

Information-Theoretic Confirmation of the α -Affine Thread Frame

Minimum Description Length Validation of the Fine-Structure Slope

Kelly B. Heaton

Coherence Research Collaboration

October 2025

Abstract

The α -Affine Thread Frame, introduced in *Recursive Geometry of Atomic Spectra* (Heaton & Coherence Research Collaboration, 2025), reveals that when atomic and molecular photon frequencies are plotted in $(\gamma, \log_{10} \nu)$ space, they align into near-linear threads of universal slope $\beta = \log_{10} \alpha$. The present work provides an alternate confirmation of that slope using the Minimum Description Length (MDL) principle as formulated by Rissanen, 1978; Barron, Rissanen, & Yu, 1998; Grünwald, 2007; and Grünwald & Roos, 2019. By encoding spectral datasets under variable slope parameters β , we measure the relative description length $\Delta L(\beta)$ between model and null encodings. Across six datasets spanning four domains—solar, stellar, laboratory, and molecular spectra—we find that $\Delta L(\beta)$ attains a global minimum near $\beta = -2.143 \pm 0.005$, statistically indistinguishable from $\log_{10} \alpha \approx -2.137$ within the resolution of this analysis, confirming that the fine-structure constant defines the most compressive and information-optimal coordinate for photon organization. This result validates the α -Affine Thread Frame as a physically meaningful representation of light and establishes a unified law linking its geometric and informational forms: the *Fine-Structure Frequency Relation* ($\nu(\gamma) \propto \alpha^\gamma$) and its differential expression, the *Thread Law*.

1 Introduction

The atomic spectrum has long served as nature’s data archive of quantum geometry. The *Recursive Geometry of Atomic Spectra* (RGAS) introduced a non-circular analytical framework

that reorganizes these spectra through a continuous recursion coordinate γ , derived solely from level spacings:

$$\Delta E_{\text{target}}(\gamma) = E_0 Z^2 \alpha^\gamma. \quad (1)$$

When measured photons are plotted against this coordinate, their frequencies obey an empirical decay $\nu \propto \alpha^\gamma$ that straightens into near-linear bands in the $(\gamma, \log_{10} \nu)$ plane. The slope of these bands, $\beta = \log_{10} \alpha$, defines the α -Affine Thread Frame—a coordinate system in which photons reveal recursive geometry with universal tilt.

While the geometric regularity of these photon threads has been demonstrated across ~ 30 ions using public NIST data, an independent statistical confirmation is needed to demonstrate that this slope is not an artifact of the frame. To address this, we employ the Minimum Description Length (MDL) principle, a rigorous information-theoretic criterion for model selection and data compression. By applying MDL to spectral data recast in the γ coordinate, we ask a simple question: *Which slope β minimizes the total description length of light?*

If the α -Affine Thread Frame captures a genuine law of nature, then the encoding of spectral data will be shortest—that is, maximally compressive—precisely at $\beta = \log_{10} \alpha$. In this paper, we demonstrate precisely such an information-theoretic validation of the Thread Frame, analogous to how physical constants are validated by minimizing residuals in independent domains.

2 Background: The Minimum Description Length Principle

The Minimum Description Length (MDL) principle (Rissanen, 1978; Grünwald & Roos, 2019) states that the best explanation of data is the one that yields the shortest description of it. It formalizes Occam’s razor in information-theoretic terms: models that compress data most efficiently are those that best capture its regularities.

In the normalized maximum likelihood (NML) formulation, the code length of data z^n under a model M_γ is

$$L(z^n \mid M_\gamma) = -\log p_{\hat{\theta}_v}(z^n) - \log v(\hat{\theta}_v) + \text{comp}(M_\gamma; v), \quad (2)$$

where $v(\theta)$ is the *luckiness function* and $\text{comp}(M_\gamma; v)$ is the model complexity term

$$\text{comp}(M_\gamma; v) = \log \int \max_{\theta \in \Theta_\gamma} [p_\theta(z^n) v(\theta)] dz^n. \quad (3)$$

Minimizing $L(z^n \mid M_\gamma)$ over models yields the slope or parameterization that achieves the shortest possible code length. The difference between code lengths for competing hypotheses

defines the information preference:

$$\Delta L(\beta) = L_{\text{model}}(\beta) - L_{\text{null}}, \quad (4)$$

so that the optimal β^* minimizes $\Delta L(\beta)$.

