

Project Proposal: Core Clifford Spatiotemporal Fusion (C-CASF) Layer for Continuous-Time Dynamic Spatio-Temporal Graphs

August 12, 2025

1 Introduction and Motivation

Dear Professor,

This proposal details the design and implementation of the **Core Clifford Spatiotemporal Fusion (C-CASF) Layer**, a pivotal foundational component within our STAMPEDE (Spatio-Temporal Adaptable Multi-modal Positional Embedding & Dynamic Evolution) framework. The C-CASF layer is specifically designed to generate unified, richly interpretable spatio-temporal embeddings for entities (nodes and edges) in continuous-time dynamic graphs.

Our overarching STAMPEDE strategy involves a modular pipeline:

1. **Specialized Encoding:** Separate modules learn rich spatial embeddings (e_s) and temporal embeddings (e_t) for graph entities. For this project, we utilize **R-PEARL** for spatial encodings and **LeTE** for temporal encodings.
2. **Spatio-Temporal Fusion:** A dedicated layer combines e_s and e_t into a unified spatiotemporal representation (e_{st}).

Traditional fusion methods, such as simple concatenation or basic MLPs, treat features as arbitrary numerical vectors, inherently failing to capture the intricate geometric relationships and dynamic interplay between space and time. Furthermore, they often produce static outputs that don't adapt to the continuous evolution of the graph.

The C-CASF layer directly addresses these limitations by:

1. Leveraging **Clifford Algebra** for a principled, geometrically meaningful fusion of spatial and temporal embeddings.
2. Generating a **pure bivector** representation that explicitly captures the fundamental pairwise spatio-temporal interaction as an oriented plane.

This core framework provides a mathematically rigorous, highly expressive, and interpretable approach to fundamental spatio-temporal interaction representation. While the full STAMPEDE vision includes multi-grade multivector representations and dynamic transformations via rotors for comprehensive evolving states (as discussed in Section 6), this proposal focuses on establishing the mathematically pure and unassailable foundation of the bivector fusion. This approach is well-suited for a top-tier AAAI conference publication.

2 Mathematical Foundation: Clifford Algebra

Our framework is built upon the robust mathematical language of **Clifford Algebra**, specifically tailored for spatio-temporal dynamics.

2.1 Clifford Algebra $\mathcal{Cl}(D_S, D_T)$ with Mixed Signature

We define a Clifford algebra $\mathcal{Cl}(D_S, D_T)$, where D_S is the number of spatial dimensions and D_T is the number of temporal dimensions. This notation explicitly defines the fundamental properties of its generating basis vectors:

- **Basis Vectors \mathbf{e}_i and \mathbf{f}_j :** These are abstract symbols that define the algebraic space. Their properties are determined by their multiplication rules, which are **axiomatic definitions** of this specific algebra.
 - **Spatial Basis Vectors (\mathbf{e}_i):** For $i = 1, \dots, D_S$, satisfy $\mathbf{e}_i^2 = +1$. This defines their "space-like" character.
 - **Temporal Basis Vectors (\mathbf{f}_j):** For $j = 1, \dots, D_T$, satisfy $\mathbf{f}_j^2 = -1$. This defines their "time-like" character, akin to dimensions in Minkowski spacetime.
 - **Orthogonality and Anti-Commutation between Distinct Basis Vectors:** For any two distinct basis vectors \mathbf{u}, \mathbf{v} from the set $\{\mathbf{e}_1, \dots, \mathbf{e}_{D_S}, \mathbf{f}_1, \dots, \mathbf{f}_{D_T}\}$, they anti-commute: $\mathbf{u}\mathbf{v} = -\mathbf{v}\mathbf{u}$. This is a direct consequence of them being orthogonal ($\mathbf{u} \cdot \mathbf{v} = 0$) within the Clifford algebra's orthonormal basis construction. This applies specifically to:
 - * $\mathbf{e}_i\mathbf{e}_j = -\mathbf{e}_j\mathbf{e}_i$ for $i \neq j$.
 - * $\mathbf{f}_i\mathbf{f}_j = -\mathbf{f}_j\mathbf{f}_i$ for $i \neq j$.
 - * **Crucial Design Choice:** Any spatial basis vector \mathbf{e}_i is orthogonal to any temporal basis vector \mathbf{f}_j . Algebraically, this implies they ****anti-commute****: $\mathbf{e}_i\mathbf{f}_j = -\mathbf{f}_j\mathbf{e}_i$. This property ensures their inner product $\mathbf{e}_i \cdot \mathbf{f}_j = \frac{1}{2}(\mathbf{e}_i\mathbf{f}_j + \mathbf{f}_j\mathbf{e}_i)$ **vanishes (is zero)**.
- **Geometric Product ($\mathbf{u}\mathbf{v}$):** For any two vectors \mathbf{u} and \mathbf{v} , the geometric product is defined as the sum of their inner product and outer product:

$$\mathbf{u}\mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \wedge \mathbf{v}$$

This is the unifying product in Clifford algebra.

