METHODS AND SYSTEMS FOR STRUCTURAL INTERFEROMETRY OF INFORMATION SYSTEMS VIA PRINCIPAL COMPONENT ABLATION AND ALGEBRAIC LATTICE DECOMPOSITION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Provisional Patent Application No. 63/983,234, filed February 14, 2026, entitled "Coherent Semantic Interferometry: A Method for Measuring Structural Sensitivity of Language Model Encoders Using Algebraic Reference Lattices," the entire disclosure of which is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to methods for characterizing the structural information geometry of vector representations produced by systems operating on structured input data, and more particularly to interferometric methods that employ principal component ablation to separate structural information from dominant variance in representation spaces. The methods require only that structured input be projectable into a vector space amenable to principal component analysis; the representation may be produced by learned systems (neural network encoders), engineered transformations (such as mel-frequency cepstral coefficient extraction), raw co-occurrence statistics, or any other mapping that yields dense vector representations. Applications include but are not limited to natural language encoders, audio and music encoders, bioacoustic signal processors, combinatorial symbol system analyzers, protein and genomic sequence encoders, multimodal representation systems, and any future system or transformation that projects structured input into a representation space.

BACKGROUND OF THE INVENTION

Trained computational systems operating on structured input data---including but not limited to natural language encoders, audio and music encoders, protein sequence models, and multimodal representation systems---produce dense vector representations (embeddings) of their inputs. These embeddings reside in high-dimensional vector spaces and are widely used in downstream tasks including classification, information retrieval, generation quality assessment, and structural analysis. The present invention addresses a fundamental question common to all such systems: the nature and distribution of information within these learned representations, and in particular whether qualitatively different types of information---structural versus distributional---occupy separable subspaces of the embedding space.

The present invention assembles its method from established analytical techniques, each with extensive independent validation. Principal component analysis (PCA) is a standard linear algebraic decomposition introduced by Pearson (1901) and Hotelling (1933). Interferometric methods---the measurement of structure through the difference between two projections of the same signal---are foundational in physics from Michelson (1887) onward. Algebraic lattice theory provides the formal framework for partially ordered sets of binary features (Birkhoff, 1933). The Ising model (Ising, 1925; Onsager, 1944) provides a statistical mechanical framework for analyzing pairwise coupling in binary systems. None of these components is novel. The novelty of the present invention resides in their combination and application to embedding spaces produced by internally coherent information systems, and in the surprising empirical finding that this combination reveals.

In the domain of natural language processing, prior work has established the "Linear Representation Hypothesis" (Park et al., 2023), which posits that high-level semantic concepts are encoded as linear directions in the representation space. However, existing approaches treat the embedding space as a monolithic object and do not distinguish between qualitatively different types of information that may reside in different subspaces. The distinction between structural information (pertaining to the logical, syntactic, and organizational properties of the input) and distributional information (pertaining to the statistical frequency, register, and surface-level identity of the input) has not been addressed at the level of subspace decomposition in any domain.

Anonymous et al. (2026), in "Symmetry in Language Statistics Shapes the Geometry of Model Representations" (arXiv:2602.15029), prove that translation symmetries in pairwise word co-occurrence statistics govern the emergence of geometric structures in learned representations, including circular geometries for cyclical concepts and smooth manifolds for continuous sequences. Their framework provides a theoretical organizing principle for the geometric phenomena measured by the present invention. Korchinski et al. (2025) further demonstrate that binary semantic attributes induce parallelogram geometry in representation space, providing theoretical support for the geometric structure of binary lattice vertices. However, neither Anonymous et al. nor Korchinski et al. address the decomposition of representation geometry into subspaces, the diagnostic question of which subspace carries structural versus distributional information, or the interferometric method of the present invention. The present invention extends beyond these theoretical accounts by operating in the penumbral subspace---the representation geometry that persists after removal of the dominant statistical axis---which is not predicted or addressed by co-occurrence symmetry models.

Mu and Viswanath (2018) demonstrated that removing dominant principal components from word embeddings improves downstream task performance, but employed this removal as an engineering optimization rather than as a diagnostic instrument for revealing information geometry. Wang et al. (2020) ("Double-Hard Debias") extend this by identifying that dominant principal components encode word frequency and removing them before debiasing. Both works observe that removing dominant components can improve representations. Neither work uses the removal as a measurement instrument. Neither work inputs a binary structural property and outputs a ternary classification of that property. Neither work computes a delta between reduced and full representations as the primary observable. Neither work defines or measures the penumbral subspace. The distinction is between intervention (prior art: remove components to improve representations) and measurement (present invention: remove a component to determine where a property lives).

Most recently, Xiong et al. (2026) demonstrated that linear attribute directions in LLM embeddings induce concept lattices via half-space intersections, providing independent confirmation that algebraic lattice structure is present in learned representations. However, the Xiong framework operates exclusively in the full embedding space without subspace decomposition, recovers semantic concept hierarchies (noun taxonomies) rather than structural morphosyntactic features, and relies on externally supplied ontological ground truth (WordNet) rather than discovering lattice structure from production statistics. Although Xiong et al. contemplate extension of their framework to grammatical features and verb classes, they do not perform such extension; their experimental validation is restricted to noun-concept taxonomies in the full embedding space. No existing method provides a principled interferometric diagnostic for determining which subspace of a system's representation carries structural information, which carries distributional information, and what the geometric relationship between these subspaces is.

Independently, recent work on the robustness of large reasoning models has revealed that structural coherence degrades under adversarial pressure in ways that existing confidence-based metrics fail to detect. Li, Krishnan, and Padman (2026) demonstrate that reasoning models exhibit five distinct failure modes under multi-turn attack---including self-doubt, social conformity, and reasoning fatigue---and that confidence-aware defense frameworks fail because the model's self-assessed confidence is decoupled from its actual structural consistency. This decoupling is consistent with the finding of the present invention that the dominant variance direction (which carries confidence and fluency signals) is orthogonal to the subspace where structural information resides.

In the domain of representation geometry, Sevetlidis et al. (2026) introduce "representation holonomy," a gauge-invariant statistic that composes Procrustes rotations around closed loops in input space to detect curvature in neural network feature spaces. The present invention is distinguished from Sevetlidis et al. in three respects. First, the loops of the present invention are defined in structural feature space (the algebraic lattice of binary structural codes) rather than in input space. Second, the present invention operates in the penumbral subspace after principal component ablation, not in the full representation space. Third, the holonomy measure of the present invention recovers product group structure (SU(2)xSU(2)) and contemplates higher symmetry groups, whereas Sevetlidis et al. measure holonomy as a scalar deviation from the identity in SO(d).

Furthermore, in systems where principal component analysis is applied to embedding spaces, the dominant principal component is widely assumed to carry the most important information, as it accounts for the largest share of total variance. The present invention discloses the surprising and counter-intuitive finding that this dominant component carries zero structural information and actively degrades both structural and fine-grained classification when included. This finding, validated across multiple system architectures trained with fundamentally different objectives, on typologically diverse input data, and on non-linguistic substrates including music and combinatorial symbol systems, has significant implications for the design of evaluation instruments and the deployment of embedding-based systems in high-stakes domains.

A critical limitation of any single-encoder measurement---whether probing, ablation, or geometric analysis---is that the resulting classification may reflect encoder-specific geometric artifacts rather than properties of the underlying structure. The present invention addresses this limitation through a multi-encoder convergence criterion (Section XI herein), establishing that a structural classification is reliable only when multiple independent encoders converge on the same verdict. This convergence criterion is itself a methodological contribution with no precedent in the prior art; existing probing and analysis methods operate on single encoders without systematic cross-encoder validation.

SUMMARY OF THE INVENTION

The present invention provides methods and systems for structural interferometry of information systems through principal component ablation (hereinafter, the "penumbra method" or "PCA ablation diagnostic"). The invention is based on the discovery that structural information and distributional information in the embedding spaces of internally coherent information systems occupy orthogonal subspaces that can be separated by principal component analysis, with structural information residing predominantly in intermediate principal components (PCs 2--50, hereinafter the "penumbral subspace") and distributional variance residing in the first principal component (PC1). The penumbral subspace exhibits topological-geometric structure---including subspace decomposition, holonomy, and product group symmetry---that is invariant under changes of the producing system's architecture, training objective, and input domain. This discovery has been validated across multiple system architectures, training objectives, input domains, and signal substrates---including natural language text, baroque music audio, and combinatorial symbol systems---demonstrating that the orthogonality of structural and distributional information is a general property of any internally coherent information system whose outputs can be accurately classified in conformity with one or more algebraic lattices, rather than an artifact of any particular architecture or input modality.