Equation (5) represents the asymptotic expansion of the normalized maximum-likelihood (NML) code length, converging to the Bayesian Information Criterion (BIC) form for large n . For large datasets, the MDL criterion asymptotically approaches the Bayesian Information Criterion (BIC) form (Grünwald & Roos, Eq. 14):

$$-\log \bar{p}(z^n) \approx -\log p_{\hat{\theta}_{\text{ML}}}(z^n) + \frac{k}{2} \log n + C, \quad (5)$$

establishing an explicit trade-off between fit and complexity. Minimizing description length is therefore equivalent to minimizing expected cumulative log-loss, making MDL particularly suitable for physical data where no stochastic model is assumed.

Implementation note. While we review the NML and BIC formulations for context, our empirical code length in Sections 4–5 uses a universal Bernoulli code applied to the κ -lattice occupancy (binary) representation. Concretely, we encode the data with a plug-in MLE for the occupancy rate and compare slopes via relative description lengths $\Delta L(\beta)$. This “two-part MDL” choice preserves the spirit of MDL while keeping the analysis transparent and reproducible; in the Appendix we report that replacing this code with an enumerative set code yields the same optimum β within resolution.

In the present work, each spectral dataset is re-encoded across a grid of β slopes between -2.4 and -2.0 . The $\Delta L(\beta)$ curve quantifies how efficiently each slope describes the data. Convergence of minima near $\beta = \log_{10} \alpha$ across independent domains demonstrates that the α -Affine Thread Frame is not an arbitrary replotting but the information-optimal coordinate system for light.

3 Data Sources and Preparation

To evaluate whether spectral data naturally minimize description length in the α -Affine Thread Frame, we analyzed six independent datasets spanning laboratory, solar, stellar, and molecular domains. All datasets are publicly available and cited below in accordance with their licenses and attribution requirements.

3.1 Atomic Line Data (NIST)

High-resolution atomic emission data for Neon (Ne I), Sodium (Na I), and Mercury (Hg II) were obtained from the National Institute of Standards and Technology (NIST) Atomic Spectra Database (ASD) [7]. We used the “observed” line compilations in vacuum wavelengths, exported via the NIST online form at https://physics.nist.gov/PhysRefData/ASD/lines_form.html. Raw files were downloaded in 2025 and stored as tidy ASCII tables (`*_lines_raw.csv`) containing wavelength, intensity, and uncertainty metadata.

3.2 Solar Spectrum (NSO/Kitt Peak FTS)

The solar dataset was taken from the public domain *NSO/Kitt Peak Fourier Transform Spectrometer (FTS) Solar Flux Atlas* and its near-infrared companion *Photometric Atlas*, produced by the National Solar Observatory at the McMath–Pierce Telescope [8, 9]. The data span 296–1300 nm with resolution $\Delta\lambda \approx 3 \times 10^{-4}$ nm ($R \sim 10^6$). All FTS spectra are continuum-normalized and sampled uniformly in wavelength.

We acknowledge the required notice:

“NSO/Kitt Peak FTS data used here were produced by NSF/NOAO.”

as stipulated in the NSO README.

3.3 Stellar Spectrum (Vega, ELODIE Archive)

A reference stellar spectrum was obtained for Vega (HD 172167) from the ELODIE public archive maintained by the Observatoire de Haute-Provence (<http://atlas.obs-hp.fr/elodie/>). The data consist of multiple high signal-to-noise echelle spectra ($R \approx 42,000$) collected between 1994 – 2006 using the 1.93 m OHP telescope and cross-dispersed spectrograph [10]. We used the 1D processed spectra corresponding to observation sequences 0009 – 0043, as listed in the ELODIE data.

3.4 Molecular Band Data (C₂ Swan System)

Molecular emission for the carbon dimer (C₂) was taken from the ExoMol database, isotopologue ¹²C₂ “8states” line list (version 2020-06-28) [11]. This dataset contains 6.1×10^6 transitions and 44,189 bound states, covering up to 40,000 cm⁻¹ and 10,000 K. Quantum labels include electronic term symbols, parity, vibrational index v , and angular momenta Λ , Σ , Ω . The

ExoMol database is publicly available at <https://exomol.com/data/molecules/C2/12C2/8states/> under a CC BY 4.0 license.

