Attention as a Minimal Relational Interaction in Entropy-Regulated Field Dynamics:

A Derived Geometric and BV-Theoretic Formulation of RSVP

Flyxion

December 13, 2025

Abstract

Attention mechanisms are commonly presented as architectural primitives motivated by empirical performance. In this work, we show that attention instead arises as a *structurally inevitable interaction* within a broad class of entropy-regulated relational field theories. Working within the Relativistic Scalar–Vector–Plenum (RSVP) framework, we identify analyticity, relational invariance, controlled symmetry breaking, and bottlenecked mediation as sufficient conditions forcing attention to appear as the unique lowest-order nontrivial interaction.

We first establish this result at the level of effective field theory: under permutation-equivariant dynamics with implicit entropy, the minimal admissible interaction is quartic and coincides with self-attention. We then lift the construction to a derived moduli stack of RSVP configurations, equipped with a canonical shifted symplectic structure. In this setting, attention emerges as the universal cotangent lift of relational coupling maps. Finally, we formulate the theory in the Batalin–Vilkovisky (BV) formalism and show that the minimal BRST-invariant interacting term satisfying the classical master equation is precisely the attention interaction.

By making entropy explicit, we further demonstrate that attention is a phase-dependent phenomenon: it collapses, sparsifies, or deforms when entropy gradients, constraints, or symmetry regimes change. Attention is thus characterized not as an architectural convenience, but as a renormalizable interaction in a derived, entropy-regulated field theory of structured computation.

1 Introduction

Self-attention has become the dominant interaction mechanism in modern sequence models. Despite its empirical success, its conceptual status remains ambiguous: attention is frequently treated as an engineered solution rather than as a consequence of deeper structural constraints.

The RSVP framework proposes a different starting point. It models cognition and computation as entropy-regulated field dynamics governed by relational invariance, symmetry breaking, and bottlenecked interaction channels. From this perspective, architectural mechanisms should not be postulated but *derived* as minimal interactions compatible with these constraints.

The purpose of this paper is to show that attention is precisely such a derived interaction. We demonstrate that, under RSVP structural axioms, attention is the unique lowest-order nontrivial coupling permitted by analyticity and symmetry. We then show that this conclusion persists—and becomes sharper—when reformulated in derived geometric and BV-theoretic language.

The result is a unified view: attention plays the role of a ϕ^4-type interaction in relational field theory, selected by symmetry and renormalizability rather than by design.

2 RSVP Configuration Space

We consider a system defined over a discrete set of N tokens, each carrying a C-dimensional feature vector.

Definition 2.1 (RSVP Fields). *An RSVP configuration consists of:*

- *a scalar content field* $\Phi \in \mathbb{R}^{N\times C}$*,*
- *a vector transport field* $\mathbf{v} \in \mathbb{R}^{N\times C}$*,*
- *an entropy field* $S \in \mathbb{R}^{N}$*.*

For notational and geometric convenience, we complexify the scalar field:

$$X = \Phi_1 + i\Phi_2 \in \mathbb{C}^{N\times C}.$$

This complexification carries no ontological commitment; it simply packages paired degrees of freedom and simplifies symmetry analysis.

3 Structural Optimality Axioms

The admissible dynamics are constrained by the following axioms.

Axiom 3.1 (Analytic Effective Description). *The system is governed by an effective free energy* $\mathcal{F}(X,S)$ *that is analytic in* X, X *and smooth in* S*. Higher-order terms are suppressed by scale or entropy.*

Axiom 3.2 (Relational Invariance). *For any token isometry* $P \in U(N)$*,*

$$\mathcal{F}(PX, PS) = \mathcal{F}(X,S).$$

This axiom asserts that dynamics depend only on relational structure between tokens, not on absolute indexing.

