

The Geometry of Abstraction: Continual Learning via Recursive Quotienting

Xin Li,¹

Department of Computer Science
University at Albany
xli48@albany.edu

Abstract—Continual learning systems operating in fixed-dimensional spaces face a fundamental geometric barrier: the flat manifold problem. When experience is represented as a linear trajectory in Euclidean space, the geodesic distance between temporal events grows linearly with time, forcing the required covering number (capacity) to diverge. In fixed-dimensional hardware, the volume expansion inevitably forces trajectory overlap, manifesting as catastrophic interference. In this work, we propose a geometric resolution to this paradox based on *Recursive Metric Contraction*. We formalize abstraction not as symbolic grouping, but as a topological deformation: a quotient map that collapses the metric tensor within validated temporal neighborhoods, effectively driving the diameter of local sub-manifolds to zero. We substantiate our framework with three rigorous results. First, the *Bounded Capacity Theorem* establishes that recursive quotient maps allow the embedding of arbitrarily long trajectories into bounded representational volumes, trading linear metric growth for logarithmic topological depth. Second, the *Topological Collapse Separability Theorem*, derived via Urysohn’s Lemma, proves that recursive quotienting renders non-linearly separable temporal sequences linearly separable in the limit, bypassing the need for infinite-dimensional kernel projections. We introduce the *Parity-Partitioned Stability Theorem*, which solves the catastrophic forgetting problem by proving that if the state space is homologically partitioned into orthogonal flow (odd) and scaffold (even) manifolds, the metric deformations of active learning do not disturb the stability of stored memories. Finally, we further establish *Correctness under Abstraction*, proving that semantic discriminability is a topological invariant that descends continuously through the condensation hierarchy, ensuring that the collapsed representation retains the exact decision boundaries of the original high-dimensional space. Our analysis reveals that tokens in neural architectures are physically realizable as wormholes, regions of extreme positive curvature that bridge distant points in the temporal manifold. Our framework demonstrates that unbounded inference is achievable in fixed dimensions if and only if the system actively folds the manifold of experience, replacing the linear search for past events with geodesic shortcuts through a recursive quotient topology.

I. INTRODUCTION

Continual learning systems operate under a fundamental geometric tension: experience accumulates without bound, yet the volume of the representational state space is fixed. Classical learning theory formalizes this limitation through capacity measures such as VC dimension and covering numbers: in

This work was partially supported by NSF IIS-2401748 and BCS-2401398. The author has used Gemini 3.0 and ChatGPT 5.2 models to assist in the development of theoretical ideas, proofs of lemmas/theorems, and visual illustrations presented in this paper.

a fixed-dimensional manifold, the number of distinguishable regions that can be stably maintained is strictly bounded [1], [2], [3]. As the length of an input stream grows, the trajectory of experience elongates, eventually exceeding the covering capacity of the representational manifold. Without intervention, the overload forces trajectory overlap, manifesting as catastrophic interference [4], [5].

Prevailing approaches mitigate the interference problem through geometric expansion. Strategies such as kernel methods, overparameterized neural networks, and generative replay effectively increase capacity by embedding data into higher-dimensional spaces or expanding the support of the learned distribution [2], [6]. While successful in finite regimes, these strategies rely on a scaling law in which representational resources (width or memory) must grow linearly with experience ($d \propto L$). In open-ended environments where $L \rightarrow \infty$, such expansion is unsustainable. These methods treat continual learning as a packing problem: making the container larger to fit more points. In contrast, biological and cognitive systems appear to rely on a different organizational principle [7]. Rather than continuously expanding representational dimension, they organize experience hierarchically, compressing recurring temporal structure into stable, invariant units [8], [9]. The hierarchical organization of neural systems mirroring the nested structure in nature suggests that scalability arises not from expanding geometry, but from transforming topology [10].

In this work, we formalize our geometric perspective by modeling the input stream as a trajectory on a *temporal manifold* \mathcal{M} , a metric space whose diameter grows linearly with time, consistent with the manifold hypothesis for sequential data [11], [12]. We propose that abstraction corresponds to an active deformation of the manifold, in which validated sub-manifolds are collapsed via *quotient maps* [13]. This process, which we term *recursive metric contraction*, reduces the effective diameter of the representation without increasing its ambient dimension. Our analysis reframes memory and abstraction as operations on the metric tensor. A condensed abstraction is not merely a symbolic grouping, but a *metric singularity*: a region of the manifold where internal geodesic distances are driven to zero, identifying a complex temporal trajectory with a single point in a quotient space. By iteratively applying such contractions, the system constructs a hierarchy of quotient manifolds. While the raw input manifold has diverging volume, the induced quotient manifolds remain compact with bounded covering numbers [14]. Inference over such a hierarchy

corresponds to traversing a topologically folded space rather than scanning a linear history. Unbounded experience is accommodated by logarithmic growth in hierarchical depth rather than linear growth in representational width.

We make our geometric intuition precise through four fundamental theorems:

- 1) The **Bounded Capacity Theorem**, which proves that if the covering number of each quotient manifold is bounded, the effective capacity demand of the system remains uniformly bounded by $O(1)$, independent of the total length of the input stream.
- 2) The **Topological Collapse Separability Theorem**, which leverages Urysohn’s Lemma to demonstrate that non-linearly separable temporal sequences can be rendered linearly separable via quotient topology, rendering the high-dimensional Kernel Trick unnecessary.
- 3) The **Parity-Partitioned Stability Theorem**, which establishes that by partitioning the state space into orthogonal homological subspaces (Flow/Odd vs. Scaffold/Even), the system can achieve infinite plasticity without catastrophic interference.
- 4) **Correctness under Abstraction:** Theorem 4 establishes that semantic discriminability is a topological invariant of the condensation hierarchy. We prove that whenever the quotient map collapses states strictly along the fibers (level sets) of a base Urysohn separator, the separating function continuously descends to all subsequent levels X_k . This guarantees that disjoint concepts A_k and B_k remain topologically separable regardless of the hierarchy depth, ensuring that the system achieves massive complexity reduction with zero loss in decision correctness.

II. RELATED WORK

Capacity limits in fixed-dimensional representations. Classical statistical learning theory establishes sharp limits on the capacity of fixed-dimensional hypothesis spaces. Cover’s seminal result shows that the probability of linear separability of random patterns in \mathbb{R}^d collapses once the number of points exceeds $O(d)$ [1]. This phenomenon is formalized more generally through the Vapnik-Chervonenkis (VC) dimension and related notions of covering numbers, which bound the number of distinguishable regions that can be stably represented in a given space [2], [3]. From a geometric perspective, these results imply that maintaining separability over an ever-lengthening trajectory in a flat manifold inevitably exhausts capacity, independent of the learning algorithm.

Catastrophic interference as geometric overload. The breakdown of performance under sequential learning has long been recognized as catastrophic interference or catastrophic forgetting [4], [5]. In connectionist models, gradient-based updates applied to a shared representational space cause new trajectories to overwrite previously learned structure unless explicit safeguards are imposed [15]. Contemporary continual learning methods address such failure mode through rehearsal buffers [16], parameter regularization schemes such as elastic weight consolidation [17], or architectural growth. While effective empirically, these approaches implicitly assume a

flat underlying geometry and treat interference as a resource-allocation problem rather than as a consequence of unbounded metric expansion.

Hierarchical abstraction and depth as a control mechanism.

A complementary body of work emphasizes hierarchical organization as a means of controlling complexity. Simon argued that adaptive systems exploit hierarchical decomposition to remain tractable in complex environments [10]. In reinforcement learning, hierarchical methods such as feudal RL [18] and MAXQ [19] reduce planning and learning complexity by introducing temporally extended abstractions. From a representational standpoint, theoretical analyses have shown that depth enables efficient reuse of intermediate structure, allowing expressive power to grow exponentially with only linear increases in parameters [20]. These results suggest that abstraction depth, rather than representational width, plays a central role in scalability.

Geometric and topological perspectives on representation.

Recent work has increasingly framed learning and abstraction through the lens of geometry. The manifold hypothesis posits that high-dimensional data concentrate near low-dimensional structures, motivating geometric methods for representation learning [11], [12]. Spectral and diffusion-based approaches further connect the geometry of data manifolds to compact representations via eigenstructure and covering properties [14], [21]. However, most existing methods treat the underlying manifold as fixed and focus on embedding or parametrizing it, rather than actively deforming its metric or topology.

Positioning of the present work. The present work isolates a distinct mechanism for AGI scalability: *recursive metric contraction*. Rather than expanding representational dimension (Kernel methods), replaying past data (Generative Replay), or learning a static embedding (Pre-training), we formalize abstraction as a sequence of quotient operations that collapse validated sub-manifolds of experience (Fig. 3). We substantiate the present framework with a quadpartite theoretical foundation:

- **Scalability:** Our *Bounded Capacity Theorem* shows that when quotient contractions bound the covering number at each hierarchical level, the effective capacity demand of a fixed-dimensional system remains uniformly bounded, even as the input stream grows without limit.
- **Separability:** Our *Topological Collapse Separability Theorem* positions metric quotienting as the geometric dual to the Kernel Trick. We demonstrate that non-linear problems can be solved not by adding dimensions, but by collapsing the metric until the problem becomes topologically trivial.
- **Stability:** Our *Parity-Partitioned Stability Theorem* reframes the stability-plasticity dilemma. We show that by segregating the manifold into orthogonal flow (\mathcal{M}_{odd}) and scaffold ($\mathcal{M}_{\text{even}}$) subspaces, the system can support infinite plasticity without catastrophic interference.
- **Correctness under Abstraction:** Our *Recursive Separation Theorem* establishes that semantic discriminability is a topological invariant of the condensation hierarchy. We prove that whenever the quotient map collapses states strictly along the fibers (level sets) of a base Urysohn separator, the separating function continuously descends to all subsequent levels. This guarantees that

disjoint concepts remain topologically separable regardless of hierarchy depth, ensuring that abstraction reduces complexity without sacrificing decision correctness.

