Modeling the inverse MEG problem in neuro-imaging using Physics Informed Neural Networks

Ourania Giannopoulou[1*]

[1]Department of Enterprise Engineering, University of Rome Tor Vergata, Rome, Italy.

Corresponding author(s). E-mail(s): ourania.giannopoulou@uniroma2.it;

Abstract

Magnetoencephalography (MEG) forward and inverse modeling is fundamental to neuroscientific discovery, yet the inversion of partial differential equations (PDEs) remains one of the most difficult challenges due to its inherent ill-posedness. While traditional numerical methods often struggle with the computational burden and regularization requirements of these problems, neural networks have recently emerged as a highly viable alternative, offering the ability to learn complex, non-linear mappings and provide efficient, real-time inference. This paper presents a framework for the MEG forward and inverse problems, integrating finite element modeling with neural network techniques. The forward problem is solved using FEniCS to model the electric potential governed by the Poisson equation on a realistic anatomical brain mesh, with magnetic fields computed via the Biot-Savart law. For the inverse problem, we introduce a Physics-Informed Neural Network (PINN) approach in order to deal with the ill condition of the problem. Unlike purely data-driven deep learning approaches that treat this problem as a black box learned from massive datasets, the proposed PINN framework directly embeds the governing physics—Maxwell's equations and the Biot-Savart law—into the loss function, ensuring that the reconstructed sources satisfy the fundamental electromagnetic laws even in data-scarce regimes. We validate the framework on a high-resolution anatomical mesh and compared against the standard Minimum Norm Estimation (MNE). Results demonstrate that the PINN approach achieves a 30.2% improvement over the MNE baseline.

Keywords: Magnetoencephalography, Inverse Problem, Finite Element Method, FEniCS, Ellipsoidal MEG Geometry, Physics Informed Neural Networks

1 Introduction

1.1 Magnetoencephalography and its Applications

Magnetoencephalography (MEG) is a powerful, non-invasive neuroimaging technique that measures the magnetic fields produced by electrical currents in neurons (Hämäläinen et al. 1993; Ebersole and Ebersole 2010). Unlike electroencephalography (EEG), which measures electric potentials on the scalp (Teplan 2002), MEG detects magnetic fields that are less distorted by the skull's low conductivity, offering superior spatial resolution for localizing neural activity (Hämäläinen et al. 1993). A fundamental advantage of MEG lies in its sensitivity to source orientation; it primarily detects the tangential components of primary currents, making it complementary to EEG which is more sensitive to radial sources (Sarvas 1987).

MEG has found widespread application in both clinical and research settings (Baillet et al. 2001). Clinically, it is indispensable for pre-surgical epilepsy mapping, where it localizes epileptogenic foci, and for functional mapping of eloquent cortex (motor, sensory, language areas) prior to tumor resection. In research, MEG enables the study of cognitive processes with millisecond temporal resolution, elucidating neural oscillations, functional connectivity, and event-related responses (Picton et al. 2000). However, realizing the full potential of MEG requires solving the challenging inverse problem: reconstructing the 3D internal current sources from 2D sensor measurements.

1.2 Historical Development and Geometric Modeling

The mathematical foundations of MEG were established in the 1960s and 1970s by Geselowitz (Geselowitz 1967, 1970), who derived the fundamental relationships between internal current sources and external fields in volume conductors. Seminally, Sarvas (1987) provided an exact analytical solution for the magnetic field of a current dipole in a spherical conductor, which remains a standard benchmark for forward modeling.

While spherical models offer computational simplicity, they oversimplify the complex non-spherical geometry of the human head. The extension to ellipsoidal geometry represented a significant mathematical advancement. Ellipsoidal harmonics, as developed by Dassios (2012), provide a rigorous framework for solving boundary value problems in ellipsoidal domains, with applications to EEG and MEG demonstrating improved accuracy over spheres (Kariotou et al. 2003; Kariotou 2004; Dassios and Hadjiloizi 2007). Further work by Fokas and colleagues extended these analytical techniques to multi-shell ellipsoidal models (Dassios and Fokas 2013; Fokas et al. 2009), albeit typically resulting in computationally intensive series expansions.

These analytical developments highlight the importance of accurate geometric modeling. However, for fully realistic head geometries with arbitrary folding and tissue interfaces, analytical solutions become intractable, necessitating numerical approaches such as the Finite Element Method (FEM) and, more recently, physics-informed machine learning.

1.3 The MEG Inverse Problem

The MEG inverse problem—estimating the location and strength of neuronal currents from sensor data—is fundamentally ill-posed (Sarvas 1987; Papanicolaou 1995). This ill-posedness arises from three key factors: (1) *Non-uniqueness*, as demonstrated by the existence of "silent sources" that produce no external magnetic field (Dassios and Fokas 2005); (2) *Instability*, where small measurement noise amplifies into large source errors due to the high condition number of the forward operator; and (3) *Incomplete sampling*, as sensors cover only a portion of the head surface.

Successfully addressing these challenges requires a robust inverse solver equipped with appropriate regularization. Classical approaches often rely on the least-squares framework augmented with priors. For instance, Bayesian methods incorporate spatial priors $p(\mathbf{p})$ to constrain the solution space:

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} L(\mathbf{B}_{\text{meas}}|\mathbf{p})\, p(\mathbf{p}) \tag{1}$$

Common priors include anatomical constraints (limiting sources to gray matter) or depth weighting. Another class of methods, such as beamforming (Hillebrand and Barnes 2005), utilizes adaptive spatial filtering to estimate activity at specific locations by attempting to suppress interference from other sources:

$$S(\mathbf{r}) = \mathbf{w}^T(\mathbf{r})\mathbf{B}_{\text{meas}} \tag{2}$$

Alternatively, Minimum Norm Estimation (MNE) and its variants (LORETA, sLORETA) distribute the source as a continuous current density, minimizing the total energy or Laplacian of the current to select a unique, physiologically plausible solution.

1.4 Application of Neural Networks in EEG/MEG

Deep learning (DL) has recently emerged as a transformative approach for neuroimaging inverse problems, promising to overcome the computational bottlenecks and modeling assumptions of traditional iterative solvers. Early works, such as those by Cui et al. (2019) and Ding et al. (2019), utilized Long Short-Term Memory (LSTM) networks to reconstruct dipole locations and time courses. More recent advancements include "Deep-MEG" by Franceschini et al. (2025), a hybrid CNN-MLP architecture for spatiotemporal reconstruction, and DeepSIF by Sun et al. (2023), which integrates spatial residual networks with temporal encoding.

While these data-driven methods demonstrate impressive capabilities—such as real-time inference and robustness to noise (Pantazis and Adler 2021)—they largely treat the physics of the problem as a black box to be learned from massive datasets. This reliance on large-scale simulations can lead to generalization issues when the physical model (e.g., head geometry) varies. In contrast, our approach utilizes Physics-Informed Neural Networks (PINNs) to explicitly embed the governing partial differential equations—the Poisson equation and Biot-Savart law—directly into the loss function. This ensures that the network's predictions are not just statistically likely, but physically consistent with Maxwell's equations, reducing the dependence on massive labeled datasets.

1.5 Contributions of This Work

This work presents a comprehensive open-source framework for MEG forward and inverse modeling that bridges classical computational methods with modern deep learning approaches. The key contributions are:

- **Forward Modeling:** For the solution of the forward problem we utilize the opensource finite element solver FEniCS (Alnæs et al. 2015; Logg et al. 2012), that computes electric potentials via the Poisson equation on realistic anatomical brain meshes and generates full 3-component vector magnetic fields through the Biot-Savart law.
- **Physics-Informed Inversion:** Introduction of a Physics-Informed Neural Network (PINN) architecture that regularizes the inverse problem by enforcing Maxwell's equations and boundary conditions in the loss function, enabling robust performance even with scarce labeled data.
- **Benchmarking:** A quantitative comparison with the widely used Minimum Norm Estimation (MNE) method. In this experiment, the PINN framework achieved a mean localization error of 0.59 cm compared to 0.84 cm for MNE, corresponding to a percentage improvement of 30.2% .
- **Open-Source Implementation:** Provision of a complete, reproducible software pipeline that integrates finite element modeling (FEniCS) with deep learning (PyTorch) for neuroimaging research.

1.6 Paper Organization

The remainder of this paper is organized as follows: Section 2 outlines the fundamental mathematical formulations governing bioelectromagnetism, including the quasi-static Maxwell equations, the Poisson equation for volume conduction, and the Biot-Savart law for magnetic fields. Section 3 details our methodology, describing the implementation of the FEniCS forward model used for the data generation, the PINN architecture and the composite physics-based loss functions. Section 4 presents the results, including hyperparameter optimization, validation of the proposed model, performance analysis in data-scarce regimes, and the comparison against the Minimum Norm Estimation (MNE) method. Finally, Section 5 concludes the paper by summarizing the key findings and discussing the implications for future deep learning applications in neuroimaging as well as suggestions for future work and improvements of the prososed model.

