

# Compression Complementarity in Dual-Frame Information Theory

A. R. Wells

December 2025

## Abstract

Modern AI systems exhibit a duality in internal representations: symbolic, discrete structures (e.g., tokens, clusters) alongside harmonic, continuous structures (e.g., embeddings, manifolds). This paper develops the information-theoretic formulation of Dual-Frame Theory (DFT), a framework for characterizing this phenomenon. We assume signals arise from a latent source admitting incompatible symbolic (S-frame) and harmonic (T-frame) representations. Key contributions include dual-frame Kolmogorov complexities and entropies, a compression complementarity principle, and a dual-frame rate-distortion theory. DFT yields falsifiable predictions for neural networks, compression, and interpretable constraints relevant to cognition, tested via empirical proxies like cluster entropy and embedding rank. This work focuses specifically on the information-theoretic structure of DFT and its consequences for representation, compression, and learning.

## 1 Introduction

Modern artificial intelligence systems exhibit an intriguing duality in their internal representations. Neural networks simultaneously develop *symbolic, cluster-like, discrete structure*—such as token identities, categories, edges, or sparse activations—and *harmonic, continuous, phase-like structure*—such as linear embedding manifolds, analogy directions, spectral coherence, and smooth interpolation. These two representational modes are both indispensable, yet fundamentally different. Despite the ubiquity of this phenomenon across modalities and architectures, there is currently no unified theoretical explanation for why both forms of structure emerge together or why they exhibit a characteristic trade-off.

Classical information theory and algorithmic complexity provide two foundational perspectives. Shannon’s framework characterizes uncertainty and compressibility in terms of probability distributions over symbolic sequences, while Kolmogorov complexity characterizes the algorithmic compressibility of individual strings. However, neither theory—on its own—captures the empirical coupling between discrete symbolic structure and continuous harmonic structure observed in modern neural representations. Symbolic entropy says nothing about phase coherence; harmonic sparsity says nothing about cluster boundaries. Existing approaches model these phenomena independently rather than as two incompatible aspects of a single representational process.

### 1.1 A Dual-Representation Perspective

Dual-Frame Theory (DFT) introduces a unifying principle rooted entirely in representation and information theory. Every observable signal  $x$  is assumed to be generated by an underlying *latent source*  $Z$ , which may be a random variable, stochastic process, or structured object, depending on context. The key assumption is not about dynamics or motion, but about representation: the same source  $Z$  admits multiple, incompatible descriptive encodings, extending earlier dual-frame formulations into an explicitly information-theoretic setting.

Two distinct representation functionals act on the same source:

- $F_S$ , which extracts *symbolic, structural, or partition-based* information (e.g., tokens, clusters, discrete states);
- $F_T$ , which extracts *harmonic, spectral, or coherence-based* information (e.g., eigenmodes, continuous embeddings, spectral structure).

These representations emphasize fundamentally different organizing principles of the same data. In general, neither representation is invertible from the other, and no single encoding can be simultaneously optimal for both. This representational incompatibility is the foundation of Dual-Frame Theory.

No assumptions are made about the physical, temporal, or dynamical nature of  $Z$ ; it may represent a random variable, a stochastic process, a dataset, or a learned internal state of a model. The latent trajectory  $\sigma(\lambda)$  denotes a one-parameter indexing of correlated observations (e.g., time, training step, or abstract phase) underlying the observable signal. No physical interpretation is assumed.

The latent trajectory  $\sigma(\lambda)$  is not asserted as an ontological primitive. It is a representational device that unifies families of correlated observations under a single generative parameterization. Its utility is assessed indirectly through the success or failure of predicted dual-frame trade-offs, not through direct observation.

This work is self-contained within information theory and representation learning. While Dual-Frame Theory also admits physical interpretations, no physical assumptions are required here.

## 1.2 Admissible and Incompatible Representation Frames

Not all pairs of representation functionals  $(F_S, F_T)$  are admissible within Dual-Frame Theory. If  $F_S$  and  $F_T$  are jointly invertible or nearly so, no nontrivial complementarity can arise.

We therefore restrict attention to *incompatible frames*, defined as follows.

**Definition 1** (Incompatible representation frames). Two representation functionals  $F_S$  and  $F_T$  acting on a latent source  $Z$  are said to be *incompatible* if there exists no encoding of  $Z$  that simultaneously attains (up to additive constants) independent of  $Z$ , a minimal symbolic description length for  $F_S(Z)$  and a minimal harmonic description length for  $F_T(Z)$ .

Operationally, incompatibility may be characterized by a nontrivial mutual coherence or uncertainty bound between the effective bases or features induced by  $F_S$  and  $F_T$ . Canonical examples include time–frequency representations, symbolic partitions versus spectral bases, or clustering versus manifold coordinates.

Trivial or degenerate choices of  $(F_S, F_T)$ —such as identity mappings or mutually invertible transforms—are explicitly excluded, as they do not satisfy this criterion and yield no meaningful dual-frame trade-off.

## 1.3 A Dual-Frame Coherence Functional

While incompatibility was defined qualitatively as the impossibility of simultaneously minimizing symbolic and harmonic description lengths, it is useful to introduce a quantitative measure of *how incompatible* two representation frames are.

Let  $Z$  be a latent source with induced frame variables

$$S = F_S(Z), \quad T = F_T(Z).$$

We define the *dual-frame coherence* of the pair  $(F_S, F_T)$  as

$$\mu_{\text{DFT}}(F_S, F_T) := \frac{I(S; T)}{\min\{H(S), H(T)\}}, \quad (1)$$

with the convention that the ratio is zero whenever the denominator vanishes.

This quantity measures the degree to which symbolic and harmonic representations share information across the source ensemble, rather than conditioned on a fixed latent state. Perfectly invertible or nearly invertible frame pairs yield  $\mu_{\text{DFT}}$  close to unity under idealized measurement and discretization, indicating substantial redundancy between symbolic and harmonic descriptions, whereas strongly incompatible frames yield  $\mu_{\text{DFT}} \ll 1$ , reflecting minimal shared information across the two representations.

By construction,  $\mu_{\text{DFT}} \in [0, 1]$ . Intuitively:

- $\mu_{\text{DFT}} \approx 1$  indicates that the two frames are nearly redundant, sharing most of their information;
- $\mu_{\text{DFT}} \approx 0$  indicates strong incompatibility, with little mutual information between symbolic and harmonic descriptions.

This functional provides a graded notion of incompatibility. Perfectly invertible or nearly invertible frame pairs yield  $\mu_{\text{DFT}} \approx 1$  and are therefore excluded from the regime where dual-frame complementarity is nontrivial. Conversely, strongly incompatible frames satisfy  $\mu_{\text{DFT}} \ll 1$ , enforcing meaningful compression trade-offs.

The definition (1) is an information-theoretic analogue of mutual coherence measures used in sparse representation theory. In the special case where  $F_S$  and  $F_T$  correspond to coefficient maps in two orthonormal bases,  $\mu_{\text{DFT}}$  reduces (up to normalization) to the standard basis coherence that appears in discrete uncertainty principles.