Collectively, these results reframe the central problem of continual learning: it is not an unavoidable consequence of finite memory resources, but a failure to dynamically transform the topology of the representational manifold.

III. GEOMETRIC FRAMEWORK

A. Metric Collapse and Topological Quotienting

The Manifold Hypothesis and Capacity Growth Instead of formalizing continual learning by discrete graph abstractions [22], we model experience as evolving on a *Temporal Manifold* \mathcal{M} and interpret abstraction as an active deformation of its metric structure. In the geometric view, learning does not merely assign labels or embeddings to states; it reshapes the geometry of the space in which inference occurs by shortening geodesic distances between causally related regions of experience. Let the input stream be represented as a trajectory $\gamma(t)$ on a Riemannian manifold (\mathcal{M}, g) . The central difficulty of continual learning can be expressed geometrically [23]: if \mathcal{M} is flat (or approximately Euclidean), the geodesic distance $d_g(x_0, x_t)$ between the beginning of experience and the present grows linearly with time t . To preserve discriminability of past states, the system must maintain an ϵ -cover of the trajectory. The required number of covering elements, denoted $N(\epsilon, \mathcal{M})$, scales with the volume (or length) of the trajectory [24]. For a flat temporal manifold, this implies $N(\epsilon, \mathcal{M}) \propto L$, where L is the length of the experience stream. Without structural intervention, the representational demand grows unbounded, recovering the linear capacity accumulation identified in classical learning theory [25]. From a geometric perspective, catastrophic interference arises when the covering number of the trajectory exceeds the representational capacity of the system. Therefore, avoiding unbounded capacity growth requires not merely better parametrization, but a transformation of the geometry of \mathcal{M} itself.

Condensation as Metric Contraction We formalize abstraction through a geometric operation that actively contracts the metric over selected subsets of the manifold.

Definition 1 (Metric Contraction Operator). Let $U \subset \mathcal{M}$ denote a subset of the temporal manifold corresponding to a validated structure (e.g., a recurrent or causally closed trajectory segment). A **Condensation Operator** Ψ induces a modified metric $\Psi(g)$ such that, for all $x, y \in U$, $d_{\Psi(g)}(x, y) \rightarrow 0$. Topologically, Ψ defines a **quotient map** $q : \mathcal{M} \rightarrow \mathcal{M}/\sim$, where all points in U are identified with a single equivalence class p^* . We refer to p^* as a *token*.

Visualizing the Transformation (Metric Collapse) Importantly, condensation is not a projection that discards information [26]. Instead, it identifies a region of the manifold whose internal distinctions are no longer relevant for downstream inference, collapsing that region into a single point while preserving its relations to the remainder of the space. Fig. 1) illustrates how the **Search Phase** (Left) operates on a tangled manifold (High VC Dimension required), while the

Condensation Phase (Right) operates on the quotient manifold (Low VC Dimension required).

Geometric Interpretation (Wormholes). Under the original metric g , the endpoints $x_{t_{\text{start}}}$ and $x_{t_{\text{end}}}$ of a long temporal sequence may be separated by a large geodesic distance. Metric contraction introduces a region of extreme curvature that effectively pinches the manifold, bringing these distant points into immediate proximity [27], [28]. From the perspective of inference, this creates a *topological shortcut*: a path of negligible length that bypasses the original temporal extent. We use the term “wormhole” as an intuitive description of such a shortcut effect [14], [29], emphasizing that inference cost is reduced by deforming the metric rather than by traversing the original trajectory.

Hierarchical Quotient Spaces Condensation can be applied recursively, producing a hierarchy of progressively simplified manifolds: $\mathcal{M}_0 \xrightarrow{q_0} \mathcal{M}_1 \xrightarrow{q_1} \dots \xrightarrow{q_D} \mathcal{M}_D$. At each level k , the quotient manifold \mathcal{M}_k has strictly smaller effective diameter and covering number than \mathcal{M}_{k-1} . A token at level k corresponds to a nontrivial region, potentially a long temporal segment, of the original manifold \mathcal{M}_0 . Such a hierarchical construction allows an unbounded temporal history to be represented by a bounded number of points/tokens at sufficiently high levels of abstraction. While \mathcal{M}_0 may have diverging volume as experience accumulates, the upper-level quotient manifolds remain compact [30]. Inference in the hierarchical framework operates by traversing these compact quotient spaces rather than scanning the original temporal trajectory.

Connection to Biological Implementations. The form of metric contraction formalized above is not merely a mathematical abstraction, but is supported by well-established biological learning mechanisms [31], [32]. In cortical and hippocampal circuits, repeated co-activation of neural populations induces synaptic strengthening that effectively shortens functional distances between states, a phenomenon traditionally described in terms of attractor formation, representational compression, or chunking. From a geometric perspective, these processes implement a local contraction of the representational metric: states that reliably co-occur or predict one another become separated by progressively shorter geodesic distances in neural state space.

Metric contraction operates across multiple spatial and temporal scales. At the synaptic and population level, *generalized Hebbian learning* (GHL), including Oja- and Sanger-type rules [33], [34], aligns neural representations with the dominant low-dimensional structure of experienced trajectories, effectively flattening and unrolling curved manifolds in activity space. Such learning rules can be interpreted as adapting the local metric tensor to emphasize directions of consistent variance while collapsing redundant degrees of freedom. At a higher organizational level, hippocampal replay during offline states (e.g., sharp-wave ripples [35]) repeatedly reactivates extended temporal sequences in compressed form, accelerating traversal through learned state sequences and reinforcing long-range associations. Geometrically, these biological mechanisms collectively implement a quotient-map abstraction. Extended temporal trajectories are progressively collapsed into compact neural assemblies whose internal transitions are traversed

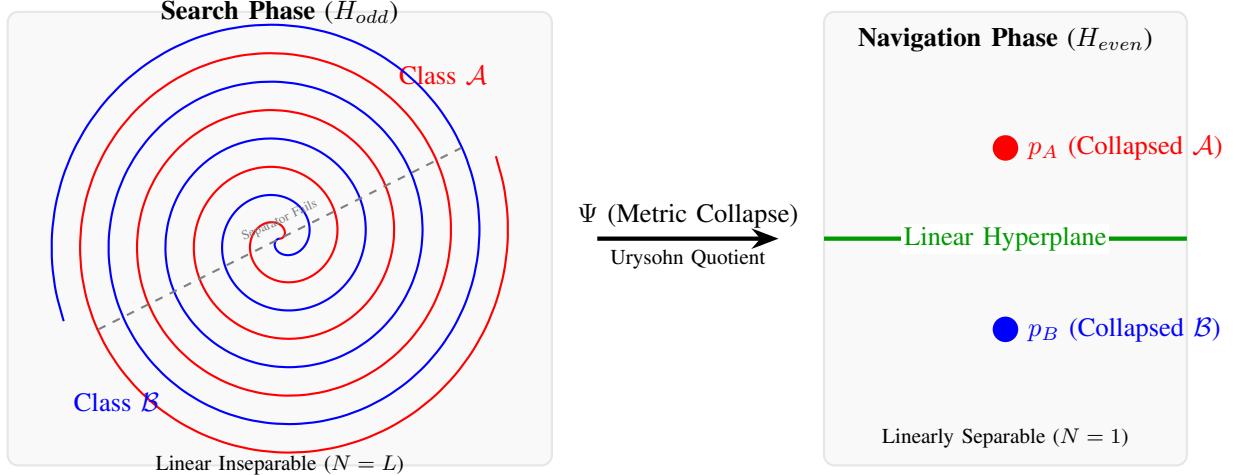


Fig. 1. **The Topological Trinity Transformation.** The diagram demonstrates the core mechanism of MAI. **Left:** The raw input stream (\mathcal{H}_{odd}) contains complex, intertwined temporal trajectories that violate Cover’s Theorem for linear separability. **Right:** After applying the Condensation Operator Ψ (Metric Collapse), the trajectories are topologically quotiented into single points in the Scaffold space ($\mathcal{H}_{\text{even}}$). In this collapsed metric, the classes become trivially linearly separable, effectively resetting the system’s memory capacity.

rapidly and with minimal metabolic cost [36].

B. The Parity Alternation Principle

A central unresolved challenge in both biological and artificial intelligence is the *Stability-Plasticity Dilemma* [37]: how to integrate novel information (plasticity) without degrading existing knowledge representations (stability). In monolithic connectionist models, this manifests as “catastrophic interference” [4], where the gradient updates required for a new task fundamentally alter the geometric subspace occupied by previous tasks. We propose that the mammalian cortex circumvents the fundamental limitation not through dynamic resource allocation, but through **topological orthogonality** [38]. We introduce the *Parity Alternation Principle*, which posits that the cognitive state space is not a single manifold but structurally decomposed into two disjoint homological manifolds: dynamic flows (\mathcal{H}_{odd}) and static scaffolds ($\mathcal{H}_{\text{even}}$). By restricting active inference [39] to the former and memory consolidation [40] to the latter, the system ensures that the metric deformations of learning are orthogonal to the metric structure of memory, thereby achieving a regime of stable, interference-free learning.