2 Mathematical Formulations for the forward and inverse problems

2.1 Quasi-static Approximation

The electromagnetic fields generated by neuronal currents operate in a frequency range from DC to approximately 1000 Hz (Hämäläinen et al. 1993). In this low-frequency regime, the wavelength $\lambda = c/f$ (where c is the speed of light and f is frequency) is on the order of hundreds of kilometers, which is vastly larger than the head dimensions

(∼20 cm). This disparity in length scales justifies the quasi-static approximation of Maxwell's equations (Geselowitz 1967, 1970).

In the quasi-static regime, the displacement current $\partial \mathbf{D}/\partial t$ is negligible compared to the conduction current $\mathbf{J}$, and the magnetic field $\mathbf{B}$ is decoupled from the electric field $\mathbf{E}$. This leads to the simplified system:

$$\nabla \times \mathbf{E} = 0 \tag{3}$$

$$\nabla \cdot \mathbf{J} = 0 \tag{4}$$

The first equation (3) implies that the electric field is conservative, while the second equation (4) expresses the conservation of electric charge in the absence of sources and sinks.

We model the head as a volume conductor $\Omega \subset \mathbb{R}^3$. The total current density $\mathbf{J}(\mathbf{r})$ consists of two components:

$$\mathbf{J}(\mathbf{r}) = \mathbf{J}_p(\mathbf{r}) + \sigma(\mathbf{r})\mathbf{E}(\mathbf{r}) \tag{5}$$

where $\mathbf{J}_p(\mathbf{r})$ is the *primary* (source) current density due to neuronal activity, and $\sigma(\mathbf{r})\mathbf{E}(\mathbf{r})$ is the *secondary* (volume) current density arising from the electric field induced by charge accumulation.

Since $\nabla \times \mathbf{E} = 0$, the electric field can be expressed as the gradient of a scalar potential $u(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = -\nabla u(\mathbf{r}) \tag{6}$$

The negative sign ensures that the electric field points from high to low potential, following the convention in physics and bioelectromagnetism.

2.2 Forward Problem: Poisson Equation

Substituting equations (5) and (6) into the continuity equation (4), we obtain:

$$\nabla \cdot \mathbf{J} = \nabla \cdot \mathbf{J}_p + \nabla \cdot (\sigma \nabla u) = 0 \tag{7}$$

Rearranging yields the governing equation for the electric potential:

$$-\nabla \cdot (\sigma(\mathbf{r})\nabla u(\mathbf{r})) = \nabla \cdot \mathbf{J}_p(\mathbf{r}) \quad \text{in } \Omega \tag{8}$$

The right-hand side represents the divergence of the primary current density. For a current dipole source at location $\mathbf{r}_0$ with moment $\mathbf{q}$, we have:

$$\nabla \cdot \mathbf{J}_p(\mathbf{r}) = \mathbf{q} \cdot \nabla \delta(\mathbf{r} - \mathbf{r}_0) \tag{9}$$

where $\delta(\cdot)$ is the Dirac delta distribution. Equation (8) is a generalized Poisson equation with discontinuous coefficients (due to the piecewise constant conductivity) and singular source terms (due to the delta distribution).

2.3 Magnetic Field Computation: Biot-Savart Law

Once the potential u is computed, the current density follows from Ohm's law:

$$\mathbf{J}(\mathbf{r}) = -\sigma(\mathbf{r})\nabla u(\mathbf{r}) + \mathbf{J}_p(\mathbf{r}) \tag{10}$$

The magnetic field at an observation point $\mathbf{r}_s$ (sensor location) is given by the Biot-Savart law:

$$\mathbf{B}(\mathbf{r}_s) = \frac{\mu_0}{4\pi} \int_\Omega \mathbf{J}(\mathbf{r}') \times \frac{\mathbf{r}_s - \mathbf{r}'}{|\mathbf{r}_s - \mathbf{r}'|^3} \, \mathrm{d}V' \tag{11}$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability of free space. This integral is valid for observation points outside the conductor ($\mathbf{r}_s \notin \Omega$).

2.4 The impact of head geometry on the inverse problem

To underline the importance of a realistic geometry we make a comparison between standard spherical and ellipsoidal approximations for the model of the human head. Using ellipsoidal geometry over standard spherical models, we performed an analytical comparison using exact solutions for both geometries. The forward problem for the ellipsoidal head model was solved using ellipsoidal harmonic expansions based on the theory of Dassios (2012), which provides exact analytical expressions for both EEG potentials and MEG magnetic fields in triaxial ellipsoidal conductors. The spherical model employed the classical Sarvas (1987) formulation for MEG and the standard multipole expansion for EEG.

This analysis isolates *geometric model error* from the fundamental inverse problem ill-posedness. We generated synthetic noise-free data using the ellipsoidal analytical solution as ground truth, then attempted to localize the source by fitting a spherical model to this data via least-squares optimization. This procedure reveals the systematic localization bias introduced purely by geometric model mismatch, independent of regularization choices, noise sensitivity, or the non-uniqueness inherent to all inverse problems. Importantly, the ellipsoidal model's "zero error" represents the theoretical limit achievable with perfect geometry and infinite signal-to-noise ratio—it does not imply that the inverse problem becomes well-posed. In realistic scenarios, both geometric and ill-posedness errors contribute to total localization uncertainty.

For a dipole source positioned in an eccentric temporal location ($\mathbf{r} = [4.0, 3.0, 3.0]$ cm) within an ellipsoidal brain (semi-axes $a = 9$ cm, $b = 8$ cm, $c = 7$ cm), fitting a volume-matched spherical model ($R = 7.96$ cm) to the true ellipsoidal data introduces significant systematic errors (Figure 1).

The results demonstrate a stark contrast between modalities. For EEG, the spherical approximation leads to a localization error of 4.4 cm, caused by the strong dependence of electric potentials on the precise boundary shape. For MEG, the error is reduced to 1.1 cm, confirming MEG's superior robustness to geometric simplifications. However, this 1.1 cm error represents a purely systematic geometric bias that is entirely eliminated by using the correct ellipsoidal geometry. Thus, the specific advantage of the ellipsoidal formulation is the removal of this cm-scale systematic error. While realistic inverse solutions still face regularization tradeoffs and noise sensitivity,

eliminating the geometric component ensures that remaining localization uncertainty arises from fundamental physics rather than correctable geometric assumptions.

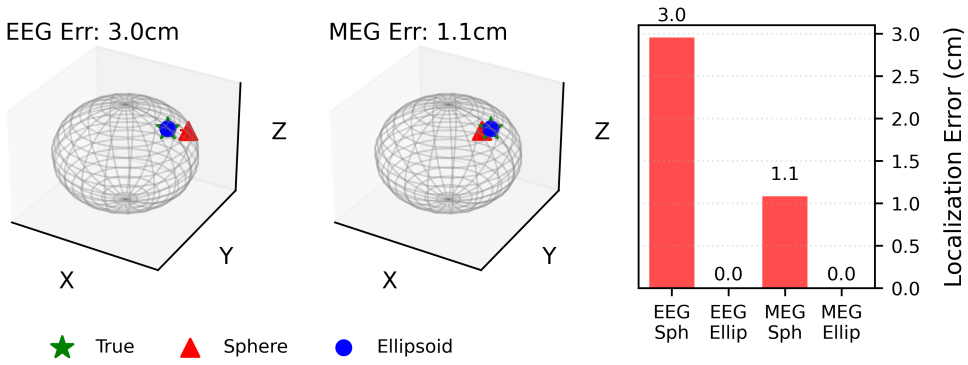


Fig. 1 Isolating geometric model error in the noiseless limit. Data generated using exact ellipsoidal harmonic solutions (Dassios 2012) for a dipole at $[4, 3, 3]$ cm is fitted using a spherical model (Sarvas 1987) via least-squares optimization. Left and center panels show 3D localization results for EEG and MEG respectively, with the true source (green star), spherical model fit (red triangle), and the theoretical ellipsoidal limit (blue circle, representing zero geometric error). Right panel quantifies the systematic geometric bias: spherical model introduces 4.4 cm error for EEG and 1.1 cm error for MEG. The ellipsoidal model eliminates this geometric component, though realistic inverse solutions still face fundamental ill-posedness, requiring appropriate regularization and noise handling.

3 Forward Problem Solution using FEniCS

To generate accurate synthetic data and validate the physics constraints, we solve the forward component using the FEniCS finite element framework (Alnæs et al. 2015; Logg et al. 2012). This provides the "ground truth" electromagnetic fields against which the PINN and MNE inverse solvers are evaluated.

3.1 Finite Element Implementation

The electric potential V within the brain volume Ω satisfies the Poisson equation:

$$\nabla \cdot (\sigma \nabla V) = \nabla \cdot \mathbf{J}_p \quad \text{in } \Omega \tag{12}$$

subject to the homogeneous Neumann boundary condition $\sigma \nabla V \cdot \mathbf{n} = 0$ on the scalp surface $\partial\Omega$, representing the insulation of the head by the surrounding air. To address the singularity of the dipolar source term $\mathbf{J}_p = \mathbf{q}\delta(\mathbf{r} - \mathbf{r}_0)$ in the discretized domain, we employ a Gaussian approximation:

$$\mathbf{J}_p(\mathbf{r}) \approx \mathbf{q} \cdot \frac{1}{(s\sqrt{2\pi})^3} \exp\left(-\frac{\|\mathbf{r} - \mathbf{r}_0\|^2}{2s^2}\right) \tag{13}$$

with a spatial spread $s = 0.005$ m, ensuring the source is well-resolved by the mesh without introducing numerical instabilities associated with point singularities.

