PROCEEDINGS OF SCIENCE

From Light to Muons: Towards a Unified Framework for Physics-based 3D Scene Reconstruction

Felix Sattler,[a,*] **Jean-Marco Alameddine,**[a] **Ángel Bueno Rodriguez,**[a] **Maurice Stephan**[a] **and Sarah Barnes**[a]

[a] *German Aerospace Center, Institute for the Protection of Maritime Infrastructures, Fischkai 1 Bremerhaven, Germany*

E-mail: felix.sattler@dlr.de

Inverse problems like computed tomography, optical inverse rendering, and muon tomography, amongst others, occur in a vast range of scientific, medical, and security applications and are usually solved with highly specific algorithms depending on the task. Approaching these problems from a physical perspective and reformulating them as a function of particle interactions, enables 3D scene reconstruction in a physically consistent manner across particle-mediated modalities. Recent developments in differentiable volumetric rendering and optical optimization techniques, such as Neural Radiance Fields, Gaussian Splatting, and Scene Representations Networks (SRN), have been used to demonstrate the feasibility of jointly estimating unknown geometry and material parameters of a 3D scene. Some works also show the feasibility of modeling refraction and multiple scattering of light using differentiable optimization.

By approaching these problems from a physical perspective and reformulating them in terms of transport and interaction of radiation or particles we can formulate a unified forward model that maps unknown scene parameters to measurements. This way we enable physically consistent 3D reconstruction for interactions that can be modeled by emission–absorption, refraction, and limited multiple scattering. Directly incorporating these interactions into a differentiable pipeline captured by a parameterized observer, allows decoupling the optimization procedure from both, the specific type of interaction and the capture mechanism. We perform a first experimental validation of our method using simulated and experimental optical scans from different sensing devices. Lastly, we explore the inter-domain capability of the new reconstruction method to other inverse problems, including muon tomography imaging.

Fifth MODE Workshop on Differentiable Programming for Experiment Design (MODE2025)
8-13 June 2025
Kolymbari, Crete, Greece

*Speaker

 ISSN 1824-8039 . Published by SISSA Medialab.

PoS(MODE2025)033

1. Introduction

A wide range of imaging and sensing techniques, from optical light fields to muon scattering tomography (MST), are traditionally treated as separate problems, each requiring its own forward models and reconstruction algorithms. However, at their core, these modalities can all be expressed within a unified physical framework: particles traverse a scene, interact with matter, and their measurements encode structural or material information. Casting such processes into a common formulation is not only conceptually elegant, but also practically beneficial. A unified view enables cross-initialization between modalities, efficient integration of heterogeneous data sources, and a shared optimization pipeline that reduces computational overhead while improving reconstruction robustness.

Muon scattering tomography, for instance, exploits naturally occurring cosmic-ray muons to probe the interiors of dense objects by measuring the angular distribution of scattering events. Light field imaging captures the spatial and angular distribution of light rays to recover geometry and appearance. Despite spanning vastly different energy scales and application domains, both are governed by similar path-based transport and can be described by aggregating interaction quantities, like scattering power or radiance, along a path and optimizing a reconstruction loss. This perspective motivates our contributions: (1) a fully differentiable pipeline for optimizing diverse physical phenomena using explicit voxel grids, (2) a demonstration that 3D reconstruction across modalities can be achieved in a common framework, and (3) a modular optimization scheme that separates path aggregation from loss definitions to broaden applicability.

The foundations of this unified view lie in classical inverse problems: specify a physics-based forward operator and infer scene parameters from data under an appropriate noise model. With advances in hardware and deep learning, large-scale optimization pipelines using gradient descent on GPUs have become practical [1]. Within computer vision and graphics, inverse rendering, which describes the process of inferring 3D scene properties, like lighting, materials, and geometry from 2D observations, has emerged as a particularly prominent domain. Neural Radiance Fields (NeRF) by Mildenhall *et al.* [2] have demonstrated high-quality novel-view synthesis through differentiable volumetric rendering. Earlier work by Sitzmann *et al.* [3] anticipated this development by showing that deep networks could recover 3D structure from multi-view images.

Concurrently, several works proposed to replace implicit neural scene representations with explicit voxel or grid-based parameterizations, optimized directly via gradient descent. These include Plenoxels [4] and ReLU fields [5], which inherit ideas from classical methods such as space carving [6] and volumetric rendering [7], but adapt them to the differentiable optimization paradigm. Over time, these methods have been extended beyond purely diffuse appearance models to incorporate reflection, refraction, and spectral imaging, ranging from thermal to full-wavelength modalities.