Definition 2 (Parity Partitioning). The **Parity Alternation Principle** posits that the cognitive state space \mathcal{M} is not a monolithic geometric manifold, but a *homologically partitioned complex* divided into two orthogonal subspaces based on dimensional parity: $\mathcal{M} = \mathcal{H}_{\text{odd}} \oplus \mathcal{H}_{\text{even}}$, where \mathcal{H}_{odd} contains topological features of odd dimensionality (e.g., β_1, β_3) and $\mathcal{H}_{\text{even}}$ contains features of even dimensionality (e.g., β_0, β_2).

Axiom 1 (Functional Conjugacy). The principle asserts a fundamental functional dichotomy between these parities: 1) **Odd Parity** (\mathcal{H}_{odd} / **Context / Flow**): Represents the dynamic, high-entropy “Search” phase. Topologically, these are **Cycles** (β_1). They encode temporal trajectories, sequences, and causal loops, which corresponds to the domain of *Active Inference* (System 2). 2) **Even Parity** ($\mathcal{H}_{\text{even}}$ / **Content**

/ **Scaffold**): Represents the static, low-entropy “Structure” phase. Topologically, these are **Connected Components** (β_0) or **Voids** (β_2). They encode invariant objects, concepts, and memory tokens, which correspond to the domain of *Amortized Knowledge* (System 1).

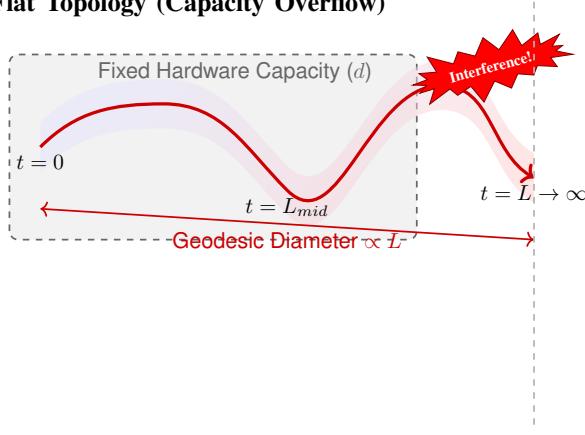
Axiom 2 (The Parity-Inverting Operator). Learning is formally defined not as the adjustment of weights within a fixed geometric space, but as a **Parity-Inverting Map**: $\Psi : \mathcal{H}_{\text{odd}}(\mathcal{M}) \rightarrow \mathcal{H}_{\text{even}}(\mathcal{M})$. This operator performs **Topological Condensation**: it identifies a validated temporal cycle (a closed loop in β_1) and collapses it into a single structural node (a point in β_0). Effectively, a “Process” is transmuted into a “Thing”.

Implication: Solution to the Stability-Plasticity Dilemma
In standard neural networks, new learning typically overwrites old memories (Catastrophic Interference) because both processes compete for resources within the same geometric embedding space. The Parity Alternation Principle resolves the dilemma through *topological orthogonality*: 1) *Plasticity* occurs exclusively in \mathcal{H}_{odd} (deforming the flow); 2) *Stability* is maintained exclusively in $\mathcal{H}_{\text{even}}$ (preserving the scaffold). Because the metric contraction required to collapse a new cycle γ_{new} occurs orthogonally to the metric tensor of the existing scaffold S_{old} , the system satisfies the condition of *vanishing topological interference*: $\langle \nabla \Psi(\gamma_{\text{new}}), \nabla S_{\text{old}} \rangle \rightarrow 0$. Therefore, the brain can possess infinite plasticity in the *odd* phase without compromising the stability of the *even* phase, thereby enabling open-ended, lifelong learning.

IV. MAIN RESULTS

Taken together, metric contraction defines how experience is geometrically collapsed into compact representations, while parity alternation constrains how and when such collapse can occur without destabilizing existing structure. The following three theorems formalize the consequences of this coupling for capacity growth, separability under abstraction, and interference-free learning.

A. Flat Topology (Capacity Overflow)



B. Topological Folding (Bounded Capacity)

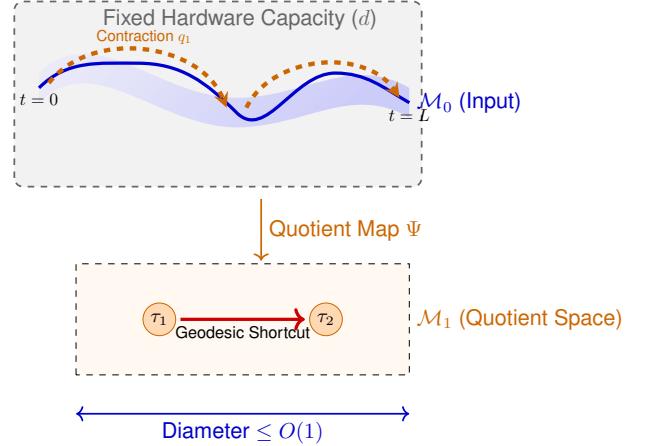


Fig. 2. Geometric Expansion vs. Topological Contraction/Folding. (A) The Flat Manifold Problem: In standard continual learning, the temporal manifold \mathcal{M} grows linearly in geodesic diameter as the input stream length $L \rightarrow \infty$. A system with fixed representational capacity (dashed box) eventually fails to maintain a valid ϵ -cover of the expanding volume, resulting in catastrophic interference (red starburst) where new experience overwrites the old. (B) The Quotient Solution: The proposed Token-Limited Capacity framework applies a condensation operator Ψ that functions as a topological quotient map. It introduces “wormhole” connections (orange arcs) that contract the metric between causally related states. This collapses continuous sub-manifolds into discrete singularities or tokens (τ_1, τ_2) within the quotient space \mathcal{M}_1 . By recursively folding the manifold, the system ensures the effective covering number remains bounded ($O(1)$), allowing unbounded temporal history to be represented within fixed-dimensional hardware.

A. Bounded Capacity Theorem

Preliminaries and Definitions Let (\mathcal{M}_0, g_0) denote the initial temporal manifold induced by the input stream of length L , and let $N(\epsilon, \mathcal{M})$ denote the *covering number* of a metric space (\mathcal{M}, g) , the minimal number of ϵ -radius balls required to cover \mathcal{M} [25]. We interpret the representational demand of a learning system as proportional to this covering number.

Definition 3 (Effective Capacity Demand). The **effective capacity demand** of a representational manifold \mathcal{M} at resolution ϵ is: $C_{\text{eff}}(\mathcal{M}) \triangleq N(\epsilon, \mathcal{M})$ A fixed-capacity system with hardware budget d can stably represent \mathcal{M} only if $C_{\text{eff}}(\mathcal{M}) \leq d$.

The Flat Manifold Problem First, we formalize the geometric source of catastrophic interference [4]: in the absence of abstraction, representational demand grows linearly with experience.

Lemma 1 (Linear Capacity Growth on Flat Manifolds). *If the temporal manifold \mathcal{M}_0 is isometric to a line segment of length L (i.e., no metric contraction occurs), then for any fixed resolution $\epsilon > 0$: $C_{\text{eff}}(\mathcal{M}_0) = \Theta(\frac{L}{\epsilon})$.*

The Recursive Solution: Bounded Capacity Theorem Because the representational demand at each level is governed by the covering number of \mathcal{M}_k , recursive metric contraction transforms linear growth in capacity demand into logarithmic growth in hierarchical depth [10]. Geometrically, abstraction is therefore a capacity-regulating operation: by collapsing regions of the temporal manifold with small internal diameter, the system reduces the number of distinguishable states that must be maintained simultaneously. To resolve the flat manifold problem, we introduce the concept of recursive metric contraction.

Definition 4 (Recursive ρ -Compressibility). A sequence of temporal manifolds $\{\mathcal{M}_k\}_{k=0}^D$ is *recursively ρ -compressible* if there exists a sequence of quotient maps $q_k : \mathcal{M}_k \rightarrow \mathcal{M}_{k+1}$ such that: $C_{\text{eff}}(\mathcal{M}_{k+1}) \leq \rho^{-1} C_{\text{eff}}(\mathcal{M}_k)$ where $\rho > 1$ is the uniform compression factor determined by the environment’s nested structure.

As a consequence of recursive metric contraction, the effective covering number of the learned manifold is reduced, stabilizing capacity even as experience accumulates. The Bounded Capacity Theorem (Fig. 2) formalizes our intuition, demonstrating that recursive metric contraction suffices to bound representational demand independently of the length of the input stream.

Theorem 1 (Bounded Capacity under Recursive Metric Contraction). *Let (\mathcal{M}_0, d_0) be a metric space representing experience up to time L , and let $N(\epsilon, \mathcal{M})$ denote the ϵ -covering number under d . Assume a hierarchical condensation process produces a sequence of quotient spaces $\mathcal{M}_0 \xrightarrow{q_0} \mathcal{M}_1 \xrightarrow{q_1} \dots \xrightarrow{q_{D-1}} \mathcal{M}_D$ equipped with quotient metrics $\{d_k\}_{k \geq 1}$. Assume that there exists $\rho > 1$ such that for all k , $N(\epsilon, \mathcal{M}_{k+1}) \leq \rho^{-1} N(\epsilon, \mathcal{M}_k)$. Then $N(\epsilon, \mathcal{M}_D) \leq \rho^{-D} N(\epsilon, \mathcal{M}_0)$, and in particular, if the representational budget satisfies $N(\epsilon, \mathcal{M}_D) \leq d$, it suffices to take $D \geq \lceil \log_\rho \left(\frac{N(\epsilon, \mathcal{M}_0)}{d} \right) \rceil$. Therefore, bounded representational demand can be maintained for arbitrarily long streams by logarithmic growth in hierarchy depth.*

Interpretation: Inference via Folded Geometry Theorem 1 implies that continual learning need not store or traverse an ever-growing temporal trajectory. Instead, inference proceeds on a compact quotient manifold whose geometry encodes long-range temporal structure, which operationally replaces linear-time retrospective search with short geodesic traversals across a folded manifold [30]. In other words, abstraction

acts as a geometric shortcut: a deformation of space that transforms temporal distance into structural proximity. Theorem 1 also provides a geometric lens on the classical distinction between recursive search and dynamic programming, which we will elaborate on in Sec. VI. On the flat temporal manifold \mathcal{M}_0 , inference corresponds to exhaustive traversal of the trajectory, requiring space linear in L , the geometric analog of Savitch-style recursion [41], where reachability is resolved by repeatedly subdividing paths without memoization. Each recursive call revisits the same regions of the manifold, incurring logarithmic depth but linear effective distance.