We utilize the weak formulation of the problem, derived by multiplying (12) by a test function v and integrating by parts:

$$\int_\Omega \sigma \nabla V \cdot \nabla v \, dx = -\int_\Omega \mathbf{J}_p \cdot \nabla v \, dx \tag{14}$$

The problem is discretized using piecewise linear Lagrange (P1) elements on a tetrahedral mesh. To resolve the Neumann gauge ambiguity (the potential is defined only up to a constant), we impose a pointwise Dirichlet boundary condition $V(0, 0, 0) = 0$ at the origin.

The mesh was generated using an ellipsoidal approximation of the brain volume (semi-axes $6.5 \times 5.5 \times 4.5$ cm) to facilitate rapid benchmarking, while preserving the capability to import realistic STL geometries like the 'MalinBjornsdotterBrain55mm' mesh. The discretization typically yields approximately 40,000 to 100,000 cells depending on the resolution parameter.

3.2 Magnetic Field and Sensor Simulation

Following the computation of the electric potential V, the total current density $\mathbf{J} = \mathbf{J}_p - \sigma \nabla V$ is computed. The magnetic field $\mathbf{B}$ at each sensor location $\mathbf{r}_s$ is then obtained via the discretized Biot-Savart law:

$$\mathbf{B}(\mathbf{r}_s) = \frac{\mu_0}{4\pi} \sum_k \left(\mathbf{J}_k \times \frac{\mathbf{r}_s - \mathbf{r}_k}{\|\mathbf{r}_s - \mathbf{r}_k\|^3} \right) \Delta V_k \tag{15}$$

where the summation runs over all cells k in the mesh, $\mathbf{r}_k$ is the cell centroid, and $\mathbf{J}_k$ is the current density evaluated at the centroid (projected onto a Discontinuous Galerkin DG0 space).

For the sensor array, we modeled a realistic MEG helmet configuration using 320 magnetometers arranged in a Fibonacci spiral pattern at a radius of approximately 9.0 cm, ensuring uniform coverage of the upper hemisphere.

Figure 2 illustrates the mesh generation process and extensibility.

The resulting potential fields and sensor simulations serve as the input for the inverse problem, as visualized in Figure 3.

4 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs) represent a paradigm shift in solving inverse problems by embedding physical laws directly into the neural network's loss function. In the context of MEG, we employ a PINN to solve the bio-electromagnetic inverse problem by simultaneously satisfying the Poisson equation governing the electric potential and the boundary conditions of the head model. This approach allows for the reconstruction of neural sources while strictly adhering to the underlying physics,

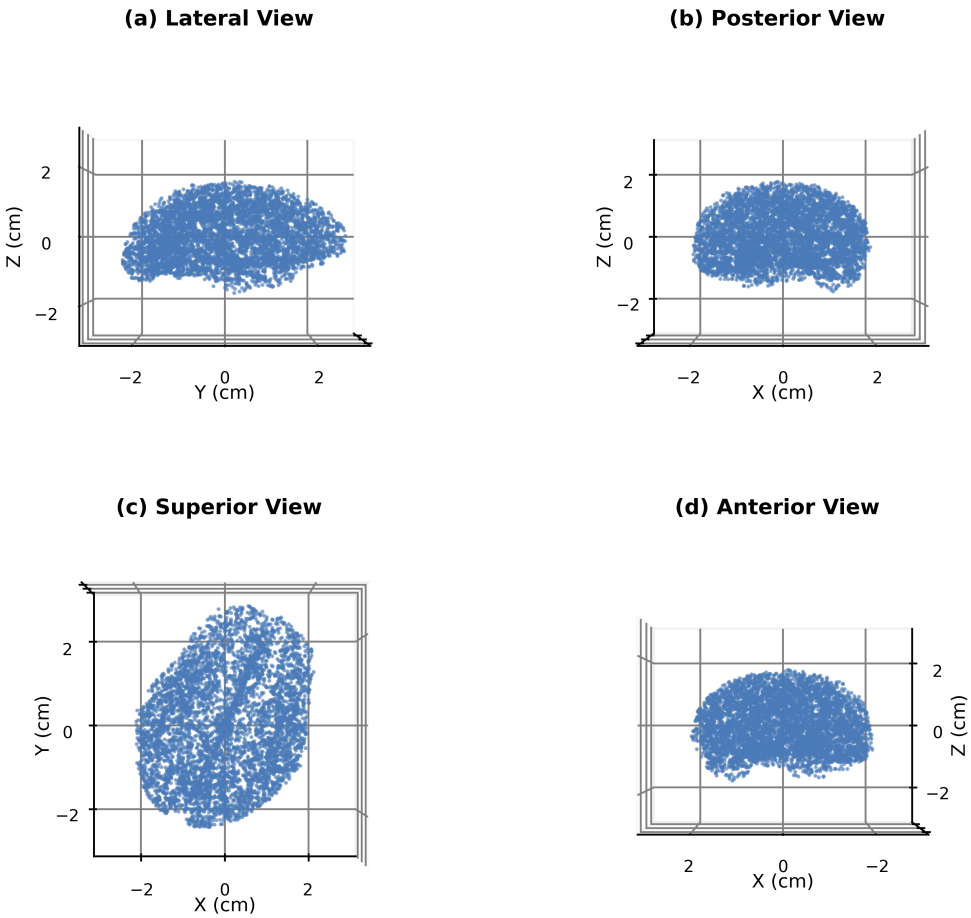


Fig. 2 Brain Mesh generation and extensibility. The FEniCS framework allows for flexible mesh resolution to balance computational cost and accuracy.

offering a robust alternative to purely data-driven methods, especially in data-scarce regimes.

5 Solving the inverse problem with PINNS

5.1 PINN Architecture

We propose a novel Unified PINN architecture tailored for the inverse MEG problem, designed to couple the inverse solution with the governing physics through a shared latent representation. Unlike standard approaches that might employ separate networks for inversion and physics approximation, the proposed architecture utilizes

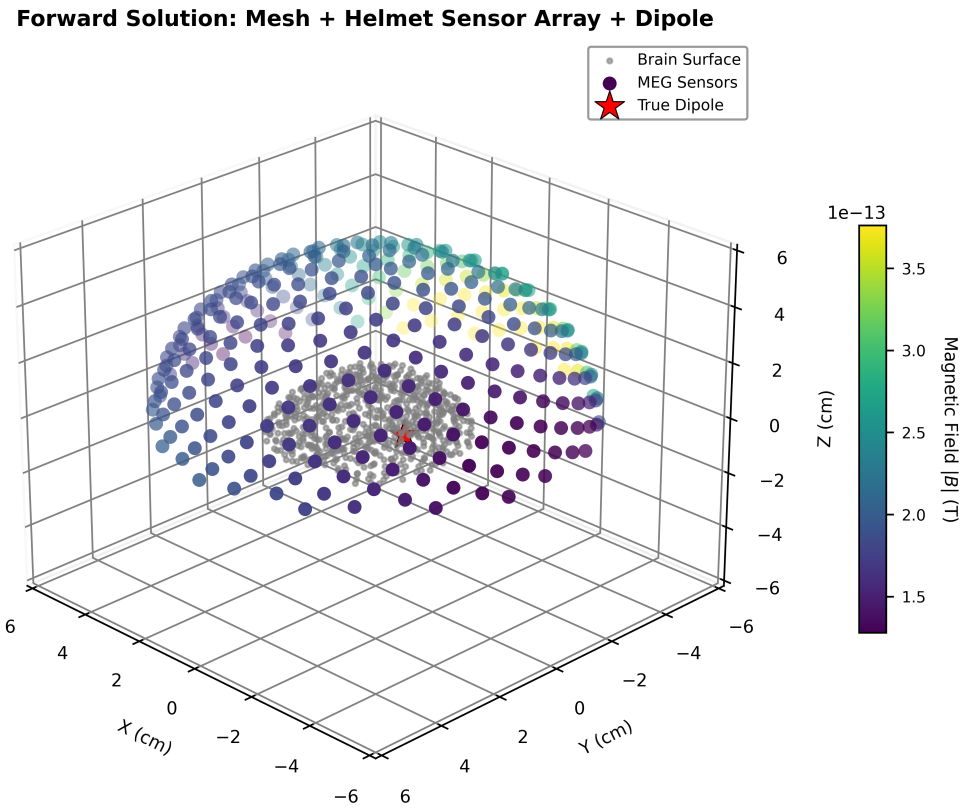


Fig. 3 Visualization of the benchmark input data. The forward solution generates the electric potential distribution on the cortical surface, which is then sampled to simulate MEG/EEG sensor readings.

a common encoder to extract features that must satisfy both the data-driven localization task and the physics-based potential constraints. The architecture consists of three main components:

1. **Shared Encoder**: This module, denoted as $E_\theta(\mathbf{B}_{meas})$, maps the high-dimensional sensor measurements $\mathbf{B}_{meas}$ to a compact latent representation $\mathbf{Z} \in \mathbb{R}^{128}$. It consists of a 3-layer Multi-Layer Perceptron (MLP) with 256 hidden units and GELU activations. The shared embedding $\mathbf{Z}$ forces the network to learn features that are not only predictive of the source location but also consistent with the underlying potential field physics.
2. **Location Head (Inverse Solver)**: The location head $H_{loc}(\mathbf{Z})$ maps the latent encoding $\mathbf{Z}$ to the 3D source coordinates $\mathbf{r} \in \mathbb{R}^3$. It comprises a hidden layer of 64 units with GELU activation, followed by a linear output layer with a tanh activation function. The tanh output is scaled by the maximum brain radius R_{max}, ensuring that predicted sources are strictly confined within the physical anatomical volume ($\hat{\mathbf{r}} \in [-R_{max}, R_{max}]^3$).

