The Relativistic Scalar-Vector Plenum (RSVP) Framework: A Field-Theoretic Approach to Interpretability and Multimodal Integration in Large Language Models

Flyxion

June 2025

Abstract

The Relativistic Scalar-Vector Plenum (RSVP) framework proposes a field-theoretic model for semantic cognition, representing meaning, dynamics, and uncertainty as coupled scalar (Φ), vector ($\vec{v}$), and entropy (S) fields evolving on derived geometric manifolds. This paper formalizes RSVP using partial differential equations (PDEs), gauge invariance, and higher category theory, with a focus on enhancing interpretability and multimodal integration in Large Language Models (LLMs). We establish precise mappings between RSVP fields and LLM components: Φ to token embeddings, $\vec{v}$ to attention flows, and S to uncertainty and stability metrics. Through category-theoretic proofs, we demonstrate RSVP's capacity to unify text, image, and audio modalities via derived stack morphisms and ensure interpretability through entropy gradients and cohomological invariants. Applications include field-inspired transformer architectures, entropic regularization for robust fine-tuning, and intrinsic ethical alignment, offering a mathematically rigorous foundation for next-generation AI systems.

1 Introduction

Large Language Models (LLMs) have achieved remarkable success in natural language processing, yet their internal representations and reasoning processes remain opaque, and integrating multimodal data (e.g., text, images, audio) poses significant challenges [?]. The Relativistic Scalar-Vector Plenum (RSVP) framework offers a physics-inspired, mathematically rigorous model for semantic cognition, representing meaning as coupled scalar (Φ), vector ($\vec{v}$), and entropy (S) fields on continuous manifolds or derived stacks. By grounding cognition in thermodynamic principles, gauge invariance, and category theory, RSVP addresses two critical LLM challenges: interpretability (understanding internal model states) and multimodal integration (unifying diverse data modalities).

This paper formalizes RSVP's mathematical structure, establishes its connections to LLMs, and proves its properties using category-theoretic tools. We focus on:

- Interpretability: Modeling attention and reasoning as field dynamics, with entropy gradients as diagnostic metrics.
- Multimodal Integration: Unifying modalities via derived stack correspondences.

The paper is organized as follows: Section 2 defines RSVP's fields and PDEs; Section 3 maps RSVP to LLMs; Section 4 provides category-theoretic proofs; Section 5 and 6 detail applications to interpretability and multimodal integration; and Section 7 discusses future directions.

2 RSVP Framework: Mathematical Foundations

2.1 Field Definitions

Let $\mathscr{M}$ be a smooth n-dimensional manifold representing the latent semantic space (e.g., embedding space). On $\mathscr{M}$, we define:

- Scalar Field $\Phi : \mathscr{M} \times \mathbb{R} \to \mathbb{R}$, representing semantic potential or "meaning intensity" at point $x \in \mathscr{M}$ and time t.
- Vector Field $\vec{v} : \mathscr{M} \times \mathbb{R} \to T\mathscr{M}$, encoding directional semantic flow (e.g., attention or inference directionality).
- Entropy Field $S : \mathscr{M} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$, measuring local uncertainty or semantic disorder.

These fields evolve on $\mathscr{M}$, which may be equipped with a derived stack structure to encode recursive or hierarchical semantics [?].

2.2 Coupled PDEs

The dynamics of Φ, $\vec{v}$, and S are governed by coupled nonlinear PDEs:

- Semantic Potential Evolution:
$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi \vec{v}) = D_\Phi \Delta \Phi - \alpha S \Phi + \mathscr{F}_\Phi, \tag{1}$$
where $\nabla \cdot$ is the divergence operator, Δ is the Laplace-Beltrami operator, $D_\Phi > 0$ is a diffusion coefficient, $\alpha > 0$ couples semantic decay to entropy, and $\mathscr{F}_\Phi$ models external inputs (e.g., prompts).
- Vector Flow Dynamics:
$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} = -\nabla p + \nu \Delta \vec{v} - \beta \nabla S + \vec{\mathscr{F}}_v, \tag{2}$$
where p is a semantic pressure enforcing normalization, $\nu > 0$ is viscosity, $\beta > 0$ couples flow to entropy gradients, and $\vec{\mathscr{F}}_v$ represents external forces (e.g., attention modulation).
- Entropy Evolution:
$$\frac{\partial S}{\partial t} + \nabla \cdot (S \vec{v}) = D_S \Delta S + \gamma \|\nabla \Phi\|^2 - \delta S + \mathscr{F}_S, \tag{3}$$
where $D_S > 0$ is entropy diffusion, $\gamma > 0$ couples entropy production to semantic gradients, $\delta > 0$ models dissipation, and $\mathscr{F}_S$ represents external entropy sources.

2.3 Gauge Invariance

The fields are invariant under gauge transformations:
$$\Phi \to \Phi + \chi, \quad \vec{v} \to \vec{v} + \nabla \chi, \tag{4}$$
for a smooth scalar $\chi : \mathscr{M} \to \mathbb{R}$, preserving semantic observables. The Lagrangian is:
$$\mathscr{L} = \frac{1}{2} \left\| \frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi \vec{v}) \right\|^2 - \frac{D_\Phi}{2} \|\nabla \Phi\|^2 + \frac{\nu}{2} \|\nabla \vec{v}\|^2 - \beta S \nabla \cdot \vec{v} - \mathscr{U}(S, \Phi), \tag{5}$$
where $\mathscr{U}(S, \Phi)$ is a potential coupling entropy and semantics. The action is:
$$\mathscr{S}[\Phi, \vec{v}, S] = \int_{t_0}^{t_1} \int_{\mathscr{M}} \mathscr{L} \, d\mu \, dt, \tag{6}$$
with $d\mu$ the volume form on $\mathscr{M}$. The PDEs (1–3) arise from the Euler-Lagrange equations.

2.4 Derived Stacks

To model recursive semantics, $\mathscr{M}$ is generalized to a derived stack $\mathscr{D}$, encoding layered knowledge structures via Postnikov towers or Goodwillie calculus [?]. Semantic transformations are morphisms in the ∞-category of derived stacks.

3 Connections to Large Language Models

3.1 Embeddings as Scalar Fields

Token embeddings in LLMs are discrete vectors in $\mathbb{R}^d$. In RSVP, $\Phi(x,t)$ assigns a continuous semantic potential to each point $x \in \mathscr{M}$, evolving with context, offering a smoother, interpretable representation.

3.2 Attention as Vector Flows

Attention mechanisms compute weighted relationships between tokens. The vector field $\vec{v}$ models these as directional flows, with PDE (2) governing how semantic influence propagates, akin to attention scores as discretizations of $\vec{v}$.

3.3 Entropy and Interpretability

The entropy field S quantifies uncertainty, aligning with model confidence or perplexity. Gradients ∇S highlight regions of high ambiguity, serving as diagnostic tools for interpretability.

3.4 Multimodal Integration

RSVP unifies modalities (text, images, audio) by mapping them to field configurations on $\mathscr{D}$. For example, text tokens contribute to Φ_{text}, image features to Φ_{image}, with morphisms defining cross-modal correspondences.

4 Category-Theoretic Formalization

We formalize RSVP using ∞-category theory, focusing on interpretability and multimodal integration.

Definition 1. Let $\mathscr{C}$ be the ∞-category of derived stacks over $\mathbb{R}$. A semantic field configuration is a triple $(\Phi, \vec{v}, S) \in \mathsf{Fun}(\mathscr{M} \times \mathbb{R}, \mathbb{R}) \times \mathsf{Fun}(\mathscr{M} \times \mathbb{R}, T\mathscr{M}) \times \mathsf{Fun}(\mathscr{M} \times \mathbb{R}, \mathbb{R}_{\geq 0})$, with dynamics given by PDEs (1–3).

Theorem 1. The RSVP action $\mathscr{S}[\Phi, \vec{v}, S]$ is invariant under gauge transformations $\Phi \to \Phi + \chi$, $\vec{v} \to \vec{v} + \nabla\chi$.

