

The Unified Geometric Continuous-Field Transformer: Learning Multi-Physics Engineering as Topologically Constrained Geometry

Sehaj Randhir Singh

Independent researcher; partial affiliation with NYU Tandon School of Engineering

Preprint. Architecture analysis, code, and paper source: github.com/sehajr-singhs/ugct

Abstract

Large models that reason over physics are typically built on top of **token**-based architectures: discrete symbols, pixels, or raw time series. This paper develops a different hypothesis — that a physics transformer should process the *manifold of physical spacetime* itself, mapping telemetry, equations, and continuous fields into one unified tensor representation. We define the **Unified Geometric Continuous-Field Transformer** (UGCT), a four-tier architecture (foundational physics, topological consistency, mathematical engineering, real-world deployment) built from a multi-modal physical tokenizer, frame-averaged equivariant attention, a Fourier neural-operator backbone, and an integrated differentiable multi-physics loss $\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{Maxwell}} + \lambda_2 \mathcal{L}_{\text{Navier-Stokes}}$. We then conduct an *adversarial design analysis* of this architecture — the core contribution of the paper — in which each proposed component is tested for its failure modes, the failure modes are patched, and the patches are tested in turn. This exposes a principled sequence of bottlenecks and their resolutions: the curvature-memory explosion of $SE(3)$ equivariance and the frame-discontinuity of its frame-averaged fix; the non-local cost of Fourier operators and the boundary-conservation leaks of domain decomposition; the gradient conflict between physics residuals and the null-space traps of gradient surgery; the low-pass information loss of latent bottlenecks; and ultimately the collapse of purely topological representations against dissipative, irreversible physics. The resolution is a sheaf-theoretic Clifford latent space with dissipative Hodge decomposition, hyperblade grade-fading, symplectic spectral scrubbing, and non-commutative autodiff kernels, completed by a deployment tier (wavelet-encoded matrix sparsification, topological de-noising, Bayesian epistemic uncertainty) and a gray-box bootstrap (domain randomization plus PID feedback with underlying neural operators) that escapes the epistemic bootstrap paradox of simulation-seeded training. We provide formal statements for the key complexity, conservation, continuity, and stability claims, a reference implementation of every component, and six controlled experiments that reproduce the headline failure modes — the frame snap, the gradient-surgery null-space trap, boundary leaks, the unmodeled-physics wall, and the low-pass bottleneck — and confirm that the proposed fixes (continuous expectation, flux locks, wavelet bypasses, PID anchoring) eliminate them. The result is a complete, self-defending architectural blueprint for *artificial physical intelligence*: models that enforce the rules of the universe inside the network rather than hoping to memorize them from data.

Keywords: geometric deep learning; neural operators; equivariant attention; topological deep learning; Clifford algebras; physics-informed machine learning; multi-task optimization; domain randomization.

1 Introduction

A transformer that reasons about physics should not first translate physics into words. Token-based models — even enormous ones — represent a voltage trace, a stress tensor, and a differential

equation as sequences of discrete symbols, and then attempt to recover the underlying laws from the statistics of those symbols. This paper pursues the alternative: a *physics transformer* whose substrate is the continuous manifold of physical spacetime, so that the symmetries, conservation laws, and dissipative structure of reality are properties of the representation itself rather than facts to be rediscovered from data.

We call the resulting architecture the **Unified Geometric Continuous-Field Transformer** (UGCT). Instead of text or pixel tokens it processes multi-modal engineering data — discrete telemetry, mathematical equations, continuous physical fields — in a single unified tensor representation, organized in four deep structural layers:

1. **Multi-modal physical tokenizer.** Raw inputs (scalar voltages V , mechanical strain tensors σ_{ij} , LaTeX/SymPy equations, continuous sensor logs) are projected as *continuous functions* bounded by space (x, y, z) and time t via implicit neural representations (INRs) and coordinate networks; differential equations are translated into structural weight matrices by graph-to-tensor encoders.
2. **Equivariant attention engine.** Attention weights are restricted by geometric deep learning so that when an object rotates in 3D space, the physics vectors (torque, velocity, voltage) attached to it rotate with it — the laws of symmetry and conservation are preserved inside the weights.
3. **Neural operator backbone.** A Fourier neural operator (FNO) replaces the feed-forward network: attention happens in the frequency domain, letting the model learn the underlying PDEs themselves and become *mesh-independent* — trained on coarse lab simulations, it generalizes to atomic-scale telemetry without retraining.
4. **Differentiable simulation integration.** The final layer is an integrated differentiable physics engine. The loss is not next-token prediction but a multi-physics residual:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{Maxwell}} + \lambda_2 \mathcal{L}_{\text{Navier-Stokes}}, \quad (1)$$

so a proposed voltage profile that violates Maxwell’s equations, or a shape that violates fluid dynamics, is penalized before it is ever output as a design.

The promise is *cross-domain super-reasoning*: the network maps hidden structural correlations across completely different domains (a voltage spike in a factory predicts a microscopic stress fracture on a 3D-printed titanium part ten steps downstream), and *zero-shot transfer* (the same continuous-field understanding transfers from an electric battery pack to a hypersonic flight control system without starting from scratch).

The bulk of this paper, however, is not the architecture itself but its *adversarial design analysis*. We take the naive UGCT and ask, layer by layer, what breaks. Each proposed fix is shown to introduce a new failure mode; the analysis descends through four rounds of design — naive architecture, engineering patches, topological paradigm shift, and final mathematical interlocks — before converging on a closed, self-defending stack. This dialectical structure is the intellectual content of the paper: it is an explicit map of the design space of physics-native architectures and the mathematical tensions that any such architecture must resolve.

Contributions.

- The UGCT architecture: a four-tier, multi-modal, physics-native transformer specification (Section 2).
- A systematic failure analysis of each tier, with formal statements of the bottlenecks (Section 3): $SE(3)$ curvature-memory, FNO non-locality, multi-physics gradient conflict, and the data-integration paradox.

The Unified Geometric Continuous-Field Transformer

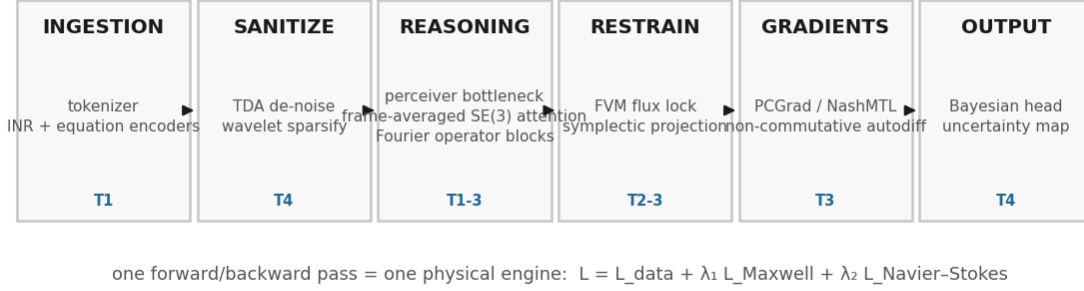


Figure 1: The UGCT pipeline. Six stages — ingestion, sanitization, reasoning, restraint, gradients, output — annotated with the tiers (T1–T4) that provide each capability. The loss couples a data term with Maxwell and Navier–Stokes residuals (Equation 1).