A closely related mathematical structure drives the recent advances in muon tomography. Geometry-based reconstruction methods such as Point-of-Closest-Approach (PoCA) [8] and Angle Statistics Reconstruction (ASR) [9] as well as statistical voxel-based models [10] all employed voxel tracing of muon scattering interactions. More recent work has pushed this further, with TomOpt [11] showing that fully differentiable voxel-based optimization pipelines can be applied directly to muon scattering data to perform end-to-end learning for detector optimization.

These developments highlight a unifying principle: For particle-based modalities, like optical light fields and muon tomography, solving the inverse problem reduces to the same computational steps. Rays or trajectories are traced through a volumetric representation, an interaction quantity is aggregated along the path, and a loss is minimized to refine the reconstruction. Building on these insights, our work unifies photon- and muon-based processes into a single differentiable optimization framework, combining developments from novel-view synthesis to statistical muon tomography.

2. Methodology

Our goal is to unify the solution of inverse problems across different physical processes. The benefits include: Conceptual and algorithmic unification of reconstruction methods, reduced computational overhead through shared optimization pipelines, improved maintenance and code reuse, scalability to new modalities, and reusability of common operators.

The foundation of our framework is the idea of volumetric rendering. First introduced in graphics by Nelson Max [7] and later popularized in differentiable form through Neural Radiance Fields (NeRF) [2], it describes how light interacts with opaque and semi-transparent media: Rays are traced through a spatial structure from which quantities are sampled (e.g., density, color) and accumulated along the ray to synthesize measurements. This process naturally generalizes to other physical interactions, making it a powerful abstraction for unifying modalities. In optical imaging, the scene is parameterized by a light field or radiance field, describing the radiance emitted in different directions from each point in space. Rays are cast from the camera into the scene, sampling the underlying volumetric representation (e.g., voxel grid). In this work we follow the implementations of Plenoxels [4] and ReLU fields [5] and use regular 3D voxel grids as a spatial structure for all volumetric tasks.

In the optical case, let a ray $r(t) = \mathbf{o} + t\mathbf{d}$, with origin $\mathbf{o}$ and direction $\mathbf{d}$, traverse a voxel grid in t steps, bounded by t_n to t_f. To estimate an aggregated quantity $C(r)$ (like color or density) per ray, Nelson Max [7] formulated the emission-absorption model as a path integral that aggregates contributions from density and emission values along the ray as:

$$C(r) = \int_{t_n}^{t_f} T(t)\,\sigma(r(t))\,c(r(t))\,dt, \quad T(t) = \exp\left(-\int_{t_n}^{t} \sigma(r(s))\,ds\right), \tag{1}$$

where σ is the density, c the color or emission term and $T(t)$ is the absorption term. This emission-absorption model was adopted by NeRFs [2] and will serve as our canonical optical case.

When inspecting Eq. 1 more closely, it becomes clear that it consists of a function $T(t)$ that describes the weighting of two sampled quantities σ and c. We evaluate the integral via numerical quadrature. We can thus rewrite Eq. 1 more generally as:

$$C(r) = \sum_{i=1}^{N} w_i \lambda_i \tag{2}$$

where N is the number of steps i along ray r between t_n and t_f, w is the weighting function, and λ denotes the sampled quantity. Karnewar *et al.* [5] perform optical 3D volume reconstruction

with this simple equation by tracing rays through voxels, aggregate per-voxel interaction terms, and comparing the result to measured data via an L_2 loss function.

Applying this same principle to muon tomography, and based on the ray approximation by Schultz *et al.* [10], we substitute the optical absorption term with the muon scattering variance:

$$C(r) = \sum_{i=1}^{N} W_i \lambda_i, \quad W_i \equiv \begin{bmatrix} L_i & L_i^2/2 + L_i T_i \\ L_i^2/2 + L_i T_i & L_i^3/3 + L_i^2 T_i + L_i T_i^2 \end{bmatrix} \tag{3}$$

where λ_i encodes the scattering power of voxel i and W_i is a geometric weight proportional to the path length L through the i-th voxel and path length T from the i-th voxel to the N-th voxel. Please note that T in this case refers to a distance and not a function as in Eq. 1. The name is kept for compatibility with the work by Schultz *et al.* [10].