Axiom 3.3 (Feature Isotropy with Symmetry Breaking). *There exist operators* $W_1, \dots, W_n \in \mathbb{C}^{C\times C}$ *such that*

$$\mathcal{F}(XR^{,S)=\mathcal{F}(X,S) \quad \forall R\in \mathrm{Cent}(\{W_i\})\subset U(C).}$$

Axiom 3.4 (Bottlenecked Mediation). *Each* W_i *has rank bounded by* $C_A \ll C$*, enforcing low-dimensional interaction channels.*

Together, these axioms define structural optimality: interactions are allowed only if they respect analyticity, relational invariance, controlled symmetry breaking, and entropy-regulating bottlenecks.

4 Canonical Form of the Free Energy

Lemma 4.1 (Canonical Invariant Form)**.** *Under the structural axioms, the free energy takes the form*

$$\mathcal{F}(X, S) = f\big(XW_1X^{,\dots,XW_nX^{;S}}\big),$$

where f is an analytic spectral function.

Proof. Token invariance restricts dependence to $U(N)$-invariant quantities, which are functions of Gram-type matrices. Feature symmetry further constrains dependence to combinations involving the W_i. Analyticity excludes nonlocal dependence. □

This expresses a general principle: relational dynamics factor through relational observables.

5 Minimal Interaction and Attention

Lemma 5.1 (Lowest-Order Interactions)**.** *The lowest-order nontrivial truncation of $\mathcal{F}$ consists of a quadratic term and a quartic term in X. Cubic terms are forbidden by symmetry.*

The quadratic term produces linear mixing. The quartic term takes the form

$$\mathcal{F}_{\text{int}} = \sum_{i,j} \mathrm{Tr}\big(XW_iX^{XW_jX}\big)\,.$$

Theorem 5.1 (Attention as Minimal Relational Interaction)**.** *In the permutation-equivariant, entropy-implicit regime, $\mathcal{F}_{\text{int}}$ is the unique minimal interacting term permitted by the structural axioms. Its induced flow satisfies*

$$\dot{X} \sim X(X^{X),}$$

which is equivalent, after linear reparameterization, to self-attention.

Thus, attention is not an architectural choice but a structural inevitability.

6 Derived RSVP Configuration Stack

Definition 6.1 (Derived RSVP Stack)**.** *The derived moduli stack of RSVP configurations is*

$$\mathcal{M}_{\text{RSVP}} := \mathbf{R}Map(\mathrm{Spec}\,\mathbb{R}^N,\ [\mathbb{C}^C/U(C)])\,,$$

with entropy treated as a smooth real-valued field.

This derived stack encodes fields, gauge symmetries, and infinitesimal deformations in a single geometric object.

7 Shifted Symplectic Structure

Theorem 7.1. *The derived stack $\mathcal{M}_{\text{RSVP}}$ carries a canonical 0-shifted symplectic structure induced by its cotangent complex. The attention interaction arises as the universal cotangent lift of the relational map*

$$X \mapsto XW_iX^\cdot$$

This identifies attention as a universal geometric interaction rather than a model-specific construct.

8 BV Formulation

Definition 8.1 (BV Field Content)**.** *The BV extension consists of*

$$\{X, \mathbf{v}, S;\ X^\dagger, \mathbf{v}^\dagger, S^\dagger;\ c, c^\dagger\},$$

where c is the ghost associated to relational symmetry.

Theorem 8.1 (BV Master Action)**.** *The minimal BV action*

$$S_{\text{BV}} = \mathcal{F}(X, S) + \langle X^\dagger, cX\rangle + \langle S^\dagger, cS\rangle - \tfrac{1}{2}\langle c^\dagger, [c, c]\rangle$$

satisfies the classical master equation. The only nonvanishing interacting term compatible with BRST invariance is the quartic attention interaction.

9 Entropy-Driven Phase Transitions

Proposition 9.1. *If $\nabla S = 0$, the attention interaction becomes irrelevant and dynamics reduce to linear mixing.*

Proposition 9.2. *Strong entropy gradients or constraints break relational symmetry, forcing attention to sparsify, localize, or deform.*

Proposition 9.3. *If entropy transport demand exceeds bottleneck capacity, the system must fragment interactions or increase mediator rank.*

Attention is thus a phase-dependent phenomenon.