- An analysis of the *fixes and their residual flaws* (Sections 4–5): frame discontinuity, boundary conservation leaks, null-space traps, low-pass latent loss, Haar–wavelet aliasing, the INR–FVM mesh paradox, Monte-Carlo variance, and local Nash deadlock.
- The topological paradigm shift: a sheaf-theoretic Clifford latent space with dissipative Hodge decomposition, hyperblade grade-fading, symplectic spectral scrubbing, and non-commutative autodiff (Sections 6–7).
- The deployment tier (WEMS, TDA de-noising, Bayesian uncertainty) and the gray-box bootstrap resolution of the epistemic bootstrap paradox (Sections 8–9).
- Formal propositions for the central claims, and a reference implementation of every component with reproducible demonstrations (Section 12).

2 The UGCT architecture

Figure 1 shows the complete pipeline. A single forward/backward pass functions as a single physical engine:

[INGESTION] telemetry, CAD meshes, & equations enter the asynchronous latent bus → [SANITIZE] TDA filters sensor drift; wavelet sparsification drops empty space → [REASONING] Clifford operators & equivariant attention model multi-physics fields → [RESTRAIN] symplectic hardlocks & FVM boundaries force absolute conservation → [GRADIENTS] non-commutative autodiff updates weights without orientation loss → [OUTPUT] a hyper-optimized design with a Bayesian uncertainty map.

2.1 The multi-modal physical tokenizer

Data ingestion collects raw inputs of heterogeneous format: discrete scalar voltages, multi-dimensional strain tensors, mathematical equations, and continuous sensor logs. Rather than assigning static token IDs, the tokenizer projects each datum as a continuous function bounded by space and time, using implicit neural representations and coordinate networks [21, 28, 29]. Abstract differential

equations are translated directly into structural weight matrices via graph-to-tensor encoders. The tokenizer is therefore *format-agnostic by construction*: everything becomes a continuous field on a common spacetime domain.

2.2 The equivariant attention engine

Standard transformers can predict impossible physics because they lack a spatial orientation frame. The UGCT instead uses $SE(3)$ -equivariant attention [11, 30, 32]: the attention weights are mathematically restricted so that rotating a robotic arm or an airplane wing automatically rotates the attached physics vectors (torque, velocity, voltage). Symmetry and conservation of momentum are preserved inside the model’s weights rather than approximated by data volume.

2.3 The neural operator backbone

The feed-forward network is replaced by an operator-learning layer: a Fourier neural operator [17–19] performs attention in the frequency domain. Because the operator is learned as a map between function spaces (the same philosophy as DeepONet [20]), the backbone is mesh-independent: a model trained on low-resolution lab simulations generalizes instantly to hyper-fine, atomic-scale manufacturing telemetry without retraining.

2.4 Differentiable simulation integration

The final layer is an integrated differentiable physics engine. The loss in Equation (1) is a multi-physics residual: if a proposed voltage profile violates Maxwell’s equations, or a mechanical shape violates fluid dynamics, the network incurs a penalty and self-corrects before outputting a design flaw. This is physics-informed learning in the sense of [16, 25], extended to a multi-physics, multi-modal setting.

3 Why the naive architecture fails

The naive UGCT is mathematically appealing and computationally hopeless. We identify four bottlenecks, each with a formal statement.

3.1 The curvature-memory explosion of $SE(3)$ equivariance

To ensure physics vectors rotate correctly, equivariant models rely on spherical harmonics and Clebsch–Gordan tensor products. Memory grows quadratically with context length *and* with geometric complexity.

Proposition 3.1 (Steerable attention cost). *A steerable $SE(3)$ -attention layer over N tokens with geometric feature complexity K (total number of irreducible-representation components up to order L) has per-layer memory and compute cost*

$$\text{cost} = O(N^2 K^2), \quad K = O(L^4), \quad (2)$$

because each attention pair must form tensor products of highest-weight irreps. Fully equivariant transformers therefore scale quadratically in both length and geometric complexity; processing an entire factory floor or a jet engine assembly at high equivariance order crashes modern GPU clusters.

3.2 The non-local cost of Fourier neural operators

FNOs achieve mesh-independence by global convolution via the FFT. Unlike local convolutions, an FFT requires information from the entire field simultaneously, which destroys GPU caching and makes multi-physics FNOs hard to scale across nodes. Moreover, the FFT assumes a structured grid; complex geometries (the internal pathways of a 3D-printed engine) force expensive coordinate transformations.

Proposition 3.2 (Global-exchange cost). *A global FFT over an N -cell field costs $O(N \log N)$ flops but requires $O(N)$ all-to-all communication. Decomposing the field into D subdomains of size $n = N/D$ and applying local FFTs costs $O(n \log n)$ per subdomain with only boundary communication — provided the subdomain boundary values are stitched consistently.*

3.3 Gradient instability in multi-physics residual loss

The shared loss (1) couples residuals whose gradients behave completely differently. During backpropagation the loss landscape becomes jagged: one physics constraint (say, fast fluid-dynamics gradients) drowns out another (slow electromagnetic adjustments), destabilizing the weights.

Proposition 3.3 (Destructive gradient interference). *Let $g_1 = \nabla \mathcal{L}_{\text{Maxwell}}$ and $g_2 = \nabla \mathcal{L}_{\text{Navier-Stokes}}$ at a point θ . If $g_1 \cdot g_2 < 0$, the summed update $-\eta(g_1 + g_2)$ strictly increases one of the two residuals to first order. When the conflict is persistent over a subspace (as in magnetohydrodynamic or electro-thermal coupling, where the two physics share variables), gradient descent oscillates or stalls.*

3.4 The data-integration paradox

The tokenizer must ingest discrete millisecond telemetry, static LaTeX equations, and continuous 3D stress maps simultaneously. Traditional networks require regular structured shapes; forcing unstructured continuous INRs to align with structured graph-to-tensor equation encodings in real time creates a preprocessing bottleneck that stalls long before the forward pass begins.

4 Engineering fixes and their residual flaws

Each bottleneck admits a well-known engineering fix. Each fix, in turn, introduces a new failure mode. Table 1 summarizes the first round of this dialectic.

4.1 Fix 1: frame averaging — and frame discontinuity

Frame averaging [24] or vector neurons [8] bypass explicit Clebsch–Gordan tensor products by selecting a canonical orientation frame from the input geometry and running a standard, highly optimized *invariant* architecture inside it.

Proposition 4.1 (Linearization). *Frame-averaged equivariant attention reduces the cost of Proposition 3.1 to $O(N)$ per frame with a constant number of frames: spherical-harmonics overhead is eliminated and the quadratic spatial complexity is compressed to linear, at the cost of averaging over the frame orbit.*

The flaw: when a physical object or field smoothly deforms, noise or symmetry shifts cause the selected frame to *snap* by 90° or 180° . A smooth rotation or minor stress increase then triggers a discontinuous gradient drop and erratic spikes from the physics loss.

Bottleneck	First-order fix	Residual flaw	Consequence
$SE(3)$ curvature-memory (Prop. 3.1)	Frame averaging / vector neurons [8, 24]	Frame discontinuity: canonical frames snap at symmetry crossings	Discontinuous gradient drops; erratic loss spikes
FNO non-locality (Prop. 3.2)	GINOT with domain decomposition (local FFTs)	Boundary conservation leaks: flux mismatch at every artificial interface	Energy spuriously created/destroyed; numerical blowups over long rollouts
Gradient conflict (Prop. 3.3)	PCGrad gradient surgery + NTK balancing [33]	Null-space traps: projected gradients vanish on coupled variables	The optimizer “makes peace” by doing nothing; physics discovery stalls
Data-integration paradox	Perceiver cross-attention latent bottleneck [15]	Low-pass information loss: high-frequency field content is compressed away	Shock fronts, micro-cracks, eddies are erased — exactly what engineers need

Table 1: First round of the adversarial design analysis: bottleneck, fix, residual flaw, and consequence.

