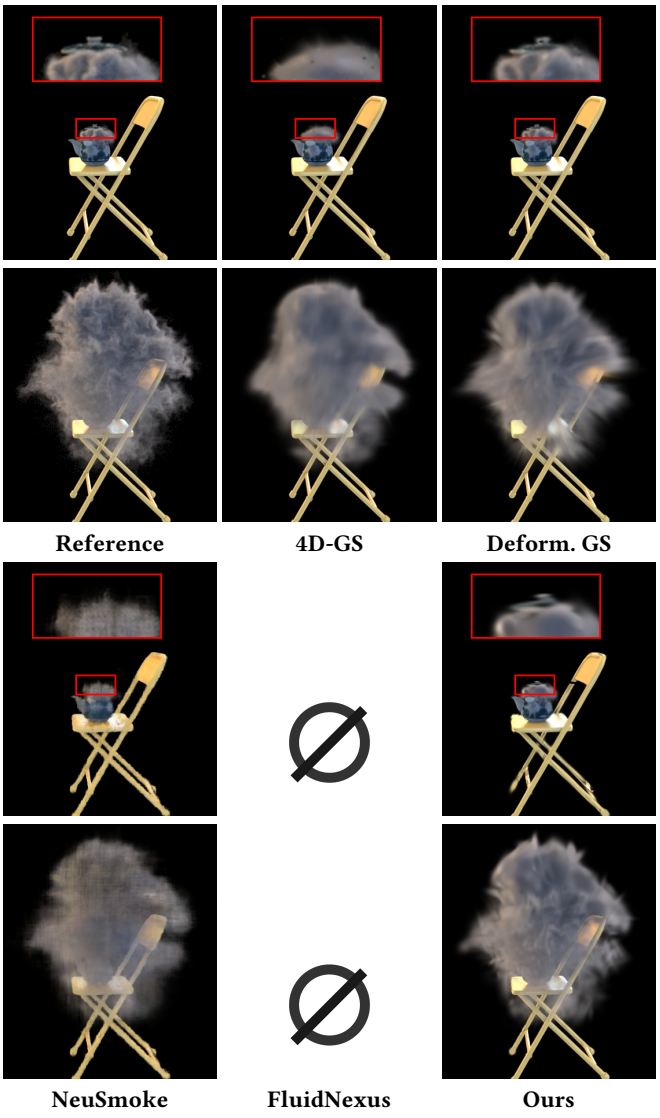


Fig. 10. **Challenges with dynamic solid objects.** The top row shows the results for the 16th frame, while the bottom row shows the results for the 60th frame. Note that *FluidNexus* fails to reconstruct the scene using the parameters and code provided by the authors.

Yibo Liu, Zhixin Fang, Sune Darkner, Noam Aigerman, Kenny Erleben, Paul Kry, and Teseo Schneider. 2025. Neural Kinematic Bases for Fluids. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25)*. Association for Computing Machinery, New York, NY, USA, Article 28, 10 pages. doi:10.1145/3757377.3763925

Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural Volumes: Learning Dynamic Renderable Volumes from Images. *ACM Transactions on Graphics* 38, 4 (2019).

Miles Macklin, Matthias Müller, Nuttapong Chentanez, and Tae-Yong Kim. 2014. Unified particle physics for real-time applications. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–12.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Sang Il Park and Myoung Jun Kim. 2005. Vortex fluid for gaseous phenomena. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*. 261–270.

Jiaxiong Qiu, Ruihong Cen, Zhong Li, Han Yan, Ming-Ming Cheng, and Bo Ren. 2024. NeuSmoke: Efficient Smoke Reconstruction and View Synthesis with Neural Transportation Fields. In *SIGGRAPH Asia 2024 Conference Papers*. 1–12.

Sheng Qiu, Chen Li, Changbo Wang, and Hong Qin. 2021. A Rapid, End-to-End, Generative Model for Gaseous Phenomena from Limited Views. In *Computer Graphics Forum*.

Chinmay Talegaonkar, Yash Belhe, Ravi Ramamoorthi, and Nicholas Antipa. 2025. Volumetrically Consistent 3D Gaussian Rasterization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 10953–10963.

