Nuclear Instruments and Methods in Physics Research A 1093 (2027) 171923

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Full Length Article

A differentiable neural-field forward model for cosmic-ray muon scattering tomography

D. Pagano *

Department of Mechanical and Industrial Engineering, University of Brescia, Italy
Istituto Nazionale di Fisica Nucleare (INFN), Pavia, Italy

ARTICLE INFO

Dataset link: MINT GitHub repository

Keywords:
Muography
Muon tomography
Multiple Coulomb scattering
Neural radiance fields
Inverse problems
Differentiable rendering
Coordinate-based representation
Hash-grid encoding

ABSTRACT

Cosmic-ray muon scattering tomography (MST) recovers the inverse radiation length $\lambda(\boldsymbol{r}) = 1/X_0(\boldsymbol{r})$ of dense or shielded objects, but classical algorithms (PoCA, MLEM) discretise the volume into a fixed voxel grid and become noise-limited at fine resolution. We introduce MINT (Muon Implicit Neural Tomography), a differentiable neural-field framework that represents λ by a coordinate-based network with multi-resolution hash encoding and optimises the per-track multiple Coulomb scattering log-likelihood end-to-end by stochastic gradient descent.

Validated on two Geant4-based benchmarks spanning nuclear security and geophysical surveying, and on real detector data, MINT consistently outperforms MLEM in structural fidelity, dense-object segmentation, and background smoothness, while matching or exceeding its intensity accuracy in both the logarithmic and the linear domain.

In a small-object detection study inside a full-scale ISO-like 20-foot container (twelve high-Z cubes of decreasing size, 10–1 cm, exposed in air and concealed inside 30-cm-side steel cubes), at a fixed 0.5% background-voxel exceedance fraction MINT detects 10 of 12 objects — down to a 1 cm tungsten cube hidden in steel — versus 4 of 12 for both PoCA and MLEM; and when the grid is refined at fixed track budget PoCA and MLEM lose the targets while MINT still localises them. In a mine-tunnel ore-body survey ($1000 \times 800 \times 400$ cm, 12.5 cm voxels, 1.6×10^6 tracks), MINT reduces the log-RMSE to 0.413 against 1.070 for MLEM and 1.895 for PoCA and the linear RMSE to 0.164 against 0.440 cm^{-1}; under grid refinement at a fixed track budget its structural-fidelity advantage over MLEM widens from $\times 1.3$ to $\times 10$ as PoCA and MLEM degrade to noise. Finally, on real data from the INFN Padova large-volume prototype (1.26×10^6 cosmic-ray tracks, six material samples from Al to W), MINT with a Gaussian single-Φ likelihood locates all six samples over a markedly cleaner background than MLEM and with slightly better reconstructed scattering densities, confirming the simulation results under real detector conditions.

1. Introduction

Atmospheric muons, produced in hadronic cascades at altitudes of 15–20 km, arrive at sea level at a flux of ≈ 1 cm^{-2}min^{-1} with a median momentum near 3 GeV/c [1,2]. They represent a free, penetrating, and omnidirectional probe that requires no active radiation source. Flux-attenuation measurements (muon radiography) have exploited this property for decades, with applications ranging from the discovery of hidden chambers inside ancient Egyptian pyramids [3,4] and the imaging of active volcano interiors [5] to the inspection of damaged nuclear reactor cores [6]. The complementary scattering-based modality, muon scattering tomography (MST), leverages the angular deflection induced by multiple Coulomb scattering (MCS) to recover the three-dimensional material distribution inside dense or shielded objects. Pioneered by Borozdin et al. [7], MST is now applied to nuclear contraband detection, nuclear industry, civil engineering and geophysics [8–11]. Because the MCS variance scales as $Z^2\rho/A$, the deflection signal is intrinsically a probe of high-Z content, an attractive feature where active interrogation is impractical.

The classical reconstruction algorithm, Point-of-Closest-Approach (PoCA), assigns each muon's scattering to a single voxel near the intersection of the incoming and outgoing tracks [12,13]. PoCA is fast and intuitive but throws away most of the spatial information of each track. Maximum-Likelihood Expectation-Maximisation (MLEM) algorithms, based on the analytic MCS covariance, address this limitation [12], but they discretise the volume into a fixed voxel grid and do

* Correspondence to: Department of Mechanical and Industrial Engineering, University of Brescia, Italy.
E-mail address: davide.pagano@unibs.it.

https://doi.org/10.1016/j.nima.2026.171923
Received 5 June 2026; Received in revised form 20 July 2026; Accepted 30 July 2026
Available online 10 August 2026

not exploit modern coordinate-based representations. Recent work has applied machine learning and differentiable programming to MST at different stages of the reconstruction chain. Bae and Chatzidakis [14] use machine learning to improve the event-level estimate of the scattering depth, whereas Pezzotti et al. [15] enhance simulation-derived muographic images with a deep image-to-image model. Differentiable approaches include TomOpt, an end-to-end framework for detector and inference optimisation [16], and the direct maximisation of a voxelised MST likelihood by automatic differentiation [17]. MINT differs from these methods by representing the inverse radiation length as a continuous multi-resolution neural field and fitting that field directly through the per-track scattering likelihood.

In an apparently unrelated field, coordinate-based neural networks have emerged as a powerful, mesh-free representation of 3D geometry and physical fields [18]. Neural Radiance Fields (NeRF) extended the idea to view-dependent radiance and have since become the de facto standard for inverse problems in scene reconstruction [19,20]. NeRF parameterises the radiance and density of a 3D scene by a coordinate-based neural network and fits its parameters by minimising the discrepancy between rendered and observed images through a differentiable volume rendering integral. The recent multi-resolution hash encoding of Müller et al. [21] reduces NeRF training times by orders of magnitude and makes it natural to handle scenes with strongly non-uniform spatial frequency content. Neural fields and their applications across visual computing and physics are surveyed by Xie et al. [22]. The same continuous, physically grounded parameterisation underlies physics-informed neural networks [23,24] and the broader programme of deep-learning-based solvers for inverse imaging problems [25,26], of which the present work can be regarded as the muon-scattering instance.

In this work we transpose the NeRF formalism to MST and call the resulting framework **MINT** (**M**uon **I**mplicit **N**eural **T**omography). The radiance/density field is replaced by the inverse radiation length $\lambda(\boldsymbol{r}) = 1/X_0(\boldsymbol{r})$, the line integral of λ along each muon path plays the role of the rendering integral, and the ground-truth images are replaced by per-track observables (angular deflections and lateral offsets) whose distribution is prescribed by Highland's MCS theory [1,27]. The forward model is closed-form differentiable in λ and the overall negative log-likelihood is therefore optimised by stochastic gradient descent.

We validate the framework on two simulated benchmarks and one experimental dataset: a small-object detection-limit study, in a full-scale ISO-like 20-foot maritime container [28] ($6.2 \times 2.6 \times 2.7$ m), quantifying the smallest exposed and concealed high-Z object each method can detect; a Geant4 mine-tunnel ore-body (MTOB) survey with a decreasing-radius ore horizon in a granite matrix, probing the ore detection limit and resolution robustness; and real cosmic-ray data acquired with the INFN Padova large-volume muon-tomography prototype [29], with six material samples from Al to sintered W.

The remainder of the paper is organised as follows. Section 2 develops the physical and algorithmic foundations of MINT: the multiple-scattering forward model and per-track likelihood, the neural-field parameterisation of λ, the differentiable line integral, and the training procedure. Section 3.1 presents the small-object detection-limit study. Section 3.2 presents the mine-tunnel ore-body benchmark. Section 4 presents the experimental validation on real detector data. Section 5 discusses the advantages and limitations of the continuous neural-field representation, and Section 6 concludes.

2. Theoretical framework and method

This section develops the physical and algorithmic foundations of MINT. We first derive the multiple-scattering forward model and the closed-form per-track likelihood that constitutes the training objective, then we describe the neural-field parameterisation of λ, the differentiable forward integral, and the full training procedure.

2.1. Multiple Coulomb scattering and path covariance

A relativistic charged particle of momentum p and velocity βc travelling a path length L through a scattering medium acquires, in the small-angle limit, a projected scattering angle that is approximately Gaussian. Let $\boldsymbol{r}(s) \in \mathbb{R}^3$ denote the position along the muon track at path length $s \in [0, L]$, and let

$$T \equiv \int_0^L \lambda(\boldsymbol{r}(s))\, ds \tag{1}$$

be the line integral of the inverse radiation length along the track; expressed in radiation lengths, $T = \int_0^L ds/X_0(\boldsymbol{r}(s))$ is dimensionless. The Highland parameterisation of multiple Coulomb scattering gives the projected angular variance as [1,27,30]

$$\sigma_\vartheta^2(T_i) = \kappa_i T_i \left[1 + \varepsilon \ln T_i\right]^2, \qquad \kappa_i \equiv \left(\frac{13.6\,\text{MeV}}{p_i \beta_i c}\right)^2, \quad \varepsilon = 0.038, \tag{2}$$

for a track i of momentum p_i and velocity $\beta_i c$. The slowly varying bracket $g(T) \equiv 1 + \varepsilon \ln T$ is the logarithmic Highland correction, whose impact is not negligible for the recovery of dense, high-Z material [30]. In what follows we restrict the analysis to the physical branch $g(T) > 0$, equivalently $T > \exp(-1/\varepsilon)$, on which the Highland variance is monotone increasing. The parameterisation is conventionally quoted for $10^{-3} \lesssim T \lesssim 100$; outside this interval we use it as an approximate extrapolation [1,30].

Because T is a line integral of λ, this forward model shares the line-integral structure of Beer–Lambert transmission [31], although here T parameterises a scattering *variance* rather than an attenuation. Crucially, T is a linear functional of λ. For any differentiable parameterisation λ_θ and any per-track objective of the form $\ell_i(\theta) = \phi_i\big(T_i(\theta)\big)$, the chain rule gives

$$\nabla_\theta \ell_i = \phi_i'(T_i) \int_0^{L_i} \nabla_\theta \lambda_\theta\big(\boldsymbol{r}_i(s)\big)\, ds.$$

Thus, at first order, the gradient is a sum of point-wise contributions along the muon path, all multiplied by the common ray-level factor $\phi_i'(T_i)$. The contributions are therefore not independent in the objective: the nonlinear likelihood couples every path point through the shared scalar T_i, and shared neural-field parameters can additionally couple distinct spatial locations. At the quadrature level this has the differentiable ray-marching structure used in NeRF: sample-wise derivatives are accumulated along a ray and then weighted by a common derivative of the rendered quantity [19].

For the scattering observables, take the z-axis along the reconstructed incoming direction extrapolated through the volume, with x and y transverse to it, and let $k \in \{x, y\}$ index either transverse projection. In each kz-plane, the projected deflection $\Delta\vartheta_k$ and the lateral offset Δr_k at the exit face are jointly Gaussian, with covariance [1,30]:

$$\mathrm{Cov}\big[(\Delta\vartheta_k,\, \Delta r_k)\big] = \sigma_\vartheta^2 \begin{pmatrix} 1 & L_i/2 \\ L_i/2 & L_i^2/3 \end{pmatrix}. \tag{3}$$

The three independent entries of this block,

$$a_i \equiv \sigma_\vartheta^2, \qquad b_i \equiv \frac{L_i^2}{3}\sigma_\vartheta^2, \qquad c_i \equiv \frac{L_i}{2}\sigma_\vartheta^2, \tag{4}$$

are all proportional to σ_ϑ^2 and hence, through Eq. (2), to the single scalar T_i. Strictly, the $L/2$ and $L^2/3$ moments are exact for scattering material distributed uniformly along the path; for a heterogeneous λ the offset moments involve depth-weighted line integrals (e.g. $\int \lambda(s)(L-s)^2\, ds$ for the offset variance). We retain the uniform-medium form so that Σ_i (defined in Eq. (6)) couples to the field through the single scalar T_i: the approximation leaves the angular variance exact and affects only how the observed scattering is shared between angle and offset. The two projections are statistically independent, so the joint observable vector — ordered by projection plane — is

$$\boldsymbol{y}_i = (\Delta\vartheta_x,\, \Delta r_x,\, \Delta\vartheta_y,\, \Delta r_y), \tag{5}$$

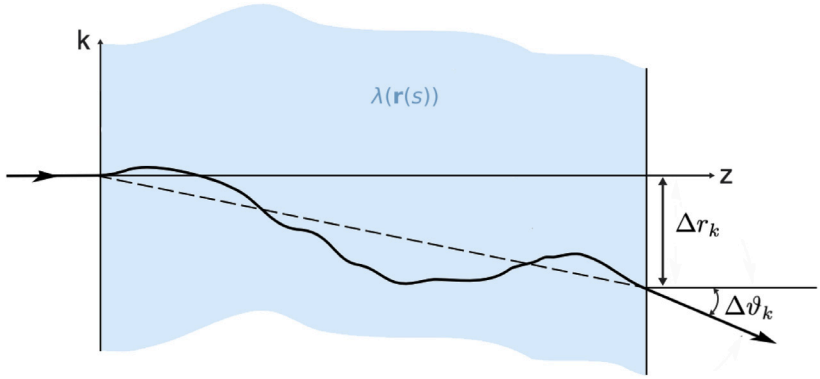

Fig. 1. Geometry of a single muon track in the kz-plane ($k \in \{x, y\}$) (inspired by Fig. 34.10 of Workman et al. [1]). The reconstructed incoming direction defines the z-axis. The muon enters the reconstruction volume (shaded) and undergoes multiple Coulomb scattering. The dashed segment is the straight chord joining the reconstructed entry and exit points; its length is L and, to the working small-angle accuracy, the same length is used in the scattering covariance. The inverse radiation length $\lambda(\boldsymbol{r}(s))$ is integrated along the chord to give $T = \int_0^L \lambda(\boldsymbol{r}(s))\,\mathrm{d}s$. At the exit face, Δr_k is the lateral displacement from the incoming z-axis and $\Delta\vartheta_k$ is the deflection angle of the outgoing direction with respect to z.

and its 4×4 covariance is block-diagonal with two identical 2×2 blocks of the form (3):

$$\Sigma_i = \begin{pmatrix} a_i & c_i & 0 & 0 \\ c_i & b_i & 0 & 0 \\ 0 & 0 & a_i & c_i \\ 0 & 0 & c_i & b_i \end{pmatrix}. \tag{6}$$

The block-diagonal structure admits closed-form expressions for the determinant and inverse of Σ_i without any matrix decomposition:

$$\det \Sigma_i = (a_i b_i - c_i^2)^2, \tag{7}$$

$$\boldsymbol{y}^\top \Sigma_i^{-1} \boldsymbol{y} = \frac{1}{a_i b_i - c_i^2} \sum_{k\in\{x,y\}} \left[b_i\, \Delta\vartheta_k^2 - 2c_i\, \Delta\vartheta_k \Delta r_k + a_i\, \Delta r_k^2 \right]. \tag{8}$$

Fig. 1 illustrates the geometry and the notation introduced above.