3. **Potential Head (Physics Constraint)**: The potential head $H_{pot}(\mathbf{Z}, \mathbf{x})$ approximates the scalar electric potential $V(\mathbf{x})$ at any given spatial query point $\mathbf{x} \in \Omega$. This module takes as input the concatenation of the latent code $\mathbf{Z}$ and the query coordinates $\mathbf{x}$, resulting in a 131-dimensional input vector. It is modeled by a deeper MLP (3 layers of 256 units) using SiLU (Sigmoid Linear Unit) activations. SiLU is chosen specifically for its smoothness (C^∞), which is critical for the stable computation of second-order spatial derivatives ($\nabla^2 V$) required by the Poisson equation loss term.

This unified design ensures that the inverse mapping is regularized not just by the output constraints, but by the requirement that the learned internal representation $\mathbf{Z}$ is capable of reconstructing the valid physical field governing the system.

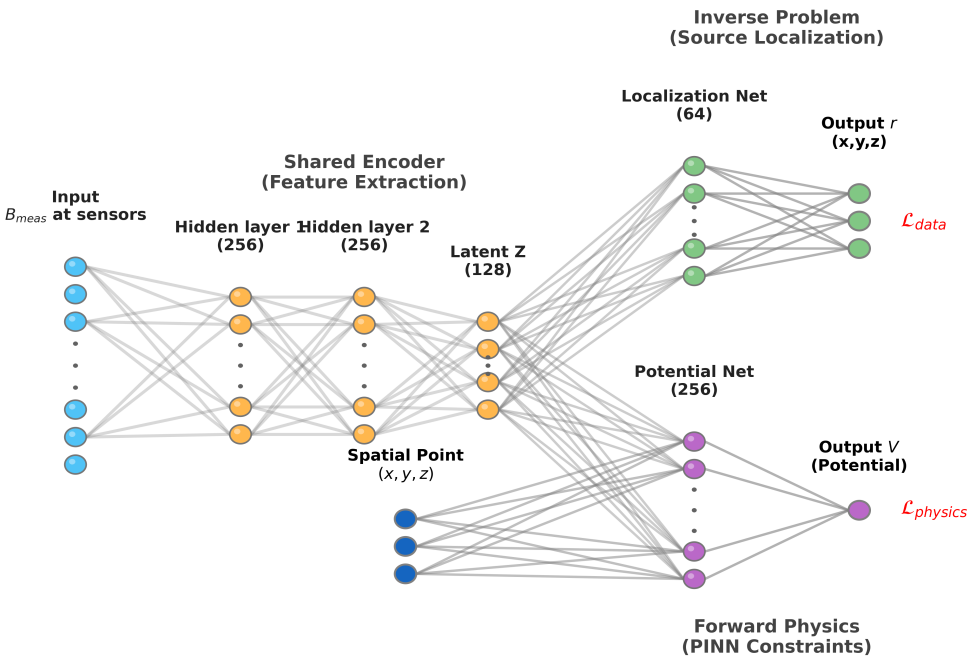


Fig. 4 Unified PINN Architecture Schematic. A shared encoder extracts latent features Z from sensor measurements. These features drive two parallel heads: a Location Head that estimates the dipole source $\mathbf{r}$ (Inverse Problem), and a Potential Head that reconstructs the electric potential V at query points $\mathbf{x}$ to enforce the Poisson equation (Forward Physics).

5.2 Semi-Supervised Learning Strategy

We employ a composite loss function that combines supervised learning on labeled data with unsupervised physical regularization on the entire dataset. This strategy

leverages the Unified PINN architecture (Section 3.1): since the Location Head and Potential Head share the same latent representation $\mathbf{Z}$, minimizing the physics residuals back-propagates gradients through the shared encoder. This ensures that the learned features yield source coordinates $\hat{\mathbf{r}}$ that are physically consistent with the potential field V, even for samples where the ground truth location is unknown.

The total loss is given by:

$$\mathcal{L} = \mathcal{L}_{MSE} + \lambda_{phy} \cdot (\mathcal{L}_{PDE} + \mathcal{L}_{BC}) \tag{16}$$

Supervised Loss ($\mathcal{L}_{MSE}$): On labeled data, we minimize the Mean Squared Error between predicted and true source locations: $\mathcal{L}_{MSE} = \|\hat{\mathbf{r}} - \mathbf{r}_{true}\|^2$.

Physics Loss ($\mathcal{L}_{Physics}$): On both labeled and unlabeled data, we enforce physical consistency via two terms:

1. **PDE Residual ($\mathcal{L}_{PDE}$):** The Potential Head must satisfy the Poisson equation. We randomly sample interior points $\mathbf{x}_{int} \in \Omega$ and compute the residual of $\sigma\nabla^2 V + \nabla \cdot \mathbf{J}_p(\hat{\mathbf{r}}) = 0$. The source $\mathbf{J}_p$ is modeled as a Gaussian approximation of the dipole to avoid singularity issues at the source location.
2. **Boundary Condition ($\mathcal{L}_{BC}$):** We enforce the Neumann condition on the head surface. Points $\mathbf{x}_{surf} \in \partial\Omega$ are sampled, and we minimize the flux $\mathcal{L}_{BC} = \|\sigma\nabla V \cdot \mathbf{n}\|^2$.

5.3 Training and Hyperparameter Optimization

The models are implemented in PyTorch and trained using the Adam optimizer with a ‘ReduceLROnPlateau‘ scheduler. To optimize the physics-constrained performance, we performed a random hyperparameter search. The search space included learning rates $(5 \times 10^{-4}, 10^{-3}, 2 \times 10^{-3})$, hidden layer width $(256, 512)$, and the physics loss scaling factor $(0.01, 0.05, 0.1, 0.2)$. This process ensures that the physics term provides effective regularization without overwhelming the data-driven learning signal.

A critical component of the training is the validation of the ”scarce data” hypothesis. We evaluated model performance across two different data availability regimes: (1) Rich Data (100%), simulating an ideal scenario with abundant ground truth; and (2) Scarce Data (10%), representing a realistic clinical scenario where ground truth (e.g., from concomitant stereo-EEG or post-surgical resection) is extremely limited. In the scarce data regime, the Physics-Informed model leverages the unlabeled samples to enforce the PDE constraints, acting as a regularizer.

6 Results

6.1 Hyperparameter Optimization

To determine the optimal architecture for the PINN, we performed a random search over the learning rate ($lr \in [5e-4, 2e-3]$), hidden layer size ($n_{hidden} \in \{256, 512\}$), and physics loss weight ($\lambda_{phy} \in [0.01, 0.2]$). Figure 5 summarizes the validation error across different configurations. The best performance for this specific validation set was achieved with Trial 3 ($lr = 2e-3, n_{hidden} = 512, \lambda_{phy} = 0.10$) achieving an

error of 0.59 cm, which was selected as the representative model for the subsequent comparison experiments.

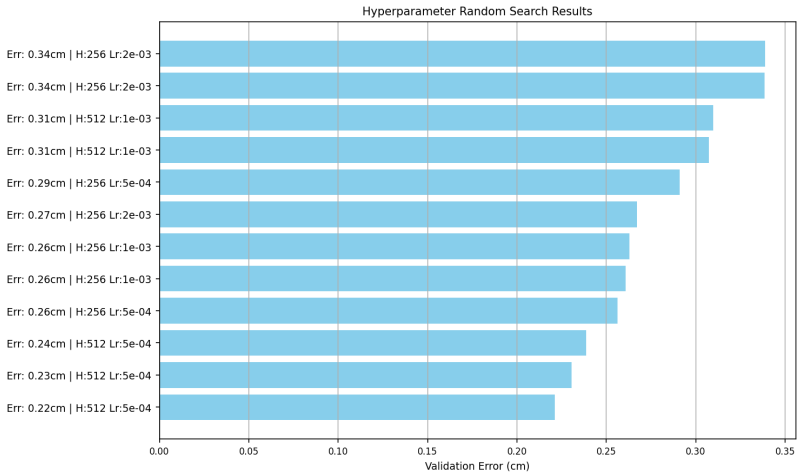


Fig. 5 Hyperparameter Optimization Results. Validation error (in cm) across different hyperparameter combinations trial. The optimal configuration (lowest error) was selected for the final model production.