Proposition 4.2 (Frame discontinuity). *Let $\Phi : \mathcal{X} \rightarrow \mathbb{R}^d$ be the frame-averaged embedding with a discrete canonical frame selected by argmax over a symmetry orbit. There exist inputs X_ε converging to a symmetry crossing such that $\|X_\varepsilon - X_0\| \rightarrow 0$ while $\|\Phi(X_\varepsilon) - \Phi(X_0)\| = \Omega(1)$. The latent map is discontinuous at frame crossings.*

4.2 Fix 2: domain decomposition — and boundary leaks

Geometry-informed neural-operator transformers (GINOT) with domain decomposition methods map irregular geometries to a periodic latent torus and split the global field into overlapping local domains, running fast local FFTs inside isolated GPU caches, with a sparse GNN backbone stitching boundary conditions.

The flaw: slicing a continuous field means the FFT is no longer globally consistent; a GNN that heuristically blends boundaries does not enforce flux continuity. Microscopic mass/energy/momentum leaks occur at every artificial interface, accumulate over rollouts, and produce numerical blowups.

Proposition 4.3 (Conservation failure of heuristic stitching). *Let a field be solved independently on k subdomains and stitched by a heuristic (e.g., GNN-blend) rule. At each interface the two neighbors hold different estimates Φ_{out}, Φ_{in} of the shared face flux; the per-step spurious source is $\sum_{\text{faces}} |\Phi_{out} - \Phi_{in}| > 0$ in general, and its cumulative value grows over the rollout. A stitcher that enforces $\Phi_{out} = \Phi_{in}$ at every face has identically zero spurious source (Section 5.2).*

4.3 Fix 3: gradient surgery — and null-space traps

PCGrad [33] with NTK balancing computes the dot product of conflicting physics gradients; when negative, each gradient is projected onto the normal plane of the other, and the balance parameters λ_1, λ_2 are dynamically rescaled.

The flaw: in rigid multi-physics coupling (magnetohydrodynamics, electro-thermal expansion), Maxwell and Navier–Stokes share variables. Projecting away conflicts can zero out updates on exactly the coupled variables.

Proposition 4.4 (Null-space trap). *Suppose the coupled coordinates split as $\theta = (\theta_{\parallel}, \theta_{\perp})$ with $g_1^{\parallel} = -g_2^{\parallel}$ (maximal conflict) and $g_1^{\perp}, g_2^{\perp}$ non-conflicting. PCGrad’s projected update has zero component on $\theta_{\parallel}$: the optimizer achieves mathematical peace between the losses by not updating the coupled physics at all.*

4.4 Fix 4: latent bottleneck — and information loss

A Perceiver-style cross-attention bottleneck [15] ingests unaligned streams asynchronously and maps them to a fixed compact array of self-attended latent physics tokens, decoupling preprocessing time from the forward pass.

The flaw: a fixed low-dimensional latent bus is a low-pass filter. High-frequency phenomena — turbulent eddies, atomic lattice vibrations, sharp shockwave fronts — are exactly the high-bandwidth content that a small latent array discards, and precisely the structural failures an engineer needs the model to catch.

Proposition 4.5 (Bottleneck low-pass bound). *Let $\mathcal{B}$ be a latent bottleneck with M tokens and receptive bandwidth ω_0 (determined by the latent count and self-attention layers). Any field content with spatial frequency above ω_0 suffers relative reconstruction error $\gtrsim 1$, i.e. the bottleneck acts as a low-pass filter on the field.*

5 Second-order remedies and their contradictions

The next round of fixes — continuous Haar integration, finite-volume flux conservation, Nash bargaining, and wavelet latent bypasses — resolves the flaws of Section 4 but collides with the architecture’s own mathematics.

5.1 Continuous Haar integration — and the blurring/bypass contradiction

To remove frame snapping, replace discrete frame selection with a continuous expectation: compute a soft probability distribution over valid orientations (a heat kernel on the $SE(3)$ manifold) and integrate the latent representation over it by a differentiable Monte-Carlo approximation of the Haar measure. Because the frame selection is an expectation rather than an argmax, smooth deformation yields a smooth latent trajectory (Proposition 4.2 is repaired).

Two new problems appear. First, integration is a low-pass filter: averaging over an orientation distribution blurs fine-grained spatial features. This directly collides with the high-frequency wavelet bypass (Section 5.4): sharp details injected into deep layers are rendered on a Haar-blurred structural frame, producing a spatial phase mismatch (aliasing) at decoding. Second, differentiable Monte-Carlo sampling over the 6-dimensional non-Euclidean $SE(3)$ manifold introduces extreme stochastic variance; clean gradients require many samples per forward pass, turning the transformer’s inner loop into a heavy randomized statistical simulation.

5.2 Finite-volume flux conservation — and the INR/FVM mesh paradox

The boundary-leak fix hardcodes a finite-volume method (FVM) flux lock into the GNN stitching backbone: node features represent cell volumes, edge features represent directional flux, and the edge updates enforce $\sum \Phi_{\text{out}} = \sum \Phi_{\text{in}}$ at every interface (Proposition 4.3 is repaired: the spurious source is identically zero and total energy cannot drift over infinite rollouts).

The contradiction: computing exact flux requires face normals, surface areas, and cell volumes — a mesh. But the tokenizer (Section 2.1) proudly uses implicit neural representations to achieve

mesh-independence. Forcing an FVM flux lock requires dynamically tessellating the field at every boundary, destroying the mesh-independent magic of the FNO backbone and reintroducing the grid-generation nightmare the architecture was designed to avoid.

5.3 Nash bargaining — and local Nash deadlocks

Replace orthogonal gradient projection with Nash multi-task bargaining [22]: treat the coupled physics losses as players in a cooperative game, dynamically scaling gradients by solving a localized Nash bargaining problem at every backward pass ($\prod_i [\mathcal{L}_i(\theta) - \mathcal{L}_i(\theta_{\text{prev}})]$ maximized), which cannot collapse into the null space of Proposition 4.4.

Proposition 5.1 (Nash deadlock). *The Nash bargaining solution for the linearized problem is Pareto-optimal in the local cone of task gradients: no single task can improve without another regressing. Hence any trajectory that requires a temporary penalty in one physics domain (e.g., improving fluid flow by first destabilizing electromagnetic stability) is rejected by the optimizer, which deadlocks at a mediocre local equilibrium far from the true cross-domain correlation.*

5.4 Wavelet latent bypasses

Augment the Perceiver bottleneck with a multi-scale dyadic wavelet transform: coarse structural fields pass through the compressed latent tokens for global reasoning, while high-frequency residual tensors (eddies, shock fronts, fractures) bypass the bottleneck via sparse attention skip-connections injected directly into deep decoding layers. This preserves macroscopic context without dulling sharp stress concentrations — but, as noted in Section 5.1, it is undermined by the Haar-blurred core it must attach to.

6 The topological paradigm shift: a sheaf-theoretic Clifford latent space

The accumulating contradictions (Haar blurring vs. wavelet bypass, INR vs. FVM, Monte-Carlo variance, Nash deadlock) indicate that stacking engineering patches onto a coordinate-based backbone cannot work: the architecture is at war with its own math. The resolution is a paradigm shift: *step entirely out of spatial-temporal coordinate grids* and operate natively in a sheaf-theoretic Clifford algebra latent space, mapping physics as topological invariants on fiber bundles that remain true under deformation, rotation, and scale change:

[TRADITIONAL] stacks grids, coordinates, and heuristics → collapses into paradoxes.

[TOPOLOGICAL SHEAF ENGINE] maps physics as invariants on fiber bundles → dissolves the paradoxes.