**Theorem 1** (Coherence-modulated compression complementarity). *Let  $(F_S, F_T)$  be admissible representation frames with  $\mu_{\text{DFT}}(F_S, F_T) < 1$ . Then for any observable  $X$  induced by  $Z$ ,*

$$H_{\text{DFT}}(X) \geq (1 - \mu_{\text{DFT}}) \max\{H(S_X), H(T_X)\}. \quad (2)$$

*In particular, strong incompatibility ( $\mu_{\text{DFT}} \ll 1$ ) enforces a nontrivial lower bound on joint symbolic-harmonic compressibility.*

*Proof sketch.* By definition of  $\mu_{\text{DFT}}$ ,

$$I(S_X; T_X) \leq \mu_{\text{DFT}} \min\{H(S_X), H(T_X)\}.$$

Substituting into

$$H_{\text{DFT}}(X) = H(S_X) + H(T_X) - I(S_X; T_X)$$

yields

$$H_{\text{DFT}}(X) \geq H(S_X) + H(T_X) - \mu_{\text{DFT}} \min\{H(S_X), H(T_X)\} \geq (1 - \mu_{\text{DFT}}) \max\{H(S_X), H(T_X)\}.$$

□

## 1.4 Core Contributions

This paper introduces a new information-theoretic framework based on this dual-frame latent trajectory model. Our contributions are:

1. We define *dual-frame Kolmogorov complexities*  $K_S(x), K_T(x)$  and the combined dual-frame complexity  $K_{\text{DFT}}(x) = K_S(x) + K_T(x)$ , demonstrating that DFT strictly extends classical Kolmogorov complexity.
2. We introduce a *dual-frame entropy*  $H_{\text{DFT}}(X) = H(S) + H(T) - I(S; T)$  that captures both symbolic and harmonic compressibility.

Table 1: S- and T-Frame Manifestations in Neural Architectures

Aspect	S-Frame (Symbolic / Discrete)	T-Frame (Harmonic / Continuous)
Language Models	Vocabularies, token sequences, syntactic trees	Continuous embeddings, attention coherence, linear analogies
Vision Systems	Class labels, segmentation masks, sparse activations	Smooth manifolds, positional encodings, phase alignments
Mathematical Fingerprints	Voronoi partitions, sparse coding, decision boundaries	Eigenvalue spectra, linear transformations, curvature

3. We derive a *compression complementarity principle*: no encoding of a signal may simultaneously minimize symbolic and harmonic complexity. Formally, we show lower bounds of the form

$$K_S(x) K_T(x) \geq C(x) > 0,$$

and corresponding entropic inequalities.

4. We develop a dual-frame rate–distortion theory and characterize the DFT rate–distortion region  $R(D_S, D_T)$ .
5. We connect these theoretical results to concrete phenomena in machine learning, compression, and neuroscience.

The framework yields falsifiable predictions, detailed in Section 9.

DFT does not claim that all successful systems operate optimally on the dual-frame frontier, nor that symbolic–harmonic complementarity explains all representation trade-offs. Rather, it asserts that for systems requiring both discrete and continuous structure, joint compressibility is constrained in a way that single-frame theories do not capture.

## 2 Dual-Frame Representations in AI Systems

Modern machine-learning systems—especially large language models, vision transformers, and multimodal embedding architectures—exhibit a consistent structural feature: internal computation proceeds in two fundamentally different representational domains. Dual-Frame Theory (DFT) provides a geometric constraint-based account of this phenomenon by identifying two complementary interpretive frames of an underlying latent trajectory: a symbolic/structural S-frame and a harmonic/phase-coherent T-frame.

### 2.1 S- and T-Frame Manifestations in Neural Architectures

In contemporary neural architectures, S-frame structure manifests as discrete, symbolic, and categorical elements, while T-frame structure emphasizes continuous geometry and relations. Table 1 summarizes key examples.

In models with learned embeddings (e.g., language and vision transformers), training induces semantic clusters in embedding space, with nearly flat submanifolds within clusters and sharp transitions at boundaries [9]. This is emergent S-frame structure inside the T-frame embedding: cluster centers as symbolic representatives, boundaries as minimal coherence points.

Transformers maintain contextual phase through attention weights, positional encodings, and normalization, leading to smooth manifolds and phase continuity. These are T-frame manifestations, where phase relations encode meaning gradients, and discontinuities signal symbolic transitions.

Empirically, neural networks never collapse to a purely symbolic or purely harmonic representation. They require symbolic discreteness for expressivity and categorical outputs, and continuous coherence for generalization and robustness. In DFT, this coexistence arises because S and T are complementary projections of the same trajectory, yielding an intrinsic trade-off: simplifying one frame complicates the other.

### 3 Dual-Frame Kolmogorov Complexity

Classical Kolmogorov complexity  $K(x)$  measures the length of the shortest program that outputs a finite object  $x$  on a fixed universal prefix Turing machine [6]. This approach implicitly assumes a single representational domain or description language: all regularities are to be captured in a single symbolic description.

In Dual-Frame Theory, observable signals  $x$  are understood as encodings of an underlying latent source  $Z$  under different representation functionals.

Related approaches such as Minimum Description Length (MDL) similarly decompose description length into model and data components, but do so within a single representational frame and a single description language [5].

#### 3.1 Setup: Dual Universal Machines

Let  $U_S$  and  $U_T$  be prefix-universal Turing machines adapted to the S- and T-frame domains, respectively:

- $U_S$  operates over a coding domain natural for symbolic sequences (e.g., token strings, parse trees, discrete labels),
- $U_T$  operates over a coding domain natural for harmonic or phase-based representations (e.g., coefficient sequences, phases, low-dimensional manifold coordinates).

The requirement of universality ensures that each can simulate any other machine within its domain up to an additive constant, preserving the usual invariance properties of Kolmogorov complexity.

For a given observable  $x$ , we associate an underlying latent trajectory  $\sigma_x$  and define the S- and T-frame projections:

$$s_x := F_S(Z_x), \quad t_x := F_T(Z_x) \quad (3)$$

These are, by construction, distinct encodings of the same ontic object.

#### 3.2 Frame-Specific Kolmogorov Complexities

**Definition 2** (S-frame Kolmogorov complexity). Let  $x$  be a finite observable with associated S-frame projection  $s_x$ . The S-frame Kolmogorov complexity of  $x$  is

$$K_S(x) := \min\{|p| : U_S(p) = s_x\}, \quad (4)$$

where  $|p|$  is the length (in bits) of the prefix-free program  $p$ .

**Definition 3** (T-frame Kolmogorov complexity). Let  $x$  be a finite observable with associated T-frame projection  $t_x$ . The T-frame Kolmogorov complexity of  $x$  is

$$K_T(x) := \min\{|q| : U_T(q) = t_x\}. \quad (5)$$

Intuitively,  $K_S(x)$  and  $K_T(x)$  quantify the shortest descriptions of the symbolic and harmonic aspects of  $x$ , respectively, relative to universal machines well-adapted to those domains.

### 3.3 Dual-Frame Kolmogorov Complexity

**Definition 4** (Dual-frame Kolmogorov complexity). The dual-frame Kolmogorov complexity of an observable  $x$  is defined as

$$K_{\text{DFT}}(x) := K_S(x) + K_T(x). \quad (6)$$

The idea is that a complete algorithmic description of  $x$  in a dual-frame world must, in general, account for both the S-frame (symbolic) and T-frame (harmonic) projections of the underlying trajectory  $\sigma_x$ . The sum  $K_S(x) + K_T(x)$  is the minimal joint description length when the two domains are treated as complementary and irreducible.

### 3.4 Relation to Classical Kolmogorov Complexity

Let  $U$  be any fixed universal prefix Turing machine used to define the classical Kolmogorov complexity  $K(x)$ . We now relate  $K_{\text{DFT}}(x)$  to  $K(x)$ .

**Theorem 2.** *For any finite observable  $x$ , up to an additive constant independent of  $x$ ,*

$$K_{\text{DFT}}(x) \geq K(x). \quad (7)$$

*Proof sketch.* Consider a classical universal machine  $U$  that, on some program  $r$ , outputs a complete dual-frame description of  $x$ ; i.e., from  $U(r)$  one can algorithmically reconstruct both  $s_x$  and  $t_x$ . Then the classical Kolmogorov complexity  $K(x)$  (defined with respect to  $U$ ) satisfies

$$K(x) \leq |r| + c, \quad (8)$$

for some constant  $c$  capturing the decoding overhead from  $U(r)$  to the pair  $(s_x, t_x)$ .