The method is grounded in an observability principle: the intrinsic dynamics of information within a representation system cannot be measured from within the system's dominant variance direction. The first principal component encodes the axis of the system's optimization objective and provides no exterior vantage for structural observation. The algebraic lattice serves as an extrinsic reference frame---a fixed structure against which information dynamics become measurable. Principal component ablation removes the intrinsic axis, positioning the measurement in the penumbral subspace where structural organization becomes visible as the interference pattern between the system's representations and the fixed lattice. This is literal interferometry: two arms (embedding signal, lattice reference), difference measurement (classification gap), fringe detection (penumbra signature).

The invention further provides a multi-encoder convergence criterion establishing that a binary structural property is classified as structural only when two or more independent encoders---differing in architecture, training objective, or computational mechanism---converge on the same penumbral verdict. This criterion was discovered empirically: adversarial validation experiments demonstrated that a known surface feature (word count in language; tempo in music) can appear penumbral on a single encoder whose dominant variance direction happens not to be aligned with that feature. Cross-encoder convergence resolves this ambiguity. The convergence criterion applies to all embodiments of the measurement method and is a necessary condition for structural inference.

The invention further provides methods for assembling verified structural properties into lattice frameworks (lattice assembly protocol), extracting selection rules from lattice coupling topology (selection rule extraction via SVD on log-odds residuals), characterizing geometric coupling between structural features via holonomy (Wilson loop analysis), iteratively resolving hierarchical variance structure to expose structural properties at successive strata (nested peeling protocol), and characterizing unknown combinatorial systems without a neural network encoder (combinatorial diagnostic battery). Each of these methods has been demonstrated on at least one substrate; several have been demonstrated on multiple substrates with cross-substrate convergence of key observables including the Ising block coupling ratio (4.9--6.3x across three substrates).

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

"Information system" refers to any system that produces dense vector representations of structured input data, including but not limited to: neural network encoders employing self-attention mechanisms (BERT, RoBERTa, ALBERT, DeBERTa, ELECTRA, XLNet), autoregressive language models (GPT-2, LLaMA), gated recurrence models (RWKV), encoder components of encoder-decoder architectures (T5-encoder, BART-encoder), audio and music encoders (wav2vec2, CLAP), engineered feature extractors (MFCC), protein and genomic sequence encoders (ESM, ProtTrans), multimodal representation systems, and any future system architecture that produces vector representations of structured input. A system is "internally coherent" for purposes of this disclosure if its vector representations can be classified against the labels of an algebraic lattice with accuracy significantly exceeding a permutation-based null distribution, as determined by the method of Section II.

"Embedding" or "vector representation" refers to the dense vector produced by an information system for a given input item, extracted at any specified processing stage of the system. For natural language encoders, this is typically the representation at the [CLS] token position or a pooled representation at a specified hidden layer. For audio encoders, this is typically the mean-pooled hidden state across the time dimension. For engineered feature extractors (e.g., MFCC), this is the concatenation of summary statistics across feature dimensions. For other information systems, the embedding is the analogous output vector at the appropriate extraction point.

"Structural feature" refers to any binary or categorical property pertaining to the logical, syntactic, organizational, or compositional structure of the input, as distinct from its distributional, topical, or surface-level identity properties. For natural language, examples include but are not limited to: verb transitivity, voice (active/passive), mood, aspect, polarity, scope, and agency. For music, examples include harmonic mode, texture (homophonic/polyphonic), and ensemble organization. For combinatorial symbol systems, examples include positional character class, token length class, and frequency class. For other domains, structural features are the analogous organizational properties appropriate to the input type.

"Algebraic lattice" refers to a partially ordered set of structural codes formed by the binary structural features assigned to input items, wherein each code corresponds to a vertex of a binary hypercube defined by the feature dimensions, and the partial ordering is induced by bitwise inclusion. When a classifier assigns N binary structural features, the resulting lattice is an N-dimensional binary hypercube with 2^N vertices. The frequency distribution of items across lattice vertices constitutes the "mass table" of the lattice and is subject to statistical mechanical analysis including Ising decomposition.

"Principal component ablation" or "PCA ablation" refers to the process of removing one or more top principal components from a set of embeddings by projecting the centered embeddings onto the subspace orthogonal to the removed components.

"Penumbral subspace" refers to the subspace of a representation space defined by principal components 2 through N (where N is typically 50), after removal of the first principal component. This subspace provides the exterior vantage point from which structural information dynamics become observable, as the dominant variance direction (PC1) is intrinsic to the system's optimization objective and affords no measurement position. The term "penumbra" reflects the interferometric principle: structure becomes visible only from outside the dominant light, in the region where the reference lattice and representation signal produce measurable interference.

"Penumbra signature" or "penumbra effect" refers to the empirical finding that classification accuracy for structural features increases when the first principal component is removed (K1 gap > K0 gap), indicating that structural information is concentrated outside the dominant variance direction and that the dominant variance direction introduces noise that interferes with structural classification. The difference Delta = K1 gap - K0 gap quantifies the penumbra effect. A positive Delta indicates PENUMBRA (structural). A negative Delta indicates ANTI-PENUMBRA (surface). A Delta near zero indicates FLAT (uninformative). This ternary verdict is the primary observable of the measurement method.

"Multi-encoder convergence" refers to the requirement that a structural classification verdict (PENUMBRA, ANTI, or FLAT) be confirmed by two or more independent information systems before it is accepted as a property of the underlying structure rather than an artifact of a particular encoder's geometry. This criterion was established empirically by the adversarial validation experiments of Sections X and XIII herein.

"Null model" refers to a permutation-based baseline in which class labels are randomly shuffled and the classification procedure is repeated multiple times (typically 1,000 permutations), generating a distribution of accuracy scores under the null hypothesis that the embeddings carry no structural information.

II. The PCA Ablation Method

A. Corpus Preparation and Structural Labeling

The method begins with a corpus of structured input items, each labeled with one or more binary structural features conforming to an algebraic lattice as defined herein. In the preferred embodiment for natural language, structural labels are assigned by a morphism classifier that analyzes the dependency parse of each sentence to determine the values of a set of binary structural bits. These bits encode features including but not limited to: phase polarity (b0), internal agency (b1), coupling transitivity (b2), external modification (b3), scope (b4), and epistemic/culmination status (b5). Together these six binary features define a 6-dimensional algebraic lattice with 64 vertices (structural codes), each representable as a two-digit hexadecimal code.

The structural labeling may alternatively be derived from Universal Dependencies morphological annotations, including Voice (Active/Passive), Mood (Indicative/Subjunctive), VerbForm (Finite/Infinitive), Tense (Past/Present), Number (Singular/Plural), Polarity (Positive/Negative), and language-specific features such as ergative case marking. For non-linguistic domains, the structural labeling is provided by a domain-appropriate classifier that assigns binary organizational features to input items (e.g., harmonic function classifiers for music, positional co-occurrence statistics for symbol systems), subject to the requirement that the classifier exhibits internal coherence as validated by the calibration methods disclosed herein.

B. Embedding Extraction

Structured input items are tokenized or otherwise preprocessed and processed through the information system. For each input item, an embedding vector is extracted at the appropriate extraction point of the system. For bidirectional encoder architectures, the embedding is typically extracted from the [CLS] token position at the final hidden layer. For autoregressive and causal architectures (including attention-based and gated recurrence models), the embedding is obtained by mean-pooling the hidden states across all token positions at the final layer. For audio encoders, the embedding is obtained by mean-pooling the hidden states across the time dimension. For engineered feature extractors, the embedding is the summary statistic vector as defined by the extraction method. The resulting embedding matrix X has dimensions N x D, where N is the number of input items and D is the hidden dimension of the information system.

C. PCA Ablation

Principal component analysis is performed on the centered embedding matrix. Let X_centered = X - mean(X). The singular value decomposition X_centered = U S V^T yields the principal components as rows of V^T, ordered by decreasing singular value. For ablation level K, the top K rows of V^T are zeroed, and the embeddings are reconstructed in the ablated subspace: X_ablated = X_centered * V_ablated^T * V_ablated, where V_ablated has its top K rows set to zero.

Three conditions are evaluated: K=0 (no ablation, baseline), K=1 (first principal component removed, the standard penumbra test), and K=-1 (projection onto the first principal component only, the negative control).

D. Classification and Null Model

In the preferred embodiment, classification is performed using a nearest-centroid classifier with cosine distance. This classifier has zero learnable parameters, preventing any possibility of overfitting and ensuring that any classification accuracy above chance is attributable to the geometric structure of the embeddings rather than to classifier expressiveness. Stratified 10-fold cross-validation is used.