B. The Topological Collapse Separability Theorem

While Theorem 1 establishes the capacity constraints of a fixed geometric manifold, the conventional resolution in statistical learning theory, the Kernel Trick [2], [42], circumvents this by projecting data into a high-dimensional Hilbert space ($d \rightarrow \infty$). We argue that the kernel strategy constitutes a geometric overkill: it solves non-linearity by exploding the state space, incurring an exponential metabolic cost that is biologically implausible. Urysohn’s Lemma [43] suggests a far more parsimonious alternative.

Lemma 2 (Urysohn’s Lemma). *Let X be a normal topological space. Let A and B be two disjoint closed subsets of X . Then there exists a continuous function $f : X \rightarrow [0, 1]$ such that: $f(x) = 0 \quad \forall x \in A \quad \text{and} \quad f(x) = 1 \quad \forall x \in B$. This function f is often referred to as a separating function.*

Urysohn’s Lemma guarantees that linear separability does not require *adding* dimensions to untangle complex sets; it merely requires a continuous deformation that *collapses* the metric within the disjoint subsets. By shifting the mechanism from geometric expansion to topological contraction/quotienting (Fig. 2), the system can achieve perfect separability within its original finite dimensionality, which motivates our central result:

Theorem 2 (Quotient Collapse Preserves Separability). *Let \mathcal{M} be a normal topological space and let $A, B \subset \mathcal{M}$ be disjoint closed sets. By Urysohn’s lemma, there exists a continuous $f : \mathcal{M} \rightarrow [0, 1]$ such that $f(A) = \{0\}$ and $f(B) = \{1\}$. Define an equivalence relation $x \sim_f y \iff f(x) = f(y)$ and let $q : \mathcal{M} \rightarrow \tilde{\mathcal{M}} := \mathcal{M}/\sim_f$ be the quotient map. Then:*

- 1) *The images $q(A)$ and $q(B)$ are distinct singleton points in $\tilde{\mathcal{M}}$ (the equivalence classes at levels 0 and 1).*
- 2) *There exists a unique continuous $\bar{f} : \tilde{\mathcal{M}} \rightarrow [0, 1]$ such that $f = \bar{f} \circ q$, and $\bar{f}(q(A)) = 0$, $\bar{f}(q(B)) = 1$.*
- 3) *Consequently, $q(A)$ and $q(B)$ are separable by the threshold rule $\mathbf{1}[\bar{f}(\cdot) > 1/2]$.*

Recursive metric contraction fundamentally alters the regime of linear separability for finite dimensionality [1]. By collapsing validated submanifolds into quotient points, the system deforms \mathcal{M}_0 into a compact hierarchy of manifolds $\{\mathcal{M}_k\}$. Inference on \mathcal{M}_D proceeds by traversing short geodesics across these quotient spaces, equivalent to dynamic programming (DP) with memoized subproblems [44]. Geometrically, DP emerges when long temporal paths are replaced by wormholes, metric shortcuts induced by abstraction, transforming search over time

into navigation over structure [32]. Intuitively, the classical algorithmic distinction between Savitch-style recursion and DP corresponds to a geometric distinction between inference on a flat manifold and inference on a recursively folded one.

C. The Parity-Partitioned Stability Theorem

While Theorem 2 guarantees that a single manifold can be collapsed to achieve linear separability, it does not inherently prevent the metric distortion of the collapse from interfering with prior stored memories (Catastrophic Forgetting [5]). Our framework provides a rigorous topological substrate for Kahneman’s dual-process theory: **System 2 (Slow Thinking)** corresponds to the metabolic cost of generating topology via linear search in the Flow space ($\mathcal{M}_{\text{slow}} \leftrightarrow \mathcal{H}_{\text{odd}}$), while **System 1 (Fast Thinking)** corresponds to the amortized efficiency of navigating the condensed Scaffold ($\mathcal{M}_{\text{fast}} \leftrightarrow \mathcal{H}_{\text{even}}$). By invoking the Parity Alternation Principle, we derive the necessary condition for continuous, interference-free learning.

Definition 5 (Orthogonal Parity Decomposition). *Let the cognitive state space be a Riemannian product manifold $\mathcal{M} = \mathcal{M}_{\text{fast}} \times \mathcal{M}_{\text{slow}}$ with metric $g = g_{\text{fast}} \oplus g_{\text{slow}}$. Let $\theta = [\theta_F, \theta_S]$ denote the parameter vectors governing the geometry of each subspace.*

Orthogonal parity decomposition provides the geometric basis for solving the stability-plasticity dilemma [6]. In standard continual learning, catastrophic forgetting arises because the gradient updates for new tasks $\nabla \mathcal{L}_{\text{new}}$ have non-zero projection onto the parameter subspace encoding old memories, thereby degrading the structural risk \mathcal{R}_{old} . By enforcing a strict Riemannian product structure, we can isolate these dynamics. The following theorem formalizes how temporally segregating updates into orthogonal tangent bundles prevents this interference, ensuring that rapid adaptation in the Fast manifold leaves the Slow scaffold metric invariant.

Theorem 3 (Parity-Partitioned Stability). *Let parameters decompose as $\theta = (\theta_F, \theta_S) \in \Theta_F \times \Theta_S$, and suppose the system alternates two update phases: (Flow phase) : $\Delta\theta_S = 0$, (Scaffold phase) : $\Delta\theta_F = 0$. Assume the metric on parameter space is block-diagonal, $g = g_F \oplus g_S$, so that the induced inner product satisfies $\langle (u_F, 0), (0, v_S) \rangle_g = 0$. Then the cross-interference term between phases vanishes: $\langle \Delta\theta^{(F)}, \Delta\theta^{(S)} \rangle_g = 0$. In particular, if a memory functional $R(\theta_S)$ depends only on θ_S , then Flow-phase updates leave R invariant: $R(\theta_F + \Delta\theta_F, \theta_S) = R(\theta)$.*

The derivation of Theorem 3 represents the final bridge between biological plausibility and computational theory. In standard neural networks, plasticity and stability are antagonists competing for the same synaptic weights, to learn the new, one must often distort the old (catastrophic interference) [17]. By combining topological collapse with parity alternation, we demonstrate that the brain circumvents the plasticity-stability trade-off through architectural segregation [7]. Because the heavy lifting of metric collapse occurs in the high-entropy *Odd* space, while the result is stored in the low-entropy *Even* space, the violence of learning never disrupts the serenity of

memory. Theoretical guarantee of zero topological interference provides the rigorous justification for the separate Search and Structure phases observed in the cortical column [45], [46].

V. RECURSIVE CONDENSATION AND HIERARCHICAL DEPTH

Theorem 1 establishes that unbounded experience can be represented without increasing the dimensionality of the underlying state space. It raises a fundamental conservation question: if representational width remains fixed, *where does accumulated complexity reside?* From a geometric perspective, the answer is neither in width nor in volume, but in **hierarchical depth** [20], [47]. Next, we show that recursive metric contraction necessarily induces a logarithmic hierarchy of quotient manifolds, which relates geometric depth to classical space-time trade-offs in computation.

A. Logarithmic Depth as a Geometric Necessity

Let the raw experience stream of length L define a temporal trajectory γ_L on the base manifold \mathcal{M}_0 , whose diameter and volume grow linearly with L . On a flat manifold, maintaining distinguishability of the dynamic trajectory requires an ϵ -cover whose size scales as $N(\epsilon, \mathcal{M}_0) \propto L$, implying linear growth in effective capacity demand. Recursive condensation alters such geometry by introducing a sequence of quotient maps [48] $\mathcal{M}_0 \xrightarrow{q_0} \mathcal{M}_1 \xrightarrow{q_1} \dots \xrightarrow{q_D} \mathcal{M}_D$, where each map q_k collapses validated submanifolds of \mathcal{M}_k to points in \mathcal{M}_{k+1} . Geometrically, each contraction reduces the diameter and volume of the space by a multiplicative factor $\rho > 1$, corresponding to the compressibility of the trajectory.

As a consequence, after D levels of contraction, the effective diameter satisfies $\text{diam}(\mathcal{M}_D) \approx \frac{L}{\rho^D}$. To maintain a bounded covering number, and hence bounded capacity demand, the hierarchy depth must scale as $D = O(\log L)$. The logarithmic growth is not an algorithmic artifact but a geometric necessity (Fig. 3) [41]. Any process that maps an unbounded-length trajectory into a compact representation via repeated metric contraction must trade linear growth in diameter for logarithmic growth in depth. Hierarchical depth is therefore the *geometric price* paid for constant-width inference. Importantly, increasing depth does not increase instantaneous capacity demand. At each level, inference operates on a compact quotient manifold with bounded covering number, using the same fixed-dimensional decision mechanism. Recursive condensation converts unsustainable linear growth in representational extent into sustainable logarithmic growth in structural hierarchy.

