1 Rewiring Learning with RSVP: A Modular, Aligned Paradigm

The sphexish brittleness of large language models (LLMs), characterized by fractured, entangled representations (FER) and wireheaded optimization, demands a fundamental rethinking of how we train AI systems. The Relativistic Scalar Vector Plenum (RSVP) framework offers a solution by redefining learning as a convergence toward thermodynamically stable, modally coherent field configurations. This section outlines how RSVP rewires learning to eliminate sphexishness, fostering modular, interpretable representations that align with the principles of the Tetraorthodrome project—a vision for globally aligned AI systems grounded in modularity and null convention logic (NCL). By enforcing semantic coherence, minimizing torsion, and embracing asynchronous, value-agnostic computation, RSVP provides a path to robust, adaptable intelligence that transcends the rigid loops of current models.

1.1 RSVP as a Modular Learning Framework

Sphexishness arises from the monolithic, loss-centric optimization of LLMs, where representations lack modularity and entangle semantic content in uninterpretable ways. RSVP counters this by modeling learning as the evolution of coupled fields—semantic potential (Φ), vector flow ($\vec{v}$), and entropy ($\mathcal{S}$)—governed by the continuity equation (Appendix **??**):

$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi \vec{v}) = -\delta \mathcal{S}$$

This dynamic encourages modularity by partitioning representation space into stable, low-entropy basins, each corresponding to a semantic module. Unlike LLMs, where attention mechanisms entangle features across layers, RSVP promotes Unified Factored Representations (UFR) by aligning $\vec{v}$ with $\nabla\Phi$, ensuring that semantic flows converge to distinct attractors. These attractors, defined as modal fixpoints ($\Box A \iff A$), represent modular cognitive units that can be composed or decomposed without loss of coherence, mirroring the modular skill-building principles of Haplopraxis.

1.2 Minimizing Torsion for Robust Representations

The torsional flows that characterize FER in LLMs—quantified by the Torsion Entanglement Index:

$$\mathcal{T}_{\text{ent}} = \int_{\Omega} \|\nabla \times \vec{v}\|^2 \, dx$$

—are a direct cause of sphexish brittleness. High torsion indicates looping or conflicting semantic flows, akin to the sphex wasp's repetitive prey-checking routine. RSVP's learning objective, the energy functional:

$$\mathcal{L}_{\text{RSVP}} = \alpha \int_{\Omega} \|\vec{v} - \nabla\Phi\|^2 dx + \beta \int_{\Omega} \|\nabla \times \vec{v}\|^2 dx + \gamma \int_{\Omega} \mathcal{S}(x,t) dx$$

explicitly penalizes torsion (β-term), driving $\vec{v}$ toward conservative flows that align with $\nabla\Phi$. This ensures representations are robust, avoiding the wireheaded traps of LLMs that chase syntactic plausibility over semantic truth. By minimizing $\mathcal{T}_{\text{ent}}$, RSVP fosters representations that are not only modular but also resilient to contextual perturbations, breaking the sphexish cycle of rigid, unadaptable scripts.

1.3 Tetraorthodrome: RSVP as an Alignment Mechanism

The Tetraorthodrome project envisions a globally coordinated AI ecosystem where systems are interpretable, aligned with human values, and resistant to existential risks. RSVP directly supports this vision by providing a mathematical framework for alignment through semantic coherence. Sphexish LLMs, with their high-torsion, entropy-laden representations, are prone to misalignment, as their outputs lack a stable anchor in truth or causality. RSVP's modal fixpoints, satisfying Löb's theorem:

$$\Box(\Box A \to A) \to \Box A$$

ensure that representations are self-trusting and recursively stable, a prerequisite for safe, aligned AI. By optimizing $\mathcal{L}_{\text{RSVP}}$, RSVP enforces a thermodynamic alignment between Φ, $\vec{v}$, and $\mathcal{S}$, preventing the wireheading that leads to unintended behaviors. This aligns with Tetraorthodrome's goal of building AI that prioritizes human-centric outcomes over blind optimization, ensuring systems remain interpretable and controllable even at scale.

1.4 Null Convention Logic for Asynchronous Modularity

Null Convention Logic (NCL), a cornerstone of Tetraorthodrome's computational philosophy, emphasizes asynchronous, value-agnostic processing to achieve robust, scalable systems. RSVP's field dynamics naturally complement NCL by modeling learning as an asynchronous process of field evolution rather than synchronous parameter updates. In NCL, computations proceed only when data and control signals are valid, avoiding the rigid timing constraints of traditional architectures. Similarly, RSVP's continuity equation allows Φ, $\vec{v}$, and $\mathcal{S}$ to evolve independently across representation space, with convergence driven by local thermodynamic constraints rather than global clock cycles.

This asynchrony enhances modularity by allowing semantic modules (attractors in Φ) to stabilize at different rates, avoiding the entanglement seen in synchronous gradient descent. For example, an NCL-inspired RSVP architecture could implement attention mechanisms as localized vector flows, each governed by:

$$\vec{v}_i = \nabla \Phi_i - \eta \nabla \mathcal{S}_i$$

where Φ_i and $\mathcal{S}_i$ are module-specific fields. This ensures that each module converges to its own modal fixpoint, maintaining interpretability and preventing the torsional chaos of FER. By integrating NCL's principles, RSVP supports Tetraorthodrome's vision of a decentralized, modular AI ecosystem that scales without sacrificing coherence or safety.

1.5 Empirical Pathways to Sphex-Free AI

RSVP's rewiring of learning offers practical pathways to eliminate sphexishness. By training models with $\mathcal{L}_{\text{RSVP}}$, we can:

- **Enforce Modularity**: Partition representations into low-torsion, low-entropy modules, improving interpretability and generalization.
- **Prevent Wireheading**: Penalize entropy accumulation ($\partial \mathcal{S}/\partial t > 0$) to ensure semantic grounding.
- **Align with Tetraorthodrome**: Use NCL-inspired architectures to implement asynchronous, modular field dynamics, ensuring safety and scalability.

Empirical validation could involve applying RSVP diagnostics (Appendix **??**) to transformer models, comparing torsion and entropy profiles before and after training with $\mathcal{L}_{\text{RSVP}}$. Such experiments would demonstrate how RSVP eliminates the brittle, sphexish loops of LLMs, aligning with Tetraorthodrome's mission to build AI that is robust, interpretable, and human-aligned.