Yuanyuan Tao, Ivan Puhachov, Derek Nowrouzezahrai, and Paul Kry. 2024. Neural Implicit Reduced Fluid Simulation. In *SIGGRAPH Asia 2024 Conference Papers*. doi:10.1145/3680528.3687628

Yecong Wan, Mingwen Shao, Yuanshuo Cheng, and Wangmeng Zuo. 2025. S2Gaussian: Sparse-View Super-Resolution 3D Gaussian Splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 711–721.

Chen Wang, Hao-Yang Peng, Ying-Tian Liu, Jiatao Gu, and Shi-Min Hu. 2025. Diffusion Models for 3D Generation: A Survey. *Computational Visual Media* 11, 1 (2025), 1–28. doi:10.26599/CVM.2025.9450452

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 2024. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 20310–20320.

Youchen Xie, Chen Li, Sheng Qiu, Zhi-Jun Wang, Chenhui Li, Yibo Zhao, Zan Gao, and Changbo Wang. 2025. FluidGS: Physics Informed Gaussian Splatting for Dynamic Fluid Reconstruction from Sparse Views. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 8438–8447.

Jingrui Xing, Bin Wang, Mengyu Chu, and Baoquan Chen. 2025. Gaussian Fluids: A Grid-Free Fluid Solver based on Gaussian Spatial Representation. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–11.

Haofei Xu, Songyou Peng, Fangjinhua Wang, Hermann Blum, Daniel Barath, Andreas Geiger, and Marc Pollefeys. 2025. Depthsplat: Connecting gaussian splatting and depth. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 16453–16463.

Chen Yang, Sikuang Li, Jiemin Fang, Ruofan Liang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, and Qi Tian. 2024b. Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. *arXiv preprint arXiv:2402.10259* (2024).

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. 2024a. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 20331–20341.

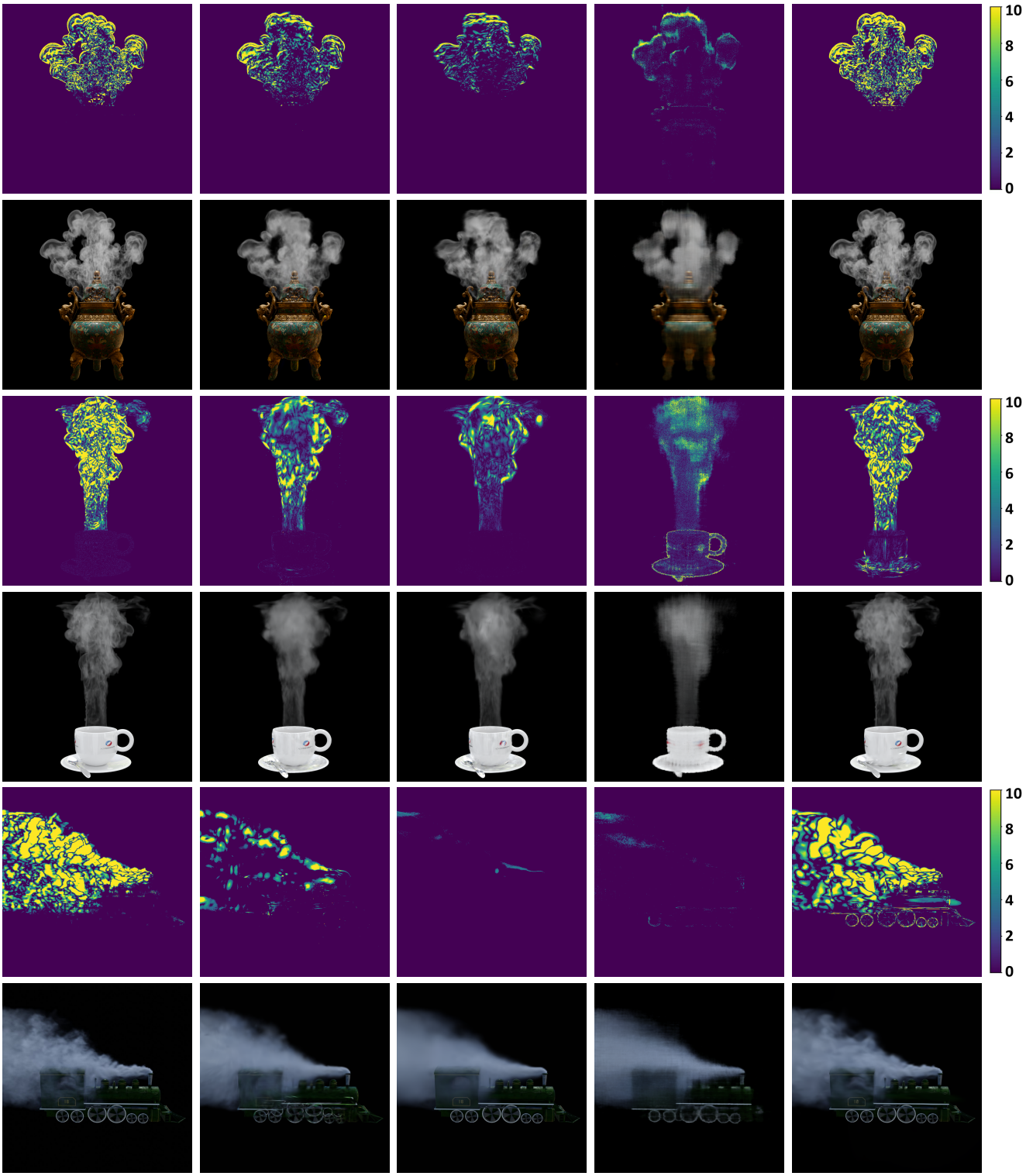
Hong-Xing Yu, Yang Zheng, Yuan Gao, Yitong Deng, Bo Zhu, and Jiajun Wu. 2023. Inferring hybrid neural fluid fields from videos. *Advances in Neural Information Processing Systems* 36 (2023), 63595–63608.