On the other hand, any such  $r$  must in particular encode descriptions sufficient to reconstruct  $s_x$  on  $U_S$  and  $t_x$  on  $U_T$ . Thus, modulo fixed simulation overheads,

$$|r| \gtrsim K_S(x) + K_T(x), \quad (9)$$

where “ $\gtrsim$ ” denotes inequality up to an additive constant that does not depend on  $x$ . Combining the two bounds and absorbing constants into a single additive term yields

$$K(x) \lesssim K_S(x) + K_T(x) = K_{\text{DFT}}(x). \quad (10)$$

Therefore, up to the usual machine-dependent additive constant,  $K_{\text{DFT}}(x) \geq K(x)$ .  $\square$

This shows that classical Kolmogorov complexity systematically underestimates the joint algorithmic structure present across incompatible representations: it ignores either the harmonic structure encoded in  $K_T$  or the symbolic structure encoded in  $K_S$ .

### 3.5 Projection-Minimal Complexities and Trade-offs

In a dual-frame ontology, the same observable  $x$  may arise from many underlying scalar trajectories  $\sigma$  consistent with  $x$  as a projection. Different  $\sigma$  can yield different balances between symbolic and harmonic structure. This motivates the notion of projection-minimal complexities.

**Definition 5** (Projection-minimal frame complexities). For a fixed observable  $x$ , define

$$K_S^{\min}(x) := \min_{\sigma: G(\sigma)=x} K_S(x; \sigma), \quad (11)$$

$$K_T^{\min}(x) := \min_{\sigma: G(\sigma)=x} K_T(x; \sigma), \quad (12)$$

where  $G$  is the physical projection that produces  $x$  from  $\sigma$ , and  $K_S(x; \sigma)$ ,  $K_T(x; \sigma)$  are the frame complexities evaluated for a specific underlying trajectory  $\sigma$  consistent with  $x$ .

Intuitively,  $K_S^{\min}(x)$  is the best possible symbolic compression of  $x$  over all admissible ontic realizations, and likewise for  $K_T^{\min}(x)$ .

We now state a qualitative “no-free-lunch” result: no trajectory can attain simultaneously minimal S- and T-frame complexities.

**Theorem 3** (Compression complementarity, informal). *For any nontrivial observable  $x$ , there exists no underlying trajectory  $\sigma$  such that both*

$$K_S(x; \sigma) = K_S^{\min}(x), \quad K_T(x; \sigma) = K_T^{\min}(x) \quad (13)$$

*are simultaneously achieved.*

*Proof idea.* At the level of functional analysis, the S- and T-frame projections can be regarded as maps into dual (or approximately dual) bases: a symbolic basis with localized support in a discrete coordinate system, and a harmonic basis with localized support in a spectral or phase domain. For a given  $\sigma$ , its projections  $F_S(\sigma)$  and  $F_T(\sigma)$  satisfy a discrete uncertainty relation of the form

$$\|F_S(\sigma)\|_0 \cdot \|F_T(\sigma)\|_0 \geq c(\sigma) > 0, \quad (14)$$

where  $\|\cdot\|_0$  denotes the number of nonzero components and  $c(\sigma)$  is a positive constant determined by the choice of dual bases and any constraints on  $\sigma$  (e.g., bandwidth or regularity).

Kolmogorov complexity is monotone (up to constants) with respect to support size in any fixed coding scheme: fewer nonzero degrees of freedom admit shorter descriptions. Thus there exist positive constants  $a_S, a_T$  such that

$$K_S(x; \sigma) \gtrsim a_S \|F_S(\sigma)\|_0, \quad K_T(x; \sigma) \gtrsim a_T \|F_T(\sigma)\|_0. \quad (15)$$

Combining these with the uncertainty inequality yields

$$K_S(x; \sigma) K_T(x; \sigma) \gtrsim a_S a_T c(\sigma) > 0. \quad (16)$$

Hence, any attempt to drive  $K_S(x; \sigma)$  to its minimum necessarily drives  $K_T(x; \sigma)$  away from its minimum, and vice versa. A single trajectory cannot saturate both minima simultaneously.  $\square$

This argument can be made fully rigorous by specifying concrete dual bases (e.g., a discrete symbolic basis and a discrete Fourier basis) and invoking standard discrete uncertainty principles [3, 4]. The essential point is that S- and T-frame compressibilities are constrained by a nontrivial product lower bound.

### 3.6 A Lower-Bound Relation for Dual-Frame Complexity

The previous ideas lead naturally to a multiplicative lower bound on the frame-specific complexities.

**Theorem 4** (Dual-frame complexity lower bound). *For any observable  $x$  arising from an admissible latent trajectory  $\sigma$ , there exists a constant  $C(x) > 0$ , depending only on the equivalence class of trajectories consistent with  $x$ , such that*

$$K_S(x; \sigma) K_T(x; \sigma) \geq C(x). \quad (17)$$

*In particular, any reduction of  $K_S$  forces a compensating increase in  $K_T$  and vice versa.*

The constant  $C(x)$  depends implicitly (in a non-universal manner) on the dual-frame coherence  $\mu_{\text{DFT}}(F_S, F_T)$ ; in the limit of nearly compatible frames ( $\mu_{\text{DFT}} \rightarrow 1$ ), the bound becomes weak, while strongly incompatible frames enforce nontrivial lower bounds.

*Proof sketch.* Fix  $x$  and consider the set  $\mathcal{S}_x$  of all trajectories  $\sigma$  such that  $G(\sigma) = x$ . For each  $\sigma \in \mathcal{S}_x$  the projections  $F_S(\sigma)$  and  $F_T(\sigma)$  satisfy a discrete uncertainty inequality as in the previous proof:

$$\|F_S(\sigma)\|_0 \cdot \|F_T(\sigma)\|_0 \geq c(\sigma). \quad (18)$$

Let  $c_{\min}(x) := \inf_{\sigma \in \mathcal{S}_x} c(\sigma)$ . Assuming nontriviality of  $x$  and mild regularity conditions on  $\mathcal{S}_x$ , one has  $c_{\min}(x) > 0$ .

Using the same monotonicity argument as above, there exist constants  $a_S, a_T > 0$  (independent of  $x$ ) such that

$$K_S(x; \sigma) \geq a_S \|F_S(\sigma)\|_0 - b_S, \quad K_T(x; \sigma) \geq a_T \|F_T(\sigma)\|_0 - b_T, \quad (19)$$

for some fixed offsets  $b_S, b_T$ . For sufficiently complex  $x$  (or by absorbing the offsets into the definition of  $C(x)$  for finite cases), one can write

$$K_S(x; \sigma) K_T(x; \sigma) \gtrsim a_S a_T \|F_S(\sigma)\|_0 \|F_T(\sigma)\|_0 \geq a_S a_T c_{\min}(x). \quad (20)$$

Defining  $C(x) := a_S a_T c_{\min}(x) > 0$  yields the claimed product lower bound, up to additive constants that can be handled in the usual way for Kolmogorov-type inequalities.  $\square$

While  $C(x)$  is not generally computable (as is standard for Kolmogorov-based quantities), the existence of such a positive lower bound is the key structural fact: *no observable can be simultaneously maximally compressible in both frames.*

### 3.7 Interpretation

The inequalities above formalize a compression complementarity: S-frame and T-frame compressibilities cannot be optimized simultaneously. This provides a unifying explanation for a wide range of empirical phenomena:

- transform codecs (e.g., JPEG, MP3) trade off symbolic and harmonic compressibility;
- neural networks exhibit both discrete clustering (S-frame) and continuous manifolds (T-frame);
- cognitive systems appear to balance explicit symbolic reasoning with implicit, phase-like intuitive processing.