A null distribution is generated by randomly permuting the structural labels and repeating the classification procedure. In the preferred embodiment, 1,000 permutations are performed, yielding a null distribution of 1,000 accuracy values. The gap (real accuracy minus null mean) and a permutation-based p-value (fraction of null accuracies exceeding real accuracy) are computed.

E. Penumbra Determination

The penumbra signature is present when all three of the following conditions are met: (i) The K0 gap (baseline accuracy minus null) is positive and statistically significant ($p < 0.05$), indicating that the full embeddings carry structural information; (ii) The K1 gap exceeds the K0 gap (K1 gap > K0 gap), indicating that removal of PC1 improves structural classification; and (iii) The PC1-only gap is at or below zero ($p > 0.05$), confirming that the first principal component alone carries no structural information.

III. Experimental Validation: English Medical Corpus

The method was validated on a corpus of 32,120 medical sentences derived from the Pile, classified into 64 hexadecimal structural categories by the morphism classifier (v2). This initial validation corpus established the method. All filing-grade results reported in Sections IV through IX are derived from a multi-register corpus of 4.4 million sentences from the Pile (seed=42, 13 source subsets), with 1,000-permutation null models, as described in Section IV. Three encoder architectures were evaluated: BERT-base-uncased (110M parameters, masked language modeling), ALBERT-base-v2 (12M parameters, factorized masked language modeling), and RoBERTa-base (125M parameters, dynamic masked language modeling). All p-values reported are based on permutation testing with 1,000 null shuffles.

A. Balanced-Bit Penumbra Results

For bits with class balance between 40% and 60% (b2, b3, b4), the penumbra signature was observed in 7 of 9 encoder-bit combinations:

Encoder	Bit	Balance	K0 Gap	K1 Gap	Delta	Verdict	p
BERT	b2	56/44	+0.098	+0.100	+0.002	Penumbra	0.000
BERT	b3	56/44	+0.103	+0.116	+0.013	Penumbra	0.000
BERT	b4	57/43	+0.118	+0.118	0.000	Flat	0.000
ALBERT	b2	56/44	+0.113	+0.102	-0.011	Anti	0.000
ALBERT	b3	56/44	+0.070	+0.098	+0.028	Penumbra	0.000
ALBERT	b4	57/43	+0.083	+0.112	+0.029	Penumbra	0.000
RoBERTa	b2	56/44	+0.085	+0.099	+0.014	Penumbra	0.000
RoBERTa	b3	56/44	+0.074	+0.094	+0.020	Penumbra	0.000
RoBERTa	b4	57/43	+0.116	+0.131	+0.015	Penumbra	0.000

The two non-penumbra results are informative rather than contradictory: BERT b4 is exactly flat (Delta = 0.000), indicating that scope information is perfectly orthogonal to PC1 in BERT, neither helped nor hurt by ablation. ALBERT b2 shows a mild anti-penumbra (Delta = -0.011), indicating that ALBERT's factorized bottleneck architecture folds some coupling information into the dominant component. These three response patterns---penumbra, flat, and anti-penumbra---constitute a richer diagnostic taxonomy than a simple binary finding.

B. PC1-Only Negative Control

When embeddings were projected onto the first principal component alone, all structural classification collapsed to or below chance across all encoders and all bits:

Encoder	b2 Gap	b3 Gap	b4 Gap	p range
BERT	-0.001	-0.001	-0.001	>0.6
ALBERT	-0.003	-0.003	-0.004	>0.9
RoBERTa	-0.004	-0.005	-0.007	>0.7

This result establishes that the first principal component of information system embeddings carries zero structural information.

IV. Extended Encoder Validation

The penumbra method was validated across eleven encoder architectures spanning five training paradigms and four fundamentally different computational mechanisms:

Family	Encoder	Training Objective	Penumbra Score
MLM bidirectional	BERT-base	Masked LM	3/3 balanced bits
MLM bidirectional	ALBERT-base	Factorized MLM	3/3 (1 anti)
MLM bidirectional	RoBERTa-base	Dynamic MLM	2/3
MLM bidirectional	DeBERTa-base	Disentangled attn	2/3
Distilled	DistilBERT	Knowledge distill.	1/3

Discriminative	ELECTRA-small	Replaced token det.	3/3
Permutation	XLNet-base	Permutation LM	3/3
Causal	GPT-2	Causal LM	1/3
Gated recurrence	RWKV	Gated linear rec.	2/3

The penumbra effect generalizes across all five training paradigms. ELECTRA (discriminative, 3/3) and XLNet (permutation-based, 3/3) demonstrate that the penumbra is not an artifact of masked language modeling. GPT-2 (causal, 1/3) demonstrates that even unidirectional autoregressive models exhibit penumbral structural information, though at reduced intensity. RWKV (gated linear recurrence, 2/3) demonstrates that the penumbra effect is not restricted to attention-based architectures; a gated recurrence mechanism without self-attention produces penumbral signatures at rates comparable to bidirectional attention models. Together, the GPT-2 and RWKV results establish that the penumbra is a property of learned representation geometry rather than an artifact of any particular computational mechanism.

V. Cross-Linguistic Validation

To determine whether the penumbra effect is language-universal or specific to English, the method was applied to monolingual encoders for French and Basque using structural labels derived from Universal Dependencies morphological annotations.

Language	Feature	K0 Gap	K1 Gap	Delta	Verdict	p (PC1)
French	Voice	+0.160	+0.241	+0.081	Penumbra	>0.6
French	Number	+0.208	+0.221	+0.013	Penumbra	>0.6
French	Gender	+0.264	+0.268	+0.004	Penumbra	>0.6
Basque	VerbForm	+0.212	+0.193	-0.019	Anti	--
Basque	Number	+0.254	+0.199	-0.055	Anti	--

French Voice (Delta = +0.081) represents the strongest penumbra effect measured in any language or encoder, exceeding all English results. Basque (BERTeus encoder) shows anti-penumbra on both tested features, replicating the ALBERT b2 pattern observed in English. The anti-penumbra is itself a diagnostic finding: it identifies architectures where the penumbral subspace is less informative.

VI. Lattice Geometry Generalization

The penumbra method was tested on higher-order lattice structures constructed from pairs of structural bits. Z3 (ternary, hexagonal lattice from b2 x b3) and Z4 (quaternary, octagonal lattice from b2 x b3 x b4) classifications were performed.

Encoder	Z3 K0	Z3 K1	Z3 Pen.	Z4 K0	Z4 K1
BERT	+0.099	+0.113	Yes	+0.140	+0.151
ALBERT	+0.062	+0.102	Yes (+65%)	+0.128	+0.144

RoBERTa	+0.077	+0.086	Yes	+0.117	+0.117 (flat)

The penumbra effect persists on higher-order lattice structures, demonstrating that the information geometry of the penumbral subspace supports ternary and quaternary classification, not just binary distinctions.

VII. Layer-Wise Structural Profiling

To characterize how structural information emerges through the processing depth of an information system, the penumbra measurement was performed at every layer of multiple encoder architectures on the Pile validation corpus with 1,000-permutation null models. Three encoder families were profiled: BERT (bert-base-uncased, 12 layers), RoBERTa (roberta-base, 12 layers), and ALBERT (albert-base-v2, 12 layers). For each encoder, the penumbra test was applied to three structural features (b2 Coupling, b4 Scope, and b5 Resolution) at each layer from L0 through the final layer, yielding nine profiling conditions per architecture.

A. Universal Layer 0 Null. At Layer 0 (raw token embeddings, prior to any contextual mixing), the penumbra test returns zero or indistinguishable-from-null gap on all structural features across all nine conditions and all three architectures. This universal L0 null establishes that the interferometer is measuring dynamics, not static properties: at L0, information has not yet flowed through the system, and there is no structural differentiation to observe.

B. Differential Layer-Wise Profiles. Different encoder architectures construct structural information at different depths and with different trajectories. BERT shows rapid onset of penumbra at L1, with peak structural signal in mid layers (L5--L8) and compression at L12. RoBERTa shows a similar onset but sustains penumbra through L12 without the terminal compression observed in BERT. ALBERT, which shares parameters across layers, shows peak penumbra at mid-to-late layers (L5--L11) with a characteristic oscillatory profile consistent with the recurrent parameter-sharing architecture.

VIII. Topic Classification Under Ablation

A surprising secondary finding is that fine-grained topic classification also improves monotonically as more principal components are removed:

Encoder	K0	K1	K5	PCs 2-50
BERT	0.169	0.178	0.200	0.115
ALBERT	0.053	0.059	0.075	0.041
RoBERTa	0.193	0.230	0.245	0.151

This demonstrates that the first principal component is not merely orthogonal to structural information but also constitutes noise with respect to fine-grained semantic discrimination within a single register. Three distinct information regimes are identified: PC1 (corpus-level noise), PCs 2--50 (structural logic), and PCs 50+ (fine-grained semantic identity).