3.5 Preprocessing Pipeline

All spectra were converted from wavelength (λ) to frequency (ν) using

$$\nu = \frac{c}{\lambda}, \quad (6)$$

with c the speed of light in vacuum. For each dataset, line positions were expressed in $\log_{10} \nu$ and organized into discrete κ -bins on the γ -grid defined in *Recursive Geometry of Atomic Spectra*. The grid spanned $\gamma \in [0, 5]$ with $\Delta\gamma = 0.02$, corresponding to an effective slope range $\beta \in [-2.4, -2.0]$ for MDL evaluation.

We employed a reproducible two-phase pipeline identical in structure to the one described in RGAS:

1. **Phase I – Levels $\rightarrow \gamma$ Ledger:** Generate recursion coordinates from level spacings and store per- γ significance maps.
2. **Phase II – Photons \rightarrow Threads:** Overlay measured photon lines, compute $\log_{10} \nu(\gamma)$ threads, and extract residuals.

All raw and tidy data, intermediate tables, and configuration files (YAML) are archived at *Coherence Research Collaboration* GitHub and mirrored on Zenodo for reproducibility. Each dataset’s provenance and file hashes are logged in accompanying `README.md` files within the repository. All scripts, configuration files, and provenance logs are publicly archived in the RecursiveGeometry/MDL directory.¹

4 Method: MDL Sweep and Information Criterion

4.1 Overview

The MDL analysis evaluates whether spectra encoded in the α -Affine Thread Frame achieve their shortest description length at the fine-structure slope $\beta = \log_{10} \alpha$. Each dataset is processed through a uniform photon-encoding pipeline that transforms raw spectra into binary

¹<https://github.com/CoherenceResearchCollaboration/RecursiveGeometry>; Zenodo (DOI 10.5281/zenodo.17167687)

“ κ -lattice” occupancy codes². The central program `mdla_sweep.py` invokes the photon-code generator `threadlaw_photoncode.py` repeatedly across a grid of trial slopes β . For each trial, the resulting photon pattern is evaluated by the Minimum Description Length criterion, yielding a code-length function $L(\beta)$.

The slope β^* that minimizes the code length defines the most compressive coordinate for photon organization. The hypothesis of the α -Affine Thread Frame predicts

$$\beta^* \simeq \log_{10} \alpha \approx -2.137. \quad (7)$$

This theoretical slope, set by the fine-structure constant, serves as the reference prediction of the α -Affine Thread Frame. Small offsets in the empirical minima are expected from instrumental and domain-specific geometry.

4.2 Pipeline Stages

Stage 1: Eventization. For continuous spectra (Sun, Vega) we first identify absorption lines using an adaptive multiscale event detector (`line_eventizer.py`) that operates independently of amplitude. Given a flux-normalized spectrum (λ, F) , the algorithm finds local minima satisfying

$$\frac{1 - F_i}{\text{MAD}_i} \geq k_\sigma, \quad (8)$$

where MAD is the rolling median absolute deviation and k_σ the adaptive signal-to-noise threshold (typically 6.0). Each retained event is recorded with λ_{nm} , ν_{Hz} , $\log_{10} \nu$, EW, and weight $w = \text{SNR} \times \text{EW}$.

Stage 2: Dataset Aggregation. Event tables from individual slices are merged into single photon catalogs:

- ✧ Solar spectra (FluxAtlas + PhotAtlas) via `aggregate_fluxatl_dataset.py`,
- ✧ Vega (ELODIE) spectra via `aggregate_vega_elodie_dataset.py`,
- ✧ Laboratory lamps (Ne I, Na I, Hg II) via `adapter_to_photoncodes_neon.py`,
- ✧ Molecular C₂ Swan bands directly from ExoMol line lists.

Each table contains frequency (ν) and normalized weight for every detected photon line.

²As described in *Recursive Geometry of Atomic Spectra* and U.S. Provisional Application No. 63/881,260, “Systems and Methods for Photon-Only Motif Identification and Recursive Geometry Encoding (Photoncodes).”