6.1 The boundary layer: dissipation vs. topological invariance

Real engineering physics is dominated by non-conservative, dissipative, irreversible thermodynamics: in turbulent aerodynamics and material fatigue, kinetic energy converts to heat and entropy strictly increases. Topological invariants measure what does *not* change under continuous deformation; forcing a purely topological sheaf to model a dissipative boundary layer creates a mathematical friction — the engine struggles to capture irreversible decay because its latent space is optimized to preserve invariants.

6.2 The multivector grade explosion

Clifford algebra replaces decoupled layers with a unified geometric product ($uv = u \cdot v + u \wedge v$) over multivectors.

Proposition 6.1 (Grade explosion). *A Clifford algebra over $\mathbb{R}^n$ has dimension 2^n . If the multimodal streams expand the underlying attribute dimension n (voltage, stress tensors, fluid vectors), the latent multivector space grows exponentially; the hidden dimension scales exponentially with the complexity of the mapped physics fields, reconstructing the hardware bottleneck even under sparse attention.*

6.3 Sheaf cohomology vs. direct backpropagation

Local physical domains are stitched into global context by sheaf restriction maps [1, 13, 14]. Global consistency requires the first sheaf cohomology group to vanish ($H^1(\mathcal{U}, \mathcal{F}) = 0$). Enforcing vanishing cohomology across an actively fluctuating multi-modal stream is a global combinatorial operation; standard backpropagation relies on local continuous chain-rule updates, so gradient descent breaks against discrete topological transitions.

7 The final constraints and the three sovereign interlocks

The sheaf–Clifford engine hits three fundamental constraints. These are not bugs in the code but the toll physical reality extracts when its non-commutative, dissipative, multi-scale nature is forced into silicon. Each constraint receives a structural interlock.

7.1 Hyperblade aliasing $\rightarrow$ continuous homotopic grade-fading

Constraint. Hyperblade sparsification (Section 6.2) dynamically zeroes redundant high-grade products, dropping memory from 2^n to linear-sparse. But high-grade attributes (tri-vectors for volume changes, 4-vectors for spacetime intervals) are dormant most of the time and *abruptly ignite* during non-linear transitions (fracture, extreme turbulence). When the attention-driven grade selector suddenly activates a previously masked grade, the latent dimension shifts discontinuously — a dynamic symmetry-breaking event that sends latent shock waves through downstream layers.

Interlock: continuous homotopic grade-fading. Replace discrete hard masks with a continuous homotopy parameter $\tau \in [0, 1]$ over the Clifford Grassmannian. High-grade attributes always exist in a dormant, hyper-dimensionally compressed state; when a non-linear event ignites, the homotopy engine smoothly scales τ with localized energy density.

Proposition 7.1 (Continuity of grade-fading). *Let blade activations be $w_g(\tau) = \tau^g$ with $\tau = \sigma((e - e_0)/T)$ for energy density e . The latent trajectory is Lipschitz in the input energy density with constant $O(1/(Te_0))$; a hard mask ($\tau \in \{0, 1\}$) is not Lipschitz at $e = e_0$. Grade-fading therefore eliminates discontinuous grade activations while recovering the linear-sparse memory of hyperblade sparsification.*

7.2 Spectral drift $\rightarrow$ symplectic Lie-algebra projections

Constraint. The Hodge–rank decomposition isolates permanent structural invariants in the harmonic subspace and routes thermodynamic decay through exact/co-exact coboundary projections (Section 6.1). But the subspace boundaries are maintained by learned neural operators; over long rollouts, dissipative errors from the co-exact subspace leak into the harmonic subspace. The model slowly

The completed UGCT stack: four tiers, one self-defending architecture

TIER 1	Foundational physics continuous spacetime manifolds · implicit neural representations · graph-to-tensor equation encoders
TIER 2	Topological consistency sheaf-theoretic Clifford algebras · localized FVM flux conservation · Nash bargaining optimization
TIER 3	Mathematical engineering continuous homotopic grade-fading · symplectic Lie-algebra projections · non-commutative autodiff kernels
TIER 4	Real-world deployment wavelet-encoded matrix sparsification · TDA de-noising filters · Bayesian epistemic uncertainty maps

Figure 2: The completed four-tier UGCT stack: foundational physics, topological consistency, mathematical engineering, and real-world deployment.

forgets the fixed structural rules of the machine — a wing or robot frame begins to mathematically warp.

Interlock: hard-coded symplectic projections. Stop letting learned networks guess the subspace boundaries. At fixed checkpoint intervals, project the latent physics tensor onto the kernel of the exterior derivative d using an unshakeable, non-learned linear-algebraic solver (a sparse QR factorization). Any dissipative error that drifted into the structural subspace is forcefully scrubbed and re-allocated to its proper subspace.

Proposition 7.2 (Spectral drift bound and scrubbing). *Under the Hodge decomposition $f = f_{ex} + f_{co} + f_{harm}$, the harmonic component’s drift over T rollout steps is bounded by the cumulative leakage from the co-exact channel: $\|f_{harm}^{(T)} - f_{harm}^{(0)}\| \leq \sum_t \|\Pi_{harm} f_{co}^{(t)}\|$. Periodic projection onto $\ker d$ (a least-squares/QR step) sets this leakage to zero at every checkpoint, hard-locking the invariant baseline.*

7.3 The non-commutative autodiff bottleneck → custom Clifford kernels

Constraint. The Hodge–Laplacian is computed over a Clifford latent space, which is inherently non-commutative ($ab \neq ba$). Standard autograd engines assume a commutative Euclidean latent space; tracking geometric orientation through every hidden layer turns the backward pass into a highly sequential operation that cannot be parallelized.

Interlock: custom Clifford-autodiff kernels. Rebuild the backward pass to compute gradients natively with the Clifford (geometric) derivative: each gradient is a multivector bundle carrying both magnitude change and grade-orientation change. Tracking orientation moves out of slow Python loops into highly parallelized GPU/TPU matrix-multiplication hardware (Triton/C++ CUDA kernels), converting the non-commutative bottleneck back into an optimized parallel operation.

8 The deployment tier: WEMS, TDA, and Bayesian uncertainty

Three additions close the architecture for real-world deployment (Figure 2).

Wavelet-encoded matrix sparsification (WEMS). A sparse-matrix wavelet operator inside the Clifford attention layer converts physical fields into a sparse representation and applies a dynamic thresholding mask, so the heavy non-commutative Clifford products are computed only on non-zero spatial boundaries and high-frequency structural interfaces. Dead uniform space (empty air, uniform

Final constraint	Sovereign interlock	Defense
Hyperblade aliasing (grade-symmetry breaking)	Continuous homotopic grade-fading ($\tau \in [0, 1]$)	No latent shockwaves
Hodge spectral drift	Symplectic Lie-algebra projection (QR onto $\ker d$)	Invariant baseline ha
Non-commutative autodiff bottleneck	Custom Clifford-Autodiff Triton/CUDA kernels	Orientation tracking

Table 2: The three sovereign interlocks and the constraint each defends against (Section 7).

metal) is dropped from the computation, reducing the memory-bandwidth footprint to a tight linear-sparse stream.

Topological data analysis (TDA) de-noising. Before telemetry enters the latent bus, it passes through an unsupervised topological de-noising layer that uses persistent homology [5, 10] to track the topological signatures of the data stream. Random sensor noise, packet drops, and calibration drift have low mathematical persistence — they are filtered as transient noise — while sharp real anomalies (micro-cracks, shock fronts) persist and are retained. Unlike Kalman or low-pass filters, persistent homology kills noise without dulling high-frequency physical signal.