IX. Wilson Loop Holonomy Analysis

To characterize the geometric coupling between structural features within the penumbral subspace, Wilson loop integrals were computed over pairs of structural bits. Wilson loop integrals were computed on a Pile-provenance sample (12,800 rows, 200/code x 64 codes, seed=42, 1,000-permutation null models, all p = 0.000). Aggregate holonomy: K0 W = 0.408, K1 W = 0.376, K5 W = 0.100.

The Wilson loop W(i,j) measures the holonomy---the path-dependent phase accumulated when traversing a closed loop defined by the binary values of bits i and j in the penumbral subspace. The per-bit-pair measure W(i,j) is computed as follows. For a pair of structural bit indices (i, j), each vertex v of the 6-bit hypercube defines a plaquette: the four vertices obtained by toggling bits i and j. Denote these vertices v_00, v_01, v_11, v_10 (subscripts indicate the values of bits i and j). Let c(v) be the centroid embedding for vertex v. Two opposite-edge vectors are computed: d_01 = c(v_01) - c(v_00) and d_32 = c(v_11) - c(v_10). Each is normalized to unit length. The plaquette holonomy is the cosine similarity between these opposite edges: W = u_01 . u_32. A value of W = 1.0 indicates flat connection (zero holonomy); W < 1.0 indicates curvature (non-parallel transport); W = 0 indicates orthogonal transport; W < 0 indicates antiparallel transport (maximum holonomy). Statistical significance is assessed by permuting the centroid-to-vertex mapping 1,000 times.

Bit Pair	K0 W	K1 W	Interpretation
(1,2) Agc.xCpl.	0.049	0.047	Highest curvature; cross
(3,4) Phs.xScp.	0.105	0.105	High curvature; within-B
(1,4) Agc.xScp.	0.133	0.135	High curvature; cross
(0,1) Pol.xAgc.	0.451	0.366	Moderate curvature; within-A
(2,5) Cpl.xRes.	0.658	0.532	Low curvature; cross
(0,5) Pol.xRes.	0.756	0.643	Near-flat; within-A

A. Product Group Structure. A clear product structure emerges: within-triad bit pairs show near-zero holonomy (locally Euclidean geometry), while cross-triad pairs show high and often growing holonomy under deeper ablation. This pattern is consistent with an SU(2)xSU(2) product group structure.

B. Ablation-Dependent Holonomy Growth. Several cross-triad pairs exhibit dramatic holonomy growth under deeper ablation, consistent with structural geometry being layered beneath multiple variance directions.

C. Higher Symmetry: Three-Body Coupling Evidence. The Ising decomposition of the 64-code mass table at the quadrupole level achieves R^2 = 0.963, capturing 96.3% of the variance. Extending to the octupole level---adding twenty three-body coupling terms---raises the fit to R^2 = 0.996. The

additional 3.3% of variance captured by three-body terms is irreducible: it cannot be decomposed into combinations of pairwise interactions. Bit b1 (Agency) is the hub of both pairwise and three-body coupling.

D. Greek Middle Voice: Empirical Confirmation of Triadic Structure. Greek BERT produced penumbral gaps of +0.118 (K0) and +0.138 (K1) at p < 0.001; XLM-R produced gaps of +0.087 (K0) and +0.094 (K1) at p < 0.001. In both encoders, Middle voice occupies its own geometric territory in the penumbral subspace---it is not decomposable into SU(2)xSU(2). English vestigial middle voice (roberta-base): K0 gap = +0.090, K1 gap = +0.158, PC1-only gap = +0.008 (at chance). The penumbral gap (+0.158) is nearly double the full-space gap, indicating that removal of the dominant semantic direction dramatically sharpens the three-way structural discrimination. The English encoder geometrically distinguishes a three-way voice contrast that the English language does not grammatically mark.

X. Adversarial Validation Battery

To establish that the penumbra measurement is not an artifact of the classification procedure and to determine the false positive rate of the instrument, an adversarial validation battery was performed on BERT-base-uncased and RoBERTa-base using the English Universal Dependencies Treebank corpus. Five label conditions were tested on the same corpus and encoder representations:

(1) Real structural labels (Tense: Past/Present). Expected verdict: PENUMBRA. (2) Shuffled labels (real Tense labels randomly permuted). Expected verdict: NULL. (3) Random binary labels (independent coin flips, no relationship to input). Expected verdict: NULL. (4) Known surface feature (word count, median-split). Expected verdict: ANTI-PENUMBRA. (5) Inverted real labels (Tense labels flipped). Expected verdict: equal magnitude to Test 1.

Each condition was run with 1,000 null permutations, 10-fold stratified cross-validation, nearest-centroid cosine classifier.

A. BERT-base Results.

Condition	K0 Gap	K1 Gap	PC1 Gap	Delta	Verdict
Real Tense	+0.234	+0.256	+0.054	+0.022	see below
Shuffled	-0.002	+0.005	--	~0	NULL
Random	+0.009	-0.004	--	~0	NULL
Word Count	+0.286	+0.169	--	-0.117	ANTI

BERT correctly classifies shuffled and random labels as null, and word count as anti-penumbral (Delta = -0.117). However, the real Tense result reveals an important subtlety. The Delta is positive (+0.022): ablation improves structural classification, and K1 > K0. But the PC1-only gap is also positive and significant (+0.054, p = 0.000), indicating that BERT carries tense information in both the penumbral subspace and in PC1. Under the three-condition penumbra definition of Section

II.E, this result fails condition (iii) and does not qualify as PENUMBRA. Tense information on BERT is split across subspaces rather than cleanly separated.

This split is itself a diagnostic finding. It demonstrates that even a genuine structural feature can fail the penumbra test on a given encoder when that encoder's PC1 partially encodes the feature. The result parallels ALBERT b2 (Section III.A), where a structural feature folds into the dominant component due to the architecture's bottleneck. The implication is the same in both cases: a single encoder's penumbral verdict is insufficient for structural inference, because the relationship between a structural feature and PC1 is encoder-specific. This finding directly motivates the multi-encoder convergence criterion of Section XI.

B. RoBERTa-base Results: Critical Finding.

Condition	K0 Acc	K0 Gap	K1 Acc	K1 Gap	Delta	Verdict
Shuffled	--	--	--	--	~0	NULL
Random	--	--	--	--	~0	NULL
Word Count	--	--	--	--	+0.093	PENUMBRA

RoBERTa correctly returns null on shuffled and random labels. However, RoBERTa classifies word count as PENUMBRA (Delta = +0.093), the opposite of BERT's verdict (Delta = -0.117) on the identical feature measured on the identical data. The same surface property appears structural on one encoder and anti-structural on another.

This divergence is not an error. It is a geometric fact: RoBERTa's PC1 is not aligned with word count. Removing PC1 in RoBERTa does not remove the surface signal; therefore the surface feature survives ablation and appears penumbral. BERT's PC1 is aligned with word count; removing it damages the surface classification and correctly identifies the feature as anti-penumbral.

C. Implication. A single-encoder penumbral reading is geometrically accurate---it correctly describes the relationship between the feature and that encoder's dominant variance direction---but semantically ambiguous. It cannot distinguish "this feature is structural" from "this feature happens to be orthogonal to this encoder's PC1." The resolution requires a second encoder. If two independent encoders, with different architectures and different PC1 alignment, both classify a feature as penumbral, the convergent verdict is robust to encoder-specific geometric artifacts. This finding motivates the multi-encoder convergence criterion of Section XI.

XI. Multi-Encoder Convergence Criterion

Based on the adversarial findings of Section X and the cross-encoder experiments of Section XIII, the present invention establishes that multi-encoder convergence is a necessary condition for structural inference from penumbral measurements. The criterion is defined as follows:

A binary structural feature is classified as STRUCTURAL with respect to an input domain if and only if it is classified as PENUMBRA (Delta > 0, $p < 0.05$) by at least two independent information systems differing in architecture, training objective, or computational mechanism.

A binary structural feature for which independent encoders produce divergent verdicts (PENUMBRA on one, ANTI or FLAT on another) is classified as AMBIGUOUS and attributed to encoder-specific geometric alignment rather than to the underlying structure of the input domain.

This criterion was established by two independent lines of evidence on two independent substrates:

(1) Language (G10): BERT classified word count as ANTI (Delta = -0.117). RoBERTa classified word count as PENUMBRA (Delta = +0.093). Same feature, same data, opposite verdict. The convergence criterion correctly rejects word count as structural.

(2) Baroque music (G13, Section XIII): MFCC classified Tempo as PENUMBRA (Delta = +0.092). wav2vec2 classified Tempo as ANTI (Delta = -0.004). Same feature, same data, opposite verdict. The convergence criterion correctly rejects Tempo as structural.