2.2. Per-track likelihood

For a given λ, the vector $\boldsymbol{y}_i$ is modelled as a zero-mean multivariate Gaussian with covariance Σ_i:

$$\mathcal{L}_i(\lambda) \;=\; p(\boldsymbol{y}_i \mid \lambda) \;=\; \frac{1}{(2\pi)^2 \sqrt{\det \Sigma_i}} \exp\!\left(-\tfrac{1}{2}\, \boldsymbol{y}_i^\top \Sigma_i^{-1} \boldsymbol{y}_i \right), \tag{9}$$

where the zero mean reflects the absence of a preferred deflection direction and Σ_i encodes the expected scattering spread. Taking $-\log$ of Eq. (9), and using $\det \Sigma_i = (a_i b_i - c_i^2)^2$ (two identical 2×2 blocks, so $\frac{1}{2} \log \det \Sigma_i = \log(a_i b_i - c_i^2)$), the per-track negative log-likelihood (NLL) reads

$$-\log \mathcal{L}_i \;=\; \tfrac{1}{2}\, \boldsymbol{y}_i^\top \Sigma_i^{-1} \boldsymbol{y}_i + \log\!\left(a_i b_i - c_i^2\right) + 2 \log 2\pi. \tag{10}$$

We now reduce both σ_ϑ^2-dependent pieces — the determinant factor and the quadratic form — to explicit functions of the per-track variance σ_ϑ^2.

Determinant factor — By using Eq. (4):

$$a_i b_i - c_i^2 \;=\; \sigma_\vartheta^2 \cdot \frac{L_i^2}{3}\, \sigma_\vartheta^2 - \left(\frac{L_i}{2}\, \sigma_\vartheta^2 \right)^2 \;=\; \frac{L_i^2}{12}\, \sigma_\vartheta^4. \tag{11}$$

Hence $\log(a_i b_i - c_i^2) = 2 \log \sigma_\vartheta^2 + \text{const}$.

Quadratic form — Inserting Eq. (4) into Eq. (8) and using Eq. (11):

$$\tfrac{1}{2}\, \boldsymbol{y}_i^\top \Sigma_i^{-1} \boldsymbol{y}_i \;=\; \frac{6}{L_i^2\, \sigma_\vartheta^2} \sum_{k\in\{x,y\}} \left(\tfrac{L_i^2}{3}\, \Delta\vartheta_k^2 - L_i\, \Delta\vartheta_k \Delta r_k + \Delta r_k^2 \right) \;\equiv\; \frac{6\, P_i}{L_i^2\, \sigma_\vartheta^2}, \tag{12}$$

where we define the *observed scattering power*

$$P_i \;\equiv\; \sum_{k\in\{x,y\}} \left(\tfrac{L_i^2}{3}\, \Delta\vartheta_k^2 - L_i\, \Delta\vartheta_k \Delta r_k + \Delta r_k^2 \right), \tag{13}$$

a positive-definite quadratic form in the measured deflections and offsets.

Combining both pieces, the NLL as a function of the per-track scattering variance alone is

$$-\log \mathcal{L}_i(\sigma_\vartheta^2) \;=\; \frac{6\, P_i}{L_i^2} \cdot \frac{1}{\sigma_\vartheta^2} \;+\; 2 \log \sigma_\vartheta^2 \;+\; \text{const.} \tag{14}$$

Through the Highland relation $\sigma_\vartheta^2 = \kappa_i T_i\, [1 + \varepsilon \ln T_i]^2$ this is an explicit, differentiable function of the line integral T_i, and hence of λ:

$$-\log \mathcal{L}_i(T_i) \;=\; \frac{6\, P_i}{L_i^2\, \kappa_i} \frac{1}{T_i\, [1 + \varepsilon \ln T_i]^2} \;+\; 2 \log T_i \;+\; 4 \log \left| 1 + \varepsilon \ln T_i \right| \;+\; \text{const.} \tag{15}$$

On the physical branch $g(T_i) > 0$, the NLL diverges at both ends: the inverse-variance term grows without bound as $T_i \downarrow \exp(-1/\varepsilon)$ and the predicted variance approaches zero, while the log-determinant grows without bound as $T_i \to \infty$. Thus, under-predicting the observed scattering produces large gradients that pull the field upward along the traversed path, whereas the logarithmic growth in the over-prediction regime prevents unbounded values of λ even without an explicit smoothness penalty.

The stationarity condition is most transparent in the variance itself. Setting $\partial(-\log \mathcal{L}_i)/\partial \sigma_\vartheta^2 = 0$ in Eq. (14),

$$-\frac{6\, P_i}{L_i^2} \cdot \frac{1}{\sigma_\vartheta^4} + \frac{2}{\sigma_\vartheta^2} \;=\; 0 \qquad \Longrightarrow \qquad \hat{\sigma}_{\vartheta,i}^2 \;=\; \frac{3\, P_i}{L_i^2}, \tag{16}$$

a closed-form maximum-likelihood estimate of the per-track scattering variance, strictly positive for any non-zero scattering observation. The corresponding line integral $\hat{T}_i$ is obtained by inverting the monotone Highland map $\sigma_\vartheta^2(T) = \hat{\sigma}_{\vartheta,i}^2$, that is, as the unique solution on the physical branch of

$$\kappa_i \hat{T}_i \left[1 + \varepsilon \ln \hat{T}_i \right]^2 = \frac{3 P_i}{L_i^2}.$$

In the leading-order limit $\varepsilon \to 0$, this equation reduces to the simple estimate

$$\hat{T}_i^{(0)} = \frac{3 P_i}{L_i^2 \kappa_i}, \tag{17}$$

The full solution is a slowly varying logarithmic correction to $\hat{T}_i^{(0)}$ and is obtained by solving the preceding scalar equation. The explicit dependence of the leading-order estimate on p_i, β_i, $\Delta\vartheta_k$, and Δr_k follows by substituting the definitions of κ_i in Eq. (2) and of P_i in Eq. (13). Because $\sigma_\vartheta^2(T)$ is monotone increasing on the physical branch $g(T) > 0$, $-\log \mathcal{L}_i$ is a proper unimodal function of T_i: the log-determinant term

prevents the unbounded growth of λ that would otherwise minimise the quadratic form alone, and yields a finite, physically grounded per-track optimum $\hat{T}_i$.

The loss in Eq. (15) is posed independently for each track. To turn it into a well-posed 3D inverse problem requires a single representation of $\lambda(\boldsymbol{r})$, whose line integrals simultaneously satisfy all per-track constraints. In the following we introduce the neural-field parameterisation adopted for this purpose.

2.3. Neural-field representation of λ

We parametrise $\lambda(\boldsymbol{r})$ with a coordinate-based neural field, built as an encoder/decoder pipeline followed by a positivity-enforcing output transform. All trainable parameters of both stages are jointly collected in the vector θ.

Encoder — The encoder $h_\theta : \mathbb{R}^3 \to \mathbb{R}^F$ is the multi-resolution hash encoding of Müller et al. [21]. The reconstruction volume is overlaid with N_ℓ axis-aligned cubic grids that share the same spatial extent but differ in resolution. Their side counts form a geometric progression $R_\ell = R_0\, b^\ell$ for $\ell = 0, \dots, N_\ell - 1$, where R_0 is a coarse base resolution and $b > 1$ a fixed per-level scale factor (e.g. $R_0 = 16$, $b = 1.5$, $N_\ell = 16$, yielding cell sizes that span almost three decades). A query coordinate $\boldsymbol{r}$ is then *simultaneously* located inside one cell on each grid, providing N_ℓ views of the same point at different spatial scales.

Each level ℓ owns an independent table $E^{(\ell)} \in \mathbb{R}^{T_{\text{hash}} \times d}$ of learnable embeddings, with table size T_{hash} fixed in advance (e.g. $T_{\text{hash}} = 2^{19}$) and feature width d typically two or four. We use the subscript "hash" to distinguish this table size from the line-integral T_i of Eq. (1). The rows of $E^{(\ell)}$ are part of the trainable parameter vector θ. The eight vertices of the enclosing cell carry integer indices $(i, j, k) \in \mathbb{Z}^3$ on the level-ℓ lattice, and each is mapped to a row of $E^{(\ell)}$ via a deterministic spatial hash

$$h(i, j, k) \;=\; \big(i\,\pi_1 \;\oplus\; j\,\pi_2 \;\oplus\; k\,\pi_3\big) \bmod T_{\text{hash}}, \tag{18}$$

with $\pi_{1,2,3}$ large fixed primes and $\oplus$ denoting bit-wise XOR [21]. The function h itself is *not* learned: only the table contents $E^{(\ell)}$ are. At coarse levels the number of distinct vertices is smaller than T_{hash}, so the hash is collision-free and acts as a plain lookup; at fine levels the number of vertices vastly exceeds T_{hash}, so many distinct vertices share the same row. This is by design: training gradients are non-zero only at vertices visited by muon tracks, and physically meaningful regions therefore dominate the optimisation of the contested rows, while the impact of hash collisions is mitigated by the multi-resolution representation.

The level-ℓ embedding at $\boldsymbol{r}$ is reconstructed by trilinear interpolation between the eight hashed rows, with weights determined by the relative position of $\boldsymbol{r}$ inside the cell. Trilinear interpolation makes the resulting feature continuous in $\boldsymbol{r}$ and differentiable almost everywhere; its spatial derivative need not be continuous across cell boundaries. The interpolation remains differentiable with respect to the learned embeddings, as required to back-propagate the line-integral loss that defines T_i. The N_ℓ per-level embeddings are then concatenated into a single feature vector $\boldsymbol{f}(\boldsymbol{r}) = \big[\boldsymbol{e}_0(\boldsymbol{r}), \dots, \boldsymbol{e}_{N_\ell - 1}(\boldsymbol{r})\big] \in \mathbb{R}^F$ with $F = N_\ell\, d$ (e.g. $F = 32$ for $N_\ell = 16$, $d = 2$), which summarises the local structure of the field at all spatial scales simultaneously. Because each level has its own fixed-capacity table, the encoder naturally allocates more representational power to regions where the field varies rapidly (for example around inclusion boundaries) and remains inexpensive where it does not (for example in homogeneous air).

Decoder — The decoder $g_\theta : \mathbb{R}^F \to \mathbb{R}$ is a small fully connected network (one hidden layer of width 128 in this work) that maps the feature vector $\boldsymbol{f}$ to a single real-valued pre-activation s. Because all spatial dependence is carried by the encoder, the decoder does not have to be deep: its goal is just to project the multi-scale descriptor onto a scalar.

Field assembly and positivity — Composing the two stages and applying a positive output transform yields the field

$$\lambda_\theta(\boldsymbol{r}) \;=\; \mathrm{softplus}\big(g_\theta \circ h_\theta(\boldsymbol{r})\big), \qquad \mathrm{softplus}(s) \;=\; \log\big(1 + e^s\big), \tag{19}$$

which is continuous, strictly positive, and differentiable *almost* everywhere with respect to both $\boldsymbol{r}$ and θ (with respect to $\boldsymbol{r}$ it is differentiable except at hash-grid cell boundaries, where the derivative of the trilinear interpolation may be discontinuous). Softplus is preferred over a plain ReLU, which would clamp λ to exactly zero on a half-space of pre-activations and stall the optimiser there, and over an exponential, whose super-linear growth amplifies large pre-activations and destabilises training. Softplus interpolates between these two extremes: it is linear in the $s \gg 0$ regime — bounding the maximum possible growth rate of λ — while approaching zero smoothly without ever reaching it.

Finally, hash encoding is preferred over alternative coordinate encodings such as sinusoidal-activation networks [32], Fourier features [19,33], dense voxel grids [34] and low-rank tensorial decompositions [35], as training time is reduced, which is critical for the near-real-time applications envisaged for this method.