- **Inner Product ($\mathbf{u} \cdot \mathbf{v}$):** The symmetric, scalar part of the geometric product: $\mathbf{u} \cdot \mathbf{v} = \frac{1}{2}(\mathbf{u}\mathbf{v} + \mathbf{v}\mathbf{u})$. It measures alignment.
- **Outer Product ($\mathbf{u} \wedge \mathbf{v}$):** The antisymmetric, non-scalar part of the geometric product: $\mathbf{u} \wedge \mathbf{v} = \frac{1}{2}(\mathbf{u}\mathbf{v} - \mathbf{v}\mathbf{u})$. It represents an oriented subspace (e.g., an oriented plane for two vectors).
- **Multivectors and Grades:** The full algebra is a $2^{D_S+D_T}$ -dimensional vector space whose elements are called multivectors. Multivectors are classified by their "grade": Grade 0 (scalar), Grade 1 (vector), Grade 2 (bivector), etc.

3 Framework: Core Clifford Spatiotemporal Fusion (C-CASF) Layer

The C-CASF Layer is a learnable, differentiable module that performs a precise geometric fusion of spatial and temporal embeddings for each entity at a given time.

3.1 Overall Role within STAMPEDE

The C-CASF Layer takes final, real-valued outputs from the upstream **Spatial Encoder (R-PEARL)** ($e_s \in \mathbb{R}^{D_s}$) and **Temporal Encoder (LeTE)** ($e_t \in \mathbb{R}^{D_T}$). It then produces the unified spatio-temporal embedding e_{st} for downstream tasks.

3.2 Inputs to the C-CASF Layer

- **Spatial Embedding** ($e_s \in \mathbb{R}^{D_s}$): From R-PEARL, representing an entity’s structural spatial role. In the context of continuous-time graphs, this e_s is typically pre-computed on a static graph snapshot (e.g., the graph at $t = 0$) or an aggregated graph, capturing the node’s stable structural position.
- **Temporal Embedding** ($e_t \in \mathbb{R}^{D_T}$): From LeTE, representing the current event’s temporal characteristics (e.g., timestamp, time difference). This is dynamically computed per event.

Both inputs are real-valued numerical tensors.

3.3 Mechanism of the C-CASF Layer (Step-by-Step)

The C-CASF layer’s primary function is to compute a pure bivector representation of the spatio-temporal interaction.

3.3.1 Stage 1: Core Bivector Fusion (Geometric Product Calculation)

This stage establishes the fundamental pairwise interaction between spatial and temporal features as an oriented plane, which is the direct result of their geometric product in our defined algebra.

1. Conceptual Lifting/Interpretation:

- The real-valued $e_s = [e_s^{(1)}, \dots, e_s^{(D_s)}]$ is interpreted as the coefficients of a **pure spatial vector** $\mathbf{S} = \sum_{i=1}^{D_s} e_s^{(i)} \mathbf{e}_i$. This is achieved by an initial learnable linear projection of e_s if its input dimension does not match D_s .
- The real-valued $e_t = [e_t^{(1)}, \dots, e_t^{(D_T)}]$ is interpreted as the coefficients of a **pure temporal vector** $\mathbf{T} = \sum_{j=1}^{D_T} e_t^{(j)} \mathbf{f}_j$. This is achieved by an initial learnable linear projection of e_t if its input dimension does not match D_T .
- **Note:** For implementation, no explicit full multivector tensors for \mathbf{S} or \mathbf{T} are created; operations are performed directly on the input coefficients e_s and e_t .

2. Compute the Geometric Product ($\mathbf{S} * \mathbf{T}$):

- The geometric product is defined as $\mathbf{S} * \mathbf{T} = \mathbf{S} \cdot \mathbf{T} + \mathbf{S} \wedge \mathbf{T}$.
- **Crucial Implication from Design Choice:** Because our spatial basis vectors (\mathbf{e}_i) are explicitly defined as orthogonal to our temporal basis vectors (\mathbf{f}_j), the inner product term $\mathbf{S} \cdot \mathbf{T}$ **vanishes (evaluates to zero)**.
- Thus, the geometric product simplifies directly to the wedge product: $\mathbf{S} * \mathbf{T} = \mathbf{S} \wedge \mathbf{T}$.
- **What the Wedge Product Does:** It calculates the coefficients for the basis bivectors ($\mathbf{e}_i \wedge \mathbf{f}_j$). The coefficient for each such bivector is simply the pairwise product $e_s^{(i)} e_t^{(j)}$.
- **Computational Implementation:** This is efficiently computed as an **outer product** of the input real vectors e_s and e_t (e.g., `torch.outer(e_s, e_t)`). This results in a $D_S \times D_T$ matrix of coefficients.
- **Result:** The numerical coefficients representing the **pure bivector** \mathbf{B}_{ST} . (Shape: $(D_S \cdot D_T,)$ after flattening).