In a dual-frame world, any single-frame Kolmogorov complexity systematically underestimates the total algorithmic structure present in real signals. The dual-frame complexity  $K_{\text{DFT}}(x)$  captures this additional structure and obeys nontrivial lower bounds linking the two frames.

## 4 Dual-Frame Entropy and Compression Complementarity

Classical Shannon entropy quantifies uncertainty over a single representation of a random variable. In Dual-Frame Theory (DFT), observable signals are understood as arising from an underlying latent trajectory  $\sigma(\lambda)$  which admits two complementary projections:

$$S = F_S(\sigma), \quad T = F_T(\sigma). \quad (21)$$

The S-frame  $S$  captures symbolic or structural degrees of freedom, while the T-frame  $T$  captures harmonic or phase-coherent degrees of freedom. In this section we extend entropy, mutual information, and rate-distortion to the dual-frame setting and formalize a compression trade-off in Shannon-theoretic terms.



## 4.1 Dual-Frame Entropy

Let  $X$  be a random variable ranging over scalar trajectories, and let

$$S_X := F_S(X), \quad T_X := F_T(X) \quad (22)$$

be the induced S- and T-frame random variables. We assume all distributions are discrete for simplicity; the continuous case can be treated with differential entropy and appropriate regularization.

**Definition 6** (Dual-frame entropy). The dual-frame entropy of  $X$  is defined as

$$H_{\text{DFT}}(X) := H(S_X) + H(T_X) - I(S_X; T_X), \quad (23)$$

where  $H(\cdot)$  denotes Shannon entropy and  $I(\cdot; \cdot)$  denotes mutual information.

The term  $H(S_X) + H(T_X)$  sums the uncertainties in each frame, while the subtraction of  $I(S_X; T_X)$  avoids double-counting shared information. Hence  $H_{\text{DFT}}(X)$  quantifies the total information content of  $X$  as seen through both S and T frames.

Although the mutual information term  $I(S_X; T_X)$  may be small for strongly incompatible frames, its inclusion is essential for consistency: it ensures that shared regularities are not double-counted when frames are partially aligned. In extreme cases of near-independence,  $H_{\text{DFT}}$  approaches  $H(S_X) + H(T_X)$ , reflecting genuinely independent symbolic and harmonic uncertainties rather than overestimation.

We can immediately relate  $H_{\text{DFT}}$  to the single-frame entropies.

**Proposition 1.** For any  $X$ ,

$$H(S_X) \leq H_{\text{DFT}}(X) \leq H(S_X) + H(T_X). \quad (24)$$

Symmetrically,

$$H(T_X) \leq H_{\text{DFT}}(X) \leq H(S_X) + H(T_X). \quad (25)$$

*Proof.* By the basic inequalities  $I(S_X; T_X) \leq H(T_X)$  and  $I(S_X; T_X) \leq H(S_X)$ , we have

$$H_{\text{DFT}}(X) = H(S_X) + H(T_X) - I(S_X; T_X) \quad (26)$$

$$\geq H(S_X) + H(T_X) - H(T_X) = H(S_X), \quad (27)$$

$$H_{\text{DFT}}(X) = H(S_X) + H(T_X) - I(S_X; T_X) \quad (28)$$

$$\geq H(T_X) + H(S_X) - H(S_X) = H(T_X), \quad (29)$$

and trivially  $H_{\text{DFT}}(X) \leq H(S_X) + H(T_X)$  since  $I(S_X; T_X) \geq 0$ .  $\square$

Thus,  $H_{\text{DFT}}(X)$  is never *smaller* than the entropy seen in either frame alone. In particular, classical single-frame Shannon entropy underestimates the joint uncertainty present in dual-frame signals whenever  $T$  carries nontrivial structure.

## 4.2 Extensions to Mutual Information and Rate-Distortion

Classical mutual information measures the reduction in uncertainty about  $X$  given  $Y$ , in one representation. In DFT, a communication or learning channel may couple both S- and T-frame aspects of the data.

**Definition 7** (Dual-frame mutual information). Let  $(X, Y)$  be jointly distributed scalar trajectories with induced frame variables  $(S_X, T_X)$  and  $(S_Y, T_Y)$ . The dual-frame mutual information is

$$I_{\text{DFT}}(X; Y) := I(S_X; S_Y) + I(T_X; T_Y). \quad (30)$$

Here  $I(S_X; S_Y)$  measures shared symbolic structure, while  $I(T_X; T_Y)$  measures shared harmonic structure.

In classical rate–distortion theory, a source  $X$  is communicated through a lossy channel under a distortion constraint  $D$  with respect to a single distortion measure  $d(x, \hat{x})$  [1]. In DFT, distortions in S- and T-frames are generally not equivalent and must both be accounted for.

Let  $d_S(s, \hat{s})$  and  $d_T(t, \hat{t})$  be distortion measures in the symbolic and harmonic domains, respectively. Given a reconstruction  $\hat{X}$  of  $X$ , define the frame-wise expected distortions

$$D_S := \mathbb{E}[d_S(S_X, S_{\hat{X}})], \quad D_T := \mathbb{E}[d_T(T_X, T_{\hat{X}})]. \quad (31)$$

**Definition 8** (Dual-frame rate–distortion region). The dual-frame rate–distortion region is the set of all triples  $(R, D_S, D_T)$  such that there exists a channel  $P_{\hat{X}|X}$  satisfying

$$I(X; \hat{X}) \leq R, \quad (32)$$

$$\mathbb{E}[d_S(S_X, S_{\hat{X}})] \leq D_S, \quad (33)$$

$$\mathbb{E}[d_T(T_X, T_{\hat{X}})] \leq D_T. \quad (34)$$

This generalizes the classical rate–distortion function by allowing the encoder to trade off symbolic and harmonic fidelity. DFT predicts that optimal designs live on a *surface* in the  $(D_S, D_T)$  plane rather than on a single scalar distortion axis.

### 4.3 Entropy-Based Compression Complementarity

Consider a source  $X$  with induced S- and T-frame variables  $(S_X, T_X)$ , and suppose we design a compressor that encodes  $X$  with two codebooks: an S-frame code for  $S_X$  at rate  $R_S$  bits per symbol, a T-frame code for  $T_X$  at rate  $R_T$  bits per symbol. Assuming asymptotically optimal codes, we have  $R_S \geq H(S_X)$  and  $R_T \geq H(T_X)$ .

In the dual-frame setting, the pair  $(S_X, T_X)$  is not arbitrary: they are projections of a common latent trajectory and therefore cannot both be made arbitrarily simple without degeneracy. This induces a trade-off at the level of achievable rates.

**Assumption 1** (Dual-frame uncertainty). There exists a positive constant  $c(X) > 0$ , depending only on the source family, such that for any admissible dual-frame representation  $(S_X, T_X)$  of  $X$ ,

$$\exp(H(S_X)) \exp(H(T_X)) \geq c(X). \quad (35)$$

Equivalently,

$$H(S_X) + H(T_X) \geq \log c(X). \quad (36)$$

This assumption is a probabilistic analogue of discrete uncertainty principles for dual bases (e.g., support-size inequalities in time and frequency).

**Theorem 5** (Entropy-based compression complementarity). *Under Assumption 1, any coding scheme that assigns frame-wise rates  $R_S$  and  $R_T$  satisfying*

$$R_S \geq H(S_X), \quad R_T \geq H(T_X) \quad (37)$$

*must obey the trade-off*

$$R_S + R_T \geq \log c(X). \quad (38)$$

*In particular:*

- reducing  $R_S$  below some threshold forces  $R_T$  to increase,
- reducing  $R_T$  below some threshold forces  $R_S$  to increase.