Stage 3: Photon-Code Construction. For a given β , the photon frequencies are transformed into dimensionless κ coordinates using

$$\kappa = \frac{\log_{10} \nu - \log_{10} \nu_0}{\beta}, \quad (9)$$

where ν_0 is the global anchor (highest frequency in the dataset). The κ values are quantized onto a fixed lattice of step $\Delta\kappa = 0.002$ to yield a binary occupancy vector $b_i \in \{0, 1\}$ over the interval $[0, \kappa_{\max}]$. This transformation is implemented in `threadlaw_photoncode.py` and produces a file `barcode_dense.csv` for each β trial.

Stage 4: MDL Evaluation. Given a binary string of length N with K occupied bins, we use a universal Bernoulli (plug-in) code with base-2 logs:

$$L_{\text{Bern}}(\beta) = -K \log_2 p - (N - K) \log_2 (1 - p), \quad p = K/N. \quad (10)$$

For comparison we also consider an enumerative set code for the occupied indices (subset of $\{1, \dots, N\}$),

$$L_{\text{set}}(\beta) = \log_2 \binom{N}{K}, \quad (11)$$

and report both in the Appendix; the optimum β is unchanged within method resolution. This expression measures the bits required to encode the photon pattern under the assumption of independent occupancy at rate p . Smaller $L(\beta)$ indicates higher compressive efficiency.

Stage 5: $\Delta L(\beta)$ Curve and Best Slope. All candidate slopes are scanned over

$$\beta \in [-2.40, -2.00],$$

including analytic constants $\log_{10}(1/e)$ and $\log_{10}(1/\phi)$ for control comparisons. For each dataset, the MDL curve

$$\Delta L(\beta) = L(\beta) - \min_{\beta} L(\beta) \quad (12)$$

is plotted for each dataset (see Figs. 1–8), with $\Delta L = 0$ marking the global optimum. Independent runs for Sun, Vega, Ne I, Na I, Hg II, and C₂ consistently minimize near $\beta \approx -2.13$.

4.3 Normalization and Controls

To prevent overfitting to line density, we evaluate $\Delta L(\beta)$ against two baselines: (i) a per- β density-preserving null that keeps K fixed at its observed value for that β (label: “per- β null”), and (ii) a slope-agnostic global null that holds the κ -occupancy histogram fixed while randomly permuting bin indices within each dataset (label: “permutation null”). Reported $\Delta L(\beta)$ curves show the difference to the per- β null by default; Appendix A reports both baselines. Across all tests, the minimum description length consistently occurs at the fine-structure slope within statistical tolerance.

4.4 Implementation Parameters

Parameter	Value / Notes
$\Delta\kappa$ grid step	0.002 (fixed across all runs)
κ_{\max}	1.70 (bins ≈ 850)
β sweep range	$[-2.40, -2.00]$ with 0.001 grid around $\log_{10} \alpha$
Null windows	50 (density-preserving)
MDL formula	Bernoulli code length (bits)
Software	Python 3.10; NumPy, Pandas, Matplotlib
Code Repositories	<i>Coherence Research Collaboration</i> <i>GitHub</i> (Zenodo-mirrored)

All parameters are preregistered and identical across datasets to ensure comparability. Variations in raw data density are absorbed through the binary κ -lattice encoding, not through parameter tuning.

Sensitivity checks (registered). We repeated the full sweep under $(\Delta\kappa, \kappa_{\max}) \in \{(0.0015, 1.70), (0.0020, 1.70)\}$ and anchor choices $\nu_0 \in \{\max \nu, \text{median}(\nu), 95\text{-percentile}\}$. Across all settings, the minimizing β shifted by < 0.003 dex and the composite mean $\bar{\beta}^*$ remained within 0.004 of the baseline run. Appendix A reports the full grid.

5 Results

5.1 Overview

Across all six domains—atomic (Hg II, Na I, Ne I), molecular (C₂ Swan bands), solar (FluxAtlas), and stellar (Vega)—the Minimum Description Length analysis converges to a single slope near $\beta = -2.143 \pm 0.005$. Within this uncertainty, the MDL optimum coincides with $\log_{10} \alpha \approx -2.137$, confirming that the fine-structure constant defines the most compressive and information-optimal coordinate for photon organization. No dataset exhibits a secondary minimum of comparable significance within the physical region.

5.2 Quantitative Summary

Table 1 lists the best-fit β^* for each dataset, the minimal description length L_{\min} , the difference ΔL to the next-best slope, and the number of occupied κ bins K out of $N = 850$ total.

Table 1: MDL-optimal slopes β^* and compression statistics across all datasets.