As described, both Eq. 1 (optical) and Eq. 3 (muon scattering) reduce to aggregating voxel-based interaction terms weighted by traversal factors. The weighting functions can be swapped out to implement many different phenomena with the sampled quantities having varying dimensions. This unification enables a single computational framework for diverse modalities.

A naive approach to perform numerical integration along a ray would be to use fixed steps to compute voxel intersections with the ray. Standard algorithms like Hierarchical Differential Digital Analyzer (HDDA) [12] can be used to precompute these intersections. While giving exact results, there are two downsides to this approach: First, performing deterministic sampling on a fixed voxel grid results in aliasing due to quantization and second, computing voxel intersections per ray leads to varying numbers of intersections per ray. The second point is crucial when performing optimization on a GPU where non-uniform data layouts cause problems with parallelization. To avoid both problems, it is better to approach voxel sampling from a Monte-Carlo perspective and sample a ray with N steps using a stochastic sampling scheme (e.g. stratified, jittered, or noise-based) [2, 13]. Using these approaches will effectively average the contributions of a ray to a voxel, performing correct down-sampling before quantizing.

As stated above, the sampled quantities can have varying dimensions. While optical rendering uses RGB and density channels, our framework generalizes to arbitrary fields, for example hyper-spectral components for full spectral 3D reconstruction. But not only the sampled quantity can have varying dimensions, also the aggregation function can return non-scalar values, as shown with per sample matrix, a variance field, introduced by the statistical muon scattering formula. Each channel is aggregated according to its physical interaction model but optimized within the same voxel structure.

To optimize the voxel grid, a loss function is required. In optical rendering, supervision often takes the form of an L_2 reconstruction loss between rendered and captured images:

$$\mathcal{L}_{\text{optical}} = \|C(r) - C^{\text{obs}}(r)\|_2^2. \tag{4}$$

with $C^{\text{obs}}(r)$ being observed value for r and $C(r)$ being the predicted one. While being elegant and simple, in other domains such as muon tomography, however, we cannot directly compare observed and predicted rays. Instead, we optimize the difference in distributions through the negative log-likelihood of scattering distributions as in Schultz *et al.* [10]:

$$\mathcal{L}_{\text{muon}} = \log|C(r)| + \frac{1}{2} D^T C(r)^{-1} D + \text{const.} \tag{5}$$

where $C(r)$ is the aggregated variance field, defined in [10] for the observed quantity D. Our framework modularizes the loss, allowing interchanging L_2, negative log-likelihood without changing the ray-tracing backbone.

We optimize voxel parameters (densities, scattering powers, etc.) via gradient descent in a parallel manner. For efficient use of the GPU, we randomly sample a batch of rays per iteration, compute the loss and back propagate, speeding up optimization and performance. In addition to this we employ multi-resolution grids, similar to Plenoxels [4], by iteratively up-sampling the volume to first learn low frequency components and then optimize higher frequency details. To reduce outliers, the loss functions are regularized by additional sparsity losses to enforce compact reconstructions. A related approach is also the use of activation functions during sampling to introduce non-linearity in the pipeline that helps convergence and aids sparsity.

3. Results

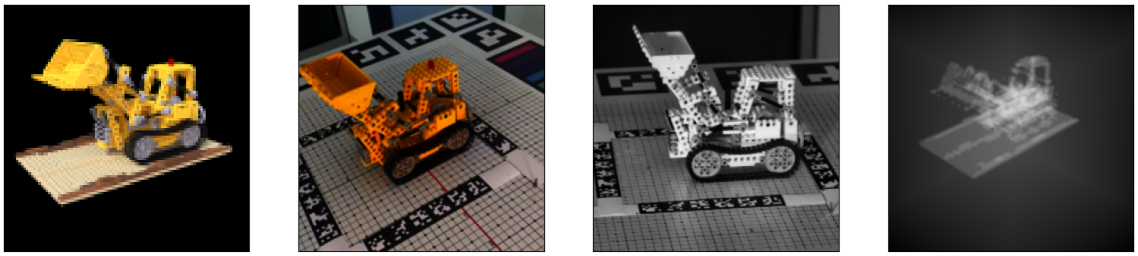
3.1 Datasets

We perform optimization on three different sensing modalities: (1) three-channel RGB optical images, (2) hyper-spectral images with 25 spectral bands for high-dimensional reconstructions of more complex physical phenomena, and (3) muon scattering tomography data to demonstrate the applicability of our framework to a fundamentally different sensing technology.