*Proof.* By optimality of source codes,  $R_S \geq H(S_X)$  and  $R_T \geq H(T_X)$ . Summing yields

$$R_S + R_T \geq H(S_X) + H(T_X). \quad (39)$$

Assumption 1 then gives

$$H(S_X) + H(T_X) \geq \log c(X), \quad (40)$$

so

$$R_S + R_T \geq \log c(X), \quad (41)$$

as claimed.  $\square$

This theorem states that, for nontrivial dual-frame sources, there is an irreducible total rate budget: one cannot arbitrarily compress both the symbolic and harmonic aspects of the data without losing representational adequacy.

#### 4.4 Relation to Uncertainty Principles and Quantum Complementarity

The dual-frame compression complementarity developed here is closely related in spirit to classical and quantum uncertainty principles, but applies to *classical* signals and *compression* rather than measurement statistics.

In the continuous-time setting, Gabor's time–bandwidth principle and its refinements state that a signal cannot be simultaneously localized arbitrarily well in time and frequency. In a Hilbert-space formulation, let  $x$  be a vector and let  $\{e_i\}$  and  $\{f_j\}$  be two orthonormal bases related by a unitary transform  $U$ . Writing

$$x = \sum_i a_i e_i = \sum_j b_j f_j, \quad (42)$$

one obtains support and entropy bounds of the schematic form

$$\|a\|_0 \cdot \|b\|_0 \geq c > 0, \quad H(a) + H(b) \geq \log c, \quad (43)$$

where  $c$  depends on mutual coherence between the bases. These express that no vector may be simultaneously sparse in both representations.

Our dual-frame setting abstracts this situation. The latent trajectory  $\sigma(\lambda)$  plays the role of a latent source, and the S- and T-frame projections

$$S = F_S(\sigma), \quad T = F_T(\sigma) \quad (44)$$

play the role of two incompatible representational bases or feature maps. We posit the following structural assumption.

**Assumption 2** (Dual-frame support uncertainty). For any latent source  $Z$  giving rise to an observable  $x$ , the S- and T-frame representations satisfy a discrete support bound

$$\|F_S(\sigma)\|_0 \cdot \|F_T(\sigma)\|_0 \geq c(\sigma) > 0, \quad (45)$$

for some constant  $c(\sigma)$  depending only on the underlying trajectory, not on the particular coding schemes used in each frame.

This is the discrete analogue of a time–frequency or position–momentum support inequality: the more localized  $x$  is in the S-frame sense (few effective degrees of freedom), the more spread it must be in the T-frame sense, and vice versa. The compression complementarity results for  $K_S, K_T$  and  $H_{\text{DFT}}$  are then derived by combining Assumption 2 with standard monotonicity links between support size and description length.

In quantum information, one considers noncommuting observables  $\hat{A}, \hat{B}$  and proves entropic lower bounds of the form [7]

$$H(\hat{A}) + H(\hat{B}) \geq \log c_q, \quad (46)$$

for outcome entropies of measurements on a quantum state. By contrast, our dual-frame framework:

- operates entirely with classical signals and classical coding schemes;
- treats  $F_S$  and  $F_T$  as purely classical, possibly nonlinear projections of a latent process  $\sigma(\lambda)$ ;
- uses entropies and Kolmogorov complexities of *codes* and *representations*, not of measurement outcomes.

The analogy is conceptual and mathematical rather than ontological or physical: in both cases one obtains lower bounds on joint compressibility in two incompatible representations, but here the setting is that of signal processing and compression, not quantum measurement.

## 4.5 Summary

Dual-frame entropy and mutual information make explicit a structural fact that is only implicit in classical information theory: real-world signals and learned representations simultaneously support symbolic and harmonic degrees of freedom that cannot both be compressed to arbitrary extent. The DFT formalism provides a quantitative measure  $H_{\text{DFT}}(X)$  of joint uncertainty across frames, a dual-frame mutual information  $I_{\text{DFT}}$  capturing shared structure in both domains, and an entropy-based compression complementarity that explains observed trade-offs in transform coding, neural embedding geometry, and clustering behavior.

## 5 Toy Latent Trajectory Models

The abstract dual-frame formalism becomes much more concrete when we exhibit simple processes in which: (i) a single latent trajectory  $\sigma(\lambda)$  suffices to generate the data, and (ii) S- and T-frame projections are manifestly incompatible, in the sense that one can be simple only at the expense of the other.

### 5.1 Rotating phase with symbolic thresholding

Consider a scalar process evolving on the unit circle:

$$\sigma(\lambda) = \omega\lambda + \phi_0 \pmod{2\pi}, \quad (47)$$

with angular frequency  $\omega$  and initial phase  $\phi_0$ . From this single trajectory we generate both a binary symbolic sequence and a continuous waveform.

**S-frame: symbolic partition.** Define a two-symbol S-frame observable by thresholding the cosine:

$$s(\lambda) = \begin{cases} 0, & \cos \sigma(\lambda) \geq 0, \\ 1, & \cos \sigma(\lambda) < 0. \end{cases} \quad (48)$$

Sampling at discrete times  $\lambda_k = k\Delta\lambda$  yields a sequence  $s_k \in \{0, 1\}$  whose structure depends on the rationality of  $\omega\Delta\lambda/2\pi$ . In many regimes this sequence is describable by a short program (a rotation on a finite-state automaton modulo a partition), hence low S-frame Kolmogorov complexity  $K_S$ .

**T-frame: harmonic description.** At the same time, the underlying continuous waveform  $x(\lambda) = \cos \sigma(\lambda)$  admits a simple harmonic representation:

$$x(\lambda) = \Re\left(e^{i(\omega\lambda+\phi_0)}\right). \quad (49)$$

Its T-frame representation is essentially one Fourier mode with frequency  $\omega$  and phase  $\phi_0$ , so  $K_T$  is minimal while the symbolic sequence  $s_k$  can look combinatorially rich (especially when  $\omega\Delta\lambda$  is incommensurate with the threshold pattern).

**Complementarity.** If we change the S-frame partition (e.g., threshold a different nonlinear function of  $\sigma$  or introduce multiple thresholds), the symbolic sequence can be made increasingly regular (lower  $K_S$ ) or irregular (higher  $K_S$ ), but this does not change the essentially single-mode harmonic T-frame description. Conversely, perturbations that complicate the harmonic content (e.g., adding small incommensurate frequencies to  $\sigma(\lambda)$ ) can render the spectral description non-sparse (higher  $K_T$ ) even when the binary sequence remains simple.

This elementary example shows explicitly how the *same* scalar trajectory can induce low S-frame complexity and high T-frame complexity, or vice versa, but not both simultaneously for generic parameter choices.

## 5.2 Switched-frequency scalar motion

As a second toy model, let  $\sigma(\lambda)$  evolve according to a piecewise-constant frequency:

$$\dot{\sigma}(\lambda) = \omega_{z(\lambda)}, \quad z(\lambda) \in \{1, \dots, M\}, \quad (50)$$

where  $z(\lambda)$  is a slow, discrete-valued mode index and each  $\omega_m$  defines a distinct oscillatory regime.

**S-frame.** Define the S-frame as the mode index sequence  $S = (z(\lambda_k))_k$ , obtained by sampling at a suitable rate. If transition dynamics between modes are simple (e.g., a low-order Markov chain or a small deterministic automaton), the S-frame complexity  $K_S$  can be low: a short program suffices to describe the switching logic.