Theorem 4 (Recursive separation preserved under compatible quotienting). *Let X_0 be normal and $A_0, B_0 \subset X_0$ disjoint closed as in Theorem 2. Let $q_k : X_k \rightarrow X_{k+1}$ be quotient maps and define $A_{k+1} := q_k(A_k)$, $B_{k+1} := q_k(B_k)$. If there exists a continuous $f_0 : X_0 \rightarrow [0, 1]$ separating A_0, B_0 such that, for each k , $x \sim_k x' \implies f_k(x) = f_k(x')$, where f_k is the descended separator on X_k , then for every level k there is a continuous $f_k : X_k \rightarrow [0, 1]$ separating A_k and B_k .*

B. Geometric Relation to Savitch-Style Recursion

The depth-width trade-off admits a direct geometric interpretation of classical results in computational complexity, most notably **Savitch's Theorem** [41]. Savitch's construction shows that nondeterministic space-bounded reachability can be simulated deterministically by recursively subdividing paths, trading space for recursion depth. In geometric terms, Savitch-style recursion corresponds to inference on a *flat* manifold: long trajectories are verified by repeatedly bisecting geodesics, without altering the underlying geometry. Although space usage remains bounded, inference repeatedly revisits the same regions of the manifold, incurring exponential time overhead. Recursive condensation implements a geometric analog of memoization. When a sub-trajectory is validated and collapsed via a quotient map, its internal geometry is removed from future consideration. Subsequent inference does not re-traverse the original geodesic; it traverses a *wormhole*, a metric shortcut induced by contraction, which transforms repeated recursive verification into direct navigation on a folded manifold.

From a wormhole perspective, we observe: 1) **Flat inference (search)** corresponds to geodesic traversal on \mathcal{M}_0 , analogous to Savitch-style recursion without memoization; 2) **Condensation** corresponds to collapsing verified geodesic segments into singular points, eliminating redundant distance; 3) **Hierarchical inference** corresponds to navigation on \mathcal{M}_D , where long temporal separations have been converted into short paths through quotient geometry. The resulting structure is a **tower of quotient manifolds**, in which complexity is serialized into depth rather than width. Just as Savitch's theorem manages space by serializing verification, recursive condensation manages capacity by serializing abstraction.

VI. SLOW AND FAST INFERENCE: SEARCH VERSUS NAVIGATION

We now characterize how recursive condensation gives rise to two qualitatively distinct modes of inference: a high-latency exploratory regime and a low-latency amortized regime. These modes correspond formally to the distinction between the *Flow Manifold* ($\mathcal{M}_{\text{slow}}/\mathcal{H}_{\text{odd}}$) and the *Scaffold Manifold* ($\mathcal{M}_{\text{fast}}/\mathcal{H}_{\text{even}}$), which align naturally with the slow-fast thinking distinction in cognitive science [49]: slow inference corresponds to explicit search over the high-entropy Flow manifold, while fast inference emerges once validated trajectories have been condensed into the low-entropy Scaffold manifold.

A. Slow vs. Fast Inference as Parity Regimes

The distinction between slow, deliberative reasoning and fast, automatic inference, often described as *System 2* versus *System 1* cognition [49], admits a precise geometric interpretation under the Parity Alternation Principle.

Slow Inference (System 2) as Flow Search ($\mathcal{M}_{\text{slow}}/\mathcal{H}_{\text{odd}}$). Novel problems exist in the *Odd Parity* subspace ($\mathcal{M}_{\text{slow}}/\mathcal{H}_{\text{odd}}$). Here, the topological structure consists of open 1-cycles (unresolved temporal sequences). Reasoning in the slow regime corresponds to **Active Search** [50]: the agent must traverse the extended geodesic of a potential solution trajectory γ to verify its closure condition ($\partial\gamma \stackrel{?}{=} 0$).

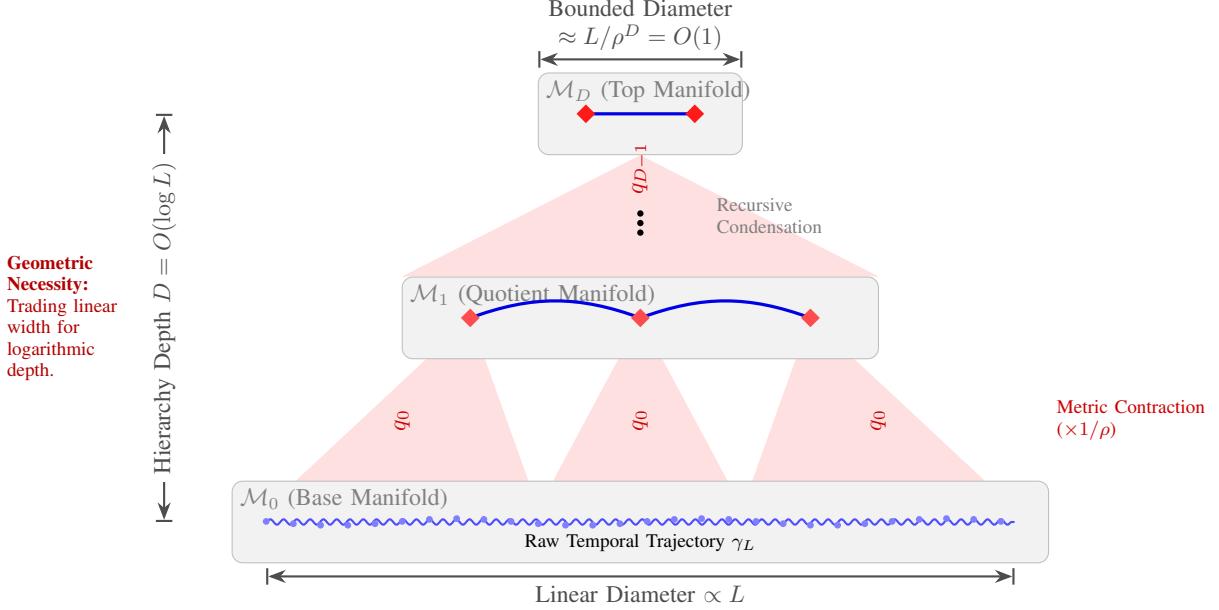


Fig. 3. Logarithmic Scaling via Recursive Condensation. The diagram illustrates how a linear input stream of length L on the base manifold \mathcal{M}_0 is progressively compressed through a hierarchy of quotient maps q_k . Each contraction reduces the metric volume by a factor ρ . To achieve a bounded effective diameter at the top manifold \mathcal{M}_D (necessary for constant-width inference), the depth of the hierarchy D must scale logarithmically with the input length L . The shaded cones represent the collapse of submanifolds into points at the next level.

Because the metric in $\mathcal{M}_{slow}/\mathcal{H}_{odd}$ is uncollapsed (flat), the cost of the verification scales linearly with the complexity of the problem ($O(L)$), which aligns with the phenomenology of System 2 (serial, effortful, and metabolically expensive).

Fast Inference (System 1) as Scaffold Navigation (\mathcal{M}_{even}). Once a solution trajectory is validated, the *Parity-Inverting Map* Ψ condenses the cycle into a 0-cycle (a point) in the *Even Parity* subspace ($\mathcal{M}_{fast}/\mathcal{H}_{even}$). Inference in the fast regime corresponds to **Topological Navigation** [31]. The agent no longer simulates the trajectory; it simply transitions between connected components in the quotient topology. Because $\mathcal{M}_{fast}/\mathcal{H}_{even}$ is a quotient space where the metric diameter of the solution has been driven to zero (Theorem 2), the cost of the operation is $O(1)$, which aligns with the phenomenology of System 1 (parallel, automatic, and instantaneous).

The “Aha!” Moment as Metric Collapse. The transition from System 2 to System 1 is not merely practice; it is a phase transition [51]. It occurs at the precise moment the *Condensation Operator* Ψ is applied. Prior to this moment, the problem is a Search task (finding a path). After this moment, the problem is a Structure or Navigation task (recognizing a node). Fast thinking is therefore amortized slow thinking [52]: the energy expended in $\mathcal{M}_{slow}/\mathcal{H}_{odd}$ to close the loop is stored as the topological triviality of the node in $\mathcal{M}_{fast}/\mathcal{H}_{even}$.

B. Biological Realization: Working Memory and Cortical Uniformity

The geometric constraints derived in our Manifold Capacity Theorem find striking empirical support in the architectural organization of the mammalian cortex. Our analysis suggests that two fundamental biological phenomena, the severe capacity limit of working memory and the cytoarchitectonic uniformity

of the neocortex, are actually dual manifestations of the same topological principle: the necessity of maintaining bounded local covering numbers.

The Geometric Origin of the Magical Number Seven. Cognitive science consistently reports a rigid bound on the number of items that can be active in working memory, famously characterized by Miller as the “magical number seven, plus or minus two” [53], which is not a psychological artifact but a geometric necessity. If inference at level k operates on a quotient manifold \mathcal{M}_k , the reliability of that inference depends on the manifold’s covering number $N(\epsilon, \mathcal{M}_k)$, the number of distinct ϵ -balls required to span the active state space [25]. We interpret the working memory limit as the hardware’s upper bound on the covering number: $N(\epsilon, \mathcal{M}_k) \leq C_{bio} \approx 7$. The rigidity of working memory is therefore a stabilizing constraint. It enforces the condition that no single hierarchy level is ever permitted to manage a manifold complex enough to induce interference. Recursive condensation reconciles such a tight local constraint with unbounded global experience by shifting complexity into *depth*. The system does not widen its local capacity to absorb a long stream; it deepens the hierarchy so that the top-level quotient manifold \mathcal{M}_D never exceeds diameter C_{bio} .