In both cases, the divergent feature is a known surface property (word count is a distributional statistic; tempo is a temporal amplitude feature). The convergence criterion correctly filters surface features that happen to be orthogonal to one encoder's PC1 but aligned with another's.

The convergence criterion has a retroactive consequence for all results reported in this specification. The Gutenberg calibration (R^2 = 0.996) is meaningful because five independent encoders (BERT, RoBERTa, GPT-2, RWKV, DeBERTa) converge on the same penumbral bits. No single encoder achieves that fit on its own terms. The convergence is the measurement.

XII. Substrate Extension: Baroque Music

To establish that the penumbra method is not specific to natural language, the method was applied to baroque music audio---a structured signal domain with well-defined organizational properties but no linguistic content.

A. Structural Lattice. A 6-bit structural lattice was defined for baroque music audio:

Bit	Property	0	1
b0	Harmonic Mode	Minor	Major
b1	Rhythmic Meter	Simple	Compound
b2	Texture	Homophonic	Polyphonic
b3	Ensemble Size	Chamber	Orchestral
b4	Tempo Register	Slow	Fast
b5	Dynamic Range	Terraced	Graduated

B. Corpus. 512 synthetic baroque audio segments, 8 per lattice code, deterministic synthesis (seed=42). Controlled harmonic templates, counterpoint simulation, amplitude envelopes. 22,050 Hz

sample rate. Each segment 2--6 seconds depending on tempo bit. Classification by zero-parameter threshold-based acoustic feature classifier: mode via chromagram template correlation, meter via onset autocorrelation, texture via spectral peak count, ensemble via spectral bandwidth, tempo via beat tracking, dynamics via RMS gradient.

C. MFCC Embedding (Encoder #1). MFCC-40 with delta and delta-delta coefficients, 7 summary statistics per coefficient. Embedding dimension: 840. No neural network. No learned weights. 200 null permutations, 10-fold stratified CV, nearest-centroid cosine classifier.

Bit	K0 Gap	K1 Gap	PC1 Gap	Delta	p(K1)	Verdict
b0:Mode	+0.334	+0.354	+0.023	+0.020	0.000	PENUMBRA
b1:Meter	-0.026	+0.019	-0.004	+0.045	0.275	FLAT
b2:Texture	+0.491	+0.084	+0.481	-0.407	0.000	ANTI
b3:Ensemble	+0.322	+0.324	+0.011	+0.002	0.000	FLAT
b4:Tempo	+0.092	+0.184	-0.010	+0.092	0.000	PENUMBRA
b5:Dynamics	+0.370	+0.372	-0.016	+0.003	0.000	FLAT

Texture is heavily ANTI (PC1 carries 98% of texture information via spectral peak count). Mode and Tempo appear PENUMBRA on MFCC.

D. Real Audio Validation (Bach Well-Tempered Clavier). The method was further validated on 600 30-second WAV chunks from the Bach Well-Tempered Clavier (Wanda Landowska recording), 38 source tracks. 5-bit quintic lattice (Mode/Tempo/Tone/Key/Chromaticism), 1,000 null permutations, MFCC-40 (160-dim embedding). 5 of 8 candidate bits showed PENUMBRA (all spectral/harmonic); 3 of 8 showed ANTI (all temporal/amplitude---an MFCC blind spot). Ising R^2 = 0.786 on the 4-bit mass table. Coupling matrix eigendecomposition revealed a triplet block (Tone-Key-Chromaticism, Frobenius norm = 0.666) and a doublet block (Mode-Tempo, Frobenius norm = 0.129), yielding a block ratio of 5.2x.

E. wav2vec2 Embedding (Encoder #2, G13). To test multi-encoder convergence on baroque, the identical synthetic corpus (seed=42, 512 segments) was embedded using facebook/wav2vec2-base (pretrained, frozen weights, mean-pooled over time, 16 kHz resampling, embedding dimension 768). Hardware: Lambda Labs A100-SXM4-40GB. Compute time: approximately 4 minutes.

Bit	MFCC Delta	MFCC Verdict	wav2vec2 Delta	wav2vec2 Verdict	Convergence
b0:Mode	+0.020	PENUMBRA	+0.029	PENUMBRA	CONVERGE
b1:Meter	+0.045	FLAT	+0.004	ANTI	--
b2:Texture	-0.407	ANTI	+0.004	PENUMBRA	DIVERGE
b3:Ensemble	+0.002	FLAT	+0.067	PENUMBRA	UPGRADE
b4:Tempo	+0.092	PENUMBRA	-0.004	ANTI	DIVERGE
b5:Dynamics	+0.003	FLAT	-0.032	ANTI	--

wav2vec2 raw measurements:

Bit	K0 Acc	K0 Gap	K1 Acc	K1 Gap	PC1 Gap	p(K1)
b0:Mode	0.688	+0.187	0.717	+0.216	-0.023	0.000
b1:Meter	0.545	+0.045	0.549	+0.048	+0.021	0.075
b2:Texture	0.895	+0.393	0.897	+0.397	+0.158	0.000
b3:Ensemble	0.670	+0.170	0.736	+0.237	+0.118	0.000
b4:Tempo	0.908	+0.410	0.904	+0.405	+0.011	0.000
b5:Dynamics	0.547	+0.041	0.514	+0.009	-0.001	0.430

F. Cross-Encoder Findings. Three bits converge as structural across both encoders: Mode (spectral/harmonic), Texture (spectral), and Ensemble (spectral). Two bits diverge: Meter and Tempo are PENUMBRA or FLAT on MFCC but ANTI on wav2vec2. Both are temporal features. MFCC's time-domain sensitivity made them appear structural; wav2vec2, which learns its own feature hierarchy, correctly places them on PC1. This is the identical pattern observed in G10 (Section X): a surface feature appearing penumbral on a single encoder. The "sector boundary" reported in subsection D (Mode resisting rotation on MFCC) was an MFCC artifact: wav2vec2 detects Mode as structural (Delta = +0.029) where MFCC saw it as FLAT.

G. Rotation Projection (MFCC only). Three lightweight projection architectures (MLP, RNN, Attention) were trained to rotate penumbral information into the dominant subspace for each structural bit. 200 epochs, CPU. 4 of 5 bits rotated successfully (Delta flipped negative). Mode resisted rotation across all three architectures: MLP Delta = -0.008 (FLAT), consistent with the Ising block partition. The architecture ordering MLP > RNN > Attention was observed, confirming that attention is not required for structural rotation. The Mode rotation resistance was subsequently explained by G13: it was an MFCC blind spot, not a structural sector boundary. On wav2vec2, Mode is PENUMBRA.

XIII. Substrate Extension: Voynich Manuscript

To establish that the measurement methods of the present invention apply to combinatorial symbol systems without any neural network encoder, the combinatorial diagnostic battery was applied to the Voynich Manuscript (Beinecke MS 408), a fifteenth-century manuscript written in an undeciphered script.

A. Method. The analysis operates on the EVA (European Voynich Alphabet) transcription of the manuscript, comprising approximately 29,000 tokens. No neural network encoder is used. The structural lattice is derived entirely from raw co-occurrence statistics---bigram frequencies, positional distributions, and token-level features.

B. Structural Lattice. A 6-bit lattice was defined from observable statistical properties of the token vocabulary:

Bit	Property	Description
b0	Initial character class	First character of token
b1	Token length class	Short/long binary split
b2	Final character class	Last character of token
b3	Position in line	Early/late
b4	Entropy class	Low/high character diversity
b5	Frequency class	Rare/common

C. Mass Table and Ising Decomposition. The 64-cell mass table (all 64 codes populated) was subjected to Ising decomposition. The quadrupole model (single-field terms plus 15 pairwise coupling constants) achieved R^2 = 0.977. Strongest coupling: length-frequency J = -0.788. Isolated block: initial-position J = -0.342. Block ratio: 4.9x.

D. Selection Rule Extraction. SVD on the log-odds residual co-occurrence matrix revealed bipartite alpha/beta charge families. Cross-family pairing is blocked. 64.6% of residual variance concentrates in the top 3 singular vectors.

E. Combinatorial Diagnostic Battery. The following diagnostics were computed:

(i) Normalized mutual information: NMI = 0.401. Shuffle baseline: 0.025. Natural language range: 0.6--0.8. The Voynich system is 16x above random but below natural language, consistent with a constrained combinatorial system.

(ii) Grid fill: 15.9% of the co-occurrence grid is populated, indicating sparse positional constraints.

(iii) Type-token ratio: TTR = 0.090, indicating high token repetition.

(iv) Triad mass: c_7 = 5.1%, far below all natural languages tested.