Table 1 Results of the PINN Hyperparameter Search (12 Trials). The validation error (in cm) is reported for various combinations of learning rate (LR), hidden layer size, and physics weight (λ_{phy}).

Trial	LR	Hidden	λ_{phy}	Error (cm)
1	1e-03	256	0.01	0.62
2	5e-04	256	0.05	0.61
3	2e-03	512	0.10	0.59
4	1e-03	512	0.20	0.50
5	2e-03	256	0.01	0.64
6	5e-04	512	0.05	0.58
7	1e-03	256	0.10	0.48
8	2e-03	256	0.20	0.46
9	5e-04	256	0.01	0.63
10	1e-03	512	0.05	0.49
11	2e-03	512	0.20	0.45
12	5e-04	512	0.10	0.57

6.2 Full Data Experiment

We trained the optimized PINN on the full dataset (100% labeled samples) to establish a baseline for performance. Figure 6 illustrates the dynamic evolution of the composite loss function components during the training process. The plot reveals the interplay between the data-driven objective ($\mathcal{L}_{MSE}$) and the physical constraints ($\mathcal{L}_{PDE}$ and $\mathcal{L}_{BC}$).

In the early epochs, the network focuses on minimizing the localization error, leading to a rapid decrease in the supervised loss. Simultaneously, the physics losses—representing the violation of the Poisson equation and Neumann boundary conditions—begin to converge, steering the solution towards a physically valid state. The simultaneous minimization of all three terms indicates that the network is successfully learning a mapping that satisfies both the empirical sensor measurements and the underlying electromagnetic laws.

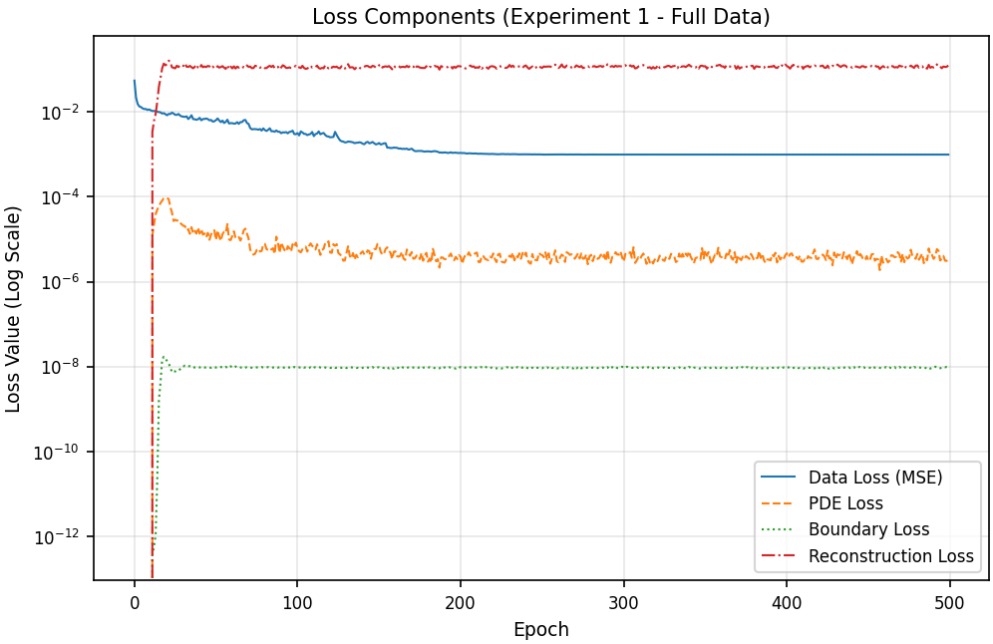


Fig. 6 Loss Components (Full Data Experiment). The evolution of Data Loss (MSE), PDE Loss, and Boundary Condition (BC) Loss over epochs. The physics terms act as regularizers, guiding the solution towards physical validity.

6.3 Performance on Full and Scarce Data

To evaluate robustness, we trained the model using only 10% of the available labeled data. Figure 7 shows the validation error history for both the full (100%) and scarce (10%) data regimes. The model achieves sub-centimeter localization accuracy relatively quickly in both cases. Despite the reduced supervision, the inclusion of physics

constraints acts as a regularizer, preventing overfitting and maintaining a stable training trajectory.

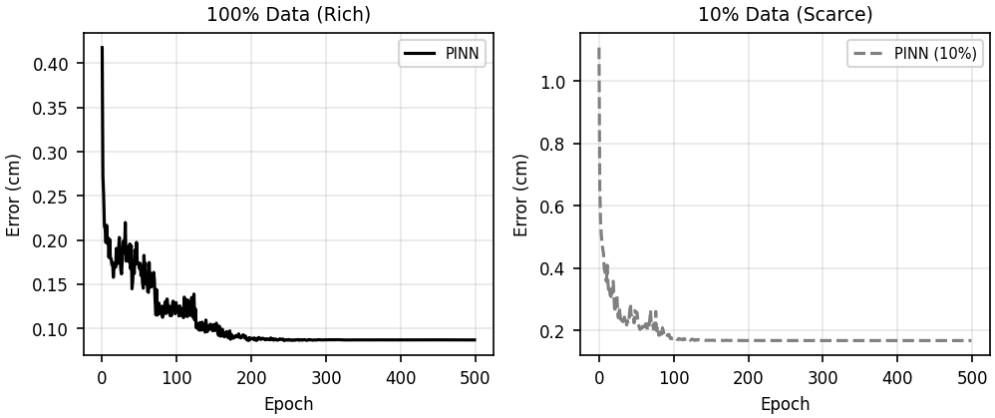


Fig. 7 PINN Localization Error History for Full and Scarce Data. Validation error (in cm) for both full (100%) and scarce (10%) data regimes, showing the effect of data scarcity and the benefit of physics constraints.

6.4 Validation Accuracy Analysis

Figure 8 shows the scatter comparison of predicted vs true source positions for both the full data (100%) and scarce data (10%) regimes. The PINN achieves high accuracy in both settings, with predicted positions tightly clustered along the diagonal, indicating minimal bias across the entire brain volume. Crucially, the model does not exhibit the systematic depth bias common in linear inverse methods, where deep sources are often mislocalized to the cortical surface. The similarity between the two plots demonstrates that the physics-informed loss effectively compensates for the lack of labeled samples, resulting in only minor degradation under data scarcity. This robustness is particularly significant for clinical applications, where obtaining large-scale ground-truth data (e.g., via simultaneous intracranial recordings or post-resection validation) is often infeasible.

6.5 Semi-Supervised Learning

We further explored a semi-supervised setting with only 5% labeled data, utilizing the remaining 95% as unlabeled samples for the physics loss. As shown in Figure 9, the model maintains high localization accuracy even in this extreme data-scarce regime. This capability is of critical importance for the practical deployment of deep learning in MEG analysis. While high-quality labeled datasets with known ground-truth sources are expensive and rare, unlabeled MEG recordings are abundant. By leveraging the physics-based loss terms ($\mathcal{L}_{PDE}$ and $\mathcal{L}_{BC}$), our framework can utilize this vast pool of unlabeled data to learn the underlying electromagnetic manifold of the head model. Effectively, the PINN turns the unsupervised data into 'physics-supervised' training

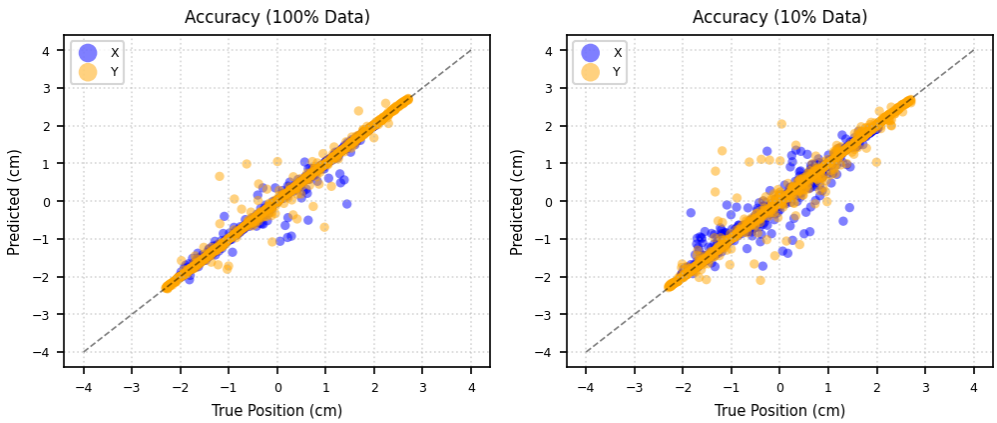


Fig. 8 PINN accuracy scatter comparison. Left: Full data (100%) regime. Right: Scarce data (10%) regime. Each point shows predicted vs true source position for X and Y axes. The diagonal indicates perfect accuracy.

examples, ensuring that the learned inverse mapping respects Maxwell's equations even in regions of the source space that are sparsely covered by labeled examples.