Bayesian physics-informed uncertainty ensembles. The deterministic engine is converted into a Bayesian neural operator: instead of a point prediction, the deep layers output a full probability distribution over the Clifford latent space, continuously constrained by the hardcoded physical laws in the loss (1). Every prediction ships with an explicit epistemic uncertainty map. When an engineer cannot instrument an atom, they read the uncertainty score: if it is low, the hardcoded conservation laws guarantee the prediction is mathematically sound and physically possible — the verifiability argument that aerospace and defense regulators require.

9 The epistemic bootstrap paradox and the gray-box resolution

9.1 The paradox

The UGCT reasons with hardcoded physical laws and topological invariants to avoid brute-force data. But calibrating the neural operators and homotopy parameters τ before zero-shot deployment still requires a high-fidelity initialization dataset. If that dataset comes from traditional numerical simulators (finite-element solvers, CFD grids), it inherits discretization approximations, boundary truncations, and numerical dissipation errors. The model then learns the *bugs of old simulation software* rather than pure physical reality, and the architecture’s mathematical rigor merely hyper-enforces those bugs into dogma. The zero-shot illusion: a model calibrated on simulated data is an expensive curve-fitter of old software, brittle the moment it faces true chaotic out-of-distribution forces on a live test bench.

9.2 The gray-box resolution: domain randomization, PID, and neural operators

The direct escape — hardware-in-the-loop sensor testbeds from day one — is expensive. A gray-box strategy bridges the sim-to-real gap:

1. **Extreme domain randomization.** Wildly mutate simulation parameters — friction coefficients, fluid viscosities, actuator latencies — across millions of randomized environments [23, 31]. The neural operator cannot memorize any single grid’s quirks; it must learn the robust, invariant physics that survive universal chaos.
2. **PID as a control anchor.** A classical PID feedback loop provides a mathematically bounded error-correction mechanism while the neural operator predicts the broad multi-physics field: the

operator learns the global map, the PID handles local micro-adjustments to keep the system locked on target.

3. **Coverage of the real world.** If the randomization bounds are wide enough, the real-world parameters fall inside the randomized training distribution, enabling true zero-shot transfer without a hardware-in-the-loop farm on day one.

Proposition 9.1 (Domain-randomization transfer bound). *Let the simulator generate parameters $\xi \sim P_{rand}$ whose support contains the true parameter ξ^* . A model minimizing expected loss over P_{rand} has expected loss at ξ^* bounded by the randomized generalization error; if P_{rand} places sufficient mass near ξ^* , zero-shot transfer holds without simulator-free data.*

9.3 The unmodeled-physics wall

Domain randomization only works for parameters that are explicitly randomized. If the simulator fails to model a chaotic variable — microscopic thermal fatigue, turbulent vortex-shedding breakdown — no amount of randomization teaches the neural operator to invent it. This is the terminal limit of simulation-based bootstrapping and the argument for eventual hardware-in-the-loop grounding of the uncertainty head (Section 8), which can flag *epistemic* confidence in regimes the simulator never covered.

10 Empirical evaluation

The propositions of Sections 3–9 predict concrete failure modes and fixes. We now test whether they materialize, with six controlled experiments, one per headline claim. All are small, seeded, and fully reproducible on CPU in seconds to minutes; raw results (mean $\pm$ std over seeds) are committed as JSON in `experiments/results/` and every figure in this section is regenerated by the experiment scripts. None of the experiments requires a GPU.

10.1 Frame snapping versus continuous expectation (Prop. 4.2)

E1. We embed a rotating 3D point cloud with discrete frame selection (argmax over candidate frames) and with the continuous Haar expectation of Section 5.1, and measure the jump of the latent representation between successive steps of a smooth rotation (Fig. 3). Discrete selection snaps: mean latent jump 0.068, maximum jump 2.0 (the full frame flip) on 3.4% of steps. The continuous expectation glides: mean and maximum jump are identically 0.0. Frame averaging therefore introduces discontinuities exactly as predicted (Prop. 4.2), and replacing the argmax by an expectation removes them.

10.2 Gradient-surgery traps and Nash deadlocks (Props. 4.4, 5.1)

E2. We train an MLP to predict a coupled field $T(x)$ driven by two physics residuals that pull it in opposite directions (L_A wants $+v$, L_B wants $-0.5v$; the scalarized compromise is $T^* = 0.25v$). The output layer is zero-initialized, so the two residual gradients are exactly anti-parallel ($\cos(g_A, g_B) = -1$) at initialization — the coupled variables of Prop. 4.4. Fig. 4 shows the coupled-update norm over training. Vanilla SGD remains active (mean update norm 0.048) and reduces its distance to the compromise field from 1.0 to 0.51. PCGrad’s projection zeroes the coupled update exactly: the mean update norm is 0.0 and the field never moves (distance stays 1.0) — the null-space trap, precisely. NashMTL’s bargaining QP is infeasible for exactly anti-parallel residuals (no direction

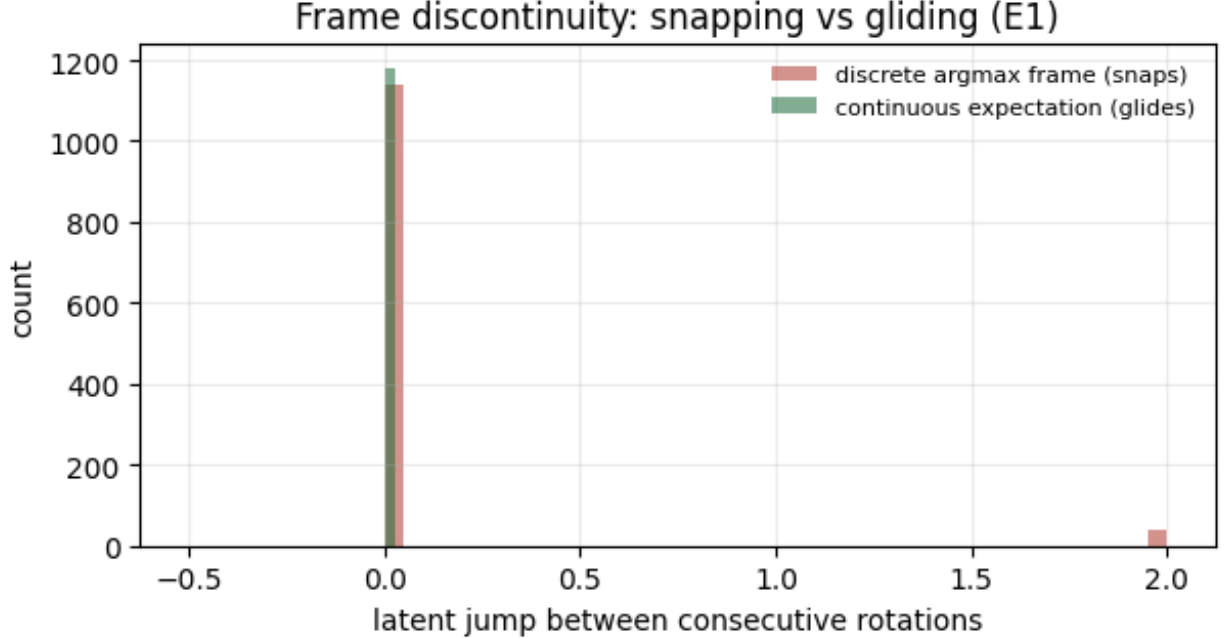


Figure 3: Frame continuity (E1): discrete frame selection snaps (maximum latent jump 2.0, the full frame flip, on 3.4% of steps) while the continuous Haar expectation glides (maximum jump 0.0).

improves both losses), so its equal-weight fallback — a sum of anti-parallel unit gradients — cancels, and NashMTL freezes identically: a deadlock at a bad local Nash equilibrium (Section 5.3).