Our pipeline is validated with both synthetic and real datasets. Synthetic data includes an optical dataset and a muon tomography dataset, while real data is available for optical and hyper-spectral setups. All scenes are normalized and bounded to the unit cube $[-1, 1]^3$ to ensure numerical stability during gradient-based optimization. As a canonical benchmark object, we use the Lego Bulldozer introduced in Mildenhall *et al.* [2]. For real experiments, we recreated this object in Lego bricks in our lab.

The synthetic optical dataset consists of 100 images at a resolution of 100×100 pixels, sampled uniformly over a hemisphere around the object. This serves as a minimal baseline to test the framework. The corresponding real optical dataset also comprises 100 images, captured at a resolution of 1280×720 pixels using a consumer smartphone camera along a hemispherical trajectory. For performance reasons the images are down-sampled 180×320 pixels for optimization. The hyper-spectral dataset consists of 34 images with 25 spectral bands, each at a resolution of 409×217 pixels, acquired with a XIMEA MQ022HG camera with the object rotated on a turntable, capturing a 360° hemisphere and illuminated with tungsten light. Finally, the synthetic muon dataset was generated using a GEANT4 simulation and the B2G4 framework [14] with approximately 1.1×10^6 events in a $(2 \times 2 \times 2)m^3$ world volume. Polypropylene was chosen as the material to approximate Lego bricks. Representative samples from all four datasets are shown in Figure 1.

Across all experiments, reconstructions converge under gradient descent optimization using the Adam [16] optimizer. For optical data, we use a batch size of 4096 randomly sampled rays from the available set of images. We optimize using a multi-resolution scheme with voxel grids of size $32, 64, 128, 256$ and iteration schedules $2^{12}, 2^{13}, 2^{14}$ to improve stability and convergence. Optimization follows the approach of Karnewar *et al.* [5] with a constant learning rate of 3×10^{-3} for all resolutions.



(a) (b) (c) (d)

Figure 1: Representative samples from our datasets. (a) synthetic and (b) real optical images, (c) hyper-spectral image averaged over 25 spectral bands, (d) volumetric rendering of the ground-truth densities for the muon experiment.

3.2 Results and Analysis

Figure 2 shows reconstructions from multiple viewpoints to illustrate reconstruction quality. Among the four datasets, the synthetic optical dataset (Figure 2, a) exhibits the cleanest reconstructions, with minimal noise and consistent clarity and detail throughout. The absence of background simplifies the reconstruction process, resulting in sharp and well-defined features.

The real optical dataset (Figure 2, b) introduces noise, particularly in texture-less regions such as the base plate beneath the bulldozer, where depth estimation is less reliable. The bulldozer itself is reconstructed with reasonable detail, though finer features are not as clearly visible as in the synthetic case. The background is largely filtered out and does not contribute to the reconstructed result. The hyper-spectral dataset (Figure 2, c) shows higher noise levels and overall less detail compared to the optical datasets. While the prominent features of the bulldozer remain visible, significant noise is present around the object. This dataset was reconstructed from only 34 images, in contrast to the 100 images used for the optical datasets, which likely contributes to the observed noisiness. Also, due to the more higher dimensional feature space the optimizer has to navigate a more complex gradient landscape. For the muon tomography dataset (Figure 2, d), reconstructions are generally even, but noise appears around the object and the features are less prominent than in the optical reconstructions. Since the muon paths were traced linearly, scattering information is missing, making a high-resolution reconstruction ill-posed. Nevertheless, the reconstructed density broadly resembles the ground-truth object, with no clusters of density or spurious noise. The object's low-density polypropylene composition also reduces scattering, contributing to a low-frequency reconstruction.

Overall the differences in detail and texture do not indicate model inconsistency, but rather the nature of transport and statistics: deterministic and dense in the optical case and stochastic and sparse for muon scattering tomography. The hyper-spectral case is closer to a true photon spectrum than the simple three-band RGB cases (a) and (b) and exhibits more noise, while still preserving global topology and appears well regularized. The smoothness of the reconstruction suggests stability of gradient descent with the optical operator. Despite its roughness, the muon tomography reconstruction, on the other hand, plausibly locates dense masses, suggesting that the transport weights and cost function adequately capture the physics involved.