Wangbo Yu, Jinbo Xing, Li Yuan, Wenbo Hu, Xiaoyu Li, Zhipeng Huang, Xiangjun Gao, Tien-Tsin Wong, Ying Shan, and Yonghong Tian. 2024b. Viewcrafter: Taming video diffusion models for high-fidelity novel view synthesis. *arXiv preprint arXiv:2409.02048* (2024).

Zehao Yu, Torsten Sattler, and Andreas Geiger. 2024a. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (ToG)* 43, 6 (2024), 1–13.

Guangming Zang, Ramzi Idoughi, Congli Wang, Anthony Bennett, Jianguo Du, Scott Skeen, William L. Roberts, Peter Wonka, and Wolfgang Heidrich. 2020. TomoFluid: Reconstructing Dynamic Fluid From Sparse View Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. 2002. EWA splatting. *IEEE Transactions on Visualization and Computer Graphics* 8, 3 (2002), 223–238.



(a) Reference (b) 4D-GS (c) Deform. GS (d) NeuSmoke (e) Ours

Fig. 11. **Qualitative comparison on hybrid scenes with complex dynamics.** We compare novel-view rendering and temporal consistency across three scenarios: *Censer* (top), *Coffee* (middle), and *Steam Train* (bottom). Odd rows show temporal difference maps ($|I_t - I_{t-1}|$); even rows show renderings. Baselines exhibit jittery noise, over-smoothed motion, or structural breaks, while our results produce sharp difference patterns matching the Reference and preserve fine smoke details without introducing artifacts on static objects.

A Closed-form Incompressibility of the Eulerian Velocity Field

Incompressibility prior and closed-form derivatives. To enforce incompressibility of the smoke velocity field, we derive closed-form expressions for its Jacobian and divergence analytically. Given the Gaussian kernel–interpolated velocity field defined in Eq. (11), these expressions allow exact evaluation at any point $\mathbf{x}$ on the Eulerian grid.

Kernel-interpolated velocity field. The interpolated velocity at $\mathbf{x}$ is defined as

$$\mathbf{u}(\mathbf{x}) = \frac{\sum_p \mathbf{u}_p\, k_p(\mathbf{x})}{\sum_p k_p(\mathbf{x})}, \qquad k_p(\mathbf{x}) = \exp\Big(-\frac{1}{2}(\mathbf{p}-\mathbf{x})^\top \Sigma_p^{-1}(\mathbf{p}-\mathbf{x})\Big), \tag{19}$$

where $\mathbf{u}_p$ and Σ_p are the velocity and covariance of Gaussian primitive p, respectively. Let $\mathbf{g}_p(\mathbf{x}) = \Sigma_p^{-1}(\mathbf{p}-\mathbf{x})$, so that

$$\nabla_{\mathbf{x}} k_p(\mathbf{x}) = k_p(\mathbf{x})\, \mathbf{g}_p(\mathbf{x}). \tag{20}$$