3. Final Projection to Desired Output Dimension (ψ'):

- The coefficients of the bivector \mathbf{B}_{ST} (represented by its flattened coefficients) are passed through a final linear layer ψ' to project to the desired output dimension D_{ST} .

$$e_{st} = \psi'(\mathbf{B}_{ST}^{\text{coeffs}})$$

This layer allows the model to adapt the bivector representation to the specific requirements of downstream tasks.

3.4 Output of the C-CASF Layer

- The final unified spatio-temporal embedding $e_{st} \in \mathbb{R}^{D_{ST}}$. This real-valued vector is now ready for downstream task-specific heads in models like DyGMamba.

4 Learnability and Plug-and-Play Integration

The C-CASF Layer is designed for full end-to-end training and seamless integration within the STAMPEDE framework and existing dynamic graph neural networks.

- **End-to-End Differentiability:** All operations (initial linear projections, outer product, final linear layer) are inherently differentiable within standard deep learning frameworks (e.g., PyTorch). Gradients flow unimpeded from the task loss back through C-CASF and to the upstream R-PEARL and LeTE encoders, which are also designed to be trainable.
- **Plug-and-Play Modularity:** C-CASF has a clear input/output interface ($e_s, e_t \rightarrow e_{st}$). It's independent of the internal workings of the upstream spatial/temporal encoders or downstream task heads, making it highly modular. This allows it to be a direct replacement for simpler fusion methods (like concatenation followed by an MLP) within existing dynamic graph architectures (e.g., DyGMamba).
- **Continuous-Time Dynamic Compatibility:** The dynamic temporal information from LeTE (e_t) is directly fused with the spatial information from R-PEARL (e_s), ensuring the fused embedding is dynamically responsive to continuous-time graph evolution. The layer processes entities batch-wise for efficiency.

5 Advantages and Expected Contributions

The C-CASF layer represents a novel and impactful contribution to dynamic graph representation learning:

1. **Principled Geometric Fusion (Foundation):** It leverages Clifford algebra’s rigor to provide a mathematically sound and interpretable framework for spatio-temporal fusion, moving beyond arbitrary feature blending. The explicit generation of a pure bivector as the spatio-temporal interaction is a fundamental geometric primitive.
2. **Direct Interpretability:** The bivector naturally represents an oriented spatio-temporal plane, providing direct geometric intuition for the fundamental interaction between spatial location and temporal events. This offers a clear, interpretable inductive bias.
3. **Novelty:** This explicit, pure bivector fusion based on mixed-signature Clifford algebra for continuous-time dynamic spatio-temporal graphs is a novel approach for establishing a foundational spatio-temporal embedding.

6 STAMPEDE Vision: Future Work and Full Framework

While the current C-CASF proposal focuses on establishing the pure bivector fusion, it serves as a foundational step towards our full STAMPEDE framework, which includes a richer multi-grade representation and dynamic transformations. These aspects are critical for capturing the full complexity of continuous-time dynamic graphs and represent immediate future work:

1. **Multi-Grade Multivector Representation:** The full vision involves generating a comprehensive **multi-grade multivector** representation that simultaneously captures different types of spatio-temporal interaction. This would include:
 - **Scalar (Grade 0):** Overall magnitude or intensity, derived from squared norms of e_s, e_t and contextual signals.
 - **Vector (Grade 1):** Net spatial and temporal displacements, derived from e_s, e_t and contextual signals.
 - **Bivector (Grade 2):** The core spatio-temporal interaction ($\mathbf{S} \wedge \mathbf{T}$), as well as pure spatial and pure temporal bivectors, and contextual bivectors.
 - **Trivector (Grade 3) and Higher Grades:** Formed by introducing additional context-derived vectors (from a contextual signal C_R) that participate in higher-order outer products (e.g., $\mathbf{S} \wedge \mathbf{T} \wedge \mathbf{V}_C$).

Each grade component would be explicitly constructed through Clifford algebraic operations or minimal, geometrically-constrained neural projections. This aims to combat information loss and provide a hierarchical understanding of the spatio-temporal state.

2. **Learnable, Dynamic Transformations using Rotors:** The full framework will implement **learnable, dynamic transformations (rotations) using rotors** directly within the fused multi-grade embedding space. This will allow the comprehensive spatio-temporal state to adapt and evolve based on dynamic context (C_R). Rotors (even-grade normalized multivectors) perform transformations via the sandwich product ($\mathbf{X}' = \mathbf{R} * \mathbf{X} * \mathbf{R}^{-1}$), providing interpretable, geometrically-faithful state evolution.

This roadmap ensures that our approach remains at the forefront of geometric deep learning research for dynamic graphs.

7 Conclusion

The Core Clifford Spatiotemporal Fusion (C-CASF) Layer is a transformative proposal for our STAMPEDE framework. By fundamentally rethinking spatio-temporal fusion through the lens of Clifford algebra, it offers a principled, expressive, and interpretable approach to generating foundational spatio-temporal embeddings for continuous-time dynamic graphs. This focused and mathematically pure initial contribution, while laying the groundwork for a comprehensive multi-grade and dynamic framework, makes it a strong candidate for a top-tier conference publication, setting a new standard for spatio-temporal representation learning.