(v) Verbose cipher test: conditional entropy H(X|X_{-1}) = 2.535 versus 2.557 for shuffled text. The 0.8% reduction is indistinguishable from noise, rejecting the verbose cipher hypothesis.

(vi) Currier A/B languages: Factor 2 of NMF applied to the positional co-occurrence matrix recovers the Currier A/B language distinction blind---without any prior information about sections or scribal hands. The coupling between positional co-occurrence features reverses sign between Herbal-A sections (J = -1.546) and Herbal-B sections (J = +1.141).

(vii) Vocabulary rank: NMF decomposition indicates a rank-4 system (9 x 2 x 5 x 3 ~ 270 core types). The 2,628 observed types in the EVA transcription represent a 9.7x expansion from the core vocabulary, consistent with a notation system that generates surface forms through constrained positional composition of a small number of structural primitives. The expansion mechanism is combinatorial: a core type appearing in different line positions with different co-occurrence contexts produces distinct observed variants.

F. Classification. The combinatorial battery classifies the Voynich system as structured, non-random, pairwise-dominated, with bipartite selection rules and rank-4 vocabulary generation. This is consistent with a notation system or constrained code and inconsistent with natural language (which would require an encoder to measure V in the penumbral subspace), a random hoax (which would not produce the coupling structure), or a verbose cipher (which is rejected at 0.8%).

XIV. Cross-Substrate Block Ratio Convergence

Three independent substrates---each analyzed with independent methods on independent data---exhibit Ising block coupling structure with convergent block ratios:

Substrate	Strong Block	Weak Block	Ratio	Ising R^2
Language (Gutenberg, 6-bit)	Agency-Coupling-Phase	Polarity-Scope-Resolution	6.3x	0.996
Baroque (WTC, 5-bit)	Tone-Key-Chromaticism	Mode-Tempo	5.2x	0.786
Voynich (EVA, 6-bit)	len-ent-fin-freq	init-pos	4.9x	0.977

Each substrate decomposes into a strongly coupled block and a weakly coupled or isolated block. The block ratio range (4.9--6.3x) is narrow given the diversity of substrates. The pattern is: every structured information domain tested exhibits a dominant coupling cluster and an isolated element, with the coupling asymmetry falling in the 5--6x range.

The Voynich strongest coupling (length-frequency, J = -0.788) is structurally analogous to Agency-Coupling in language (the strongest coupling in that domain). The Voynich isolated block (initial-position) is structurally analogous to Mode in baroque (the isolated element in that domain before G13 convergence testing). Whether this cross-substrate convergence of block topology reflects a universal property of structured information systems or a coincidence of the substrates tested is an open empirical question that the present methods are designed to investigate.

XV. Rotation Projection and Sector Boundary Analysis

To determine whether structural information, once identified in the penumbral subspace, can be projected into the dominant subspace by a lightweight learned transformation, rotation experiments were performed on both language and baroque substrates.

A. Method. A lightweight neural network (MLP, RNN, or Attention architecture) is trained to predict structural bit labels from encoder embeddings, with the training objective being binary cross-entropy on each structural bit. After training, the penumbra measurement is repeated on the network's output representations. If the Delta flips from positive (penumbral) to negative, the structural information has been successfully rotated from the penumbral subspace into the dominant subspace. Bits that resist rotation across all three architectures are identified as occupying a different geometric sector.

B. Language Results (G1 and G1b). Two rotation experiments were performed on language. In G1, three Universal Dependencies morphological features (Tense, Number, VerbForm) were tested on two encoders (BERT-base, RoBERTa-base) across three architectures (MLP, GRU, Attention), yielding 18 conditions. All 18 showed successful rotation. BERT Tense: Delta dropped from +0.022 to -0.473 (MLP). BERT Number: Delta dropped from +0.013 to -0.496 (MLP).

In G1b, the original 6-bit structural lattice (Agency, Coupling, Phase, Polarity, Scope, Resolution) was tested. All six bits rotated on all three architectures. Strongest rotation: b4_scope MLP Delta = -0.453. Weakest: b2_coupling MLP Delta = -0.065. No sector boundary was found in language encoders across either experiment.

C. Baroque Results (MFCC). On the 5-bit baroque lattice, 4 of 5 bits rotated. Mode resisted rotation across all three architectures (MLP Delta = -0.008, FLAT). Architecture ordering: MLP > RNN > Attention on MFCC, confirming that attention is not required for structural rotation.

D. Sector Boundary Dissolution (G13). The Mode rotation resistance on MFCC was subsequently tested on wav2vec2 (Section XII.E). wav2vec2 classifies Mode as PENUMBRA (Delta = +0.029). The rotation resistance observed on MFCC was an artifact of the MFCC feature extraction, not a property of the baroque structural lattice. MFCC cannot represent harmonic mode in its cepstral statistics with sufficient resolution for the penumbra method to detect it as structural; wav2vec2, which learns its own feature hierarchy, resolves Mode as structural.

This finding has a general implication: sector boundaries observed on a single encoder should be treated as hypotheses pending multi-encoder convergence testing. When a bit resists rotation on one encoder but is classified as PENUMBRA on another, the resistance is attributed to the encoder's representational limitations, not to a structural property of the lattice.

XVI. Nested Peeling Protocol

When a corpus contains hierarchical variance structure---multiple binary surface variables that dominate the embedding space at different scales---the penumbra method applied directly may classify all candidate structural bits as aliases of the dominant surface split. The nested peeling protocol resolves this by iteratively removing surface variance layers.

A. Method. (1) Apply the penumbra battery (Section II) to all candidate structural bits. (2) If multiple bits converge to the same K0 accuracy, this diagnostic convergence indicates that a single dominant surface variable is being detected by all bits simultaneously. (3) Filter the corpus to one side of the dominant binary split (e.g., retain only orca vocalizations, discarding background noise). (4) Re-apply the penumbra battery to the filtered corpus. (5) Repeat until penumbral bits emerge or the battery returns null.

B. Feasibility Demonstration (Orca Bioacoustics). The nested peeling protocol was designed for application to orca bioacoustic recordings, where hierarchical variance structure is expected: the dominant split between orca vocalizations and background noise, followed by within-vocalization

splits (clicks vs. tonal calls), followed by structural properties within call types. Preliminary exploration confirmed the expected hierarchical structure: an initial round of the penumbra battery identified a dominant binary surface split (orca vs. noise) with multiple candidate structural bits converging to the same baseline accuracy, diagnostic of an unresolved surface variable. After filtering to orca vocalizations only, a second surface split (clicks vs. tonal) emerged as the new dominant variable. Full experimental validation of the peeling protocol on bioacoustic data, with filing-grade null models and multi-encoder convergence, is deferred to a continuation.

C. Diagnostic Significance. The convergence of multiple bits to the same K0 is itself a diagnostic: it identifies an unresolved surface variable that must be peeled before structural measurement can proceed. The peeling protocol is the method for resolving hierarchical variance structure in substrates where multiple binary surface variables dominate the embedding space.

XVII. Cross-Lingual Factorization Diagnostic

To characterize how the penumbral subspace interacts with cross-lingual transfer, experimental runs were performed across multiple encoder architectures and validation sets spanning typologically diverse languages.

A. Method. For each encoder, monolingual structural validity preservation was measured on English and on typologically distant validation sets. Cross-lingual grafting experiments measured whether structural information factorized from distributional information degrades when the encoder processes typologically distant input.

B. Failure Mode Taxonomy. Three distinct failure modes were identified, each corresponding to a training objective class:

(i) Contrastive encoders (e.g., multilingual sentence transformers): signal collapse. The penumbral signature attenuates toward null, as the contrastive objective aligns all structural information along the dominant similarity direction.

(ii) Discriminative encoders (e.g., mBERT): entanglement. Structural and distributional information partially merge, producing intermediate Delta values that are neither cleanly penumbral nor cleanly anti-penumbral. A notable anomaly was observed: mBERT's cross-lingual structural factorization was stronger than its monolingual factorization, attributed to mBERT's equal-sampling training regime which prevents any single language's distributional statistics from dominating PC1.

(iii) Generative encoders (e.g., mT5): explosive coupling. On logographic input, the validity vector fuses with PC1 rather than remaining orthogonal. The structural signal does not disappear; it merges with the dominant variance direction, making the penumbra method inapplicable in this configuration.

C. Language Typology Hypothesis. Alphabetic writing systems produce strong structural factorization. Logographic writing systems produce systematic degradation. The writing system, through its effect on tokenization and co-occurrence statistics, constrains the encoder's ability to

separate structural from distributional information. Full quantitative characterization of these failure modes with filing-grade null models across ten or more typologically diverse languages is deferred to a continuation.