2.4. Forward model

Eq. (19) delivers a continuous, differentiable field $\lambda_\theta(\boldsymbol{r})$ that can be evaluated at any point in space, but the per-track NLL of Eq. (15) couples to it only through the scalar line integral T_i defined in Eq. (1). The forward model, described in this section, bridges the two representations: it specifies how each per-track contribution $T_i(\theta)$ is computed from λ_θ, in a way that remains differentiable in θ so that gradients of the NLL can propagate back to every entry of the hash tables and to every weight of the decoder via the chain rule.

For each track i we approximate the muon trajectory inside the reconstruction volume by the straight segment from the $\boldsymbol{r}_{\text{in},i}$ entry point to the $\boldsymbol{r}_{\text{out},i}$ exit point, both reconstructed from the upstream and downstream trackers. Its length is $L_i = \|\boldsymbol{r}_{\text{out},i} - \boldsymbol{r}_{\text{in},i}\|$, consistently with the geometric convention introduced above. This straight-line approximation is correct to leading order in the scattering angle; a more accurate model based on a stochastic path-integral over Brownian bridges conditioned on the entry and exit directions $(\boldsymbol{u}_{\text{in}}, \boldsymbol{u}_{\text{out}})$ measured by the trackers is a natural extension and is left to future work. The line integral in Eq. (1) is evaluated by *stratified jittered Monte Carlo* quadrature with N samples (typically $N = 96$ in this work). The path parameter is first rescaled to the unit interval through the change of variables $s = L_i\, t$, so that $T_i = L_i \int_0^1 \lambda_\theta(\boldsymbol{r}(t))\, \mathrm{d}t$ with $\boldsymbol{r}(t) = \boldsymbol{r}_{\text{in},i} + t\,(\boldsymbol{r}_{\text{out},i} - \boldsymbol{r}_{\text{in},i})$. The unit interval is then partitioned into N equal strata of width $1/N$, and at every training iteration one sample is drawn independently and uniformly inside each stratum,

$$t_n \;\sim\; \mathcal{U}\Big(\tfrac{n-1}{N}, \tfrac{n}{N}\Big), \qquad n = 1, \dots, N. \tag{20}$$

The integral is approximated by the Monte Carlo average

$$T_i(\theta) \;\approx\; L_i \cdot \frac{1}{N} \sum_{n=1}^{N} \lambda_\theta\big(\boldsymbol{r}_{\text{in},i} + t_n(\boldsymbol{r}_{\text{out},i} - \boldsymbol{r}_{\text{in},i})\big). \tag{21}$$

This scheme combines two complementary ideas: (i) *stratification* ensures that every $1/N$-fraction of the track contributes exactly one sample, guaranteeing uniform coverage of the path and yielding lower variance than a plain Monte Carlo estimator with the same number of samples; (ii) *jittering*, i.e. randomising t_n inside each stratum at every iteration instead of fixing it at the stratum midpoint, breaks the periodic coincidence between sample positions and the regular hash-grid vertices that would otherwise inject aliasing artefacts into the gradient, and adds a mild stochastic perturbation that acts as an implicit regulariser on θ. The estimator is unbiased for any N and its gradient factorises as

$$\nabla_\theta T_i \;=\; L_i \cdot \frac{1}{N} \sum_{n=1}^{N} \nabla_\theta\, \lambda_\theta\big(\boldsymbol{r}_{i,n}\big), \tag{22}$$

flowing through all N stratified sample positions simultaneously. A single muon track therefore adjusts the hash-table representation along its entire path through the volume, coupling features at all N_ℓ resolution levels.

2.5. Loss function and training

Combining the per-track NLL of Eq. (10) with the differentiable forward model of Section 2.4 turns the negative log-likelihood into an explicit, differentiable function of the network parameters θ, which can in principle be minimised directly by stochastic gradient descent. On its own, however, the data term leaves the field under-determined in two respects: (i) regions of the volume not crossed by any muon; (ii) high capacity of the hash encoder can be exploited to fit observation noise with high-frequency artefacts in regions that are crossed only sparsely. We address both pathologies by augmenting the data term with two coordinate-space regularisers. The complete training objective is

$$\mathcal{J}(\theta) = -\log \mathcal{L}(\theta) + \alpha\, \mathcal{R}_{\mathrm{TV}}(\theta) + \beta\, \mathcal{R}_{\mathrm{bg}}(\theta), \tag{23}$$

where the regularisers $\mathcal{R}_{\mathrm{TV}}$ and $\mathcal{R}_{\mathrm{bg}}$ are defined in the following. At each iteration, N_{reg} base points $\boldsymbol{r}_j$ are sampled uniformly in the reconstruction volume Ω. With $\boldsymbol{e}_k$ the Cartesian unit vectors and a finite-difference step δ, we define

$$D_k^{\delta}\lambda_\theta(\boldsymbol{r}) \equiv \frac{\lambda_\theta(\boldsymbol{r}+\delta\boldsymbol{e}_k) - \lambda_\theta(\boldsymbol{r})}{\delta}, \qquad k \in \{x, y, z\}. \tag{24}$$

The stochastic anisotropic total-variation regulariser $\mathcal{R}_{\mathrm{TV}}$ [36] is the sample mean

$$\mathcal{R}_{\mathrm{TV}}(\theta) = \frac{1}{N_{\mathrm{reg}}} \sum_{j=1}^{N_{\mathrm{reg}}} \sum_{k\in\{x,y,z\}} \left| D_k^{\delta}\lambda_\theta(\boldsymbol{r}_j) \right|. \tag{25}$$

This edge-preserving smoothness prior penalises small spatial fluctuations while allowing sharp material transitions. The background regulariser $\mathcal{R}_{\mathrm{bg}}$ is a soft L_2 prior pulling λ toward a problem-specific value λ_{bg} at the same random points,

$$\mathcal{R}_{\mathrm{bg}}(\theta) = \frac{1}{N_{\mathrm{reg}}} \sum_{j=1}^{N_{\mathrm{reg}}} \left(\lambda_\theta(\boldsymbol{r}_j) - \lambda_{\mathrm{bg}}\right)^2, \tag{26}$$

combating hallucinations in regions not crossed by any track. The choice of λ_{bg} is scenario-specific: in open-air scenes (Section 3.1) $\lambda_{\mathrm{bg}} = \lambda_{\mathrm{air}}$; for heterogeneous, multi-material scenes (Section 3.2) the zero-background limit $\lambda_{\mathrm{bg}} = 0$ is preferred and turns $\mathcal{R}_{\mathrm{bg}}$ into a global zero-centred L_2 shrinkage prior on λ. Where the data gradient dominates, the relative contribution of $\mathcal{R}_{\mathrm{bg}}$ is small; its main role is to anchor the field in regions reached by few or no muons. We use $\delta = 1$ cm throughout. The two Geant4 benchmarks use $N_{\mathrm{reg}} = 2\,048$ random points per iteration, while the real-data reconstruction uses $N_{\mathrm{reg}} = 16\,384$; in all cases both regularisers are normalised as sample means, so their scale does not depend directly on N_{reg}.

The objective function (23) is minimised by Adam [37] with a cosine-annealed learning rate (warm-up for the first 5% of iterations, minimum ratio 0.05) and gradient clipping at norm 1. Each iteration draws a random mini-batch of muon tracks, computes T_i for each track via Eq. (21) with stratified-jittered quadrature, and back-propagates through the analytic NLL of Eq. (10). The code has been implemented in PyTorch [38].

3. Geant4-based benchmarks

In this section we validate MINT against Point-of-Closest-Approach (PoCA) and maximum-likelihood expectation-maximisation (MLEM) on two benchmarks spanning nuclear security and geophysical surveying: (i) a small-object detection-limit study in a full-scale ISO-like 20-foot maritime container ($6.2 \times 2.6 \times 2.7$ m), with high-Z cubes of decreasing size exposed and concealed inside 30-cm-side steel cubes; (ii) a mine-tunnel ore-body (MTOB) survey with a small air shaft and a decreasing-radius ore horizon (uraninite/galena) in a granite matrix.

Table 1
Material properties used in the Geant4 simulation. X_0: radiation length; $\lambda = 1/X_0$: target field for the muon channel.

Material	ρ (g/cm^3)	X_0 (cm)	λ (cm^{-1})
Air	1.21×10^{-3}	30 420	3.3×10^{-5}
Iron	7.87	1.757	0.5692
Lead	11.35	0.561	1.782
Tungsten	19.30	0.3504	2.854
HEU (93% ^{235}U)	18.70	0.317	3.155

All benchmarks are based on GEANT4 [39] and ECOMUG [40], as detailed below.

3.1. Small-object detection limit

This first benchmark targets the operationally decisive question for security screening: how small an object can each method actually *detect*, including when it is deliberately concealed.

3.1.1. Phantom geometry and setup

The scene is a full-scale ISO-like 20-foot maritime shipping-container proxy ($620 \times 260 \times 270$ cm outer, rounding up the nominal ISO 668 external dimensions of $606 \times 244 \times 259$ cm [28]; 10 mm uniform steel walls, a simplified model of the real corrugated shell and frame; $\lambda_{\mathrm{Fe}} = 0.569$ cm^{-1}) placed in air, with four ideal tracker planes at its faces. Inside we place twelve high-Z cubes of *decreasing* full side (10, 6, 4, 3, 2, 1 cm), well separated along the container axis (Fig. 3, ground truth):

- an *exposed* row of six cubes sitting in air;
- a *hidden* row of six cubes (same size series), each centred inside a 30-cm-side steel cube, so that the small high-Z target must be distinguished from the strongly scattering steel around it.

The target materials are mixed across the high-Z range relevant to interdiction — highly-enriched uranium (HEU, 93% ^{235}U, $\lambda_{\mathrm{HEU}} = 3.16$ cm^{-1}), tungsten ($\lambda_{\mathrm{W}} = 2.85$ cm^{-1}), and lead ($\lambda_{\mathrm{Pb}} = 1.78$ cm^{-1}) — cycled over the objects so the result is not specific to a single material. The relevant material constants are listed in Table 1. Cosmic-ray (CR) muons were generated with ECOMUG, a parametric CR muon generator with realistic momentum and angular distributions, and propagated through GEANT4; entry and exit four-momenta were recorded on four tracker planes at the container faces. After requiring a full four-plane traversal, we retain 5.98×10^5 muon tracks, used identically by all reconstruction algorithms for a fair data-budget comparison.

The reconstruction volume spans $[-310, +310] \times [-130, +130] \times [-135, +135]$ cm. The voxel methods (PoCA and MLEM) reconstruct on a regular grid. We use two grids: a 5 cm grid (0.35 million voxels), on which the detection comparison is scored because it gives the voxel methods enough crossings per voxel to converge, and an ultra-fine 1.25 cm grid (22 million voxels) used solely as a resolution stress test to expose the voxel-size dependence of the voxel methods at a fixed track budget. MINT is queried as the continuous field and, for like-for-like comparison, also resampled onto the same grids.

The network uses a 16-level multi-resolution hash grid with $T_{\mathrm{hash}} = 2^{19}$ entries per table and $d = 2$ features per entry (16.8 M total parameters in the hash grid). With base resolution 16 and per-level scale 1.5, the grid cell size spans from ≈39 cm (level 0) to ≈0.09 cm (level 15), providing sub-millimetre capacity that is limited in practice only by muon-track density and hash-table collisions at the finest levels. These network and training settings are reused by the later benchmarks.

Training uses the objective (23) with $\alpha = 2 \times 10^{-3}$ and $\beta = 5 \times 10^{-3}$. We train using Adam with an initial learning rate 3×10^{-3} and gradient clipping at norm 1; random mini-batches of 8 192 muon tracks (96 stratified samples each) per iteration, for 6 000 iterations on the GPU of an Apple M3 Max via Metal Performance Shaders.

3.1.2. Resolution stress test

Before presenting the detection comparison, we stress a fundamental constraint that applies to all voxel-based methods (PoCA and MLEM) but not to MINT: the voxel size must be chosen so that each voxel accumulates statistically sufficient track crossings, even when this choice is dictated by the available statistics rather than by the intrinsic scale of the features of interest. This is best appreciated by refining the reconstruction grid at a *fixed* track budget. At 5 cm voxels the 5.98×10^5 tracks populate each voxel with enough crossings for MLEM to converge; refining to 1.25 cm multiplies the voxel count by 64 while leaving the data fixed. The resulting starvation is method-specific but severe in every case: PoCA condenses each track into a single vertex, so its per-voxel statistics fall by the full factor of 64; for MLEM, which spreads each track over the whole column of voxels it crosses, the per-voxel crossing count falls with the transverse voxel area (a factor of 16), but each crossing constrains the voxel only through a chord length that shrinks with the voxel side, so the per-voxel Fisher information falls even faster, by a factor of ∼256 (Section 5). Fig. 2 shows the consequence at 1.25 cm with all methods on the same grid: PoCA collapses to a sparse field of isolated spikes and MLEM to near-zero noise, the objects becoming irrecoverable, whereas MINT — whose shared multi-resolution hash-grid parameters aggregate gradient signal from all tracks crossing *any* point in a node's receptive field — still localises them. The voxel size is therefore a *statistics-driven lower bound* for the voxel methods, not a free choice; MINT is not tied to the output sampling grid. The same constraint recurs in the mine-tunnel benchmark, where the voxel size was likewise set by the available track budget.