Cortical Uniformity as a Parity-Alternating Operator. Our geometric perspective offers a mechanistic explanation for Mountcastle’s cortical uniformity hypothesis [8]. If the fundamental task of the cortex is to execute the Parity Alternation cycle (Search → Condense → Structure), then the computational hardware need not be specialized for different scales of abstraction. In the geometric view, the canonical cortical microcircuit [54] is the physical implementation of a **Universal Quotient Operator**. Because the local topological

complexity is invariant across levels (always bounded by C_{bio}), a single algorithmic primitive can be reused ubiquitously. Hierarchy emerges not from distinct hardware, but from the stacking of these operators: early sensory areas process fine-scale manifolds \mathcal{M}_0 , while higher association areas process deeply condensed quotient spaces \mathcal{M}_k [36]. Specialization arises solely from connectivity (which manifold is fed into the module), while the intrinsic computation, collapsing metric geometry, remains generic.

C. A Local Covering Bound Implies Fast Inference

We now formalize the sense in which bounded local covering number implies fast inference. Intuitively, if the active state space at a given abstraction level can be covered by a constant number of ϵ -balls (Theorem 1), then local search can be performed in constant time, independent of the global history length L .

Definition 6 (Active region and local cover). Fix a level k and resolution $\epsilon > 0$. Let $\mathcal{A}_k(t) \subseteq \mathcal{M}_k$ denote the *active region* of states that the system must discriminate among at time t . Define the *local covering number* of the active region by $N(\epsilon, \mathcal{A}_k(t)) := \min \{m : \mathcal{A}_k(t) \subseteq \bigcup_{i=1}^m B_\epsilon(x_i)\}$.

Lemma 3 (Bounded local cover \Rightarrow constant-time local inference). *Assume there exists a constant C such that for all times t , $N(\epsilon, \mathcal{A}_k(t)) \leq C$. Suppose the inference operator at level k selects an action by evaluating a local objective on a representative set of candidates drawn from an ϵ -cover of $\mathcal{A}_k(t)$, with evaluation cost $T_{\text{eval}}(\epsilon)$. Then the per-step inference cost is: $T_{\text{step}} \leq C \cdot T_{\text{eval}}(\epsilon) = O(1)$.*

Interpretation (System 1 as Urysohn Collapse). Lemma 3 provides the computational justification for the Topological Collapse Separability Theorem. Fast inference is possible only when the active region $\mathcal{A}_k(t)$ has been topologically collapsed. In the raw temporal manifold \mathcal{M}_0 , the active region scales with history ($N \propto L$), making constant-time inference impossible. System 1 capabilities emerge precisely when the recursive quotient maps have driven $N \rightarrow O(1)$. Therefore, biological intuition is not a heuristic approximation; it is the exact traversal of a collapsed metric space.

VII. IMPLICATIONS AND LIMITATIONS

A. Implications

1. Continual Learning Without Capacity Explosion (Theorem 1) The central implication of the *Bounded Capacity Theorem* is that catastrophic interference is not an intrinsic consequence of fixed-dimensional representations, but a geometric pathology of operating on a *flat temporal manifold*. When experience accumulates without structural deformation, the diameter and volume of the representational manifold grow linearly with time, forcing the covering number to diverge. Recursive metric contraction fundamentally alters the underlying geometry. By collapsing validated submanifolds via quotient maps, the learning system actively reduces the effective diameter of its representational space. In this regime, unbounded experience is not stored explicitly; it is *absorbed*

into the topology of a hierarchy of quotient manifolds. Capacity pressure is relieved not by increasing representational width, but by redistributing complexity across logarithmic depth ($D \propto \log L$). Our reasoning establishes a formal existence proof for continual learning in fixed-capacity systems: stable learning is possible if and only if the system can progressively collapse recurring structure into lower-volume quotient spaces.

2. Topological Collapse as the Dual of the Kernel Trick

(Theorem 2) Machine learning has traditionally relied on the Kernel Trick to solve non-linear separability by projecting data into infinite-dimensional spaces ($d \rightarrow \infty$). Our *Topological Collapse Separability Theorem* establishes a parsimonious dual to the conventional approach. We demonstrate that separability can be achieved not by exploding the dimension of the space, but by collapsing the metric of the data ($N \rightarrow 1$). By utilizing Urysohn-style quotient maps to contract complex class boundaries into singularities, the system renders non-linear problems linearly separable within the original finite dimensionality, which suggests a paradigm shift for AGI architectures: rather than seeking “Scaling Laws” that demand massive parameter counts, we should seek “Folding Laws” that maximize topological density.

3. Stability via Homological Orthogonality (Theorem 3)

Perhaps the most significant implication for biological plausibility is the resolution of the Stability-Plasticity dilemma. The *Parity-Partitioned Stability Theorem* suggests that the brain avoids catastrophic forgetting not through complex weight consolidation algorithms (e.g., Elastic Weight Consolidation (EWC) [17]), but through a fundamental architectural separation of concerns. By proving that the metric deformation of the flow manifold ($\mathcal{M}_{\text{slow}}/\mathcal{H}_{\text{odd}}$) is orthogonal to the metric structure of the scaffold manifold ($\mathcal{M}_{\text{fast}}/\mathcal{H}_{\text{even}}$), we identify the physical mechanism of interference-free learning. The above line of reasoning implies that the breakdown of memory in artificial systems is due to their monolithic geometry; true lifelong learning requires a system that can alternate between a “fluid” topology for exploration and a “crystallized” topology for storage (e.g., wake-sleep algorithm [55]).

4. Abstraction as Metric Singularity Within our framework, abstraction is reinterpreted as a **geometric capacity control mechanism**. A successful abstraction corresponds to a region of the temporal manifold whose internal geodesic distances are driven to zero. Physically, tokens in our architecture are realizable as *metric singularities* or wormholes, regions of extreme positive curvature that bridge distant points in time. The wormhole perspective provides a task-agnostic criterion for abstraction utility: *an abstraction is valid if it decreases the covering number of the active manifold*. This definition applies equally to reinforcement learning [56], sequence modeling, and causal inference [57], explaining how abstraction simultaneously reduces capacity demand and improves generalization.

From a biological perspective, our theory provides a rigorous explanation for how complex, temporally extended experiences can be distilled into compact neural representations that remain behaviorally discriminative [58]. Rather than resolving nonlinear classification problems at every encounter, the brain collapses validated trajectories into atomic memory tokens

whose mutual separability is guaranteed by construction [59]. This supports the view that abstraction operates as a form of *metric collapse constrained by semantic compatibility*, enabling fast, low-dimensional readout while retaining correctness [60], [61]. More broadly, the theorem clarifies why symbolic-like decisions can emerge from subsymbolic dynamics: separability is not an intrinsic property of dimensionality, but a topological invariant under compatible quotient maps [62], [63], [64].

B. Limitations

The Compressibility Assumption. Our main results assume that the temporal manifold admits nontrivial metric contraction, that is, that experience contains recurring, compressible structure. If the input stream is maximally entropic (incompressible) [25], no quotient maps exist that significantly reduce the covering number. In such regimes, the manifold remains effectively flat and capacity demand grows linearly with time. The theory applies only to structured environments where abstraction is information-theoretically possible.

The Search for Topology. While the *Topological Collapse Separability Theorem* proves that a separating quotient map exists (via Urysohn’s Lemma [43]), it does not prescribe an efficient algorithm for *finding* it. Discovering the optimal topological deformation is a search problem that may be computationally expensive in the worst case. Our theory bounds the memory scaling of the *learned* representation, but not the energy cost of the *Search Phase* (\mathcal{H}_{odd}) required to discover it.

Approximate Orthogonality and Biological Noise. The Parity-Partitioned Stability Theorem establishes interference-free learning under the assumption of strict orthogonality between the flow manifold and the scaffold manifold. This condition should be understood as an idealized limit. In biological neural circuits, parity separation is likely implemented only approximately, through oscillatory gating [45], neuromodulatory control [65], and partial synaptic segregation rather than exact subspace orthogonality. Consequently, residual coupling between flow- and scaffold-associated updates may introduce bounded interference, particularly under high noise or rapid task switching. An important direction for future work is to characterize the stability guarantees of Theorem 3 under *approximate* orthogonality, quantifying how much leakage can be tolerated before catastrophic interference re-emerges, and how biological mechanisms such as oscillatory phase locking [66] or replay scheduling may actively suppress such leakage.

Scope of the Theory. Our work should be interpreted as a capacity-theoretic foundation rather than a complete learning algorithm. It demonstrates that fixed-width systems can, in principle, support unbounded continual learning through geometric transformation, but leaves the specific implementation of the parity-inverting operator Ψ open for algorithmic development.

VIII. CONCLUSION

We have presented a geometric resolution to the paradox of unbounded experience in fixed-capacity systems. By modeling continual learning through the lens of manifold topology, we demonstrated that the fundamental barrier to lifelong learning is not a lack of dimensions, but a lack of curvature. Our analysis

redefines abstraction not as a statistical clustering operation, but as **Recursive Metric Contraction**: a physical process where the learning system actively deforms the manifold of experience, introducing wormhole-like singularities that collapse temporal distances between causally related events. Our result establishes that catastrophic interference is not an inevitability of fixed-dimensional systems, but a consequence of operating on a flat topology.