XVIII. Training Validation

To validate the penumbra-aware training method, three training conditions were evaluated on BERT-base (110M parameters) using 70,000 structurally classified Pile sentences over 3 epochs on a single A100 GPU. The PCA projection defining the penumbral subspace was pre-computed from vanilla BERT-base and held fixed during training.

A. Structural Classification at the Lattice Level. Continued pre-training with masked language modeling plus a six-bit structural classification head operating on PCs 2--50. SFI Delta improved from +0.043 to +0.046. The number of structurally engaged bits (PEN) increased from 3/6 to 4/6.

B. Weakness-Weighted Structural Classification. Same architecture with each bit's contribution to the classification loss weighted inversely to the model's current accuracy on that bit. SFI Delta improved to +0.048. b3 gap reached +0.128 versus +0.121 under uniform weighting.

C. Geometric Angle Targeting. The loss function targeted the precipice angle theta_1 directly. theta_1 moved from 79.9 degrees to 82.1 degrees over 3 epochs. However, SFI Delta dropped from +0.043 to +0.023. The structural content of the representations degraded even as the geometric summary statistic improved. The precipice angle is a diagnostic observable, not a causal lever.

D. Resolution: The Lattice Is the Training Signal. The three conditions establish that the effective training signal operates at the level of per-feature structural classification within the lattice, not at the level of geometric summary statistics. Training on the lattice directly (conditions A and B) improves structural sensitivity. Training on the angle (condition C) degrades it.

E. The Saturated/Frontier Partition. Three bits (b0, b1, b5) are saturated: vanilla BERT-base achieves 79--86% accuracy. Three bits (b2, b3, b4) are frontier: vanilla BERT-base achieves 50--61% accuracy. All training gains concentrate on frontier bits. The partition provides a diagnostic for lattice refinement.

F. Penumbra Sensitivity and Benchmark Performance. Across eleven validated architectures, penumbra sensitivity correlates with both GLUE (Spearman rho = 0.836, p = 0.001) and BLiMP (Spearman rho = 0.836, p = 0.003, n = 9). Model size alone shows no significant correlation (r = 0.255, n.s.). XLNet-base (permutation LM, SFI Delta = +0.082) achieves the highest penumbra sensitivity, consistent with the finding that structural factorization, not model capacity, drives penumbral separation.

XIX. Contemplated Applications

A. Structural Coherence Evaluation Platform. The disclosed methods may be embodied in an automated system for evaluating the structural coherence of any structured input processed by an

information system. In such a system, a reference corpus of structurally vetted input items is embedded and the penumbral subspace is characterized. New input items are embedded, projected into the penumbral subspace, and compared against the reference distribution. Items exhibiting anomalous structural signatures are flagged for review. Such a system would have particular utility in domains where structural coherence failures have material consequences, including legal (brief logic), medical (clinical note structure), financial (regulatory filing integrity), and scientific (experimental report organization). In a further contemplated embodiment, the method is applied to monitor the structural coherence of reasoning model outputs across successive turns of a multi-turn interaction, detecting the onset of structural degradation (reasoning fatigue, self-doubt, social conformity) before the model produces an overtly incorrect response.

B. Information System Selection and Architecture Evaluation. The penumbra profile---comprising the per-bit penumbra gaps, the stage-wise structural fingerprint, and the holonomy curvature map---constitutes a comprehensive structural characterization of an information system architecture. Practitioners selecting an information system for deployment in a structurally sensitive application may use the disclosed method to compare candidate architectures.

C. Cross-Linguistic Structural Diagnostics. The demonstrated cross-linguistic applicability enables structural evaluation of multilingual and cross-lingual systems. The convergence of penumbral geometry across typologically diverse languages and independently trained monolingual encoders suggests the existence of universal structural invariants in the penumbral subspace.

D. Extension to Non-Linguistic Information Domains. The disclosed method is applicable to any domain in which an information system produces dense vector representations of structured input and a domain-appropriate structural classifier assigns binary features conforming to an algebraic lattice. Demonstrated non-linguistic applications include baroque music audio (Section XII) and combinatorial symbol systems (Section XIII). Contemplated extensions include protein and genomic sequence analysis, bioacoustic signal processing (with feasibility demonstrated in Section XVI.B), and multimodal systems.

E. Thermodynamic Characterization of Structural Domains. The Ising decomposition method enables characterization of any structured input domain through the coupling topology of its structural classification. Different input domains are expected to exhibit different coupling topologies, different gatekeeper dimensions, and different starvation patterns. The cross-substrate convergence of block ratios (4.9--6.3x, Section XIV) provides a benchmark for this characterization.

F. Higher-Order Symmetry in Coupled Information Systems. The three-body coupling evidence (Section IX.C) and the triadic voice classification evidence (Section IX.D) demonstrate that the geometric content of even a single system exceeds the SU(2)xSU(2) product group. In contemplated extensions involving coupled information systems---for example, an encoder coupled with an active agent that modifies its input in response to structural feedback---the coupling between systems may introduce additional degrees of freedom.

G. Future Experimental Directions. The following experimental extensions are contemplated: (i) monolingual encoder validation across ten or more typologically diverse languages to establish the universality of penumbral geometry as a candidate geometric invariant of natural language structure; (ii) stage-wise profiling across all validated architectures to complete the structural fingerprint catalog; (iii) Wilson loop holonomy computation on class-balanced structural features; (iv) extension of the morphism classifier to additional structural dimensions beyond the current six-bit scheme; (v) application of the interferometric method to non-linguistic information systems; (vi) investigation of the relationship between penumbral geometry and the Lattice Representation Hypothesis (Xiong et al., 2026); (vii) expansion of the vestigial middle voice finding to additional languages and architectures; (viii) penumbra-aware fine-tuning, validated experimentally in Section XVIII, with the iterative refinement cycle (measure, train, diagnose, refine the lattice, repeat) constituting a method for progressive improvement of structural fidelity; and (ix) application of the interferometric method to canonical texts with demonstrated cultural persistence as a lattice validation protocol.

CLAIMS

What is claimed is:

1. A method for structural interferometry of an internally coherent information system, the method comprising:

(a) obtaining a plurality of vector representations from the information system, each vector representation corresponding to a structured input item in a corpus, wherein each input item is associated with a binary structural feature label conforming to an algebraic lattice;

(b) performing principal component analysis on the plurality of vector representations to identify principal components ordered by decreasing variance;

(c) generating an ablated representation for each vector representation by projecting the vector representation onto a subspace orthogonal to at least the first principal component;

(d) classifying the ablated representations against the structural feature labels using a classifier to obtain an ablated classification accuracy;

(e) generating a null distribution by repeating the classification of step (d) with randomly permuted labels;

(f) determining that structural information resides in the ablated subspace when the ablated classification accuracy significantly exceeds the null distribution.

2. The method of claim 1, further comprising:

(g) generating a PC1-only representation for each vector representation by projecting the vector representation onto only the first principal component;

(h) classifying the PC1-only representations against the structural feature labels to obtain a PC1-only classification accuracy;

(i) confirming that the first principal component carries no structural information when the PC1-only classification accuracy is at or below the null distribution mean.

3. The method of claim 2, further comprising determining that a penumbra signature is present when: (i) the ablated classification accuracy of step (d) exceeds a baseline classification accuracy obtained from the unablated vector representations; and (ii) the PC1-only classification accuracy of step (h) is at or below chance level.

4. The method of claim 1, wherein the classifier of step (d) is a nearest-centroid classifier using cosine distance, having zero learnable parameters.

5. The method of claim 1, wherein the method is performed on a plurality of distinct information system architectures, including systems employing different computational mechanisms, and wherein a structural feature is classified as structural only when the penumbra signature is present across the plurality of architectures.

6. The method of claim 1, wherein the binary structural feature labels are determined by automated analysis of the dependency parse of each sentence, the analysis comprising identification of the root verb and classification of morphological properties including one or more of: transitivity, voice, mood, aspect, polarity, and scope.

7. The method of claim 1, wherein the binary structural feature labels are derived from Universal Dependencies morphological annotations associated with each sentence.

8. A method for generating a stage-wise structural profile of an information system, the method comprising:

(a) for each processing stage in the information system from the first stage through the final stage, extracting a vector representation for each input item in a corpus at that stage;

(b) for each processing stage, performing the method of claim 1 to obtain a penumbra measurement for one or more structural features;

(c) constructing a stage-wise profile comprising the penumbra measurement for each structural feature at each processing stage;

(d) wherein different structural features may exhibit different characteristic depth profiles across the processing stages, constituting a structural fingerprint of the information system architecture.