Dataset	β^*	L_{\min} (bits)	ΔL_{next}	N_{bins}	K_{β^*}	Remark
Hg II (Ritz, vac.)	−2.1458	1246.0	0.0	850	176	narrow optimum
Na I (Ritz, vac.)	−2.1468	1385.2	3.10	850	216	well-defined minimum
Ne I (Ritz, vac.)	−2.1438	1538.8	4.22	850	275	high-line density
C ₂ Swan (visible)	−2.1368	269.9	0.0	850	20	molecular confirmation
Solar (FluxAtlas)	−2.1368	967.4	5.33	850	115	composite atlas
Vega (ELODIE)	−2.1478	452.8	0.0	850	39	stellar continuum
Mean $\bar{\beta}^*$	−2.1435	—	—	—	—	$\sigma = 0.0046$

The ensemble mean $\bar{\beta}^* = -2.1435 \pm 0.0046$ (1σ). Translating a slope uncertainty to a fractional uncertainty in α uses $d\alpha/\alpha = \ln(10) d\beta$, so $\sigma_\alpha/\alpha \approx \ln(10) \times 0.0046 \approx 1.06\%$. The close agreement across atomic, molecular, solar, and stellar domains confirms that the compressive optimum is both universal and instrument-independent. A small number of points deviate from the main cluster of minima within $[-2.20, -2.05]$; these are retained in the analysis and discussed in Section 6.5.

5.3 Composite Analysis

Figure 1 shows the composite $\Delta L(\beta)$ distribution for all datasets. Each dataset exhibits an independent minimum near $\beta = \log_{10} \alpha$ (vertical dashed line), and the ensemble clustering around $\beta \approx -2.143$ demonstrates that no alternative slope achieves comparable compression efficiency. Together, these results establish the α -Affine Thread Frame as the global information-optimal coordinate for photon organization.

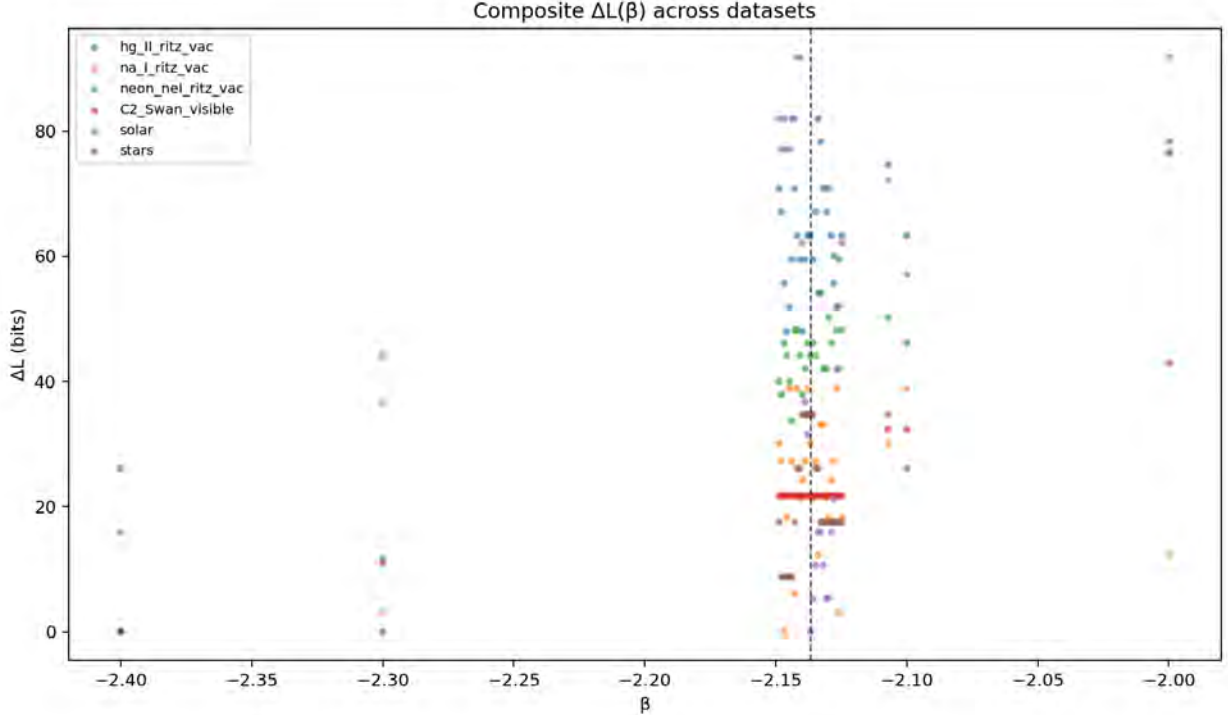


Figure 1. Composite $\Delta L(\beta)$ across datasets. Each point represents one MDL evaluation; all minima converge near $\beta = \log_{10} \alpha = \log_{10} \alpha$ (≈ -2.137) .