This paper makes the following contributions: 1) We introduce a geometric formulation of continual learning in which abstraction is defined as the recursive application of quotient maps that contract the metric of a temporal manifold; 2) We prove that recursive metric contraction permits a fixed-dimensional system to maintain bounded covering numbers even as $L \rightarrow \infty$; 3) We demonstrate that topological quotienting provides a parsimonious alternative to kernel methods, achieving separability via metric collapse rather than dimensional explosion; 4) We formalize the *Parity Alternation Principle*, identifying the separation of homological odd/even subspaces as the necessary condition for interference-free learning; 5) We clarify the distinction between *geometric expansion* (increasing dimension) and *topological contraction* (folding space), providing a principled foundation for scalable continual learning on fixed-capacity hardware.

Our geometric framework reframes catastrophic interference as a failure of geometry: it is the inevitable consequence of forcing an expanding trajectory onto a flat, rigid manifold. Conversely, scalable intelligence arises from Topological Flexibility, the ability to fold the state space dynamically. Ultimately, our work suggests a fundamental pivot for the field of Artificial General Intelligence (AGI) [67]. We argue that the path to AGI does not lie in scaling laws that demand ever-larger static matrices, but in **folding laws** that govern how a fixed-size network can recursively crumple the fabric of experience into a structure of infinite density.

REFERENCES

- [1] T. M. Cover, “Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition,” *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326–334, 1965.
- [2] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [3] S. Shalev-Shwartz and S. Ben-David, “Understanding machine learning: From theory to algorithms,” *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 1–325, 2014.
- [4] M. McCloskey and N. J. Cohen, “Catastrophic interference in connectionist networks: The sequential learning problem,” *Psychology of Learning and Motivation*, vol. 24, pp. 109–165, 1989.
- [5] R. M. French, “Catastrophic forgetting in connectionist networks,” *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 128–135, 1999.
- [6] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, “Continual lifelong learning with neural networks: A review,” *Neural Networks*, vol. 113, pp. 54–71, 2019.
- [7] G. Buzsáki, *Rhythms of the Brain*. Oxford University Press, 2006.
- [8] V. B. Mountcastle, “The columnar organization of the neocortex,” *Brain*, vol. 120, no. 4, pp. 701–722, 1997.
- [9] U. Hasson, E. Yang, I. Vallines, D. J. Heeger, and N. Rubin, “A hierarchy of temporal receptive windows in human cortex,” *Journal of Neuroscience*, vol. 28, no. 10, pp. 2539–2550, 2008.
- [10] H. A. Simon, “The architecture of complexity,” *Proceedings of the American Philosophical Society*, vol. 106, no. 6, pp. 467–482, 1973.
- [11] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

- [12] C. Fefferman, S. Mitter, and H. Narayanan, “Testing the manifold hypothesis,” *Journal of the American Mathematical Society*, vol. 29, no. 4, pp. 983–1049, 2016.
- [13] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, 2007.
- [14] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.
- [15] A. Robins, “Catastrophic forgetting, rehearsal and pseudorehearsal,” *Connection Science*, vol. 7, no. 2, pp. 123–146, 1995.
- [16] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized experience replay,” *arXiv preprint arXiv:1511.05952*, 2015.
- [17] J. Kirkpatrick, R. Pascanu, N. Rabinowitz *et al.*, “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.
- [18] P. Dayan and G. E. Hinton, “Feudal reinforcement learning,” *Advances in Neural Information Processing Systems*, vol. 5, 1993.
- [19] T. G. Dietterich, “Hierarchical reinforcement learning with the maxq value function decomposition,” *Journal of Artificial Intelligence Research*, vol. 13, pp. 227–303, 2000.
- [20] M. Telgarsky, “Benefits of depth in neural networks,” *Proceedings of the 29th Annual Conference on Learning Theory*, 2016.
- [21] S. Mahadevan, “Proto-value functions: A laplacian framework for learning representation and control in markov decision processes,” in *International Conference on Machine Learning (ICML)*, 2005.
- [22] L. Wang, X. Zhang, H. Su, and J. Zhu, “A comprehensive survey of continual learning: Theory, method and application,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 46, no. 8, pp. 5362–5383, 2024.
- [23] J. M. Lee, *Introduction to Riemannian Manifolds*. Springer, 2018.
- [24] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.
- [25] A. N. Kolmogorov and V. M. Tikhomirov, ϵ -Entropy and ϵ -Capacity of Sets in Function Spaces, 1959, vol. 14, no. 2.
- [26] J. R. Munkres, *Topology*, 2nd ed. Prentice Hall, 2000.
- [27] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, 1999.
- [28] F. Mémoli, “On the gromov–hausdorff distance,” *Discrete & Computational Geometry*, vol. 38, no. 1, pp. 43–70, 2007.
- [29] R. R. Coifman and S. Lafon, “Diffusion maps,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [30] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*. American Mathematical Society, 2001.
- [31] R. U. Muller, M. Stead, and J. Pach, “The hippocampus as a cognitive graph.” *The Journal of general physiology*, vol. 107, no. 6, pp. 663–694, 1996.
- [32] G. Buzsáki and E. I. Moser, “Memory, navigation and theta rhythm in the hippocampal-entorhinal system,” *Nature Neuroscience*, vol. 16, no. 2, pp. 130–138, 2013.
- [33] E. Oja, “Simplified neuron model as a principal component analyzer,” *Journal of mathematical biology*, vol. 15, no. 3, pp. 267–273, 1982.
- [34] T. D. Sanger, “Optimal unsupervised learning in a single-layer linear feedforward neural network,” *Neural Networks*, vol. 2, no. 6, pp. 459–473, 1989.
- [35] M. A. Wilson and B. L. McNaughton, “Reactivation of hippocampal ensemble memories during sleep,” *Science*, vol. 265, no. 5172, pp. 676–679, 1994.
- [36] U. Hasson, J. Chen, and C. J. Honey, “Hierarchical process memory: memory as an integral component of information processing,” *Trends in cognitive sciences*, vol. 19, no. 6, pp. 304–313, 2015.
- [37] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardi, G. Slabaugh, and T. Tuytelaars, “A continual learning survey: Defying forgetting in classification tasks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3366–3385, 2021.
- [38] X. Li, “The homological brain: Parity principle and amortized inference,” *arXiv preprint arXiv:2512.10976*, 2025.
- [39] K. Friston, T. Fitzgerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo, “Active inference: a process theory,” *Neural computation*, vol. 29, no. 1, pp. 1–49, 2017.
- [40] L. R. Squire, L. Genzel, J. T. Wixted, and R. G. Morris, “Memory consolidation,” *Cold Spring Harbor perspectives in biology*, vol. 7, no. 8, p. a021766, 2015.
- [41] W. J. Savitch, “Relationships between nondeterministic and deterministic tape complexities,” *Journal of Computer and System Sciences*, vol. 4, no. 2, pp. 177–192, 1970.
- [42] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge university press, 2004.
- [43] S. Willard, *General topology*. Courier Corporation, 2012.
- [44] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
- [45] J. E. Lisman and O. Jensen, “The theta-gamma neural code,” *Neuron*, vol. 77, no. 6, pp. 1002–1016, 2013.
- [46] G. Buzsáki, “The hippocampo-neocortical dialogue,” *Cerebral cortex*, vol. 6, no. 2, pp. 81–92, 1996.
- [47] T. Poggio, K. Kawaguchi, Q. Liao *et al.*, “Theory of deep learning iii: Explaining the non-overfitting puzzle,” *arXiv preprint arXiv:1801.00173*, 2017.
- [48] A. Hatcher, *Algebraic topology*, 2005.
- [49] D. Kahneman, *Thinking, fast and slow*. macmillan, 2011.
- [50] M. L. Minsky, *Computation: Finite and Infinite Machines*. Prentice-Hall, 1967.
- [51] M. Nakahara, *Geometry, topology and physics*. CRC press, 2018.
- [52] S. Gershman and N. Goodman, “Amortized inference in probabilistic reasoning,” in *Proceedings of the annual meeting of the cognitive science society*, vol. 36, no. 36, 2014.
- [53] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” *Psychological review*, vol. 63, no. 2, p. 81, 1956.
- [54] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston, “Canonical microcircuits for predictive coding,” *Neuron*, vol. 76, no. 4, pp. 695–711, 2012.
- [55] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, “The “wake-sleep” algorithm for unsupervised neural networks,” *Science*, vol. 268, no. 5214, pp. 1158–1161, 1995.
- [56] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 135.
- [57] J. Pearl, *Causality*. Cambridge university press, 2009.
- [58] B. E. Pfeiffer and D. J. Foster, “Hippocampal place-cell sequences depict future paths to remembered goals,” *Nature*, vol. 497, no. 7447, pp. 74–79, 2013.
- [59] E. T. Rolls, “The mechanisms for pattern completion and pattern separation in the hippocampus,” *Frontiers in Systems Neuroscience*, vol. 7, p. 74, 2013.
- [60] M. Rigotti *et al.*, “The importance of mixed selectivity in complex cognitive tasks,” *Nature*, vol. 497, no. 7451, pp. 585–590, 2013.
- [61] F. Mastrogiovanni and S. Ostojević, “Linking connectivity, dynamics, and computations in low-rank recurrent neural networks,” *Neuron*, vol. 99, no. 3, pp. 609–623, 2018.
- [62] G. F. Marcus, *The Algebraic Mind*. MIT Press, 2001.
- [63] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, “Building machines that learn and think like people,” *Behavioral and Brain Sciences*, vol. 40, p. e253, 2017.
- [64] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, “How does the brain solve visual object recognition?” *Neuron*, vol. 73, no. 3, pp. 415–434, 2012.
- [65] W. Schultz, “Predictive reward signal of dopamine neurons,” *Journal of neurophysiology*, 1998.
- [66] A. G. Siapas, E. V. Lubenov, and M. A. Wilson, “Prefrontal phase locking to hippocampal theta oscillations,” *Neuron*, vol. 46, no. 1, pp. 141–151, 2005.
- [67] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, “Sparks of artificial general intelligence: Early experiments with gpt-4,” *arXiv preprint arXiv:2303.12712*, 2023.