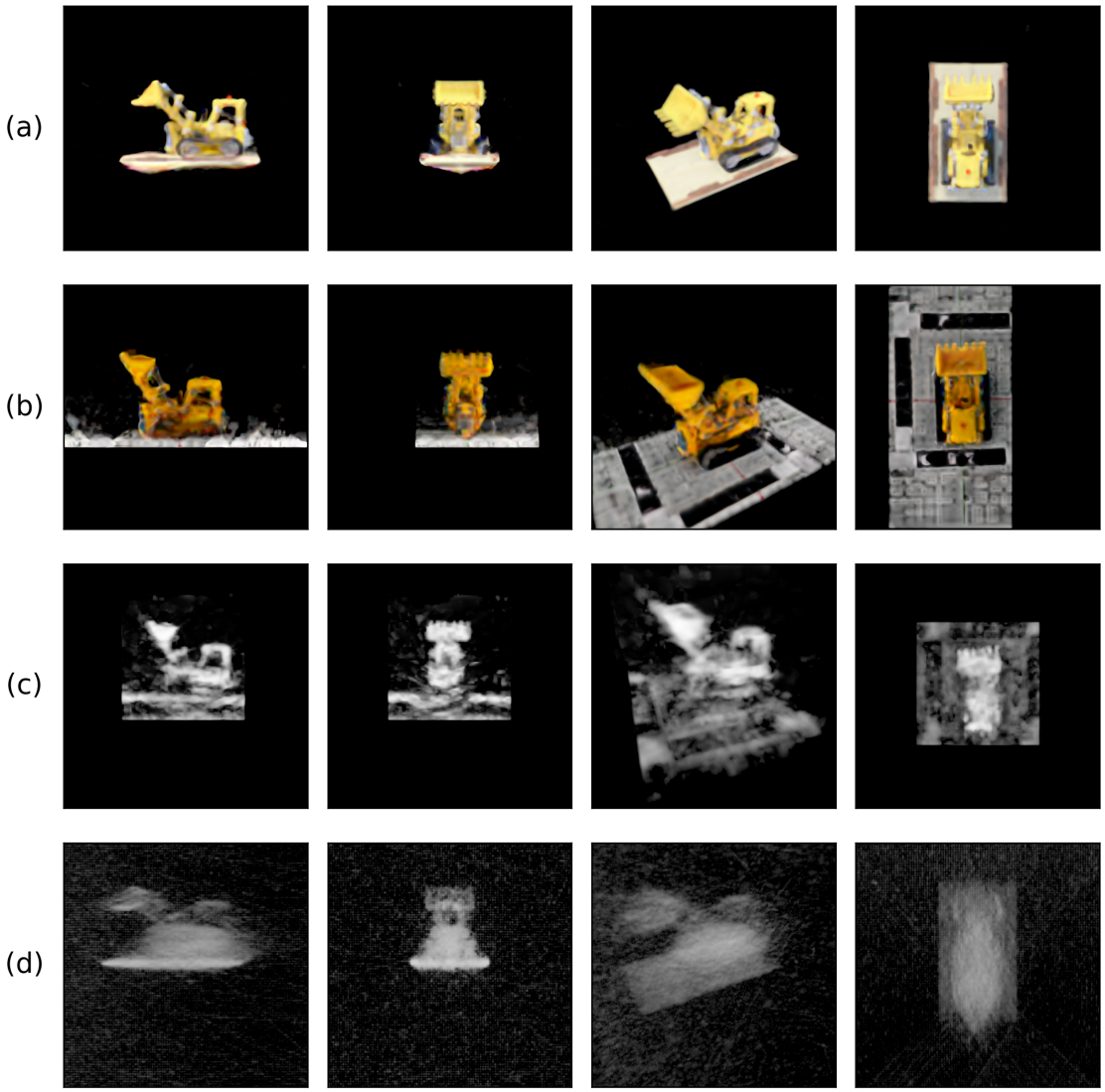


Figure 2: Reconstruction from left, front, orthographic, and top perspective across all datasets, cropped to a volume of interest around the objects. (a) Synthetic dataset, (b) optical dataset, (c) hyper-spectral dataset, and (d) muon scattering tomography dataset. To visualize the hyper-spectral data, a mean over all 25 bands has been computed for rendering. The results in (d) show the density only, as no color information is present in muon data.