3.1.3. Detection criterion and results

The detection comparison is therefore scored on the 5 cm grid, where the voxel methods are given their best chance (MINT is evaluated as the continuous field). A target is counted as *detected* when the peak reconstructed λ within the object volume exceeds the 99.5th percentile of the reconstructed λ over all same-medium background voxels (air for the exposed row, steel for the hidden row), a criterion that fixes the background-voxel exceedance fraction per method and per medium at 0.5%. This is deliberately self-calibrating: a method with a noisy background (e.g. PoCA, whose air-background 99.5th percentile reaches 3.0 cm^{-1} and whose steel percentile reaches 21 cm^{-1}) is required to clear a correspondingly higher bar, so spurious high-λ speckle does not count as detection.

Fig. 3 shows the xz slice through the two object rows. We emphasise that the 5 cm voxelisation of MINT in this figure is adopted *purely for display*, so that all three methods are shown at the same granularity and the comparison is visually fair; it does *not* enter any quantitative result. All metrics reported in this paper treat MINT as the underlying continuous field — here, the detection score below queries the field directly, so MINT's sub-voxel peaks are never diluted by the 5 cm averaging used only to render this panel. Even at this identical display granularity, PoCA is dominated by diffuse high-λ speckle that swamps the smaller targets and MLEM by per-voxel noise, whereas MINT yields compact, low-noise object voxels over a clean background: the difference is therefore one of reconstruction quality at fixed resolution, not merely of nominal resolution. Fig. 4 summarises the outcome: MINT detects 10 of the 12 objects, against 4 of 12 for both PoCA and MLEM. MINT recovers every exposed target down to 2 cm (missing only the 1 cm cube) and five of the six concealed targets, including a 1 cm tungsten cube centred inside a 30-cm-side steel cube. Crucially, MINT's detections degrade *monotonically* with size (all large objects found, only the very smallest missed), whereas the PoCA and MLEM detections are erratic — driven by background noise spikes that happen to clear the threshold rather than by genuine localisation, so that they miss 4 and 3 cm objects yet occasionally flag a 1 cm one. The advantage is largest exactly where it matters for screening: small and concealed high-Z objects, where the continuous field localises the sub-voxel scattering excess that the fixed-grid methods dilute into a voxel average and lose in noise. The fixed 0.5% criterion is one operating point of a threshold sweep over the background-voxel exceedance fraction (Fig. 5). MINT yields more target detections throughout the sweep, detecting 9/12 objects already at a 0.1% background-voxel exceedance fraction and all 12 at 2%, whereas PoCA and MLEM reach 10/12 and 9/12 only near 5%.

3.2. Mine-tunnel ore-body survey benchmark

The second benchmark evaluates MINT in a geometrically and contrast-wise distinct scenario: the passive survey of ore-bearing bodies in a mine-tunnel rock wall using cosmic-ray muons. This test case probes spatial resolution and material discriminability at scales of tens of centimetres in a moderate-Z matrix, complementing the high-Z container scenario of Section 3.1.

3.2.1. Phantom geometry and setup

The benchmark simulates a shallow mine-tunnel survey in which cosmic-ray muons detect ore-bearing zones embedded in a rock wall [41].

The rock matrix is a $1000 \times 800 \times 400$ cm block of a granite proxy ($X_0 = 11.55$ cm, $\lambda_{\mathrm{rock}} = 0.0866$ cm^{-1}) centred at the origin; this block coincides with the reconstruction volume $[-500, +500] \times [-400, +400] \times [-200, +200]$ cm. We orient the frame so that x is the lateral (cross-gallery) coordinate, y runs along the mine gallery, and z is the muon-propagation (depth) axis. Every inclusion — the air shaft and the eight ore bodies — is modelled as a cylinder with its axis along y. With reference to the panel *Ground truth* in Fig. 6, the inclusions are:

- **Air shaft** (small mine tunnel): a single cylinder of radius $r = 50$ cm ($\lambda \approx 0$), placed off-axis at $(x, z) = (0, -120)$ cm.
- **Ore horizon** at $z = +80$ cm: eight cylinders of *decreasing* radius $r = 60, 45, 30, 20, 15, 10, 7, 5$ cm, laid out along the horizon at $x = -380, -200, 0, 150, 280, 380, 450, -470$ cm respectively — the seven largest in size order from left to right, with the smallest ($r = 5$ cm) placed apart at the far negative-x end. Two minerals of strongly different contrast are interleaved along the row: uraninite UO_2 ($\lambda = 1.648$ cm^{-1}, 19× rock) for the $r = 60, 30, 15, 10, 5$ cm bodies and galena PbS ($\lambda = 1.085$ cm^{-1}, 12.5× rock) for the $r = 45, 20, 7$ cm bodies.

The decreasing-radius series probes the smallest ore body each method can recover; at the 12.5 cm reconstruction grid the $r = 5$ cm body (10 cm diameter) is sub-voxel.

Cosmic muons were generated with EcoMug and propagated through the rock by Geant4; entry and exit muons were recorded on four ideal tracker planes at $z = \pm 205, \pm 220$ cm. After requiring a full four-plane traversal we retain 1.60×10^6 muon tracks, used identically by all reconstruction algorithms.

For this multi-material scene the training loss for MINT is

$$\mathcal{J}(\theta) = -\log \mathcal{L}(\theta) + \alpha\, \mathcal{R}_{\mathrm{TV}} + \beta\, \mathcal{R}_{\mathrm{bg},0}, \tag{27}$$

with $\alpha = 2 \times 10^{-3}$ and $\beta = 5 \times 10^{-3}$. We adopt the zero-background limit $\lambda_{\mathrm{bg}} = 0$ of the background prior of Eq. (26),

$$\mathcal{R}_{\mathrm{bg},0}(\theta) = \mathbb{E}_{\boldsymbol{r}}\left[\lambda_\theta(\boldsymbol{r})^2\right], \tag{28}$$

rather than the λ_{rock}-centred variant: with several distinct material classes spread across more than a decade in λ, no single non-zero background value is appropriate. The resulting global zero-centred L_2 shrinkage term helps suppress the shaft region and uninformed background voxels without biasing the ore-body reconstruction (whose λ values are maintained by the data NLL). All other hyperparameters match those of Section 3.1: Adam with $\mathrm{lr}_0 = 2 \times 10^{-4}$, cosine schedule, batch 8 192 tracks × 96 samples, 4 500 iterations. The reconstruction grid is $80 \times 64 \times 32$ (12.5 cm voxels); PoCA and MLEM are evaluated on it directly, while MINT is queried as the continuous field.

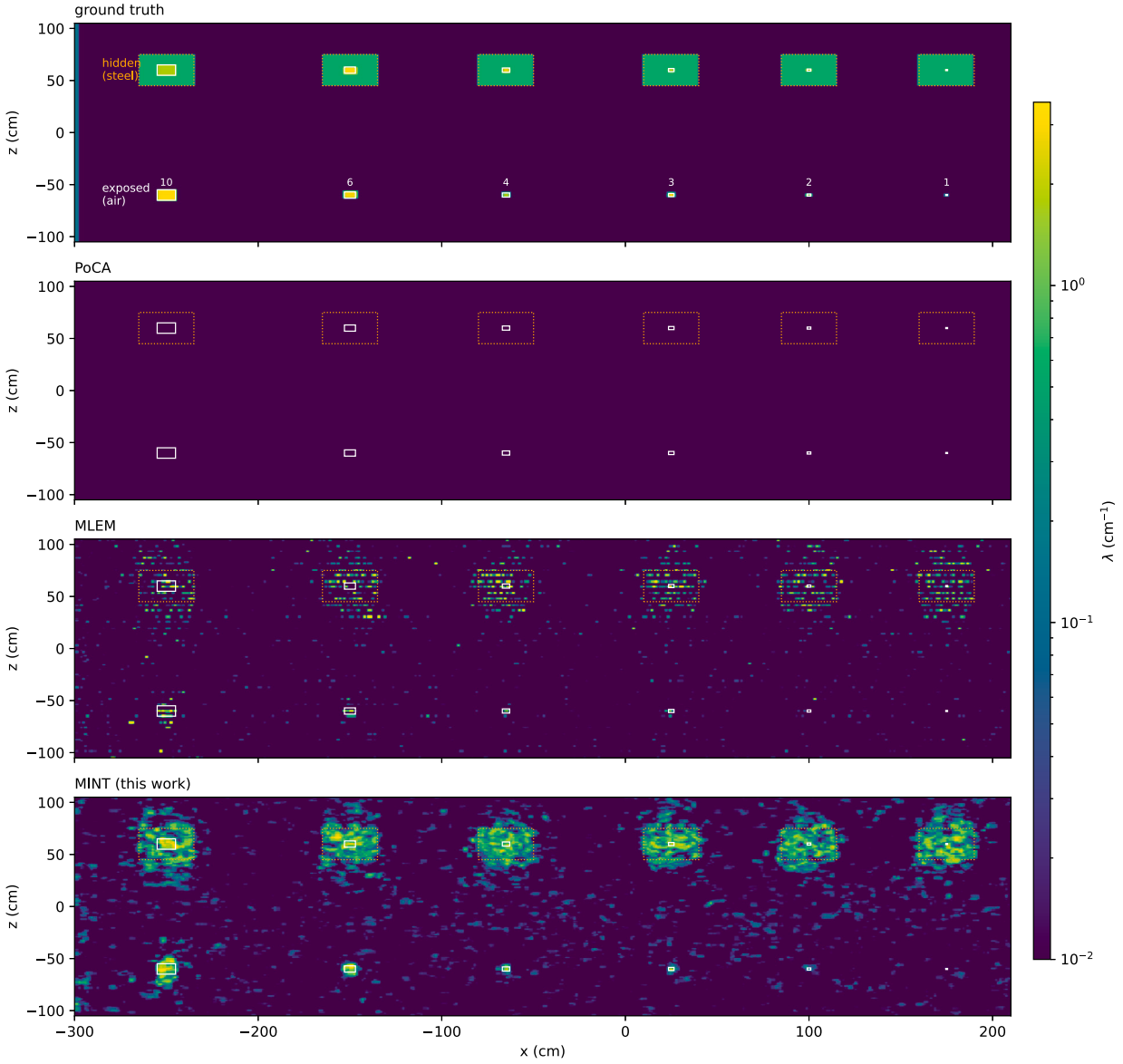

Fig. 2. Resolution stress test — xz slice ($y = 0$) with *all* methods reconstructed on the same ultra-fine 1.25 cm grid ($\approx 2.2 \times 10^7$ voxels, track budget $N/V \sim 0.03$ tracks/voxel). From top: ground truth, PoCA, MLEM, MINT. At a fixed track budget the voxel methods starve — PoCA collapses to isolated spikes, MLEM to near-zero noise — and the objects vanish, while MINT still localises the exposed and hidden targets. This is the voxel-count-driven instability that fixes the operating grid of PoCA/MLEM in every benchmark.

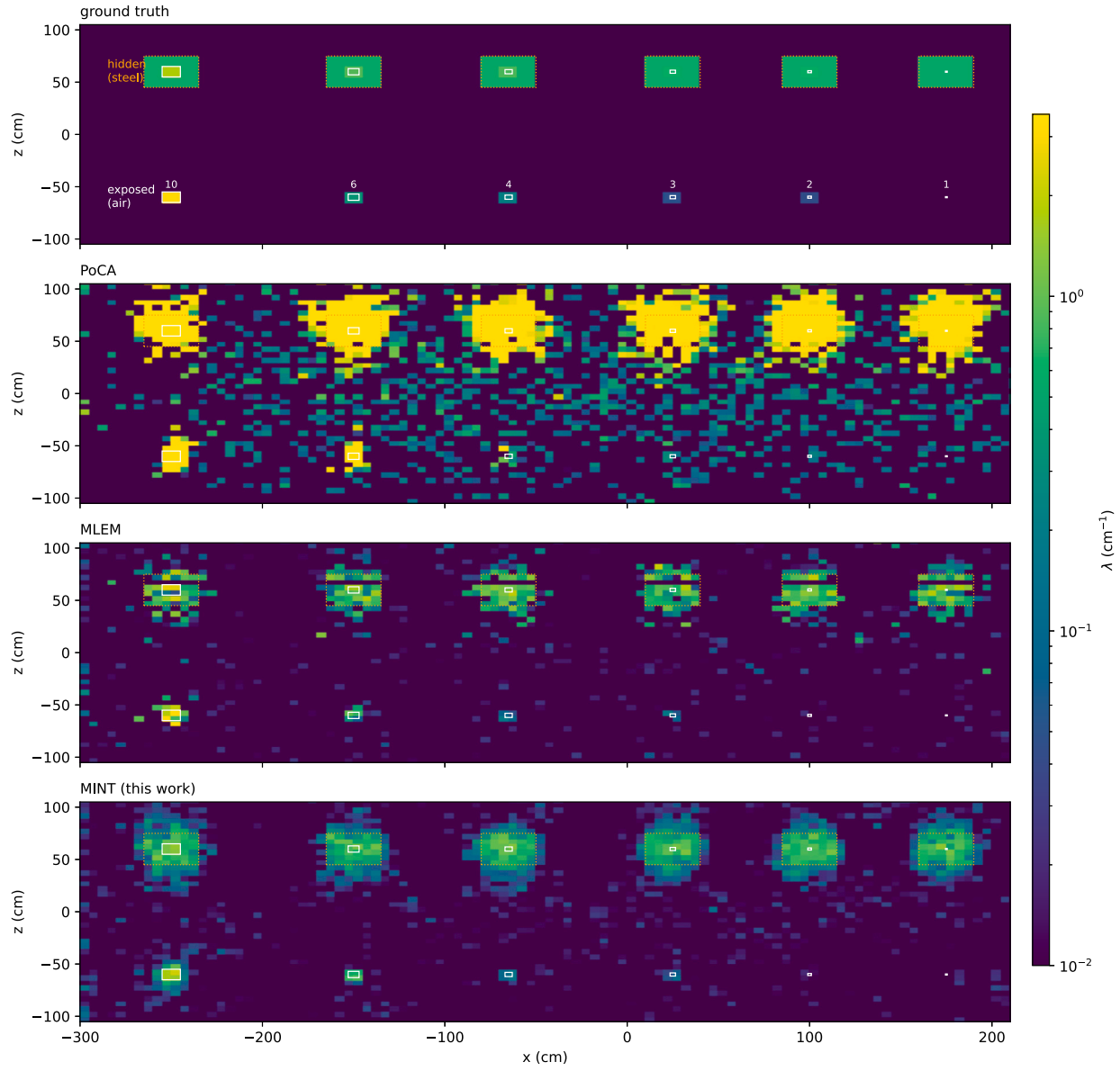

Fig. 3. Small-object detection, xz slice at $y = 0$ through both object rows, with all methods shown on the same 5 cm grid for a visually fair comparison. The 5 cm voxelisation of MINT is for display only: every quantitative metric in this paper uses MINT as the continuous field (here resampled by sub-voxel mean, as in Fig. 2; logarithmic colour scale). From top: ground truth, PoCA, MLEM, MINT. White squares mark the high-Z targets (full sides 10/6/4/3/2/1 cm); the exposed row sits in air ($z = -60$ cm), while the hidden targets are centred inside 30-cm-side steel cubes (dotted orange, $z = +60$ cm). Even at the same voxel size, MINT yields compact, low-noise object voxels over a clean background, whereas PoCA is dominated by speckle and MLEM by per-voxel noise.