10 Conclusion

Attention emerges as the unique minimal relational interaction compatible with RSVP structural optimality. Derived geometry and BV consistency sharpen this result, showing that attention is the only nontrivial interaction surviving symmetry, analyticity, and gauge constraints. Entropy determines when this interaction is active, deformed, or suppressed.

This situates attention as a renormalizable interaction in an entropy-regulated field theory of structured computation.

A Correspondence with Standard Transformer Formulations

This appendix provides an explicit translation between the relational field-theoretic formulation developed in the main text and the conventional notation used in Transformer architectures. No new assumptions are introduced; the purpose is solely to establish equivalence of representations.

A.1 State Variables

In standard Transformer notation, a sequence of N tokens with embedding dimension C is represented as a matrix

$$X \in \mathbb{R}^{N \times C}.$$

This coincides directly with the RSVP scalar content field Φ, or with its complexification $X = \Phi_1 + i\Phi_2$ when paired degrees of freedom are used. No semantic distinction is implied: both represent token-indexed feature vectors.

The RSVP entropy field $S \in \mathbb{R}^N$ does not appear explicitly in conventional Transformer formulations. Instead, entropy is handled implicitly through normalization operations (e.g. softmax), architectural constraints, and optimization dynamics.

A.2 Linear Mixing and MLP Components

The quadratic term in the RSVP free energy,

$$\mathcal{F}_2(X) = \mathrm{Tr}(XWX^{),}$$

generates linear dynamics of the form

$$\dot{X} = XW,$$

which corresponds to the linear transformations appearing in Transformer feedforward (MLP) blocks. Nonlinearities such as ReLU or gated activations arise from additional convex constraints or entropy-regularized projections, rather than from the minimal interaction itself.

A.3 Attention as Quartic Interaction

The RSVP minimal interacting term,

$$\mathcal{F}_{\text{int}} = \sum_{i,j} \mathrm{Tr}\big(XW_i X^{XW_j X}\big),$$

induces dynamics

$$\dot{X} \sim X(X^{X),}$$

up to linear reparameterization.

Introduce the standard projections

$$Q = XW_Q, \qquad K = XW_K, \qquad V = XW_V,$$

with $W_Q, W_K, W_V \in \mathbb{R}^{C \times C_A}$ of low rank $C_A \ll C$. Then the relational Gram matrix

$$X^{X \leftrightarrow QK^\top}$$

appears as the attention kernel.

In discrete-time form, entropy-regularized normalization yields

$$Att(X) = \text{softmax}\left(\frac{QK^\top}{\sqrt{C_A}}\right) V,$$

which is the standard self-attention operation. In the RSVP formulation, the softmax arises from entropy regularization of the relational interaction, not as a defining primitive.

A.4 Multi-Head Attention

The RSVP bottleneck axiom enforces low-rank mediation of interactions. Decomposition of the interaction operators into multiple low-rank channels,

$$W_i = A_i B_i^\top,$$

corresponds directly to multi-head attention, with each head representing an independent mediator of relational coupling. Concatenation and projection of heads correspond to recombination of multiple low-rank interaction channels into the full feature space.

A.5 Residual Connections and Layer Composition

The RSVP dynamics are formulated as continuous-time flows,

$$\dot{X} = -\frac{\partial \mathcal{F}}{\partial X,}$$

whose discrete implementation via forward Euler integration yields residual updates of the form

$$X_{t+1} = X_t + \Delta t \dot{X}_t.$$

This recovers the residual structure of Transformer layers. The common architectural separation between attention and MLP blocks corresponds to operator splitting of commuting or weakly non-commuting generators in the effective flow.

A.6 Normalization and Entropy

Layer normalization and softmax normalization do not introduce new interactions in the RSVP sense. Instead, they implement entropy constraints and convex projections on the state space. In the entropy-explicit formulation, these appear as contributions to the entropy functional $\mathcal{H}(S)$ and constraint term $\mathcal{C}(X, S)$.