4. Outlook & Future Work

The present study illustrates how a unified optimization strategy can recover volumetric structure across very different sensing configurations. We described how different inverse problems can be formulated as a general ray-tracing problem and provided two case studies: optical and muon scattering tomography. We evaluated reconstructions on four datasets in a qualitative way to showcase the presented framework. Synthetic optical data performed best, while real optical, hyper-spectral, and muon datasets showed reduced fidelity due to noise, lighting, or sensing limits.

While the results already point to a wide range of applications, there is still room to refine how measurements and scene properties are represented, as well as how the learning process is regularized. Future efforts could explore richer descriptions of measurement physics and material behaviour, together with improved priors that balance detail preservation and robustness to noise. It may also prove useful to separate different components of the signal or to incorporate complementary sources of information, which could shorten convergence and improve stability. Finally, grounding the formulation more explicitly in physical quantities would strengthen its interpretability and facilitate comparisons across experimental domains.

References

[1] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization Methods for Large-Scale Machine Learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018, doi: 10.1137/16M1080173. [Online]. Available: https://doi.org/10.1137/16M1080173

[2] 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 *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 405–421.

[3] V. Sitzmann, M. Zollhöfer, and G. Wetzstein, "Scene representation networks: continuous 3D-structure-aware neural scene representations," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Red Hook, NY, USA: Curran Associates Inc., 2019, Art. no. 101, 12 pp.

[4] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 5491–5500.

[5] A. Karnewar, T. Ritschel, O. Wang, and N. Mitra, "ReLU fields: The little non-linearity that could," in *ACM SIGGRAPH 2022 Conf. Proc.*, Vancouver, BC, Canada, 2022, Art. no. 27, 9 pp.

[6] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 1, 1999, pp. 307–314.

[7] N. Max, "Optical models for direct volume rendering," *IEEE Trans. Vis. Comput. Graph.*, vol. 1, no. 2, pp. 99–108, Jun. 1995.

[8] L. Schultz, *Cosmic Ray Muon Radiography*, Ph.D. dissertation, Portland State Univ., Portland, OR, USA, 2003.

[9] M. Stapleton, J. Burns, S. Quillin, and C. Steer, "Angle statistics reconstruction: a robust reconstruction algorithm for muon scattering tomography," *J. Instrum.*, vol. 9, no. 11, p. P11019, 2014.

[10] L. J. Schultz, G. S. Blanpied, K. N. Borozdin, A. M. Fraser, N. W. Hengartner, A. V. Klimenko, C. L. Morris, C. Orum, and M. J. Sossong, "Statistical reconstruction for cosmic ray muon tomography," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 1985–1993, Aug. 2007.

[11] G. C. Strong, M. Lagrange, A. Orio, A. Bordignon, F. Bury, T. Dorigo, A. Giammanco, M. Heikal, J. Kieseler, M. Lamparth, *et al.*, "TomOpt: Differential optimisation for task- and constraint-aware design of particle detectors in the context of muon tomography," *Mach. Learn.: Sci. Technol.*, vol. 5, no. 3, p. 035002, 2024.

[12] K. Museth, "Hierarchical digital differential analyzer for efficient ray-marching in OpenVDB," in *ACM SIGGRAPH 2014 Talks* (SIGGRAPH '14), Vancouver, Canada, 2014, Art. no. 40, 1 p., doi: 10.1145/2614106.2614136.

[13] Barron, J., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R. & Srinivasan, P. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. *Proceedings Of The IEEE/CVF International Conference On Computer Vision (ICCV)*. pp. 5855-5864 (2021,10)

[14] Bueno Rodriguez, A., Sattler, F., Perez Prada, M., Stephan, M. & Barnes, S. B2G4: A synthetic data pipeline for the integration of Blender models in Geant4 simulation toolkit. *Journal Of Advanced Instrumentation In Science*. **2024** (2024,5), https://jais.andromedapublisher.org/index.php/JAIS/article/view/476

[15] T. Zhou, R. Tucker, J. Flynn, G. Fyffe, and N. Snavely, "Stereo magnification: learning view synthesis using multiplane images," *ACM Trans. Graph.*, vol. 37, no. 4, Art. no. 65, Jul. 2018.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015.