**T-frame.** Define the T-frame as the Fourier or wavelet representation of  $x(\lambda) = \cos \sigma(\lambda)$ . Because  $\sigma(\lambda)$  switches between multiple frequencies, the spectrum of  $x$  contains multiple peaks, side-bands, and nonstationary features. For rich switching patterns this leads to a higher T-frame complexity  $K_T$ .

By varying the switching structure (and thus the S-frame simplicity) one can trace out empirical curves in the  $(K_S, K_T)$  plane for this toy model, providing a concrete playground for the compression complementarity: there is no choice of switching law and sampling rate for which both the mode sequence and the harmonic content are simultaneously maximally compressible.

## 5.3 Role of toy models

These toy systems are deliberately simple, but they exhibit the core structure that DFT attributes to more complex real-world data:

- there exists a latent trajectory  $\sigma(\lambda)$ ;
- S- and T-frame observables are incompatible projections of this trajectory;
- making one projection “simpler” (lower description length) generally makes the other more complex.

Table 2: Example dual-frame functionals for a natural image source. Here  $\sigma(\lambda)$  is a latent path through an image manifold (e.g., underlying scene parameters),  $F_S$  extracts symbolic / structural features, and  $F_T$  extracts harmonic content.

Component	Interpretation	Example
$X$	Observable source	Image $x(u, v)$ on a grid
$\sigma(\lambda)$	Latent trajectory	Path in scene/texture space
$F_S$	S-frame functional	Edge map, segmentation, objects
$F_T$	T-frame functional	DCT / wavelet coefficients

In later sections we argue that modern neural networks, compression algorithms, and even cognitive systems appear to inhabit precisely this dual-frame regime.

#### 5.4 Concrete dual-frame functionals for natural images

The non-metaphysical reading of  $\sigma(\lambda)$  as a latent trajectory becomes particularly clear for complex sources such as natural images. Table 2 summarizes one plausible instantiation.

For instance, one can model a video or a sequence of related images as arising from a smooth trajectory  $\sigma(\lambda)$  through a manifold of scenes (camera pose, lighting, object configuration). The S-frame projection  $F_S(\sigma)$  may correspond to a Canny edge map or semantic segmentation; the T-frame projection  $F_T(\sigma)$  may correspond to blockwise DCT or wavelet coefficients. In this view, classical image codecs can be reinterpreted as particular choices of  $F_S$  and  $F_T$ , together with a practical operating point on the dual-frame rate–distortion region.

## 6 Dual-Frame Rate–Distortion Theory

### Clarification: Two Dual-Frame Rate Notions

Two related but distinct rate notions appear in this paper.

The first, used in Sections 4 and 9, is a *frame-wise mutual information rate*

$$R_{\text{frames}} := I(S_X; S_{\hat{X}}) + I(T_X; T_{\hat{X}}),$$

which decomposes the information budget explicitly into symbolic and harmonic contributions. This quantity is directly measurable in representation-learning systems and is therefore used as an operational proxy in empirical settings.

The second is the classical rate

$$R := I(X; \hat{X}),$$

which measures the total information transmitted about the source.

In general these quantities need not coincide. However, when the projections  $F_S$  and  $F_T$  are deterministic and jointly sufficient for reconstruction, one has

$$I(X; \hat{X}) \leq I(S_X; S_{\hat{X}}) + I(T_X; T_{\hat{X}}),$$

so  $R_{\text{frames}}$  provides an upper bound on the true rate. Throughout this paper we therefore treat  $R_{\text{frames}}$  as a dual-frame operational surrogate for  $R$ .

We emphasize that  $R_{\text{frames}}$  is used as a diagnostic and comparative quantity rather than a claim about true optimal coding rates; its role is to expose symbolic–harmonic trade-offs even when  $I(X; \hat{X})$  is difficult to estimate directly.

Given a joint distortion constraint  $\mathbb{E}[d_{\text{DFT}}(X, \hat{X})] \leq D$ , the dual-frame rate–distortion function is defined as

$$R_{\text{DFT}}(D) = \inf_{P_{\hat{X}|X}: \mathbb{E}[d_{\text{DFT}}] \leq D} I(X; \hat{X}), \quad (51)$$

where  $I(X; \hat{X})$  is the usual mutual information. This reduces to the classical  $R(D)$  when one frame is trivial or ignored.

More informatively, one can define a *rate-distortion region*:

$$\mathcal{R} = \left\{ (R, D_S, D_T) : \exists P_{\hat{X}|X} \text{ with } I(X; \hat{X}) \leq R, \mathbb{E}[d_S] \leq D_S, \mathbb{E}[d_T] \leq D_T \right\}. \quad (52)$$

For a fixed rate budget  $R$ , the region describes all pairs  $(D_S, D_T)$  of achievable distortions. Dual-frame complementarity predicts that the boundary of this region is strictly trade-off shaped: pushing  $D_S$  toward its minimum forces  $D_T$  upward, and vice versa, for any nondegenerate source and incompatible pair  $(F_S, F_T)$ .

In practical coding schemes (e.g., transform codecs, autoencoders), one can measure empirical approximations to  $\mathcal{R}$  by sweeping compression hyperparameters and plotting observed triplets  $(R, D_S, D_T)$ . Systems that approach the theoretical frontier in this space are, in the dual-frame sense, optimally balancing symbolic and harmonic distortions.

## 7 Related Work and Conceptual Positioning

Dual-Frame Theory intersects several established research directions but is not reducible to any single one.

**Information Bottleneck and Representation Learning.** The information bottleneck framework studies trade-offs between compression and prediction in learned representations. DFT differs in that it does not posit a single compressed representation, but instead asserts the necessity of two incompatible representational frames whose joint compressibility is constrained.

**Minimum Description Length and Two-Part Codes.** Minimum Description Length (MDL) and two-part coding frameworks decompose description length into model and data costs, providing a principled trade-off between model complexity and data fit [5]. DFT generalizes this idea by treating symbolic and harmonic descriptions as irreducible rather than hierarchical, and by deriving uncertainty-style bounds linking their complexities.

**Uncertainty Principles and Sparse Representations.** Classical uncertainty principles bound joint localization in dual bases (e.g., time/frequency or position/momentum). DFT abstracts this structure from bases to arbitrary representation functionals and connects it directly to entropy and Kolmogorov complexity.

**Neural Representation Geometry.** Prior empirical work has observed trade-offs between clustering, smooth embeddings, and spectral structure in neural networks. DFT provides a unified information-theoretic explanation for these observations and yields falsifiable predictions about achievable trade-offs.

**Neuro-symbolic and hybrid AI.** A related but distinct body of work explores the integration of symbolic reasoning with neural representations under the banner of neuro-symbolic or hybrid AI [2]. These approaches typically combine discrete logical structures with continuous neural models to improve interpretability or reasoning (e.g., survey work by d’Avila Garcez et al. on neuro-symbolic AI, and discussions by Marcus on the limitations of purely neural systems).

While such systems explicitly juxtapose symbolic and neural components, they do not generally derive information-theoretic constraints on the joint compressibility of discrete and continuous representations. By contrast, Dual-Frame Theory does not prescribe architectural hybrids or logical formalisms. Instead, it posits a structural complementarity between symbolic and harmonic representations that arises even within a single learned model, independent of explicit symbolic modules.

**Distinction from Information Bottleneck and MDL.** The information bottleneck framework optimizes a single representation under competing objectives, typically via a Lagrangian trade-off between compression and prediction. MDL similarly decomposes description length into model and data costs. In contrast, DFT asserts that two *simultaneously required but incompatible* representations coexist, and that their joint compressibility is constrained even when prediction accuracy is held fixed. The complementarity arises structurally, not from an optimization objective.

## 8 Applications to Machine Learning, Cognitive Systems, and Neuroscience

The dual-frame information picture captures empirical regularities across compression algorithms, neural representations, and cognitive phenomenology.