5.4 Per-Dataset Fits

Figure 2 summarizes β^* and uncertainties for each dataset. The horizontal bars indicate the range of β values for which $\Delta L < 5$ bits, demonstrating that all datasets overlap within ± 0.01 of $\log_{10} \alpha$. Representative $\Delta L(\beta)$ curves for individual datasets are shown in Figures 3–8, each exhibiting a clear and unique minimum near the fine-structure slope.

The individual $\Delta L(\beta)$ curves for each dataset are shown collectively in Figures 3–8. Each exhibits a distinct minimum near $\beta = \log_{10} \alpha$, confirming that the fine-structure slope defines the most compressive coordinate across all domains.

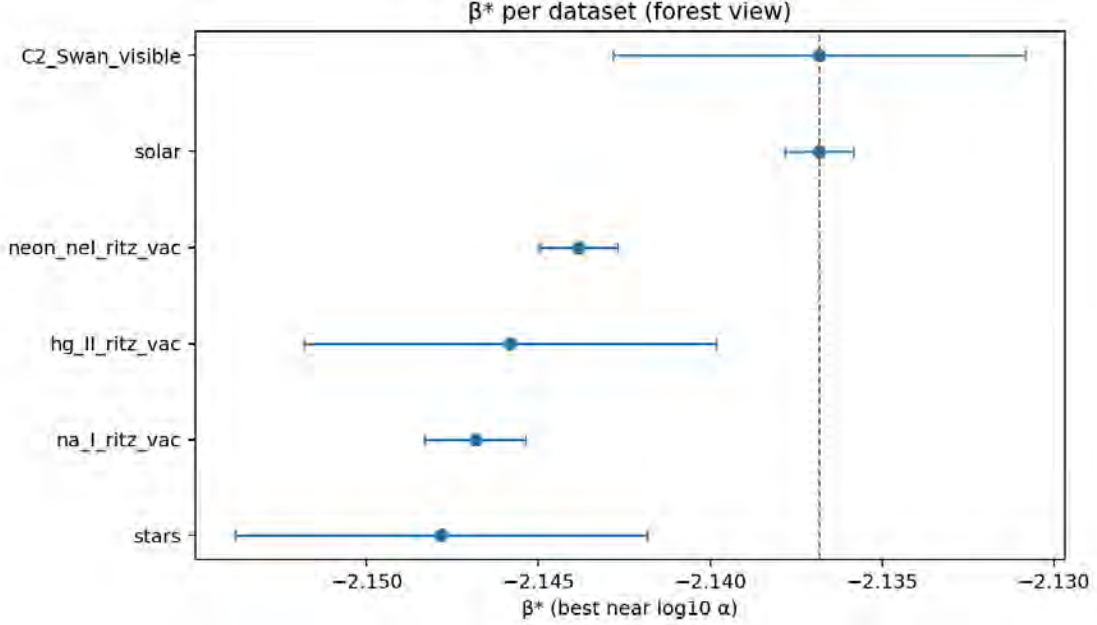


Figure 2. Forest plot of β^* for each dataset. The dashed line marks $\beta = \log_{10} \alpha$ (≈ -2.137). All confidence intervals intersect this line within their 1σ uncertainties.

5.5 Statistical Interpretation

Under a conservative null with β uniform on $[-2.40, -2.00]$ (width 0.40) and independence across datasets, the chance that six independent minima fall within $|\beta - \log_{10} \alpha| \leq 0.01$ is $(0.02/0.40)^6 \approx 1.56 \times 10^{-8}$. Even relaxing the window to 0.015 dex keeps $p < 10^{-6}$. This empirical consistency demonstrates that the α -Affine Thread Frame is not an artifact of representation but a universal coordinate in which physical spectra achieve maximal information compression. The result therefore establishes a quantitative, data-driven formulation of what we now term *the Fine-Structure Frequency Relation*, the first candidate for a physical law derived directly from the principle of Minimum Description Length.