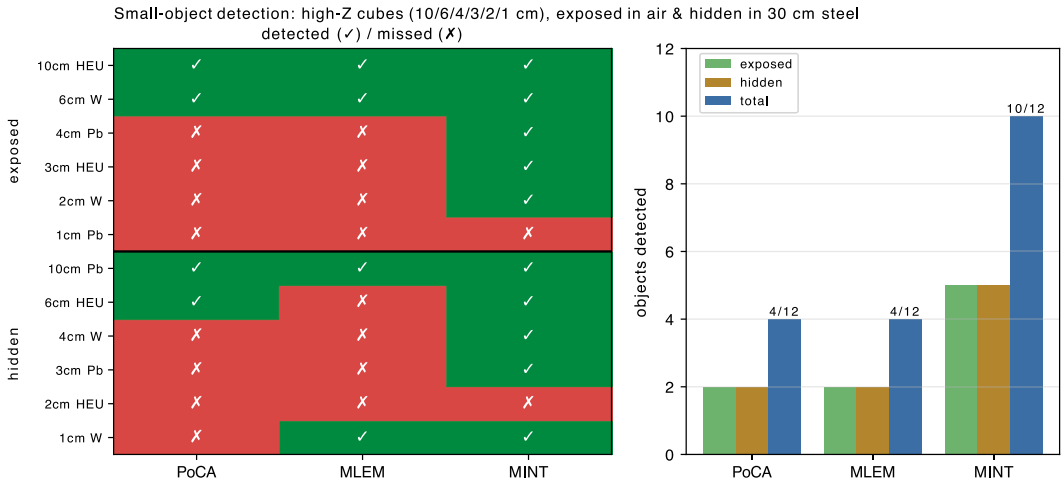

Fig. 4. Detection outcome for the twelve targets. *Left*: detection matrix (green = detected, red = missed) under the 0.5% background-voxel-exceedance criterion. *Right*: number of objects detected per method, split into exposed and hidden. MINT detects 10/12 (5 exposed, 5 hidden) versus 4/12 for both PoCA and MLEM.

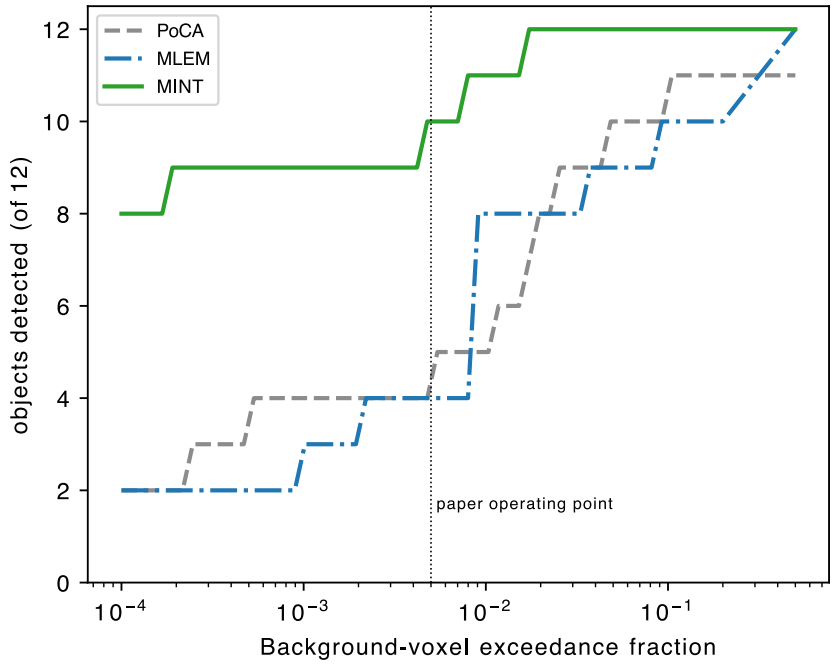

Fig. 5. Threshold sweep for the twelve-target detection task: objects detected versus background-voxel exceedance fraction, obtained by sweeping the per-medium detection threshold over the background quantiles (same detection statistic and backgrounds as Fig. 4; the dotted line marks the 0.5% operating point used in the text). MINT dominates PoCA and MLEM over the entire range.

Table 2
Pixel-wise reconstruction metrics on the 12.5 cm xz voxel slice of the Geant4 mine-tunnel benchmark: log-domain (RMSE, bias, Spearman, SSIM) and linear-domain (λ in cm^{-1}) errors. **Bold**: best per column.

	$\log_{10}\lambda$ domain				Linear	
Method	RMSE	Bias	Spearman	SSIM	RMSE_λ	Bias_λ
PoCA	1.895	−0.371	0.274	0.030	9.635	5.354
MLEM	1.070	−0.416	0.358	0.133	0.440	0.066
MINT	**0.413**	**+0.113**	**0.470**	**0.612**	**0.164**	**0.009**

3.2.2. Results

The quantitative comparison is performed on the xz mid-plane slice ($y = 0$) at 12.5 cm grid spacing (grid 80 × 64 × 32), shown in Fig. 6.

For the log-domain metrics, the scattering-density field is transformed as

$$\mathcal{T}(\lambda) = \log_{10}\left[\text{clip}\left(\lambda, 10^{-4}, 1\right)\right], \tag{29}$$

where λ is expressed in cm^{-1}, so the lower and upper clipping bounds are 10^{-4} and 1 cm^{-1}, respectively. The same transformation is applied to the ground truth and to every reconstruction; we denote the resulting arrays by $G = \mathcal{T}(\lambda_{\text{GT}})$ and $R = \mathcal{T}(\lambda_{\text{rec}})$. The linear-domain RMSE and bias are computed directly from the untransformed, unclipped scattering-density fields. The clipping in Eq. (29) is applied only to the log-domain metrics. We report four complementary pixel-wise quantities: the root-mean-square error

$$\text{RMSE} = \sqrt{\langle (R-G)^2 \rangle}, \tag{30}$$

which measures the error magnitude; the mean bias $\langle R-G \rangle$ (systematic over- or under-estimation); the Spearman rank-correlation coefficient, which measures monotone agreement and is robust to the logarithmic scaling; and the structural similarity index SSIM [42], which measures local agreement in luminance, contrast, and spatial structure.

Table 2 shows that MINT gives the closest overall agreement with the ground truth, winning every metric. In the log domain it more than halves the RMSE of MLEM (0.413 vs. 1.070) and cuts the absolute bias by a factor of almost four (+0.113 vs. −0.416, the negative MLEM value reflecting a systematic underestimation of the scattering density). At the same time it raises the Spearman correlation (0.470 vs. 0.358) and the SSIM (0.612 vs. 0.133), so the gain is not confined to the global intensity scale but extends to the spatial organisation of the reconstructed structures. In the linear domain, which instead weights the dense ore bodies, MINT again more than halves the RMSE of MLEM (0.164 vs. 0.440 cm^{-1}). PoCA gives the largest RMSE in both domains and a large positive bias in the linear one, consistent with the diffuse high-λ speckle that dominates its slice.

Ore-body detection. Following the criterion of Section 3.1, a body is counted as *detected* when its peak reconstructed λ exceeds the 99.5th percentile of the reconstructed λ over the rock background (a 0.5% background-voxel exceedance fraction per method). At 12.5 cm all three methods recover the five largest bodies, down to the $r = 15$ cm (30 cm diameter) cylinder. Of the three smallest bodies ($r = 10, 7, 5$ cm), only the $r = 10$ cm one (20 cm diameter) is detected, and only by MINT — the sole method to raise a significant excess above the rock background there; PoCA and MLEM miss all three. The $r = 7$ cm body (14 cm diameter) is faintly visible in MINT's reconstruction panel but does not clear the detection threshold: at $r \leq 7$ cm the cylinders are comparable to or smaller than the voxel size and embedded in 4 m of rock, so their per-track scattering excess is diluted below the noise floor at this track budget.

3.2.3. Resolution stress test

As in the small-object study (Section 3.1), refining the reconstruction grid at a *fixed* track budget exposes the Fisher-information deficit of the voxel methods: from 12.5 to 6.25 to 3.125 cm the voxel count grows ×8 at each step, so the track budget per voxel, N/V, falls from ∼10 to ∼0.15. N/V equals the mean per-voxel vertex count of PoCA; the mean per-voxel *crossing* count relevant to MLEM is larger (∼310, ∼78 and ∼20 at the three grids) and falls only with the transverse voxel area, but the information carried by each crossing falls with the squared chord length inside the voxel, so MLEM's per-voxel Fisher information degrades even faster — as the fourth power of the voxel side (Section 5). Table 3 reports the structural fidelity (Spearman) and log-RMSE across the sweep.

Table 3
Resolution stress test on the Geant4 mine-tunnel slice: Spearman correlation and log-RMSE as the grid is refined at a fixed 1.6×10^6-track budget. **Bold**: best per row block.

Voxel	N/V	PoCA	MLEM	MINT
Spearman ↑				
12.5 cm	∼10	0.274	0.358	**0.470**
6.25 cm	∼1.2	0.291	0.163	**0.445**
3.125 cm	∼0.15	0.073	0.043	**0.430**
log-RMSE ↓				
12.5 cm	∼10	1.895	1.070	**0.413**
6.25 cm	∼1.2	2.671	2.271	**0.435**
3.125 cm	∼0.15	2.979	2.691	**0.450**

As the grid is refined, MLEM collapses toward per-voxel noise: its Spearman correlation falls from 0.358 to 0.043 (structure essentially lost), and PoCA likewise drops to 0.073. MINT degrades far more gracefully, retaining a Spearman of 0.430 at 3.125 cm, so its advantage over MLEM widens from ×1.3 at 12.5 cm to ×10 at 3.125 cm; in log-RMSE it stays a factor of 2.6 to 6.6 better than the voxel methods across the sweep. MINT is remarkably close to scale-invariant here — its log-RMSE grows only from 0.41 to 0.45 across the sweep — because the shared hash-grid parameterisation keeps the reconstruction coherent where the per-voxel estimators starve, the same continuous-representation dividend analysed in Section 5. Fig. 7 makes this divergence visually explicit: across the three grids MLEM dissolves into per-voxel speckle while the MINT panels keep the ore horizon coherent.

4. Experimental validation on real detector data

The two benchmarks above are simulation-based. To verify that MINT's advantages survive contact with real detector data (non-Gaussian instrumental resolution, imperfect alignment, unknown per-track momentum, etc.) we conclude with a reconstruction of an experimental dataset acquired with the INFN Padova large-volume muon-tomography prototype presented in Pesente et al. [29]. A subset of data used in that work was kindly provided by the authors.

4.1. Apparatus, dataset, and track selection

The prototype [29] consists of two spare CMS Muon Barrel drift chambers of 300 × 250 cm^2 active area and ∼200 μm single-hit resolution, mounted horizontally with a 160 cm vertical gap. Each chamber measures one projection (the Φ view) with eight points from its two outer superlayers and the orthogonal projection (the Θ view) with four points from the central superlayer. As shown in Fig. 8, six material samples — Al, Fe, brass, Cu, Pb and W cubes of ∼10 cm side — rest on a light support inside the gap, spaced by about 25 cm, spanning nominal scattering densities $\lambda_0 = 0.11$–2.7 cm^{-1} (Table 1 of [29]).

We analyse a single run of 9.2×10^6 triggers. For each event we require a full-chamber (eight-point) Φ-view segment and a four-point Θ-view segment in both chambers, with fit-quality cuts and the same $|\Delta\vartheta| < 0.5$ rad fiducial cut adopted in the original analysis; 1.26×10^6 tracks survive.