We emphasize that the applications discussed in this section are intended as interpretive mappings rather than mechanistic derivations. DFT does not claim to replace domain-specific models of attention, cognition, or neural dynamics, but to provide a unifying information-theoretic constraint that such models must respect if they simultaneously support symbolic and harmonic structure.

### 8.1 Classical Compression Algorithms

Many widely used compression schemes instantiate a primitive S/T separation.

#### 8.1.1 Transform Coding (JPEG, MP3, AAC)

In block-based transform codecs:

- The S-frame corresponds to spatial or temporal arrangement plus coarse structural features (edges, contours, onsets).
- The T-frame corresponds to transform coefficients (DCT, MDCT, wavelets), capturing harmonic and phase-coherent structure.

These codecs reduce T-frame entropy by quantizing high-frequency coefficients, accepting increased  $D_T$ , while keeping S-frame distortion  $D_S$  small by preserving salient features. The entropy-based complementarity implies that aggressive T-frame compression increases S-frame unpredictability, and vice versa. Empirically, codecs perform best when balancing these, navigating the dual-frame rate trade-off.

#### 8.1.2 Model-Based vs. Data-Driven Codecs

Model-based codecs (hand-designed transforms) implicitly fix  $F_T$  and adjust  $F_S$ ; learned codecs jointly adapt  $F_S$  and  $F_T$  to the source family, discovering the dual-frame trade-off.

### 8.2 Machine Learning Representations

In deep neural networks:

- **S-frame:** discrete tokens, cluster assignments, decision boundaries, sparse feature codes.
- **T-frame:** continuous embeddings, manifolds, phase-like coherence (directions encoding analogies, style).

The coexistence of heterogeneous, mixed-selectivity features in neural populations and in trained networks is consistent with this dual-frame picture [8].



### 8.2.1 Embedding Geometry and Clustering Trade-offs

Good embeddings show cluster structure (low S-frame entropy) while supporting smooth interpolations (organized T-frame geometry). Over-regularized embeddings (low T-frame entropy) lose cluster separation; classification-focused embeddings (minimal S-frame uncertainty) become brittle. The entropy-based complementarity explains the need to balance discrete and continuous structure.

### 8.2.2 Self-Supervised and Contrastive Learning

Contrastive objectives shape  $I_{\text{DFT}}(X; Y)$  between views. Augmentations preserving semantics probe S-frame robustness; those preserving coherence probe T-frame. DFT predicts strong performance requires preserving both S- and T-frame mutual information.

During training, networks redistribute complexity between frames: early high T-frame complexity, mid formation of clusters, late stabilization of trade-off. DFT predicts overemphasizing one frame leads to brittle or incoherent outputs; optimal models sit near the uncertainty boundary.

Model compression (quantization, pruning) can be reinterpreted: S-frame compression (fewer tokens) demands richer T-frame structure to preserve performance, explaining why quantized models retain high-dimensional embeddings.

## 8.3 Cognitive Systems and Neuroscience

Human cognition can be described in terms of S-like modes (explicit reasoning, language) and T-like modes (intuition, pattern recognition). Under DFT, these are complementary projections of underlying dynamics. Neural networks replicate this by balancing frames.

Neuroscientific studies of social interaction show inter-brain phase synchronization (T-frame) and shared narratives (S-frame). "We-space" can be modeled as high  $I_{\text{DFT}}(X_1; X_2)$  between agents. Complementarity suggests rigid symbolic structure suppresses flexible dynamics, while pure synchrony lacks coordination.

## 8.4 Summary

Across domains, signals carry symbolic and harmonic structure that cannot both be maximally compressed. Successful systems balance these modes, operating near an optimal frontier in the  $(H(S_X), H(T_X))$  plane. DFT provides a common language for characterizing why certain schemes, architectures, and strategies work—and fail when over-optimized.

## 9 Empirical Proxies, Testable Predictions, and Experimental Design

To make DFT quantitatively testable, we relate abstract quantities to measurable properties.

### 9.1 Empirical Proxies for Dual-Frame Entropy

For a model and dataset:

**Proxy for  $H(S_X)$  (symbolic entropy).** Let  $Z$  be a discrete latent representation (e.g., quantized codes or token clusters via  $k$ -means). Define  $\hat{H}_S = H(\hat{P}_Z)$ , the Shannon entropy of the empirical code distribution. Refinements include MDL estimates.

**Proxy for  $H(T_X)$  (harmonic entropy).** Let  $E$  be a continuous embedding. Compute the empirical covariance eigenvalues  $\{\lambda_i\}$ ; effective rank  $r_{\text{eff}} = \exp(H_{\text{spec}})$  where  $H_{\text{spec}} = -\sum_i p_i \log p_i$ ,  $p_i = \lambda_i / \sum_j \lambda_j$ . Take  $r_{\text{eff}}$  (or its log) as proxy; more spread-out spectra indicate higher harmonic complexity.

## 9.2 Testable Predictions

DFT yields concrete predictions:

- (P1) **S-frame minimization (strong clustering or token regularization) must degrade T-frame structure.** Overemphasis on symbolic structure reduces phase coherence, degrading analogy structure and linear embedding geometry.
- (P2) **T-frame minimization (manifold smoothing or low-rank penalties) must degrade S-frame structure.** Overly smooth embeddings weaken cluster boundaries and symbolic accuracy.
- (P3) **No model can simultaneously minimize symbolic and harmonic distortion.** A Pareto frontier exists due to the dual-frame uncertainty relation

$$H(S_X) + H(T_X) \geq \log c(X).$$

- (P4) **Joint regularization outperforms single-frame optimization.** Balancing S- and T-frame reconstruction errors yields optimal performance in VAEs, transformers, and diffusion models.
- (P5) Over-clustered embeddings (low  $H(S_X)$ ) will exhibit worse linear analogy performance and poorer interpolation quality.
- (P6) Extremely smooth or low-rank embeddings (low  $H(T_X)$ ) will exhibit degraded discrete discrimination (e.g., class separability or symbolic fidelity).
- (P7) Models trained with balanced regularization on both S- and T-frame proxies will lie closest to the empirical rate-distortion frontier in  $(R, D_S, D_T)$  and yield the best joint performance on tasks that require both discrete and continuous structure.
- (P8) For large language models, aggressive quantization (reducing S-frame precision) should increase effective embedding dimensionality or geometric richness (higher T-frame complexity).

These connect to proxies like cluster entropy (S-frame), effective rank (T-frame), intrinsic dimension, and spectral coherence.

DFT would be challenged if, within a controlled experimental setting with fixed architectures, datasets, and complexity proxies, representations could be constructed that simultaneously achieve minimal symbolic and harmonic complexity without degrading task performance.

Conversely, consistent empirical observation of a trade-off—across architectures, training regimes, and domains—would constitute evidence in favor of the dual-frame complementarity principle articulated here.

## 9.3 A Controlled Regularization Sweep

As a testbed, consider a convolutional VAE on natural images (e.g., CIFAR-10), with discrete codebook (S-frame) and continuous latent (T-frame).

Three regimes:

**S-regime.** Strong entropy penalty on codes:  $\mathcal{L}_S = \lambda_S H(\hat{P}_Z)$ , reduces  $\hat{H}_S$  but increases T-frame complexity.

**T-regime.** Low-rank penalty on latents:  $\mathcal{L}_T = \lambda_T \sum_i \lambda_i^2$ , reduces effective rank but pushes structure to S-frame.

**DFT-regime.** Combined  $\mathcal{L}_{\text{DFT}} = \lambda_S \mathcal{L}_S + \lambda_T \mathcal{L}_T$ , yields intermediate values.