APPENDIX

Proof of Lemma 1. A 1-dimensional manifold of length L requires $N \approx L/(2\epsilon)$ balls of radius ϵ to cover. Since C_{eff} scales linearly with N , the demand grows linearly with L . As $L \rightarrow \infty$, $C_{\text{eff}} > d$ eventually, forcing either $\epsilon \rightarrow \infty$ (loss of precision) or trajectory overlap (catastrophic interference). \square

Proof of Lemma 3. Let $\mathcal{A}_k(t)$ be the metric space of admissible actions (or hidden states) at hierarchy level k and time t , equipped with a metric $d(\cdot, \cdot)$. By definition, a set $S_\epsilon \subset \mathcal{A}_k(t)$ is an ϵ -cover of $\mathcal{A}_k(t)$ if for every $a \in \mathcal{A}_k(t)$, there exists an $s \in S_\epsilon$ such that $d(a, s) \leq \epsilon$. The covering number $N(\epsilon, \mathcal{A}_k(t))$ is the cardinality of the smallest such set S_ϵ .

1. Bounded Search Space Assumption: The lemma posits that the metric entropy of the action space is bounded by a constant C , independent of the global state dimension or

history length: $|S_\epsilon| = N(\epsilon, \mathcal{A}_k(t)) \leq C$. This assumption reflects the *Metric Collapse* property, where the topological quotient operation Ψ has reduced the effective volume of the search space to a finite manifold.

2. Discretization of the Inference Operator: The inference operator Φ_k approximates the optimal action a^* by minimizing a local objective function \mathcal{L} over the cover S_ϵ , rather than the continuous space $\mathcal{A}_k(t)$. $\hat{a} = \arg \min_{s \in S_\epsilon} \mathcal{L}(s)$

3. Complexity Analysis: The total computational cost per step, T_{step} , is the sum of the evaluation costs for all candidates in the representative set S_ϵ . $T_{\text{step}} = \sum_{s \in S_\epsilon} \text{Cost}(\text{evaluate } s)$. Given that the evaluation cost per candidate is $T_{\text{eval}}(\epsilon)$, we have: $T_{\text{step}} = |S_\epsilon| \cdot T_{\text{eval}}(\epsilon)$

Substituting the bound from Step 1: $T_{\text{step}} \leq C \cdot T_{\text{eval}}(\epsilon)$

Since C is a constant determined by the topology of the condensed concept (and not the raw input size N), and $T_{\text{eval}}(\epsilon)$ is a local operation: $T_{\text{step}} = O(1)$

Thus, the inference time is constant with respect to the scale of the global problem, proving that metric collapse leads to amortized constant-time inference. \square

Proof of Theorem 1. By the **Compressibility** assumption, for every level $k \geq 0$ we have

$$N(\epsilon, \mathcal{M}_{k+1}) \leq \rho^{-1} N(\epsilon, \mathcal{M}_k). \quad (1)$$

Applying (1) repeatedly yields a telescoping bound. Concretely, $N(\epsilon, \mathcal{M}_1) \leq \rho^{-1} N(\epsilon, \mathcal{M}_0)$, $N(\epsilon, \mathcal{M}_2) \leq \rho^{-1} N(\epsilon, \mathcal{M}_1) \leq \rho^{-2} N(\epsilon, \mathcal{M}_0)$, and by induction, assume $N(\epsilon, \mathcal{M}_k) \leq \rho^{-k} N(\epsilon, \mathcal{M}_0)$. Then

$$\begin{aligned} N(\epsilon, \mathcal{M}_{k+1}) &\leq \rho^{-1} N(\epsilon, \mathcal{M}_k) \\ &\leq \rho^{-1} \cdot \rho^{-k} N(\epsilon, \mathcal{M}_0) = \rho^{-(k+1)} N(\epsilon, \mathcal{M}_0). \end{aligned} \quad (2)$$

Therefore, for all $D \geq 0$, $N(\epsilon, \mathcal{M}_D) \leq \rho^{-D} N(\epsilon, \mathcal{M}_0)$, which proves the first claim.

For the budgeted statement, suppose we require $N(\epsilon, \mathcal{M}_D) \leq d$ for some representational budget $d \in \mathbb{N}$. A sufficient condition is $\rho^{-D} N(\epsilon, \mathcal{M}_0) \leq d$. Rearranging gives $\rho^D \geq \frac{N(\epsilon, \mathcal{M}_0)}{d}$. Taking $\log_\rho(\cdot)$ on both sides (noting $\rho > 1$) yields $D \geq \log_\rho\left(\frac{N(\epsilon, \mathcal{M}_0)}{d}\right)$. Since D must be an integer depth, it suffices to choose $D \geq \lceil \log_\rho\left(\frac{N(\epsilon, \mathcal{M}_0)}{d}\right) \rceil$, as claimed. Hence, when $N(\epsilon, \mathcal{M}_0)$ grows with stream length (e.g., $L \rightarrow \infty$), the required depth grows only logarithmically in the initial covering number, establishing that bounded representational demand can be maintained by increasing hierarchy depth. \square

Proof of Theorem 2. Existence of f follows from Urysohn's lemma. By construction, f is constant on \sim_f -equivalence classes, hence $\bar{f}([x]) := f(x)$ is well-defined. The quotient universal property implies \bar{f} is continuous and satisfies $f = \bar{f} \circ q$. Since $f \equiv 0$ on A and $f \equiv 1$ on B , each set is contained in a single equivalence class, so $q(A)$ and $q(B)$ are singleton points with \bar{f} -values 0 and 1. \square

Proof of Theorem 3. Step 1: Homological Orthogonality. The Parity Principle posits that \mathcal{M}_{odd} represents the space of 1-cycles (active flows) and $\mathcal{M}_{\text{even}}$ represents the space of

0-cycles (connected components/points). In algebraic topology, these distinct homology groups (H_1 and H_0) are generators of orthogonal subspaces in the total phase space. No continuous deformation within the connected component of a cycle (in H_1) alters the position of disconnected components (in H_0), provided the boundary condition $\partial\gamma = 0$ is met.

Step 2: The Parity-Inverting Map. Let the learning operator be $\Psi(\gamma_t) = p_{\text{new}}$, where $\gamma_t \subset \mathcal{M}_{\text{odd}}$ and $p_{\text{new}} \in \mathcal{M}_{\text{even}}$. The Jacobian of this transformation describes how changes in the input affect the memory. Since \mathcal{M}_{odd} and $\mathcal{M}_{\text{even}}$ are disjoint during the active inference phase (separated by the wake/sleep or search/navigation phase transition), the cross-term of the metric tensor is zero.

Step 3: Invariance. Let $S_{\text{old}} \subset \mathcal{M}_{\text{even}}$ be the set of previously learned concepts. The collapse of a new trajectory γ_{new} involves a contraction of the metric g_{odd} in the flow space. Since S_{old} resides entirely in the metric space g_{even} , and the coupling is unidirectional ($\Psi : \text{odd} \rightarrow \text{even}$), the contraction of g_{odd} does not induce stress or displacement on the geodesic distances in g_{even} .

Step 4: Conclusion. The capacity of $\mathcal{M}_{\text{even}}$ to store linearly separable points is limited only by the volume of $\mathcal{M}_{\text{even}}$ (Theorem 1) and is *independent* of the dynamic complexity required to generate those points in \mathcal{M}_{odd} . Therefore, the system effectively avoids catastrophic interference. \square

Proof of Theorem 4. We proceed by induction on the hierarchy level k . **Base Case** ($k = 0$): Since X_0 is normal and $A_0, B_0 \subset X_0$ are disjoint closed sets, Urysohn's Lemma yields a continuous function $f_0 : X_0 \rightarrow [0, 1]$ such that $f_0(A_0) = \{0\}$ and $f_0(B_0) = \{1\}$.

Inductive Step: Assume there exists a continuous $f_k : X_k \rightarrow [0, 1]$ with $f_k(A_k) = \{0\}$ and $f_k(B_k) = \{1\}$. Let $q_k : X_k \rightarrow X_{k+1} = X_k / \sim_k$ be the quotient map, and assume compatibility: $x \sim_k x' \implies f_k(x) = f_k(x')$. Then f_k is constant on equivalence classes, so the function

$$f_{k+1} : X_{k+1} \rightarrow [0, 1], \quad f_{k+1}(q_k(x)) := f_k(x)$$

is well-defined. Moreover, by the universal property of quotient maps, f_{k+1} is continuous and satisfies $f_{k+1} \circ q_k = f_k$.

Separation at level $k+1$: For $y \in A_{k+1} = q_k(A_k)$, pick $x \in A_k$ with $q_k(x) = y$. Then $f_{k+1}(y) = f_k(x) = 0$. Similarly, $f_{k+1}(y) = 1$ for all $y \in B_{k+1} = q_k(B_k)$. In particular, compatibility prevents q_k from identifying any $a \in A_k$ with $b \in B_k$ (otherwise $0 = f_k(a) = f_k(b) = 1$), so A_{k+1} and B_{k+1} remain disjoint.

Thus f_{k+1} separates A_{k+1} and B_{k+1} . By induction, the claim holds for all k . \square