Jacobian of the velocity field. Writing $\mathbf{u}(\mathbf{x}) = \mathbf{N}(\mathbf{x})/D(\mathbf{x})$ with $\mathbf{N}(\mathbf{x}) = \sum_p \mathbf{u}_p k_p$ and $D(\mathbf{x}) = \sum_p k_p$, the quotient rule gives

$$\nabla_{\mathbf{x}}\mathbf{u}(\mathbf{x}) = \frac{\nabla_{\mathbf{x}}\mathbf{N}(\mathbf{x})}{D(\mathbf{x})} - \frac{\mathbf{N}(\mathbf{x})}{D^2(\mathbf{x})}\nabla_{\mathbf{x}} D(\mathbf{x}), \tag{21}$$

where $\nabla_{\mathbf{x}}\mathbf{N} = \sum_p \mathbf{u}_p \otimes \nabla_{\mathbf{x}} k_p$ and $\nabla_{\mathbf{x}} D = \sum_p \nabla_{\mathbf{x}} k_p$. Simplifying yields the closed-form Jacobian:

$$\nabla_{\mathbf{x}}\mathbf{u}(\mathbf{x}) = \frac{1}{\sum_q k_q} \sum_p k_p\, (\mathbf{u}_p - \mathbf{u}(\mathbf{x})) \otimes \mathbf{g}_p. \tag{22}$$

Divergence of the velocity field. The divergence is obtained as the trace of the Jacobian. Using $\mathrm{tr}(\mathbf{a}\otimes\mathbf{b}) = \mathbf{a}^\top\mathbf{b}$, we have

$$\nabla\cdot\mathbf{u}(\mathbf{x}) = \frac{1}{\sum_q k_q} \sum_p k_p\, (\mathbf{u}_p - \mathbf{u}(\mathbf{x}))^\top \Sigma_p^{-1}(\mathbf{p}-\mathbf{x}). \tag{23}$$

B Analytic Ray-Gaussian Integration

To obtain a truly continuous volumetric representation, we derive a closed-form solution for the line integral of a 3D Gaussian along a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$. Substituting the ray equation into the Gaussian exponent and completing the square, the contribution of the i-th Gaussian $G_i(\mathbf{r}(t))$ can be re-parameterized as:

$$G_i(\mathbf{r}(t)) = \exp\left(-\frac{1}{2}\left[\frac{(t-\gamma_i)^2}{\beta_i^2} + K_i\right]\right), \tag{24}$$

where the geometric parameters are defined by

$$\beta_i^2 = \frac{1}{\mathbf{d}^\top\Sigma_i^{-1}\mathbf{d}}, \quad \gamma_i = \frac{(\boldsymbol{\mu}_i - \mathbf{o})^\top\Sigma_i^{-1}\mathbf{d}}{\mathbf{d}^\top\Sigma_i^{-1}\mathbf{d}}, \tag{25}$$

and the transverse distance factor

$$K_i = (\mathbf{o}-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}(\mathbf{o}-\boldsymbol{\mu}_i) - \frac{((\mathbf{o}-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}\mathbf{d})^2}{\mathbf{d}^\top\Sigma_i^{-1}\mathbf{d}}. \tag{26}$$

Here, γ_i is the depth of the Gaussian center projected onto the ray, β_i is the effective scale along the view direction, and K_i accounts for the normalized perpendicular distance from the ray to the Gaussian center $\boldsymbol{\mu}_i$. Integrating over $t \in [-\infty, +\infty]$, the analytic segment transmittance for the i-th Gaussian is

$$\hat{T}_i = \exp\left(-\kappa_i \exp\left(-\frac{K_i}{2}\right)\sqrt{2\pi}\,\beta_i\right). \tag{27}$$

This closed-form formulation provides exact gradients with respect to the Gaussian's shape (Σ_i) and position ($\boldsymbol{\mu}_i$), enabling physically-consistent optimization of 3D structures from sparse 2D projections.