DFT predicts DFT-regime achieves better joint performance (classification + reconstruction) and lies closer to  $(R, D_S, D_T)$  frontier for fixed rate.

Plot  $(\hat{H}_S, H_{\text{spec}})$  against metrics across sweeps to map trade-offs; sweet spots should correspond to balanced allocations.

Such experiments would validate DFT as a principle for learned representations.

## 10 Conclusion and Limitations

We have introduced Dual-Frame Theory (DFT), an information-theoretic framework that formalizes the empirical coexistence of symbolic/discrete and harmonic/continuous representations in modern learning systems. By defining dual-frame Kolmogorov complexities, entropies, and rate-distortion trade-offs, DFT shows that no signal or representation can be simultaneously maximally compressible in both frames.

The central result is a compression complementarity: improving symbolic simplicity necessarily degrades harmonic coherence, and vice versa. This principle unifies observations from transform coding, neural embedding geometry, and cognitive systems under a single structural constraint.

**Limitations.** DFT depends on the choice of representation functionals  $F_S$  and  $F_T$ ; poorly chosen or nearly invertible pairs may weaken the complementarity bounds. The constants appearing in Kolmogorov and entropic inequalities are generally non-computable, as is standard in algorithmic information theory. Empirical proxies for frame entropies are model- and dataset-dependent. In practice, estimating  $\mu_{\text{DFT}}$  may itself require proxy measures, and its empirical behavior across learned representations remains an open question for future validation.

**On the looseness of bounds.** The complementarity bounds derived in this paper are structural rather than tight. As with classical Kolmogorov complexity and entropic uncertainty relations, the constants involved are generally loose and non-computable, and the theory does not claim near-saturation in typical systems. The significance of these bounds lies in their existence and robustness across admissible frame pairs, not in numerical sharpness.

**Outlook.** Future work includes rigorous characterization of admissible dual-frame pairs, tightening of rate-distortion bounds, and large-scale empirical validation in transformers and multimodal models. DFT may also provide a foundation for principled regularization strategies in representation learning. Ongoing work includes controlled VAE and transformer studies measuring empirical  $(H_S, H_T)$  frontiers, which will be reported in a subsequent revision.

## A Rigorous Complexity Complementarity in a Discrete Setting

In this appendix we make the dual-frame complexity bound precise in a standard finite-dimensional setting. Let  $\mathcal{H} \cong \mathbb{C}^N$  be an  $N$ -dimensional Hilbert space with canonical (symbolic)

basis  $\{e_n\}_{n=0}^{N-1}$  and discrete Fourier (harmonic) basis  $\{f_k\}_{k=0}^{N-1}$ ,

$$f_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} e^{2\pi i k n / N} e_n. \quad (53)$$

Any vector  $x \in \mathcal{H}$  admits expansions

$$x = \sum_n a_n e_n = \sum_k b_k f_k, \quad (54)$$

with coefficient vectors  $a = (a_n)$  and  $b = (b_k)$ .

### A.1 Discrete support uncertainty

Define the supports

$$\text{supp}(a) = \{n : a_n \neq 0\}, \quad \text{supp}(b) = \{k : b_k \neq 0\}, \quad (55)$$

and their cardinalities

$$M_S(x) = \|\text{supp}(a)\|_0, \quad M_T(x) = \|\text{supp}(b)\|_0. \quad (56)$$

A standard discrete uncertainty result (see, e.g., Donoho–Stark type theorems [3, 4]) states that there exists a constant  $c_N > 0$  depending only on  $N$  such that for all nonzero  $x \in \mathcal{H}$ ,

$$M_S(x) \cdot M_T(x) \geq c_N. \quad (57)$$

For the canonical/Fourier pair considered here one has, in particular,  $c_N \geq N$ ; in words,  $x$  cannot be simultaneously sparse in both the symbolic and Fourier domains.

### A.2 From support to Kolmogorov complexity

We now connect support sizes to description lengths. Fix two prefix-free universal Turing machines  $U_S$  and  $U_T$  used to define  $K_S$  and  $K_T$ , respectively, by

$$K_S(x) = \min\{|p| : U_S(p) = a\}, \quad K_T(x) = \min\{|q| : U_T(q) = b\}, \quad (58)$$

where  $a$  and  $b$  are the coefficient vectors of  $x$  in the  $e_n$  and  $f_k$  bases, written in some fixed rational or finite-precision encoding.

**Lemma 1** (Support–complexity lower bounds). *There exist positive constants  $c_S, c_T > 0$  and additive constants  $C_S, C_T$  (depending only on the chosen machines and encoding scheme) such that, for all  $x \in \mathcal{H}$ ,*

$$K_S(x) \geq c_S M_S(x) - C_S, \quad K_T(x) \geq c_T M_T(x) - C_T. \quad (59)$$

*Proof.* Given a coefficient vector  $a$  with support size  $M_S(x)$ , any program that reconstructs  $a$  must, at minimum, specify:

- (i) which  $M_S(x)$  indices are active;
- (ii) the values of the corresponding coefficients up to the chosen precision.

Even under an optimal encoding, this requires a number of bits that grows at least linearly with  $M_S(x)$  up to an additive constant. Since Kolmogorov complexity is a lower bound on effective description length,  $K_S(x) \geq c_S M_S(x) - C_S$  for some  $c_S > 0$ . The same reasoning applies to  $K_T(x)$  and  $M_T(x)$ .  $\square$

### A.3 Complexity product bound

Combining (57) and (59) yields the advertised complexity complementarity.

**Theorem 6** (Rigorous dual-frame complexity complementarity). *For any nonzero  $x \in \mathcal{H}$ , there exists a constant  $C(x) > 0$  such that*

$$K_S(x) K_T(x) \geq C(x), \quad (60)$$

*up to multiplicative and additive constants depending only on the choice of universal machines and encoding convention. In particular, no representation of  $x$  can simultaneously minimize  $K_S$  and  $K_T$ : increasing symbolic compressibility necessarily reduces harmonic compressibility, and vice versa.*

*Proof.* By Lemma 1,

$$K_S(x) K_T(x) \geq (c_S M_S(x) - C_S)(c_T M_T(x) - C_T). \quad (61)$$

For nonzero  $x$  and sufficiently large  $N$  one may absorb the additive constants into a rescaled bound; in particular there exist  $c'_S, c'_T > 0$  and  $C' > 0$  such that

$$K_S(x) K_T(x) \geq c'_S c'_T M_S(x) M_T(x) - C'. \quad (62)$$

Using the discrete uncertainty bound (57) gives

$$K_S(x) K_T(x) \geq c'_S c'_T c_N - C'. \quad (63)$$

Defining  $C(x) := c'_S c'_T c_N - C'$  yields the claim (noting that  $c_N$  depends only on  $N$  and the chosen basis pair, not on  $x$ ).  $\square$

This appendix makes explicit that the dual-frame complexity bound is a direct corollary of well-known discrete uncertainty principles once one identifies the S-frame with a symbolic basis and the T-frame with a harmonic basis.

## References

- [1] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. Wiley, 2006.
- [2] Artur d’Avila Garcez, Luis C. Lamb, and Dov M. Gabbay. Neuro-symbolic ai: The state of the art. *Artificial Intelligence Review*, 54:1–42, 2019.
- [3] David L Donoho and Philip B Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989.
- [4] Michael Elad and Alfred M Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 48(9):2558–2567, 2002.
- [5] Peter D. Grünwald. *The Minimum Description Length Principle*. MIT Press, 2007.
- [6] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 2008.
- [7] Hans Maassen and Jos B M Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103–1106, 1988.
- [8] Mattia Rigotti, Daniel Ben Dayan Rubin, Xin-Jing Wang, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497:585–590, 2013.
- [9] Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.