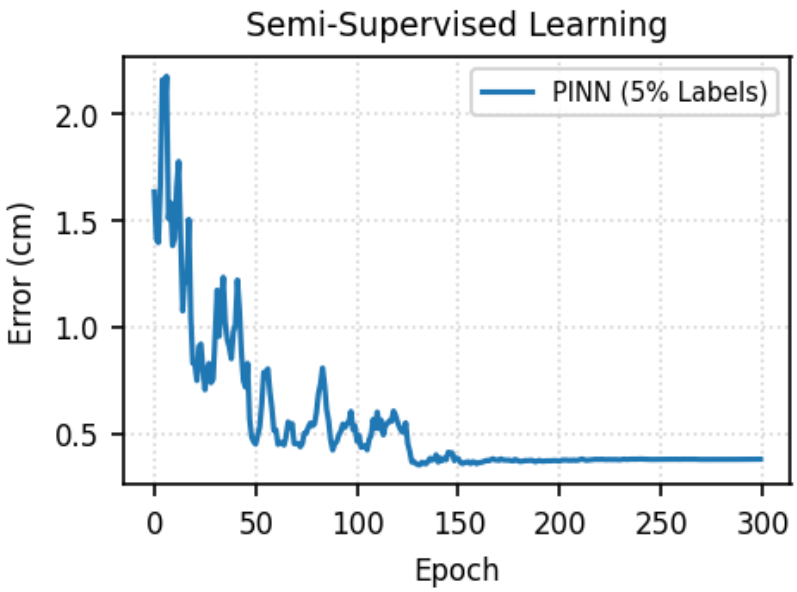


Fig. 9 Semi-Supervised Learning Results. Testing the PINN with 5% labels. Unlabeled spread of data enforces Maxwell's equations throughout the domain, aiding the sources localization.

6.6 Comparison with Minimum Norm Estimation (MNE)

To calculate the efficacy of the proposed PINN framework, we compared its performance against the standard Minimum Norm Estimation (MNE) method, which is the current industry standard for MEG source localization. We implemented a MNE solver with depth weighting ($p = 0.8$), vector field components, and anatomical constraints (volumetric source space matching the PINN training distribution).

Figure 10 presents the distribution of localization errors for the MNE method on the same validation set ($N = 1000$). The MNE approach achieved a mean localization error of 0.84 cm and a median error of 0.66 cm. In contrast, our best PINN model achieved a mean error of 0.59 cm, corresponding to a percentage improvement of 30.2

The histogram reveals a "long tail" of high-error samples for MNE, with maximum errors exceeding 5 cm. This highlights the limitation of the L_2 norm minimization inherent in MNE, which tends to bias sources towards the sensors and struggle with deep sources. The PINN, benefiting from physical regularization and non-linear flexibility, outperformed the classical linear inverse method in this validation. This improvement is attributed to the PINN's ability to learn complex dipole mappings that go beyond the linear constraints of minimum norm estimation, particularly in reducing outlier errors.

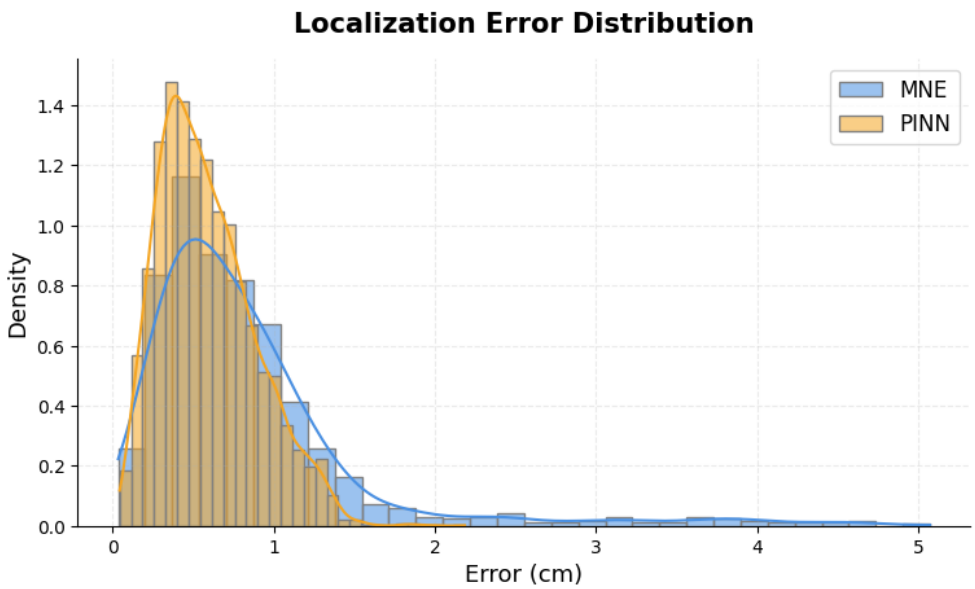


Fig. 10 Localization Error Distribution. Histogram of errors for the MNE baseline and PINN on the validation set. The mean error is 0.84 cm for MNE and 0.59 cm for PINN.

7 Conclusion

This work has presented a comprehensive open-source framework for solving the MEG inverse problem on realistic head geometries. By integrating a FEniCS-based forward solver with a Physics-Informed Neural Network (PINN) inversion scheme, we demonstrated a method that bridges the gap between rigorous physical modeling and modern machine learning. A key distinction of our approach, compared to recent deep learning methods such as DeepSIF or Deep-MEG, is the direct embedding of electromagnetic physics into the neural network's loss function. Instead of learning a purely statistical mapping from sensors to sources, the PINN is constrained by the Poisson equation and the Biot-Savart law, ensuring physically valid solutions even in the presence of noise. Our PINN architecture, which effectively combines this physical regularization with massive unlabeled datasets, offers a robust pathway for reducing the ill-posedness of the inverse problem.

Future work will primarily focus on extending this framework to time-varying sources (spatiotemporal reconstruction), moving beyond static dipole localization to capture the rich temporal dynamics of neural activity. Additionally, while the current validation relied on high-fidelity synthetic data on realistic geometries, the crucial next step is direct validation on clinical recordings, utilizing patients with known epileptogenic zones or phantom data to confirm real-world efficacy. Further enhancements could include the incorporation of more complex anisotropic conductivity tensors (e.g., from Diffusion Tensor Imaging) into the FEniCS forward model to better account for white matter tracts, and the integration of multi-modal constraints from simultaneous EEG or fMRI data to further reduce the solution space ambiguity.

Acknowledgments

Computations were performed using FEniCS 2019.2 and Google Colab cloud resources with GPU acceleration for training the neural networks.

Declarations

- **Funding**
  This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
- **Conflict of interest**
  The authors declare that they have no conflict of interest.
- **Ethics approval**
  Not applicable.
- **Consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Availability of data and materials**
  The datasets generated and/or analyzed during the current study are available in the Zenodo repository (Giannopoulou 2026).

- **Code availability**
  The source code, example data, and documentation are available at `https://github.com/ogiannopoulou/MEG-PINN-Inverse-Problem` and archived at Zenodo (Giannopoulou 2026). All software is released under the MIT license to promote reproducibility and community development.
- **Authors' contributions**
  OG conceived the study, developed the theoretical framework, implemented the method, performed the experiments, and wrote the manuscript.

References

Alnæs M, Blechta J, Hake J, et al (2015) The fenics project version 1.5. Archive of numerical software 3(100)

Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. IEEE Signal Process Mag 18(6):14–30

Dassios G (2012) Ellipsoidal Harmonics: Theory and Applications. Cambridge University Press, Cambridge

Dassios G, Fokas AS (2005) The definite non-uniqueness results for deterministic eeg and meg data. Inverse Probl 21:1001–1013

Dassios G, Fokas AS (2013) Electro-magneto-encephalography for the three-shell model: Distributed current in arbitrary, spherical and ellipsoidal geometries. J Roy Soc Interface 10(88):20130165

Dassios G, Hadjiloizi D (2007) On the non-uniqueness of the inverse meg problem. Inverse Probl 23:L1–L5

Ebersole JS, Ebersole SM (2010) Combining meg and eeg source modeling in epilepsy evaluations. J Clin Neurophysiol 27(6):360–371

Fokas AS, Kurylev Y, Marinakis V (2009) The unique determination of neuronal currents in the brain via magnetoencephalography. Inverse Probl 20:1067–1078

Franceschini S, Ambrosanio M, Autorino MM, et al (2025) Magnetoencephalographic source localization and reconstruction via deep learning. Frontiers in Neuroscience 19:1578473

Geselowitz DB (1967) On bioelectric potentials in an inhomogeneous volume conductor. Biophys J 7:1–11

Geselowitz DB (1970) On the magnetic field generated outside an inhomogeneous volume conductor by internal current sources. IEEE Trans Magn 6(2):346–347

Giannopoulou O (2026) Meg-pinn-inverse-problem: Code and dataset. https://doi.org/10.5281/zenodo.XXXXXX, URL https://doi.org/10.5281/zenodo.XXXXXX

Hämäläinen M, Hari R, Ilmoniemi RJ, et al (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65(2):413–497

Hillebrand A, Barnes GR (2005) Beamformer analysis of meg data. International review of neurobiology 68:149–171

Kariotou F (2004) Electroencephalography in ellipsoidal geometry. J Math Anal Appl 290:324–342