9. The method of claim 8, wherein the stage-wise profile reveals that the penumbra measurement at the initial processing stage is zero or indistinguishable from the null distribution for all structural

features, establishing that structural information is constructed by the system's learned transformations and is not present in the static input representation.

10. A method for characterizing geometric coupling between structural features in an information system's representation, the method comprising:

(a) obtaining ablated representations as in claim 1;

(b) for each pair of structural features (i, j), computing a holonomy measure by evaluating the path-dependent geometric phase accumulated when traversing a closed loop defined by the binary values of features i and j in the ablated subspace;

(c) identifying pairs of structural features with high holonomy as geometrically coupled and pairs with near-zero holonomy as geometrically independent;

(d) wherein the pattern of coupling and independence across structural feature pairs constitutes a curvature map of the information system's structural representation.

11. The method of claim 1, wherein the corpus comprises input items in a non-English natural language, an audio signal domain, or a combinatorial symbol system, and wherein the structural feature labels are derived from domain-appropriate structural classifiers, thereby enabling cross-domain structural diagnosis.

12. The method of claim 1, wherein the structural feature labels are multi-class labels constructed from combinations of binary structural features, enabling diagnosis of structural information content for ternary, quaternary, or higher-order classification tasks in the ablated subspace.

13. The method of claim 1, further comprising classifying the ablated representations against topic labels to determine that fine-grained topic discrimination improves under principal component ablation, thereby establishing that the dominant principal component constitutes noise with respect to both structural and fine-grained semantic classification within a single register.

14. The method of claim 13, wherein the results identify three distinct information regimes within the embedding space: a first regime comprising the first principal component and encoding corpus-level distributional properties, a second regime comprising intermediate principal components and encoding structural information, and a third regime comprising remaining principal components and encoding fine-grained semantic identity.

15. A system for evaluating structural coherence of structured input, the system comprising:

(a) an information system configured to produce vector representations for structured input items;

(b) a principal component ablation module configured to remove at least the first principal component from the vector representations;

(c) a structural classifier configured to classify the ablated representations against structural feature labels;

(d) a null model module configured to generate a permutation-based null distribution;

(e) a coherence evaluation module configured to compare the structural classification of the input against a reference distribution derived from a corpus of structurally vetted input items and to generate a structural coherence score.

16. The method of claim 1, wherein the ablation of step (c) removes K principal components, where K is greater than 1, and wherein increasing K reveals additional structural information, indicating that structural content extends beyond the immediately adjacent components to the dominant variance direction.

17. A method for characterizing the coupling topology of a structural classification scheme applied to a corpus of structured input items, the method comprising:

(a) classifying each input item in the corpus according to a set of N binary structural features, yielding a structural code for each item;

(b) computing the frequency distribution of structural codes across the corpus to obtain a mass table over the 2^N vertices of the algebraic lattice;

(c) fitting a statistical mechanical model to the log-frequency distribution, the model comprising at least single-field terms and pairwise coupling terms;

(d) extracting from the fitted model the field strengths corresponding to the marginal bias of each structural feature and the pairwise coupling constants corresponding to the interaction strength between each pair of structural features;

(e) identifying thermodynamic gatekeeper dimensions as those structural features whose field values partition the lattice into high-mass and low-mass regions;

wherein the coupling topology, field strengths, and gatekeeper dimensions constitute a characterization of the structural classification scheme that is independent of any particular information system and diagnostic of the structural organization of the input domain.

18. The method of claim 17, wherein the statistical mechanical model further comprises three-body coupling terms representing simultaneous interactions among triples of structural features, and wherein the explanatory power of the model including three-body terms significantly exceeds that of the pairwise-only model, indicating the presence of irreducible triadic structure in the structural classification scheme.

19. The method of claim 1, wherein the structural feature labels include a three-valued classification comprising three grammatically or organizationally distinct categories that are not

reducible to binary composition, and wherein the penumbral subspace measurement detects geometric structure corresponding to the three-valued classification that is not decomposable into a product of binary discriminations.

20. The method of claim 1, further comprising adjusting one or more parameters of the information system based on the penumbra measurement, wherein the adjusting is configured to increase the magnitude of the Delta, thereby sharpening the separation of structural information from distributional information within the penumbral subspace.

21. The method of claim 20, wherein the adjusting comprises computing, for each of a plurality of structural feature labels, a per-feature classification performance metric within the penumbral subspace and applying a training signal derived from the per-feature metrics, wherein the training signal operates at the level of individual structural feature classification rather than at the level of a geometric summary statistic derived from the penumbral subspace.

22. A method for validating structural classifications produced by the method of claim 1, the method comprising:

(a) applying the method of claim 1 to a binary structural feature using a first information system to obtain a first penumbral verdict;

(b) applying the method of claim 1 to the same binary structural feature on the same corpus using a second information system, the second system differing from the first in architecture, training objective, or computational mechanism, to obtain a second penumbral verdict;

(c) accepting the structural feature as structural only when both the first and second penumbral verdicts are PENUMBRA;

(d) classifying the structural feature as ambiguous when the first and second verdicts diverge;

wherein the convergence of independent information systems constitutes a necessary condition for structural inference and filters encoder-specific geometric artifacts from structural measurements.

23. The method of claim 22, further comprising:

(e) applying the method of claim 1 to a known surface feature of the corpus using the first and second information systems;

(f) confirming that at least one information system classifies the known surface feature as ANTI-PENUMBRA, establishing that the first principal component of that system is aligned with the surface feature;

(g) wherein divergence between information systems on the known surface feature validates the multi-encoder convergence criterion by demonstrating that single-encoder measurements are insufficient for structural inference.

24. A method for iteratively resolving hierarchical variance structure in a corpus, the method comprising:

(a) applying the method of claim 1 to a plurality of candidate structural features;

(b) identifying a dominant surface variable when a plurality of candidate features converge to substantially the same baseline classification accuracy;

(c) partitioning the corpus by the dominant surface variable;

(d) applying the method of claim 1 to the partitioned corpus;

(e) repeating steps (b) through (d) until penumbral structural features emerge or the method returns null across all candidates;

wherein each iteration removes one layer of surface variance, exposing structural information at successively deeper strata.

25. A method for characterizing an unknown combinatorial symbol system without a neural network encoder, the method comprising:

(a) computing a co-occurrence matrix from positional statistics of symbols in the system;

(b) computing normalized mutual information between symbol positions;

(c) computing grid fill fraction and type-token ratio;

(d) performing singular value decomposition on log-odds residuals of the co-occurrence matrix to extract charge families;

(e) performing non-negative matrix factorization to determine the effective rank of the vocabulary generation system;

(f) computing conditional entropy to test for verbose cipher structure;

wherein the battery of diagnostics (a) through (f) classifies the symbol system as natural language, cipher, notation system, or hoax based on the pattern of results across the battery.

26. The method of claim 25, further comprising:

(g) constructing an N-bit structural lattice from the statistical properties identified in steps (a) through (f);

(h) computing a mass table over the lattice vertices;

(i) fitting an Ising model to the mass table to extract coupling topology;

wherein the coupling topology characterizes the structural organization of the symbol system independently of its semantic content.

27. The method of claim 1, wherein the information system produces vector representations from audio signal input, and wherein the structural feature labels encode organizational properties of the audio signal including one or more of: harmonic mode, rhythmic meter, texture, ensemble organization, tempo register, and dynamic range.

28. The method of claim 1, wherein the method is applied to the same corpus using both a learned encoder and an engineered feature extractor, and wherein divergence between the learned encoder and the engineered extractor on a given structural feature identifies the feature as an artifact of the feature extraction method rather than a property of the underlying structure.

ABSTRACT

A method for structural interferometry of internally coherent information systems via principal component ablation and algebraic lattice decomposition. The method employs an algebraic lattice as an extrinsic reference frame, enabling measurement of information dynamics that are otherwise intrinsic and unobservable from within the system's dominant variance direction. The method discovers that structural information and distributional information in the vector representations produced by such systems occupy orthogonal subspaces separable by principal component analysis. Structural information resides in intermediate principal components (PCs 2--50, the "penumbral subspace"), while the first principal component carries zero structural information and actively degrades structural classification when included. The method is validated across eleven encoder architectures spanning five training paradigms and four computational mechanisms (bidirectional self-attention, permutation self-attention, causal self-attention, and gated linear recurrence), multiple natural languages, baroque music audio, and a combinatorial symbol system (the Voynich Manuscript). A multi-encoder convergence criterion establishes that structural inference requires agreement across independent encoders, filtering encoder-specific geometric artifacts demonstrated on both language and music substrates. Three substrates exhibit convergent Ising block coupling structure with ratios in the range 4.9--6.3x. The method enables automated structural coherence evaluation of structured input in domains including legal, medical, financial, scientific, musical, and cryptographic applications.