Proof. Consider the Lagrangian (5). Under $\Phi \to \Phi + \chi$, $\vec{v} \to \vec{v} + \nabla\chi$, compute the transformed terms:

- $\frac{\partial \Phi}{\partial t} \to \frac{\partial \Phi}{\partial t} + \frac{\partial \chi}{\partial t}$, and $\nabla \cdot (\Phi\vec{v}) \to \nabla \cdot ((\Phi + \chi)(\vec{v} + \nabla\chi))$. Since $\nabla\chi$ is divergence-free on $\mathscr{M}$ (assuming appropriate boundary conditions), the kinetic term remains invariant.
- $\nabla\Phi \to \nabla\Phi + \nabla\chi$, but $\mathscr{U}(S, \Phi)$ depends only on Φ's magnitude, preserving invariance.
- $\nabla \cdot \vec{v} \to \nabla \cdot (\vec{v} + \nabla\chi) = \nabla \cdot \vec{v}$, as $\nabla \cdot \nabla\chi = 0$.

Thus, $\mathscr{L}$ and $\mathscr{S}$ are invariant, ensuring physical observables are gauge-independent. □

Proposition 1. Multimodal integration is a functor $F : \mathscr{C}_{\text{text}} \times \mathscr{C}_{\text{image}} \times \mathscr{C}_{\text{audio}} \to \mathscr{C}$, mapping modality-specific stacks to a unified derived stack $\mathscr{D}$.

Proof. Define $\mathscr{C}_{\text{text}}$, $\mathscr{C}_{\text{image}}$, $\mathscr{C}_{\text{audio}}$ as subcategories of $\mathscr{C}$ encoding modality-specific field configurations. The functor F assigns $(\Phi_{\text{text}}, \vec{v}_{\text{text}}, S_{\text{text}}) \mapsto (\Phi, \vec{v}, S)$, where $\Phi = \sum_i w_i \Phi_i$, $\vec{v} = \sum_i w_i \vec{v}_i$, and $S = \sum_i w_i S_i$, with weights w_i determined by cross-modal attention (discretized $\vec{v}$). Functoriality follows from preservation of morphisms (semantic transformations) under field composition. □

Corollary 1. Interpretability metrics (e.g., coherence of Φ, gradients ∇S) are invariants in the homotopy type of $\mathscr{D}$.

5 Interpretability via RSVP

RSVP enhances LLM interpretability by:

- Attention as Field Dynamics: Attention weights are solutions to (2), with $\vec{v}$'s streamlines revealing semantic pathways (e.g., key-token influence).
- Entropy Diagnostics: High ∇S regions indicate uncertainty, guiding model debugging or pruning.
- Cohomological Obstructions: Forgetting or reasoning failures correspond to non-trivial cohomology classes in $\mathscr{D}$, identifiable via obstruction theory.

For example, a numerical scheme discretizing (1–3) on a lattice can simulate attention flows, with ∇S visualized as a heatmap to pinpoint ambiguous regions.

6 Multimodal Integration

RSVP unifies modalities by mapping text, image, and audio embeddings to a shared field configuration on $\mathscr{D}$. The functor F (Proposition 1) ensures coherent integration, with $\vec{v}$ mediating cross-modal attention. For instance, in a vision-language model, Φ_{image} (from pixel features) and Φ_{text} (from tokens) combine via:

$$\Phi = w_{\text{image}} \Phi_{\text{image}} + w_{\text{text}} \Phi_{\text{text}}, \tag{7}$$

with weights derived from $\vec{v}$'s flow. This enables seamless reasoning across modalities, improving generalization.

7 Conclusion

The RSVP framework provides a mathematically rigorous, physically grounded model for LLMs, enhancing interpretability through entropy diagnostics and unifying multimodal data via derived stack morphisms. Future work includes:

- Simulating RSVP PDEs to model attention dynamics.
- Designing field-inspired transformers with intrinsic alignment.
- Exploring neuromorphic hardware for field computations.

RSVP bridges symbolic, sub-symbolic, and ethical AI, promising robust, interpretable, and multimodal systems.