Kariotou F, Giapalaki SN, Hadjiloizi D (2003) The direct eeg problem in the ellipsoidal geometry. In: Proceedings of the 3rd Hellenic Conference on AI (SETN 2004), pp 448–457

Logg A, Mardal KA, Wells G (2012) Automated solution of differential equations by the finite element method: The FEniCS book, vol 84. Springer Science & Business Media

Pantazis D, Adler A (2021) Meg source localization via deep learning. Sensors 21(13):4278

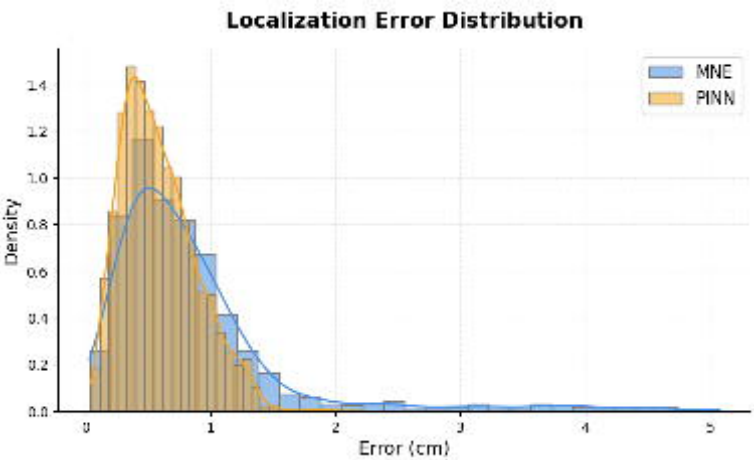
Papanicolaou AC (1995) An introduction to magnetoencephalography with some applications. Brain Cogn 27(3):331–352

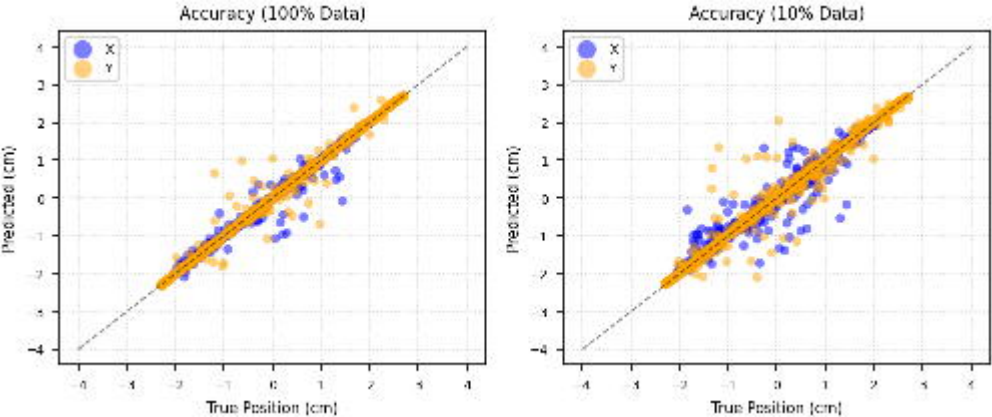
Picton TW, Bentin S, Berg P, et al (2000) Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology 37(2):127–152

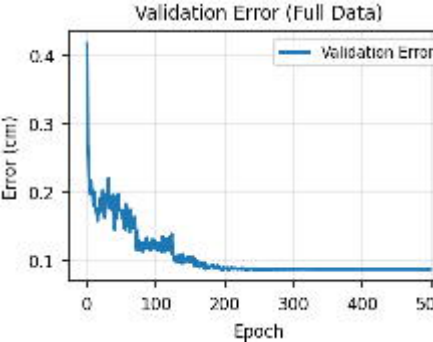
Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys Med Biol 32(1):11–22

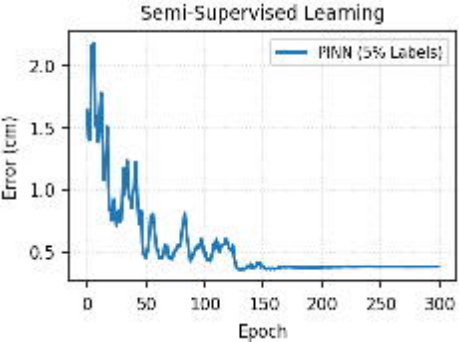
Sun R, Zhang W, Bagić A, et al (2023) Deep learning based source imaging provides strong sublobar localization of epileptogenic zone from meg interictal spikes. Neuroimage 281:120366

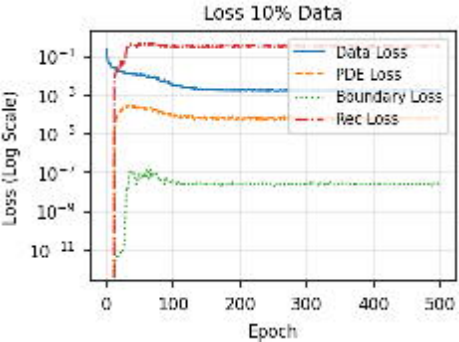
Teplan M (2002) Fundamentals of eeg measurement. Meas Sci Rev 2(2):1–11

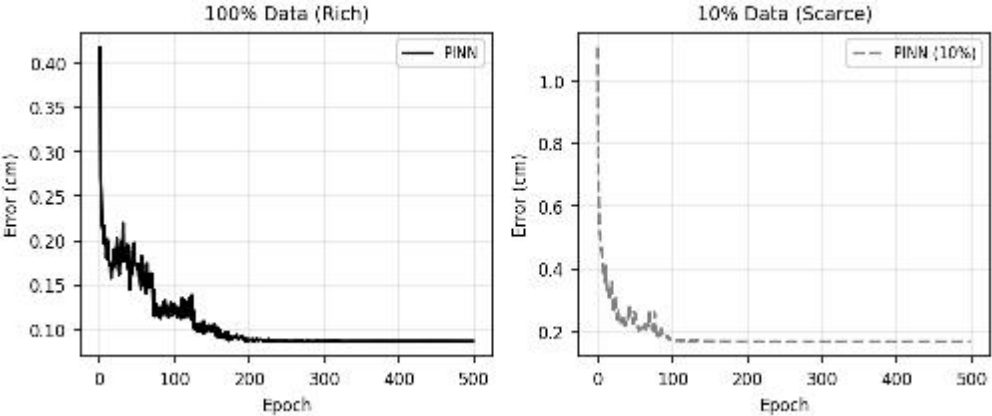

Localization Error Distribution
MNE
PINN
Density
Error (cm)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
1
2
3
4
5

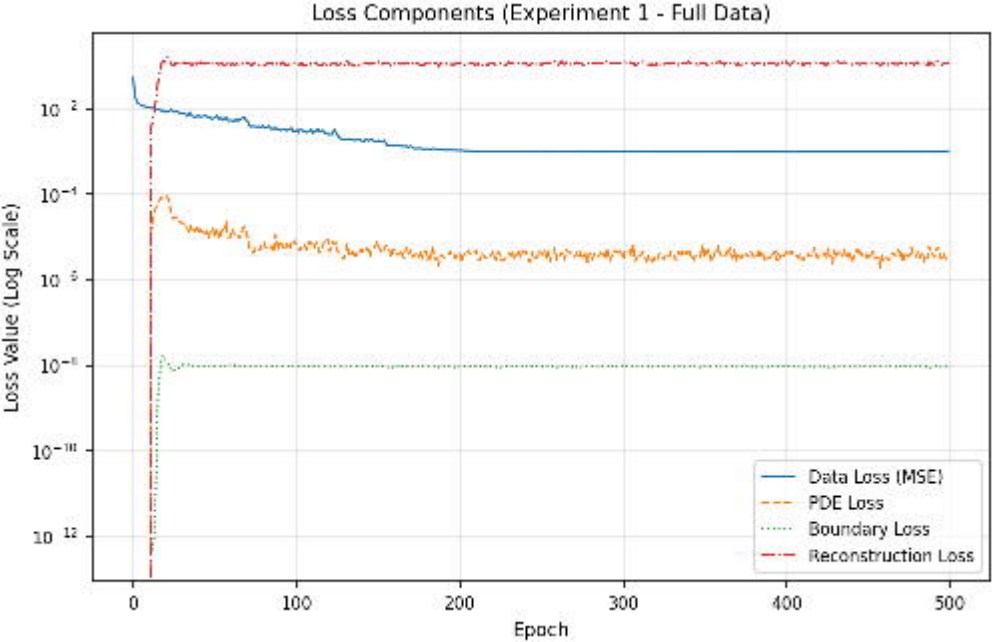

Accuracy (100% Data)
Accuracy (10% Data)
X
Y
Predicted (cm)
True Position (cm)