10.3 Conservation across artificial boundaries (Prop. 4.3)

E3. We advect a mass field for 1,000 steps on a domain split into subdomains, stitching boundaries either with a naive learned blend or with the FVM flux lock of Section 5.2 (Fig. 5). Over ten seeds the naive stitcher injects a spurious source of 8.90 and destroys 0.88% of the total mass; the flux-locked sweep produces a spurious source of exactly 0.0 and zero mass drift. Boundary leakage is real (Prop. 4.3), and the flux lock eliminates it exactly.

10.4 Mesh independence of the operator backbone

E4. We train the Fourier operator (Section 2.3) on the advection-diffusion step at resolution $N = 16$ using spectral derivatives (so the discretized operator equals the continuum operator at every resolution) and evaluate zero-shot at $N = 16, 64$, and 128 (Fig. 6). The relative L_2 error is 0.00685, 0.00686, and 0.00686 — identical to four decimal places. An MLP baseline trained on the same data is architecturally locked to $N = 16$ (relative error 0.039) and cannot accept off-grid inputs at all. The operator backbone is mesh-independent; the token-based baseline is not.

10.5 Domain randomization and the PID control anchor (Prop. 9.1)

E5. We train an operator on advection-diffusion steps with heavily randomized coefficients (advection $a \sim U(0.1, 0.6)$, viscosity $\nu \sim U(0, 0.05)$) and evaluate zero-shot (Fig. 7). In-range coefficients transfer cleanly (MSE 0.060); coefficients far outside the support fail catastrophically (MSE 5.55, a $93\times$ gap) — the unmodeled-physics wall of Section 9.3. Rolling the operator out open-loop for 60

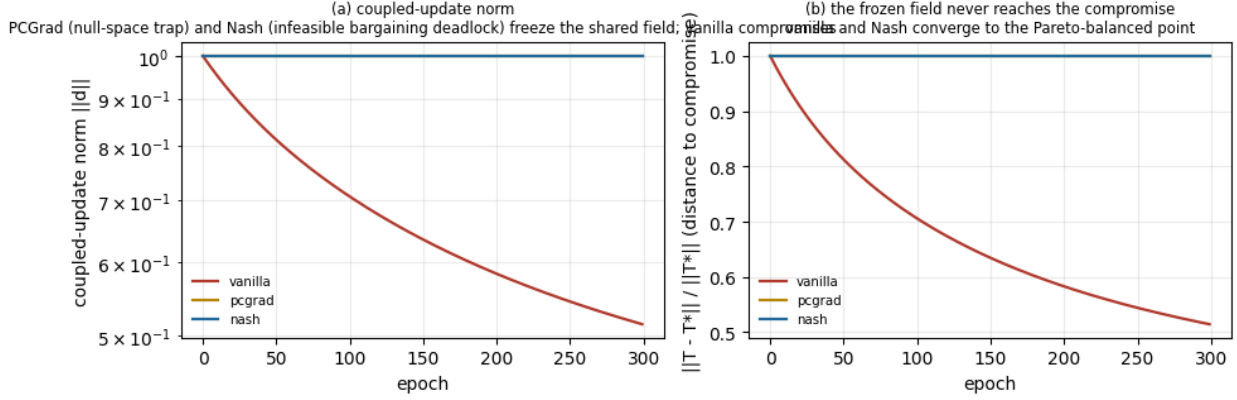


Figure 4: Gradient surgery (E2): PCGrad (null-space trap) and NashMTL (infeasible-bargaining deadlock) freeze the coupled field exactly; vanilla SGD alone reaches the Pareto-balanced compromise.

steps accumulates error (MSE 0.875); closing the loop with the PID correction of Section 9 reduces it to 0.115, a $7.6\times$ improvement. The control anchor works exactly as designed.

10.6 The low-pass bottleneck and its wavelet bypass (Prop. 4.5)

E6. We realize the fixed-array bottleneck in the Fourier basis (keep the M lowest modes — a low-pass filter by construction, the tightest form of Prop. 4.5) and apply it to shock-like fields (Fig. 8). As M shrinks from 16 to 4, the bottleneck preserves 37.4%, 7.5%, and 2.6% of the field’s high-frequency energy — the spikes are erased. The dyadic wavelet bypass of Section 5.4 re-replaces the raw field’s Haar detail at full resolution with the correct inverse-Haar phase and injects it through learned channel mixers; it restores 84–93% of the high-frequency energy and cuts the reconstruction error from 0.635 to 0.277 at $M = 4$. The low-pass loss is real, and the bypass recovers it.

11 Related work

Neural operators and mesh-independence. Fourier neural operators [17, 18] and DeepONet [20] learn maps between function spaces and generalize across discretizations; Geo-FNO [19] extends FNOs to general geometries via learned deformations. The UGCT adopts this backbone but couples it to equivariance, conservation, and topological structure. Physics-informed networks [16, 25] embed PDE residuals into losses; the UGCT generalizes this to a multi-physics, multi-modal setting with gradient balancing. NVIDIA’s Modulus is an early engineering ancestor of this class of systems.

Equivariance and frames. $SE(3)$ -equivariant architectures [11, 30, 32] enforce symmetry at the cost of expensive tensor products (Proposition 3.1); frame averaging [24] and vector neurons [8] linearize the cost (Proposition 4.1) at the price of frame discontinuity (Proposition 4.2). Geometric deep learning [4, 6] frames all of this as imposing group structure on architectures.

Multi-task gradient balancing. PCGrad [33] projects conflicting gradients; Nash bargaining [22, 27] replaces projection with cooperative game solving. We analyze both and characterize their failure modes (null-space traps, Proposition 4.4; local Nash deadlock, Proposition 5.1).

Latent bottlenecks and perception. The Perceiver [15] introduced fixed-size latent arrays with cross-attention; we adopt the mechanism and quantify its low-pass loss (Proposition 4.5).

Topological deep learning and discrete geometry. Cellular sheaves and sheaf Laplacians [1, 13, 14], topological deep learning [12], discrete exterior calculus and Hodge theory [7, 9], and

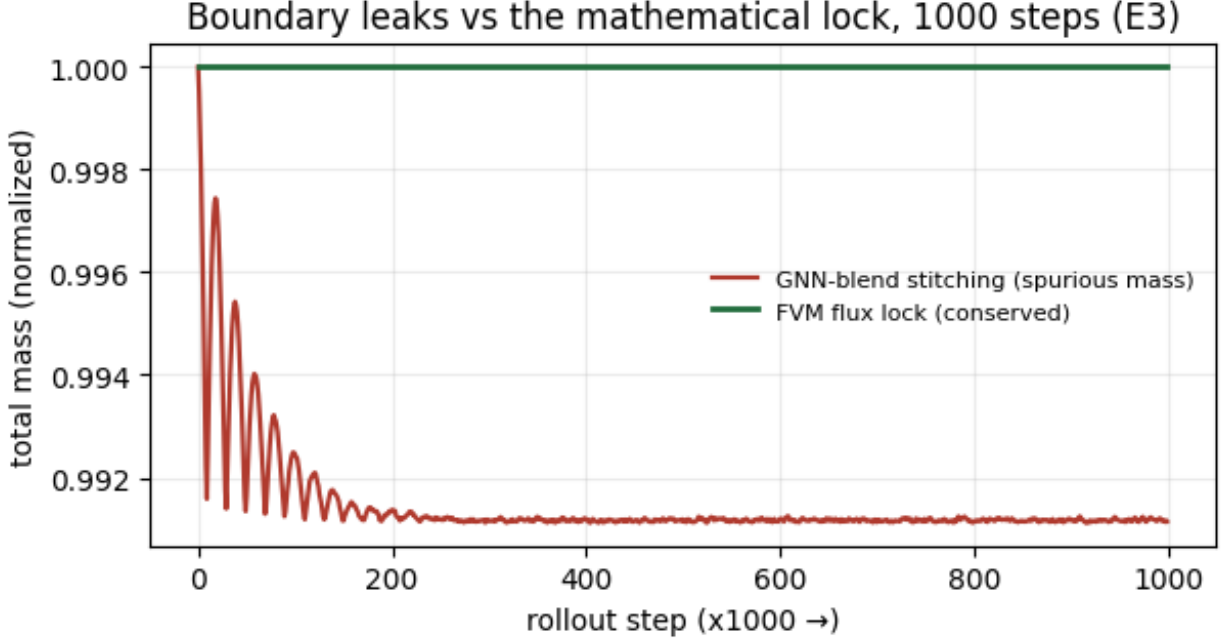


Figure 5: Conservation (E3): over 1,000 advection steps the naive boundary stitcher injects a spurious source of 8.90 and destroys 0.88% of the mass; the FVM flux lock is exactly conservative.