Following the original analysis, which reconstructs "using the scattering in the Φ-view alone, using the Θ-view only to determine the

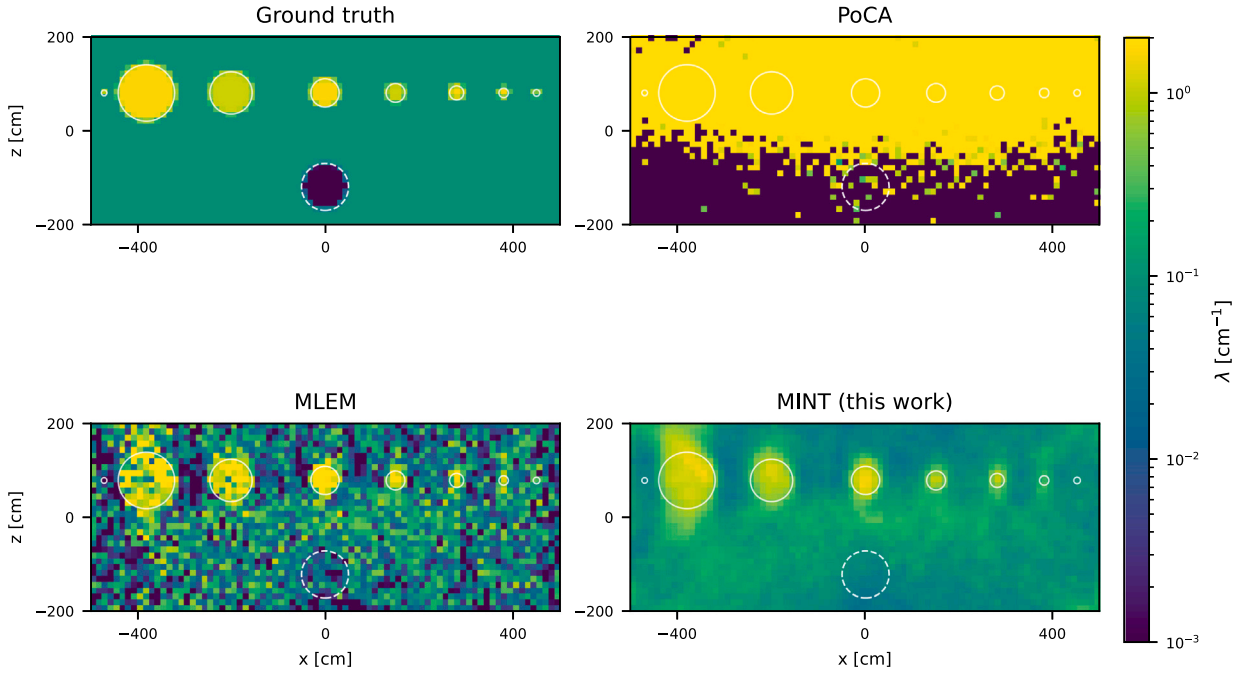

Fig. 6. xz slice (y=0) at 12.5 cm voxel spacing of the Geant4 mine-tunnel benchmark (1.6×10^{6} tracks). Panels: ground truth (top-left), PoCA (top-right), MLEM (bottom-left), MINT (bottom-right, this work). Log-scale colour map spans 10^{-3}–2 cm^{-1}. The ore horizon ($z=+80$ cm; decreasing-radius cylinders, white circles) and the small air shaft ($z=-120$ cm, dashed circle) are resolved by MINT over a smooth, uniform rock background, whereas MLEM reproduces the bright ore amplitudes but with severe per-voxel noise spikes throughout the rock, and PoCA is dominated by diffuse speckle.

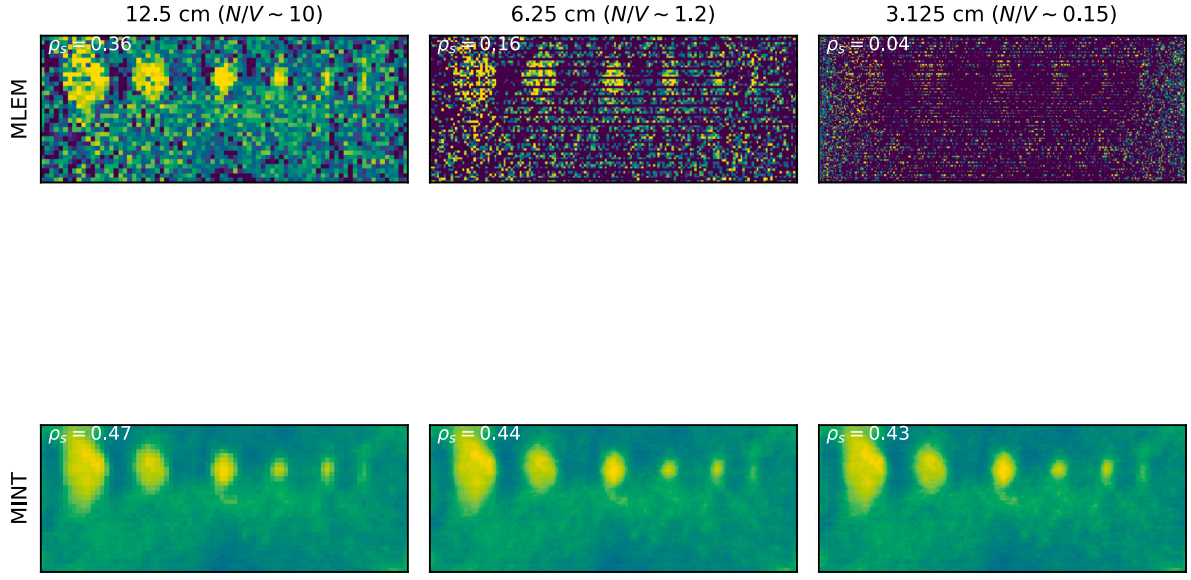

Fig. 7. Resolution stress test on the Geant4 mine-tunnel slice at a *fixed* 1.6×10^{6}-track budget. Top row: MLEM; bottom row: MINT; columns: 12.5, 6.25 and 3.125 cm voxels (left to right), with the per-voxel track budget N/V in each header and the Spearman correlation ρ_s against the ground truth annotated in each panel. Refining the grid starves the per-voxel MLEM estimator, which collapses to noise (ρ_s from 0.36 to 0.04); MINT is nearly unaffected, keeping the ore horizon coherent (ρ_s from 0.47 to 0.43). Same colour scale as Fig. 6.

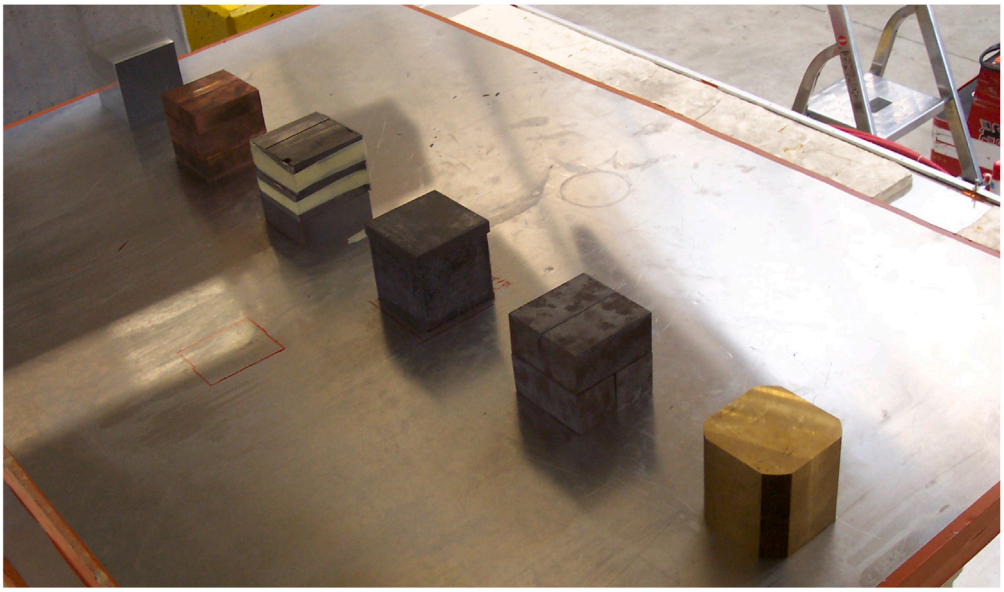

Fig. 8. Six material samples — Al, Fe, brass, Cu, Pb and W cubes of ∼10 cm side — are placed between two CMS Muon Barrel drift chambers of 300 × 250 cm^2 active area. Photo and data are kindly provided by the authors of Pesente et al. [29].

muon trajectory" [29], we do the same: the Θ angle is far less precise (four points over a 3.9 cm lever arm, against eight points over 27.5 cm in Φ). The real-data likelihood therefore uses the scalar angular deflection $\Delta\vartheta_{\Phi,i}$; the Θ view is retained only to determine the three-dimensional muon trajectory. All methods assume a fixed effective momentum ($p_{\mathrm{eff}} = 4$ GeV/c). For each method m, a single post-reconstruction multiplicative factor c_m is determined from the three medium-Z samples (Fe, brass, and Cu), as in the original work, by imposing a unit mean reconstructed-to-nominal ratio,

$$\frac{1}{3}\sum_{s\in\{\mathrm{Fe,brass,Cu}\}}\frac{c_m\,\bar{\lambda}_{m,s}}{\lambda_s^{\mathrm{nom}}} = 1. \tag{31}$$

The same factor c_m is then applied without further adjustment to Al, Pb, and W.

4.2. Reconstruction settings

MLEM is run on a 1 × 1 × 5 cm grid. The same single-Φ angular deflection and effective momentum are used, so both methods receive identical input information. PoCA is not shown: only 0.27% of the tracks place their closest-approach vertex inside the sample slab, so at any grid fine enough to resolve the samples the per-voxel occupancy falls below the minimum required by the median estimator — the same per-voxel starvation seen in Section 3.1. It should be noted that the LMIA algorithm used by Pesente et al. [29] is conceptually close to our MLEM implementation, differing primarily in the optimisation method employed to maximise the likelihood. Consequently, both techniques yield similar results.

For MINT, the encoder capacity is matched to the physical scale: with the default 16 hash-grid levels the finest levels (resolution ∼0.3 mm) are unconstrained between tracks and free to host sub-centimetre artefacts that evade both the stochastic total-variation probes and the quadrature of neighbouring tracks; truncating to 8 levels (finest ∼0.7 cm) removes them.

The per-track variance used in the real-data likelihood is

$$\sigma^2_{\Phi,i} = \sigma^2_{\vartheta}(T_i; p_{\mathrm{eff}}) + \sigma^2_{\mathrm{det}}, \qquad \sigma_{\mathrm{det}} = 9.1\ \mathrm{mrad}, \tag{32}$$

where $\sigma_{\vartheta}(T_i; p_{\mathrm{eff}})$ is given by Eq. (2). The detector term is estimated from air-only control tracks. Their Φ-angle distribution has a Gaussian core with $\sigma = 3.4$ mrad, whereas non-Gaussian tails increase the RMS of the full distribution to 9.1 mrad. We conservatively use the latter as σ_{det} to account approximately for these tails, the fixed-momentum approximation, and residual track-reconstruction errors. The corresponding scalar single-Φ negative log-likelihood is

$$-\log\mathcal{L}_i^{(\Phi)} = \frac{1}{2}\left[\frac{\Delta\vartheta^2_{\Phi,i}}{\sigma^2_{\Phi,i}} + \log\left(\sigma^2_{\Phi,i}\right)\right] + \mathrm{const.} \tag{33}$$

Table 4
Calibrated sample-averaged scattering density λ_0 (cm^{-1}) on real data, against the nominal values of Pesente et al. [29]. Fe, brass, and Cu set each method's global calibration; Al, Pb, and W are independent. Ratios are reconstructed/nominal.

		MLEM		MINT (Gaussian)	
Sample	λ_0^{nom}	λ_0	ratio	λ_0	ratio
Al	0.11	0.162	1.47	0.121	1.10
Fe	0.57	0.724	1.27	0.664	1.16
brass	0.67	0.523	0.78	0.650	0.97
Cu	0.70	0.693	0.99	0.686	0.98
Pb	1.80	1.153	0.64	1.296	0.72
W	2.70	1.247	0.46	1.542	0.57

4.3. Results

Fig. 9 shows the horizontal plane through the sample centres for MLEM and MINT. All six samples are recovered at the correct positions (∼25 cm spacing) by both methods. Table 4 reports the calibrated sample-averaged scattering densities against the nominal values, which are also shown in the right plot of Fig. 9.