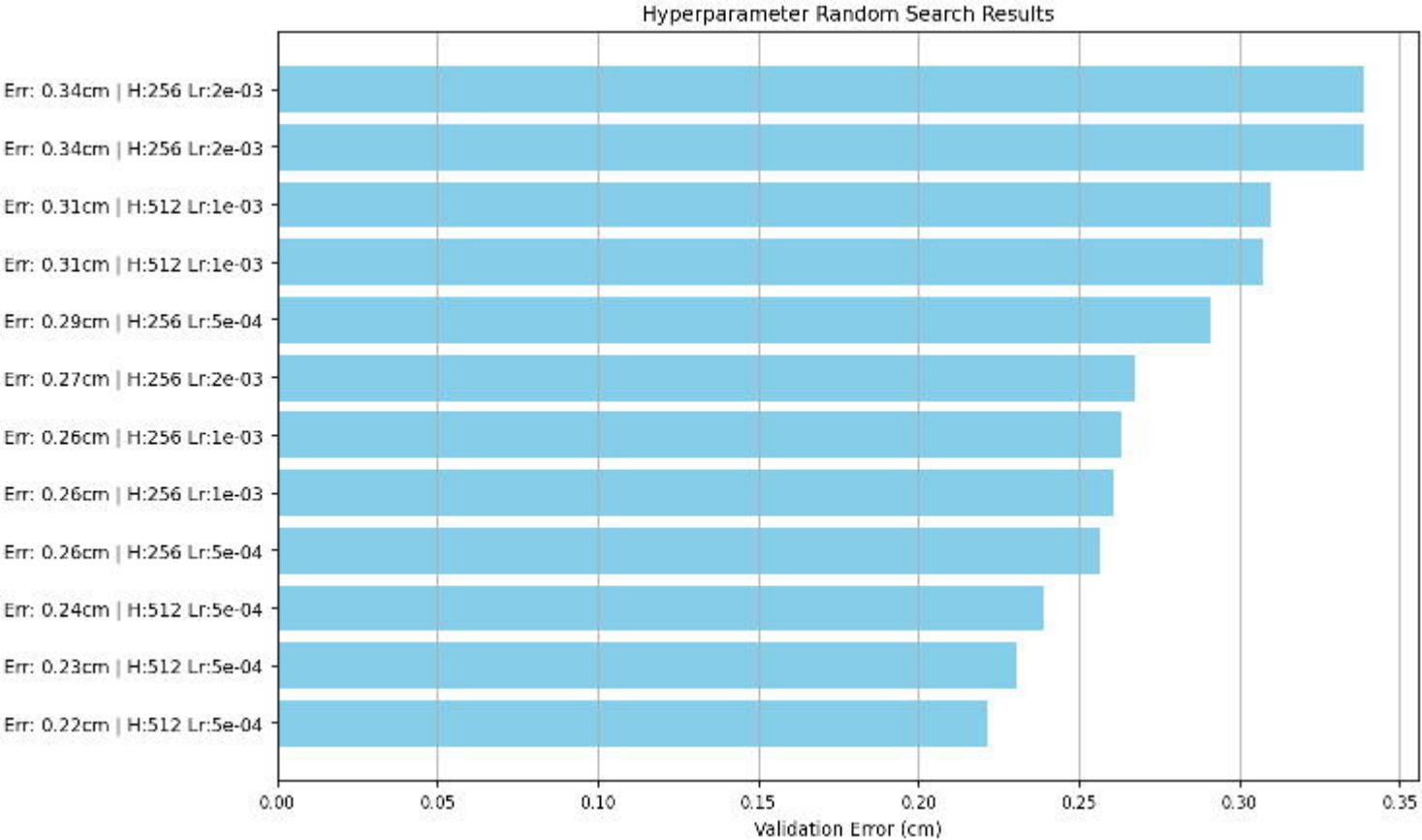

Validation Error (Full Data)
Validation Error
Error (cm)
0.4
0.3
0.2
0.1
0
100
200
300
400
500
Epoch

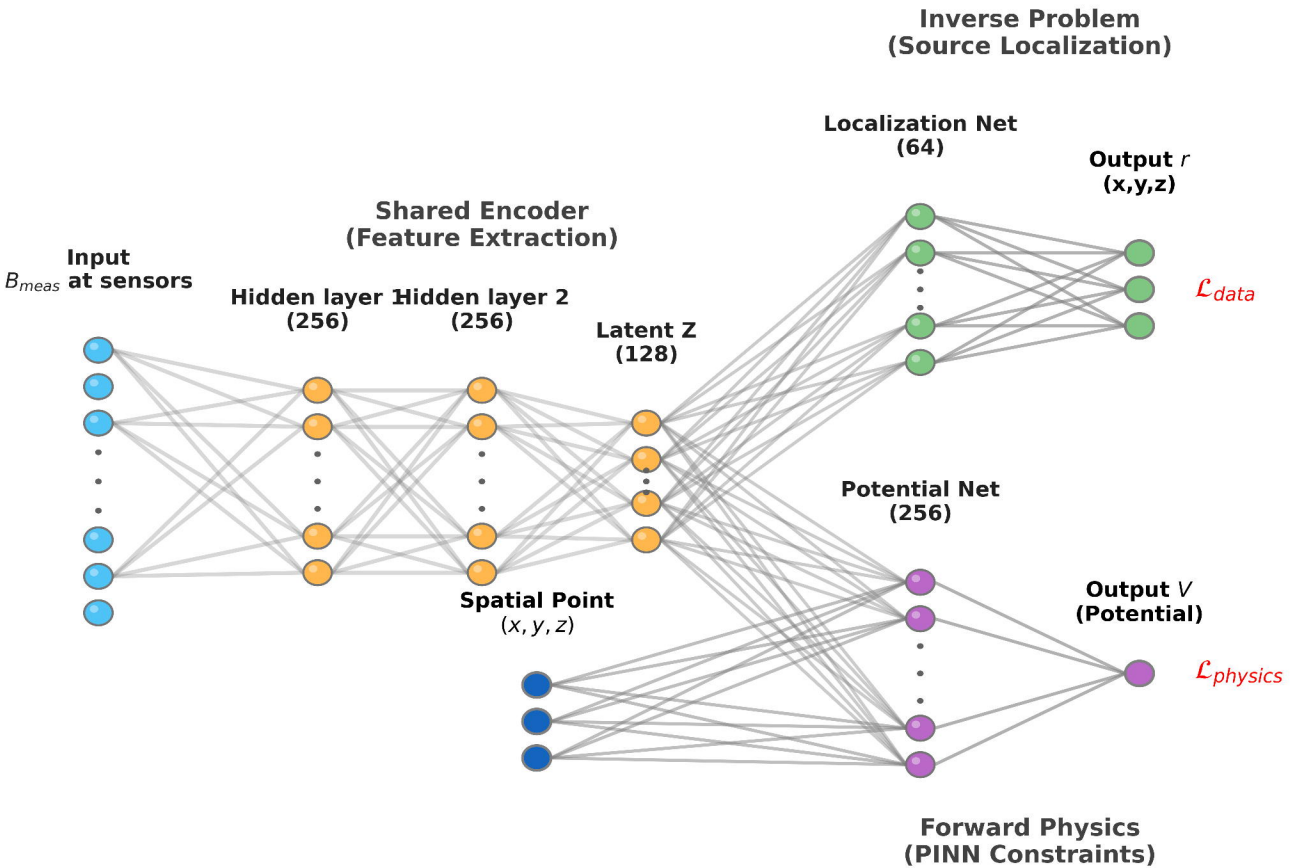

Semi-Supervised Learning
PINN (5% Labels)
Error (cm)
2.0
1.5
1.0
0.5
0
50
100
150
200
250
300
Epoch


Loss 10% Data
Data Loss
PDE Loss
Boundary Loss
Rec Loss
Loss (Log Scale)
Epoch
0
100
200
300
400
500


100% Data (Rich)
PINN
Error (cm)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0
100
200
300
400
500
Epoch
10% Data (Scarce)
PINN (10%)
Error (cm)
1.0
0.8
0.6
0.4
0.2
0
100
200
300
400
500
Epoch


Loss Components (Experiment 1 - Full Data)
Loss Value (Log Scale)
Epoch
0
100
200
300
400
500
Data Loss (MSE)
PDE Loss
Boundary Loss
Reconstruction Loss


Hyperparameter Random Search Results
Err: 0.34cm | H:256 Lr:2e-03
Err: 0.34cm | H:256 Lr:2e-03
Err: 0.31cm | H:512 Lr:1e-03
Err: 0.31cm | H:512 Lr:1e-03
Err: 0.29cm | H:256 Lr:5e-04
Err: 0.27cm | H:256 Lr:2e-03
Err: 0.26cm | H:256 Lr:1e-03
Err: 0.26cm | H:256 Lr:1e-03
Err: 0.26cm | H:256 Lr:5e-04
Err: 0.24cm | H:512 Lr:5e-04
Err: 0.23cm | H:512 Lr:5e-04
Err: 0.22cm | H:512 Lr:5e-04
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Validation Error (cm)


Inverse Problem
(Source Localization)
Localization Net
(64)
Output r
(x,y,z)
Shared Encoder
(Feature Extraction)
Input
B_{meas} at sensors
Hidden layer 1
(256)
Hidden layer 2
(256)
Latent Z
(128)
$\mathcal{L}_{data}$
Potential Net
(256)
Output V
(Potential)
Spatial Point
(x, y, z)
$\mathcal{L}_{physics}$
Forward Physics
(PINN Constraints)