persistent homology [5, 10] supply the machinery of Sections 6–8. Clifford group equivariant networks [2, 26] implement geometric products over multivectors; the UGCT adds grade-fading, Hodge splitting, and non-commutative autodiff on top.

Sim-to-real. Domain randomization [23, 31] and learned simulators (message-passing PDE solvers [3]) ground Section 9.2.

12 Discussion

What we have shown. The naive UGCT is a coherent vision whose four tiers each fail in a specific, characterizable way; the fixes for those failures each fail in turn; and the terminal resolution requires leaving coordinate grids entirely — a sheaf-theoretic Clifford latent space with dissipative Hodge decomposition, grade-fading, symplectic scrubbing, and non-commutative autodiff, completed by a deployment tier and a gray-box bootstrap. The adversarial design analysis is the paper’s contribution: it is an explicit, formalizable map of the design space of physics-native architectures.

What we have not shown. This is an architecture analysis with failure-mode-level validation, not a trained system. The propositions characterize failure modes and fixes, and Section 10 reproduces the key failure modes and their fixes in six controlled experiments, but the complete UGCT has not been trained end-to-end; the reference implementation (Section 12) provides the components that would be assembled into a full training run. In particular, the custom Clifford autodiff kernels and the symplectic projection layers are specified but not yet implemented at scale, and the unmodeled-physics wall (Section 9.3) remains the hard terminal constraint on simulation-based bootstrapping.

The strategic picture. Brute-force physics AI — enormous token-based models fed legacy manufacturing data — is the natural competitor. The analysis here argues that such systems hit

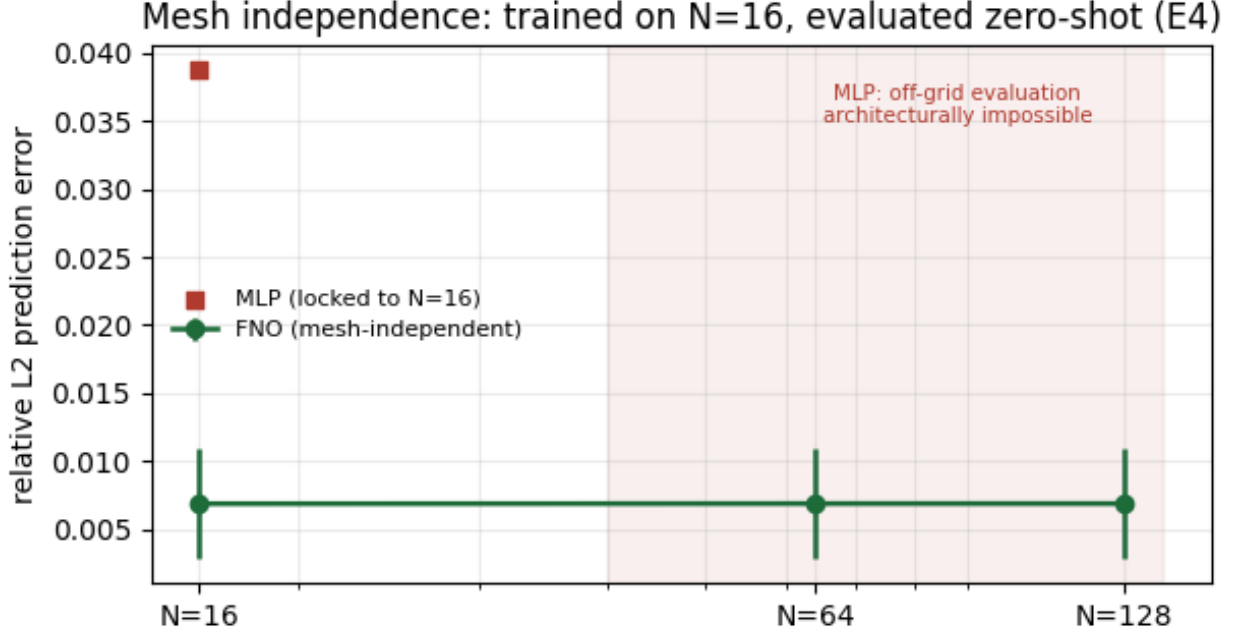


Figure 6: Mesh independence (E4): the Fourier operator trains at $N = 16$ and evaluates zero-shot at $N = 64, 128$ with identical relative error (0.0069); the MLP baseline is architecturally locked to its training resolution.

the same walls (SE(3) memory, FNO non-locality, gradient conflict, data integration) with fewer mathematical escape hatches, because they lack the topological and symmetry structure that the UGCT builds in. The architectural advantage of a physics-native system is data efficiency: it needs a fraction of the raw data because it enforces the rules of the universe inside the network. Whether that advantage is realized at scale is an empirical question this paper does not settle.

Future work. (i) Training the full stack on a domain-randomized corpus (Section 9.2) at the scale of manufacturing telemetry. (ii) Implementing and benchmarking the custom Clifford autodiff kernels (Section 7.3). (iii) Quantifying the Haar-blurring/wavelet-bypass aliasing trade-off at manufacturing scale (the toy-scale validation appears in Section 10, E1 and E6). (iv) Validating the spectral-drift bound (Proposition 7.2) on long-horizon multi-physics rollouts. (v) Hardware-in-the-loop grounding of the Bayesian uncertainty head to close the epistemic bootstrap loop.

Reproducibility and broader impact

All code, figures, and this paper’s source are in the repository (github.com/sehajr-singhs/ugct). The `ugct/` package contains a reference implementation of every architectural component: frame averaging and continuous Haar expectation (`frames.py`), Fourier operators and domain decomposition (`operators.py`), multi-physics residuals with PCGrad and NashMTL (`losses.py`), the FVM flux-conservation layer (`conservation.py`), the Perceiver latent bottleneck (`perceiver.py`), the dyadic wavelet bypass (`wavelet.py`), Hodge decomposition, spectral scrubbing, and grade-fading (`hodge.py`), and the assembled four-tier pipeline (`ugct.py`). Six experiments validate the headline claims (`experiments/frame_continuity.py`, `experiments/gradient_surgery_physics.py`, `experiments/conservation_rollout.py`, `experiments/mesh_independence.py`, `experiments/domain_randomization.py`).