Four observations follow. First, MINT produces a much cleaner map: all six samples appear as compact, well-separated clusters over a near-perfect background — including the weakly scattering Al cube, whose excess is comparable to the core angular resolution — whereas MLEM remains affected by per-voxel speckle and scattered outliers. Second, reconstructed scattering densities from MINT are closer to the nominal values than those from MLEM. Third, on the calibration samples MINT is internally more consistent than MLEM (mean absolute deviation from nominal 10%, against 17% for MLEM). Fourth, both methods underestimate Pb and W: this is the known non-linearity caused by the absorption of the low-momentum component of the cosmic spectrum in

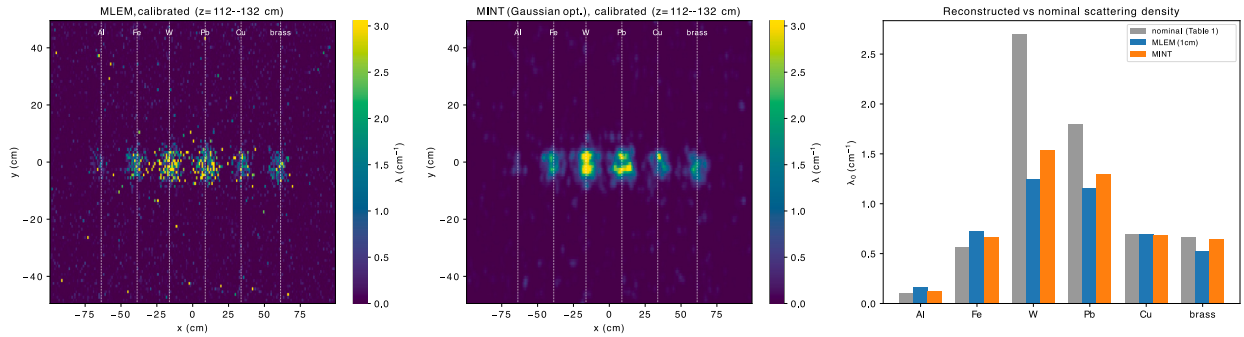

Fig. 9. Experimental data (1.26×10^6 tracks): horizontal plane through the sample centres. Left: MLEM ($1 \times 1 \times 5$ cm voxels). Centre: MINT with the Gaussian single-Φ likelihood of Eq. (33), averaged on the same grid and transversely Gaussian-averaged over 1 cm. Right: calibrated sample-averaged λ_0 against the nominal values (Table 4). Dotted lines mark the six sample positions. All six samples, including the weakly scattering Al cube, appear as compact clusters over a near-perfect background.

dense samples, reported and explained in the original analysis of this apparatus [29], and without event-by-event momentum it is irreducible at the reconstruction level. The real-data validation thus confirms the qualitative conclusion of the Geant4 benchmarks: the continuous field matches the voxel baseline's quantitative accuracy while producing markedly cleaner reconstructions.

To assess the sensitivity of the estimates to track sampling, we split the dataset into four disjoint quarter-statistics subsets and repeated the entire chain — reconstruction and per-method calibration — independently on each, the same procedure used in the original analysis [29]. The resulting dispersion is a descriptive whole-pipeline repeatability measure at one-quarter of the full statistics, covering track-sample and optimisation variability jointly. The median relative standard deviation of the calibrated λ_0 is 6% for MINT against 22% for MLEM, indicating that MINT is less sensitive to which tracks enter the reconstruction. This behaviour is consistent with the parameter sharing of the continuous field. Because only four subsets are available and each fit uses one quarter of the data, however, these dispersions are not confidence intervals and should not be interpreted as the uncertainty of the full-data estimates or as direct evidence of greater information per track.

5. Discussion

The two Geant4 benchmarks collectively show that MINT's continuous neural-field representation consistently outperforms both PoCA and MLEM in spatial resolution and material-contrast recovery, and the real-data validation of Section 4 confirms that these advantages survive genuine detector conditions. These statements are scoped to the standard, unregularised PoCA and MLEM implementations evaluated here, under the fixed-track-budget conditions of our benchmarks; they are not a universal claim about all voxel-based approaches. Regularised or Bayesian voxel variants [10,43] would temper the noise growth under grid refinement, although not the underlying per-voxel information deficit; a systematic comparison with such methods, and with emerging learned reconstruction approaches, is beyond the scope of this work. Section 5.3 quantifies, conversely, how much of MINT's own margin survives without its explicit regularisers. We now place these results in a broader context, examining the formal analogy with Neural Radiance Fields, the structural advantages of a continuous representation, and the current limitations of the approach.

5.1. Connection with NeRF

The mapping between the algorithm presented here and a standard NeRF [19] is intentionally tight. Table 5 summarises the correspondence: in both frameworks a continuous field is parameterised by a coordinate-based network, the observable is expressed as a differentiable line integral of the field, and the parameters are fit by stochastic gradient descent on the log-likelihood of the observed samples. The two key differences are (i) the physical meaning of the field and (ii) the shape of the likelihood: in NeRF the likelihood is an L_2 (or robust) photometric loss, in MST it is the analytic Gaussian of Eq. (10). Read in the more general framework of physics-informed machine learning [24], MINT is therefore a forward model whose "physics" is supplied by Highland's MCS theory rather than by a PDE residual, and whose constraint enters through an analytic likelihood rather than a soft penalty. We stress that the correspondence is architectural, not physical: NeRF solves a view-synthesis problem in which the field is only a means of reproducing images, whereas MST is a genuine physical inverse problem — λ itself is the quantity of interest, there is no emission or occlusion model, and each "measurement" is a single stochastic draw from a scattering distribution whose variance, not amplitude, carries the signal.

Table 5
Mapping between Neural Radiance Fields and the present algorithm.

NeRF	MINT
density $\sigma(\boldsymbol{r})$	$\lambda(\boldsymbol{r}) = 1/X_0(\boldsymbol{r})$
radiance $c(\boldsymbol{r}, \boldsymbol{d})$	— (not needed)
ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$	muon path entry→exit
volume rendering integral	line integral of λ
photometric L_2 loss	MCS Gaussian NLL, Eq. (10)
camera image	per-track $(\Delta\vartheta_x, \Delta r_x, \Delta\vartheta_y, \Delta r_y)$

5.2. Continuous-representation dividend

Across the benchmarks presented, MINT is the most accurate overall, leading on the structural, segmentation, and roughness-based metrics: it appear to be better at preserving coherent object geometry and at suppressing voxel-local fluctuations. The mechanism behind this stability — and behind MINT's resilience precisely where MLEM degrades under grid refinement — can be made precise.

In MLEM, voxel estimates are coupled through the tracks that cross them: the predicted scattering covariance for track i depends on the line integral

$$T_i = \sum_u \ell_{iu} \lambda_u, \tag{34}$$

where ℓ_{iu} is the chord length of track i through voxel u. Thus, each estimate depends on the other voxels crossed by the same tracks. A simple local scaling argument illustrates why the reconstruction becomes more noise-sensitive as the grid is refined. Neglecting inter-voxel correlations, the information available for a cubic voxel of side a scales approximately as

$$I_{vv} \propto n_v \bar{\ell}_v^2, \qquad n_v \propto a^2, \quad \bar{\ell}_v \propto a, \qquad \Longrightarrow \qquad I_{vv} \propto a^4, \tag{35}$$

Table 6
Single-knob ablations on the mine benchmark (12.5 cm protocol of Table 2). The fully unregularised model also reproduces the entire resolution sweep of Table 3 (Spearman 0.471/0.444/0.428 and log-RMSE 0.412/0.435/0.450 at 12.5/6.25/3.125 cm, against 0.470/0.445/0.430 and 0.413/0.435/0.450 for the baseline).

Configuration	log-RMSE	Bias	Spearman	SSIM	RMSE_λ
MINT (paper)	0.413	+0.113	0.470	0.612	0.164
no TV ($\alpha = 0$)	0.412	+0.115	0.473	0.619	0.164
no bg prior ($\beta = 0$)	0.413	+0.116	0.470	0.618	0.163
no TV, no bg	0.412	+0.112	0.471	0.614	0.164
8 hash levels	0.417	+0.100	0.466	0.575	0.162
hash table 2^{15}	0.414	+0.098	0.466	0.557	0.173
MLEM	1.070	−0.416	0.358	0.133	0.440
PoCA	1.895	−0.371	0.274	0.030	9.635

where n_v is the number of tracks crossing the voxel and $\bar{\ell}_v$ is their typical chord length inside it; the scaling assumes a smooth, approximately uniform ray density. This argument is heuristic because the full uncertainty also depends on inter-voxel correlations, but it captures how refinement at a fixed track budget reduces both the number of crossings and the sensitivity per crossing.

For PoCA, which assigns each event to a single reconstructed vertex, the expected per-voxel occupancy instead scales approximately with the voxel volume, $n_v^{\mathrm{PoCA}} \propto a^3$, provided that the PoCA vertex density is locally smooth.

The neural field breaks this per-voxel isolation by *sharing* its hash-table parameters θ across overlapping receptive fields. At hash level ℓ (grid resolution R_ℓ), the feature vector stored at node $\boldsymbol{n}$ enters $\lambda_\theta(\boldsymbol{r})$ for every $\boldsymbol{r}$ in its trilinear support — a cube of side $\sim 1/R_\ell$ — so the gradient that trains $\boldsymbol{n}$ accumulates from *all* tracks crossing that cube, not from a single voxel. At coarse levels ($R_\ell \ll N_{\mathrm{vox}}$) one node therefore pools far more tracks than any fine voxel, and its effective Fisher information per feature greatly exceeds $\mathcal{I}_v$. Equivalently, information flows laterally from well-sampled to poorly-sampled regions through the shared parameters, at no additional computational cost. This is the *continuous-representation dividend*.

The TV regulariser reinforces it: $\mathcal{R}_{\mathrm{TV}}$ couples the field at each sampled point to its three forward Cartesian neighbours through the same network, adding an explicit smoothness prior that further damps independent per-voxel fluctuations at fine scales. MLEM lacks both MINT's multi-scale parameter sharing and its explicit spatial prior. Although the voxel updates are coupled through the line integrals of shared tracks, MLEM assigns one independent parameter to each voxel and has no learned representation that pools gradients across neighbouring locations. Consequently, MINT's advantage widens exactly in the noise-limited, fine-resolution regime, where finer grids expose the voxel methods' Fisher-information deficit while leaving the neural-field estimator essentially unaffected. Consistently, the slice comparisons show this advantage most clearly in the structural, linear-domain, and dense-object metrics.

5.3. Regularisation and architecture ablations

To disentangle the contribution of the explicit regularisers from that of the continuous representation itself, we retrained the mine-benchmark model with every training choice held fixed (full Highland likelihood, learning rate, schedule, seed) and a single knob ablated at a time: the total-variation term removed ($\alpha = 0$), the background prior removed ($\beta = 0$), both removed (fully unregularised), the hash pyramid halved (8 levels), and the hash table reduced sixteen-fold (2^{15} entries). Table 6 reports the paper-protocol metrics.

In this single-seed comparison, the fully unregularised model is numerically very similar to the baseline on every reported metric. Because the field is continuous, the same trained model can be re-evaluated at any grid; it also reproduces the full resolution sweep, retaining Spearman 0.428 at 3.125 cm where MLEM falls to 0.043. These results are consistent with MINT's margin over the voxel baselines arising primarily from the shared continuous representation and the per-track likelihood, but they do not establish statistical equivalence between configurations without repeated optimisation seeds. In particular, the log-determinant is part of the likelihood normalisation and ensures a finite per-track optimum; it is not a spatial regulariser. On the mine benchmark, the explicit TV and background terms have little effect on the reported metrics. Their specific contribution to the visual cleanliness of the real-data reconstruction was not isolated by this ablation and is therefore not inferred here. Likewise, halving the encoder pyramid or shrinking the hash table sixteen-fold changes the metrics by at most a few percent in this single-seed comparison.

For completeness we report the computational cost, taking the mine benchmark (1.6×10^6 tracks; Apple M3 Max laptop) as a representative example of the typical running times. Reconstruction: PoCA takes a few seconds; on CPU, MLEM takes ∼12 minutes for 120 iterations; MINT training takes ∼40 minutes on the integrated GPU for the 4 500 iterations of the mine model and 15 s to evaluate the trained field on the full grid. The memory footprint is modest: 16.8 M parameters, 0.17 GiB of allocated GPU tensors during training, and 104 MB for the track tensors. Because each iteration draws a fixed mini-batch, training time is set by the iteration count, not by the dataset size; the track tensors grow linearly at ∼65 MB per million tracks, while MLEM's per-iteration cost itself scales linearly with the track count.

5.4. Limitations and extensions

The straight-line forward model of Eq. (21) is the single largest source of model bias. Its magnitude can be estimated: conditioned on the measured entry and exit points, the r.m.s. transverse excursion of the true trajectory from the chord scales as $\sigma_\perp \simeq \vartheta_0 D/\sqrt{48}$, where ϑ_0 is the total projected scattering angle and D the tracker separation [44]. For the strongest-scattering tracks of the container benchmark ($\vartheta_0 \approx$ 25 mrad through the 30-cm-side steel cube and its central target, $D =$ 2.6 m) this gives $\sigma_\perp \lesssim 1$ cm — comfortably below both the 5 cm working voxel and the shield size, so the residual error mainly blurs the attribution of scattering within the object neighbourhood rather than the detection itself. Because $\sigma_\perp$ grows linearly with the accumulated deflection, the approximation degrades for very thick, strongly scattering geometries, which we address below. The same chord approximation underlies PoCA and MLEM, so the method comparison is unaffected. A natural improvement is the Brownian-bridge path integral, in which the actual muon trajectory is sampled from the Gaussian process consistent with the entry and exit four-momenta and the line integral (1) is averaged over K such sample paths. The estimator remains unbiased and differentiable through the reparameterisation trick.

A natural extension of the single-field architecture presented here is a multi-modal backbone in which the same hash-grid encoder feeds independent output heads for different physical observables (e.g. the inverse radiation length λ for muon scattering and a gamma-emission rate s for spontaneous fission sources). Because the heads share a common spatial feature representation, structure resolved by one modality is implicitly transferred to the other, enabling material discrimination that neither channel alone can achieve.