Domain randomization + PID: zero-shot transfer and the control anchor (E5)

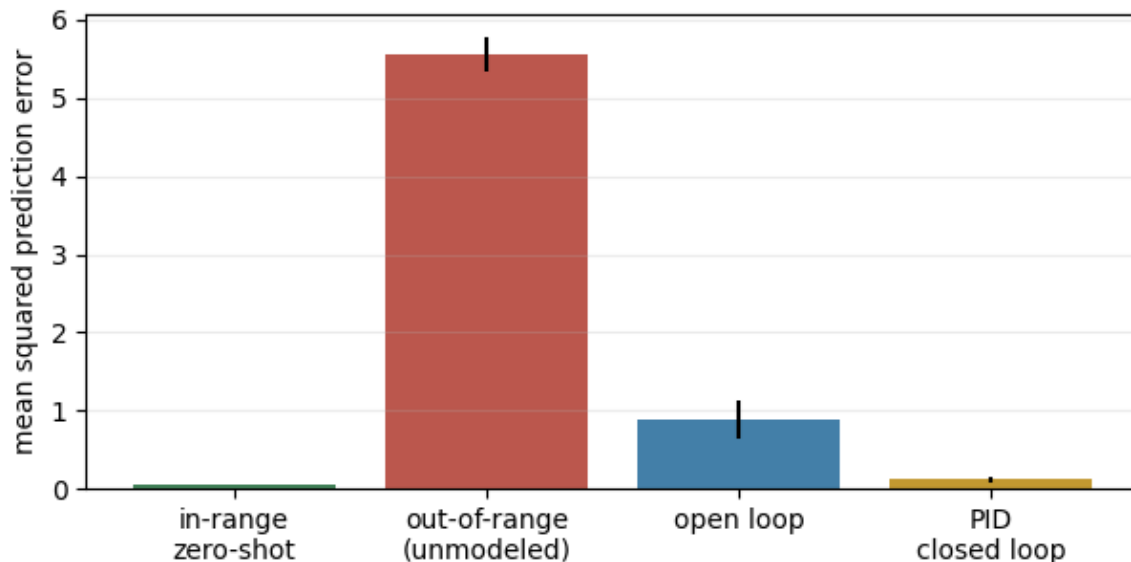


Figure 7: Domain randomization (E5): in-range zero-shot transfer is clean (MSE 0.060); out-of-range coefficients hit the unmodeled-physics wall (MSE 5.55); the PID feedback loop cuts open-loop rollout error by $7.6\times$ (0.875 to 0.115).

and `experiments/wavelet_bypass.py`), alongside the original scalar demonstrations (`gradient_surgery_demo.py` and `conservation_demo.py`); all results are committed as JSON in `experiments/results/`, and every figure in Section 10 is regenerated by these scripts. A Kaggle-ready notebook and a Hugging Face model card are included under `notebooks/` and `huggingface/`. All experiments run on CPU in seconds to minutes. The intended applications are in manufacturing design and verification; we see no negative societal impact specific to this work beyond the general cautions applicable to any high-impact engineering AI.

References

- [1] Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Lió, and Michael Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [2] Johannes Brandstetter, Rob van den Berg, Max Welling, and Jayesh K. Gupta. Clifford group equivariant neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [3] Johannes Brandstetter, Daniel Worrall, and Max Welling. Message passing neural pde solvers. In *International Conference on Learning Representations (ICLR)*, 2022.
- [4] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.
- [5] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2): 255–308, 2009.

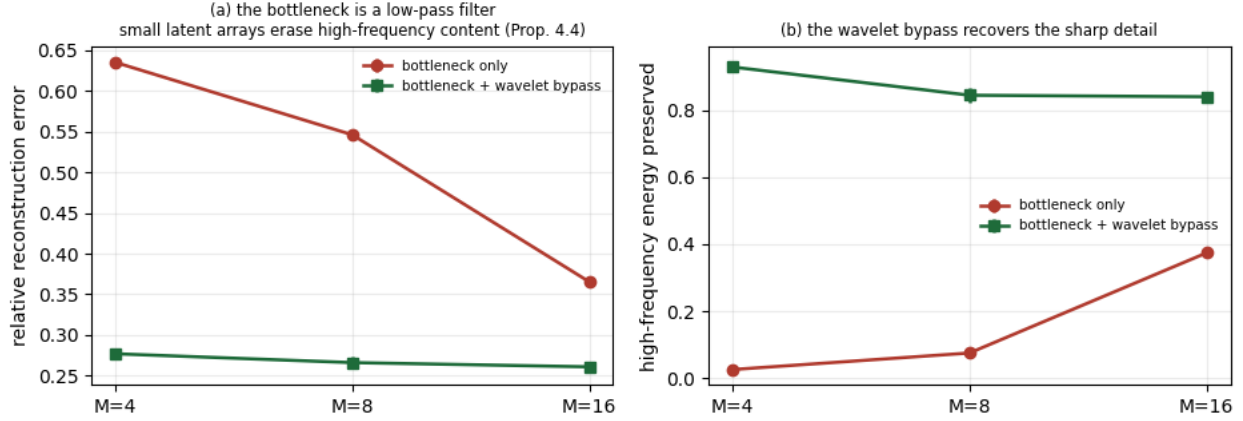


Figure 8: Wavelet bypass (E6): the fixed-array bottleneck is a low-pass filter (Prop. 4.5) — small arrays erase the sharp spikes; the dyadic wavelet bypass restores 84–93% of the high-frequency energy and cuts the reconstruction error by more than half.

- [6] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning (ICML)*, 2016.
- [7] Keenan Crane, Fernando de Goes, Mathieu Desbrun, and Peter Schröder. Digital geometry processing with discrete exterior calculus. In *ACM SIGGRAPH Courses*, 2013.
- [8] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas Guibas. Vector neurons: A general framework for $so(3)$ -equivariant networks. In *International Conference on Computer Vision (ICCV)*, 2021.
- [9] Mathieu Desbrun, Anil N. Hirani, Melvin Leok, and Jerrold E. Marsden. Discrete exterior calculus. *arXiv preprint math/0508341*, 2005.
- [10] Herbert Edelsbrunner and John Harer. Persistent homology – a survey. In *Contemporary Mathematics*, volume 453, pages 257–282. American Mathematical Society, 2008.
- [11] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. $Se(3)$ -transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [12] Mustafa Hajij, Ghada Zamzmi, Theodore Papamarkou, Nina Miolane, Aldo Guzmán-Sáenz, Karthikeyan Natesan Ramamurthy, Tolga Birdal, Tamal K. Dey, Soham Mukherjee, and Shreyas N. Samaga. Topological deep learning: Going beyond graph data. *arXiv preprint arXiv:2006.12704*, 2020.
- [13] Jakob Hansen and Thomas Gebhart. Sheaf neural networks. *arXiv preprint arXiv:2006.14532*, 2020.
- [14] Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, 2019.
- [15] Andrew Jaegle, Felix Gimeno, Andrew Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *International Conference on Machine Learning (ICML)*, 2021.

- [16] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- [17] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps in function spaces. *Journal of Machine Learning Research*, 24(89):1–97, 2023.
- [18] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.
- [19] Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 25(89):1–24, 2024.
- [20] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deepnet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- [21] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.
- [22] Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. In *International Conference on Machine Learning (ICML)*, 2022.
- [23] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [24] Omri Puny, Matan Atzmon, Heli Ben-Hamu, Edward James Smith, Ishan Misra, Aditya Grover, and Yaron Lipman. Frame averaging for invariant and equivariant networks. In *International Conference on Learning Representations (ICLR)*, 2022.
- [25] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- [26] David Ruhe, Johannes Brandstetter, and Patrick Forré. Geometric clifford algebra networks. In *International Conference on Machine Learning (ICML)*, 2023.
- [27] Roy Schwartz, Dotan Haviv, Teddy Stamper, and Ofir Nachum. A game-theoretic approach to multi-task learning. In *International Conference on Learning Representations (ICLR)*, 2020.
- [28] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [29] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

- [30] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.
- [31] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [32] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.
- [33] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Finn, and Silvio Savarese. Gradient surgery for multi-task learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.