The hash encoding may exhibit grid-aligned artefacts in regions of extremely sparse data; mitigation strategies inspired by the Mip-NeRF family [45,46] — in particular conical-frustum sampling and distortion losses — translate directly to the present setting and are a promising avenue of investigation. Implicit-surface variants such as VolSDF [47] provide an alternative route that swaps the density-style field for a signed-distance function and may improve the sharpness of high-Z/low-Z interfaces by representing them explicitly as zero-level sets.

6. Conclusion

We have presented MINT, a differentiable neural-field framework for muon scattering tomography, in which a continuous inverse-radiation-length field $\lambda(\boldsymbol{r})$ is reconstructed by minimising the analytic Gaussian log-likelihood of per-track angular and lateral observables prescribed by Highland's MCS theory. A multi-resolution hash encoding makes training fast on commodity hardware, and the closed-form MCS covariance avoids any matrix decomposition.

Validated on two Geant4 benchmarks spanning nuclear security and geophysical surveying, and on real detector data, the method shows scenario-dependent behaviour explained by a single mechanism: the neural field is regularised globally and continuously, while MLEM assigns one independent parameter to each voxel, without a shared spatial representation and becomes increasingly sensitive to local sampling noise as resolution increases. In a small-object detection study inside a full-scale ISO-like 20-foot container (twelve high-Z cubes 10–1 cm, exposed in air and concealed inside 30-cm-side steel cubes), at a fixed 0.5% background-voxel exceedance fraction MINT detects 10 of 12 objects — every exposed target down to 2 cm and five of six concealed targets, including a 1 cm tungsten cube hidden in steel — against 4 of 12 for both PoCA and MLEM; and when the reconstruction grid is refined at a fixed track budget PoCA and MLEM degrade sharply (PoCA collapsing to isolated vertices, MLEM starving per voxel) while MINT continues to localise the targets. In a Geant4 mine-tunnel survey phantom ($1000 \times 800 \times 400$ cm, 1.6×10^6 tracks through a granite matrix with a small air shaft and a decreasing-radius ore horizon spanning $\lambda = 0$–1.65 cm^{-1}), voxel-array metrics from the 12.5 cm slice show that the neural field gives the best overall agreement with the ground truth: the log-domain RMSE is 0.413 for MINT, compared with 1.070 for MLEM and 1.895 for PoCA, and the linear-domain RMSE is 0.164 cm^{-1}, compared with 0.440 and 9.635. Under grid refinement at a fixed track budget the per-voxel estimates of PoCA and MLEM degrade toward noise (Spearman correlation falling below 0.08 at 3.125 cm) while MINT retains structural fidelity (Spearman 0.43), widening its advantage from $\times 1.3$ to $\times 10$. Finally, the method was validated on real data from the INFN Padova large-volume prototype (1.26×10^6 cosmic-ray tracks through six material samples from Al to W): MINT with a Gaussian single-Φ likelihood locates all six samples at the correct positions over a markedly cleaner background than MLEM and with slightly better reconstructed scattering densities.

The framework is ready for refinements toward full path-integral forward models, multi-modal extensions, learned priors, and GPU-accelerated training for near-real-time operation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge the authors of Pesente et al. [29] for granting access to the experimental data used for the validation in this work.

Data availability

The source code of MINT is openly available in the MINT GitHub repository. The repository includes the complete Python implementation, command-line tools for model training and evaluation, data-format specifications, and a minimal working example. The software is distributed under the MIT License.

References

[1] R.L. Workman, et al., Particle Data Group Collaboration, Review of particle physics, Prog. Theor. Exp. Phys. 2022 (2022) 083C01.

[2] L.I. Dorman, Cosmic Rays in the Earth's Atmosphere and Underground, Springer, 2004, http://dx.doi.org/10.1007/978-1-4020-2113-8.

[3] L.W. Alvarez, J.A. Anderson, F. El Bedwei, J. Burkhard, A. Fakhry, A. Girgis, A. Goneid, F. Hassan, D. Iverson, G. Lynch, Z. Miligy, A.H. Mossallem, M. Moussa, M. Sharkawi, Search for hidden chambers in the pyramids, Science 167 (3919) (1970) 832–839, http://dx.doi.org/10.1126/science.167.3919.832.

[4] K. Morishima, et al., Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons, Nature 552 (2017) 386–390, http://dx.doi.org/10.1038/nature24647.

[5] H.K.M. Tanaka, T. Nakano, S. Takahashi, J. Yoshida, M. Takeo, J. Oikawa, T. Ohminato, Y. Aoki, E. Koyama, R. Tsuji, K. Niwa, High resolution imaging in the inhomogeneous crust with cosmic-ray muon radiography: the density structure below the volcanic crater floor of Mt. Asama, Japan, Earth Planet. Sci. Lett. 263 (1–2) (2007) 104–113, http://dx.doi.org/10.1016/j.epsl.2007.09.001.

[6] H. Miyadera, K.N. Borozdin, S.J. Greene, Z. Lukić, K. Masuda, E.C. Milner, C.L. Morris, J.O. Perry, Imaging Fukushima Daiichi reactors with muons, AIP Adv. 3 (5) (2013) 052133, http://dx.doi.org/10.1063/1.4808210.

[7] K.N. Borozdin, G.E. Hogan, C. Morris, W.C. Priedhorsky, A. Saunders, L.J. Schultz, M.E. Teasdale, Surveillance: Radiographic imaging with cosmic-ray muons, Nature 422 (2003) 277, http://dx.doi.org/10.1038/422277a.

[8] L. Bonechi, R. D'Alessandro, A. Giammanco, Atmospheric muons as an imaging tool, Rev. Phys. 5 (2020) 100038.

[9] G. Bonomi, P. Checchia, M. D'Errico, D. Pagano, G. Saracino, Applications of cosmic-ray muons, Prog. Part. Nucl. Phys. 112 (2020) 103768, http://dx.doi.org/10.1016/j.ppnp.2020.103768.

[10] S. Procureur, Muon imaging: Principles, technologies and applications, Nucl. Instrum. Methods A 878 (2018) 169–179, http://dx.doi.org/10.1016/j.nima.2017.08.004.

[11] R. Kaiser, Muography: Overview and future directions, Philos. Trans. R. Soc. A 377 (2137) (2019) 20180049, http://dx.doi.org/10.1098/rsta.2018.0049.

[12] L.J. Schultz, G.S. Blanpied, K.N. Borozdin, A.M. Fraser, N.W. Hengartner, A.V. Klimenko, C.L. Morris, C. Orum, M.J. Sossong, Statistical reconstruction for cosmic ray muon tomography, IEEE Trans. Image Process. 16 (8) (2007) 1985–1993.

[13] C.L. Morris, et al., Tomographic imaging with cosmic ray muons, Sci. Glob. Secur. 16 (2008) 37–53.

[14] J. Bae, S. Chatzidakis, Momentum-dependent cosmic ray muon computed tomography using a fieldable muon spectrometer, Energies 15 (7) (2022) http://dx.doi.org/10.3390/en15072666.

[15] L. Pezzotti, D. Cifarelli, D. Corradetti, J.P. Costa, G. Gabrielli, L. Galante, A. Gallerati, I. Gnesi, A. Jouve, A. Marrani, A new method for structural diagnostics with muon tomography and deep learning, J. Instrum. 20 (06) (2025) P06034, http://dx.doi.org/10.1088/1748-0221/20/06/P06034.

[16] G.C. Strong, M. Lagrange, A. Orio, A. Bordignon, F. Bury, T. Dorigo, A. Giammanco, M. Heikal, J. Kieseler, M. Lamparth, P. Martínez Ruíz del Árbol, F. Nardi, P. Vischia, H. Zaraket, TomOpt: Differential optimisation for task- and constraint-aware design of particle detectors in the context of muon tomography, Mach. Learn.: Sci. Technol. 5 (3) (2024) 035002, http://dx.doi.org/10.1088/2632-2153/ad52e7, arXiv:2309.14027.

[17] J.-M. Alameddine, F. Sattler, M. Stephan, S. Barnes, Gradient-descent-based reconstruction for muon tomography based on automatic differentiation in PyTorch, 2025, http://dx.doi.org/10.48550/arXiv.2511.05226, Contribution to the Fifth MODE Workshop on Differentiable Programming for Experiment Design (MODE2025). arXiv:2511.05226.

[18] J.J. Park, P. Florence, J. Straub, R. Newcombe, S. Lovegrove, DeepSDF: Learning continuous signed distance functions for shape representation, in: Proc. IEEE Conf. Computer Vision Pattern Recognition, CVPR, 2019, pp. 165–174.

[19] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, NeRF: Representing scenes as neural radiance fields for view synthesis, in: Proc. European Conf. Computer Vision, ECCV, 2020, pp. 405–421.

[20] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, Y. Wang, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Nießner, J.T. Barron, G. Wetzstein, M. Zollhöfer, V. Golyanik, Advances in neural rendering, Comput. Graph. Forum 41 (2) (2022) 703–735, http://dx.doi.org/10.1111/cgf.14507.

[21] T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding, ACM Trans. Graph. 41 (4) (2022) 102.

[22] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, S. Sridhar, Neural fields in visual computing and beyond, Comput. Graph. Forum 41 (2) (2022) 641–676, http://dx.doi.org/10.1111/cgf.14505.

[23] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, http://dx.doi.org/10.1016/j.jcp.2018.10.045.

[24] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (2021) 422–440, http://dx.doi.org/10.1038/s42254-021-00314-5.
[25] G. Ongie, A. Jalal, C.A. Metzler, R.G. Baraniuk, A.G. Dimakis, R. Willett, Deep learning techniques for inverse problems in imaging, IEEE J. Sel. Areas Inf. Theory 1 (1) (2020) 39–56, http://dx.doi.org/10.1109/JSAIT.2020.2991563.
[26] N. Chen, L. Cao, T.-C. Poon, B. Lee, E.Y. Lam, Differentiable imaging: A new tool for computational optical imaging, Adv. Phys. Res. 2 (6) (2023) 2200118.
[27] V.L. Highland, Some practical remarks on multiple scattering, Nucl. Instrum. Methods 129 (1975) 497–499.
[28] International Organization for Standardization, ISO 668:2020 — Series 1 Freight Containers — Classification, Dimensions and Ratings, Standard, International Organization for Standardization, Geneva, Switzerland, 2020, URL https://www.iso.org/standard/76912.html.
[29] S. Pesente, S. Vanini, M. Benettoni, G. Bonomi, P. Calvini, P. Checchia, E. Conti, F. Gonella, G. Nebbia, S. Squarcia, G. Viesti, A. Zenoni, G. Zumerle, First results on material identification and imaging with a large-volume muon tomography prototype, Nucl. Instrum. Methods A 604 (3) (2009) 738–746, http://dx.doi.org/10.1016/j.nima.2009.03.017.
[30] G.R. Lynch, O.I. Dahl, Approximations to multiple Coulomb scattering, Nucl. Instrum. Methods B 58 (1991) 6–10.
[31] D.F. Swinehart, The Beer–Lambert law, J. Chem. Educ. 39 (7) (1962) 333, http://dx.doi.org/10.1021/ed039p333.
[32] V. Sitzmann, J.N.P. Martel, A.W. Bergman, D.B. Lindell, G. Wetzstein, Implicit neural representations with periodic activation functions, in: Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 7462–7473.
[33] M. Tancik, P.P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J.T. Barron, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains, Adv. Neural Inf. Process. Syst. 33 (2020) 7537–7547.
[34] C. Sun, M. Sun, H.-T. Chen, Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2022) 5459–5469.
[35] A. Chen, Z. Xu, A. Geiger, J. Yu, H. Su, TensoRF: Tensorial radiance fields, in: Proc. European Conf. Computer Vision, ECCV, 2022, pp. 333–350.
[36] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1–4) (1992) 259–268, http://dx.doi.org/10.1016/0167-2819(92)90242-F.
[37] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, Int. Conf. Learn. Represent. (ICLR) (2015).
[38] A. Paszke, et al., PyTorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019) 8024–8035.
[39] S. Agostinelli, et al., GEANT4 Collaboration Collaboration, GEANT4—a simulation toolkit, Nucl. Instrum. Methods A 506 (2003) 250–303.
[40] D. Pagano, G. Bonomi, A. Donzella, A. Zenoni, G. Zumerle, N. Zurlo, EcoMug: An efficient COsmic MUon Generator for cosmic-ray muon applications, Nucl. Instrum. Methods A 1014 (2021) 165732.
[41] D. Pagano, A. Lorenzon, Civil and industrial applications of muography, in: Cosmic Ray Muography, World Scientific, 2023, pp. 253–291, http://dx.doi.org/10.1142/9789811264917_0010.
[42] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.
[43] G. Wang, L.J. Schultz, J. Qi, Bayesian image reconstruction for improving detection performance of muon tomography, IEEE Trans. Image Process. 18 (5) (2009) 1080–1089.
[44] W.T. Scott, The theory of small-angle multiple scattering of fast charged particles, Rev. Modern Phys. 35 (2) (1963) 231–313, http://dx.doi.org/10.1103/RevModPhys.35.231.
[45] J.T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, P.P. Srinivasan, Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields, in: Proc. IEEE Int. Conf. Computer Vision, ICCV, 2021, pp. 5855–5864.
[46] J.T. Barron, B. Mildenhall, D. Verbin, P.P. Srinivasan, P. Hedman, Mip-NeRF 360: Unbounded anti-aliased neural radiance fields, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2022) 5470–5479.
[47] L. Yariv, J. Gu, Y. Kasten, Y. Lipman, Volume rendering of neural implicit surfaces, Advances in neural information processing systems 34 (2021) 4805–4815.