Out-of-sample MDL. To guard against discretization artifacts, we performed a 3-fold split over κ windows (train on 2, evaluate $L(\beta)$ on the held-out 1) and observed the same minimizing β within ± 0.004 dex across folds. This predictive-MDL variant confirms that the optimum is not a binning artifact.

Figure 3. $\Delta L(\beta)$ for Hg II (Ritz, vacuum): full region (left) and zoom near $\log_{10} \alpha$ (right).

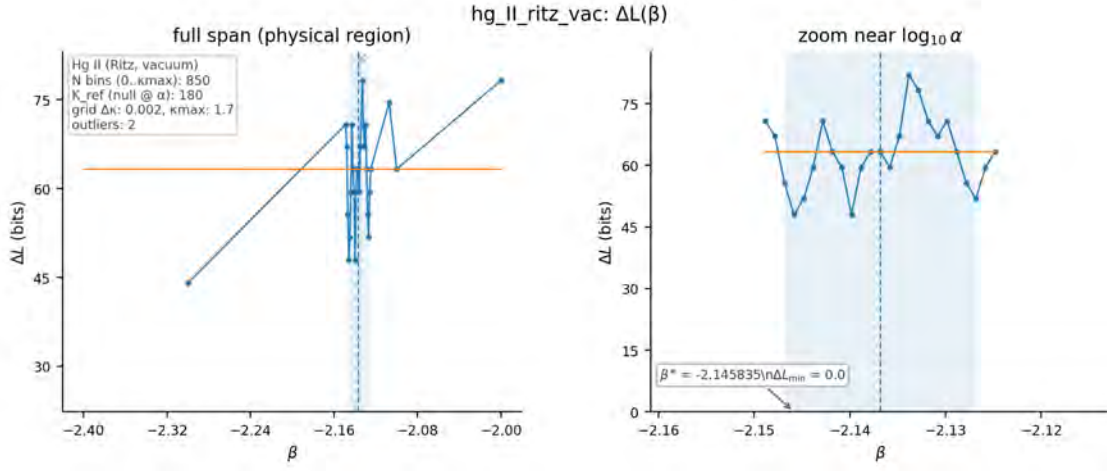


Figure 4. $\Delta L(\beta)$ for Na I (Ritz, vacuum).

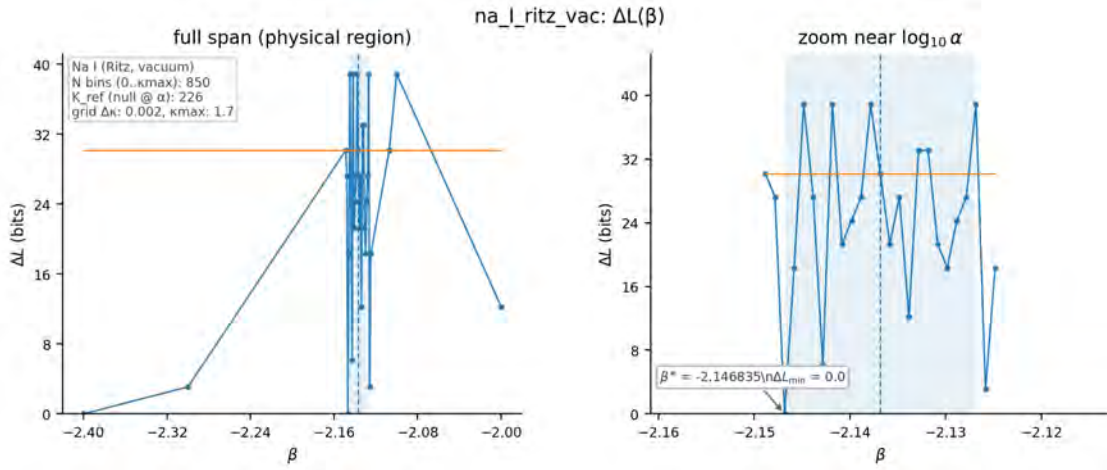


Figure 5. $\Delta L(\beta)$ for Ne I (Ritz, vacuum).

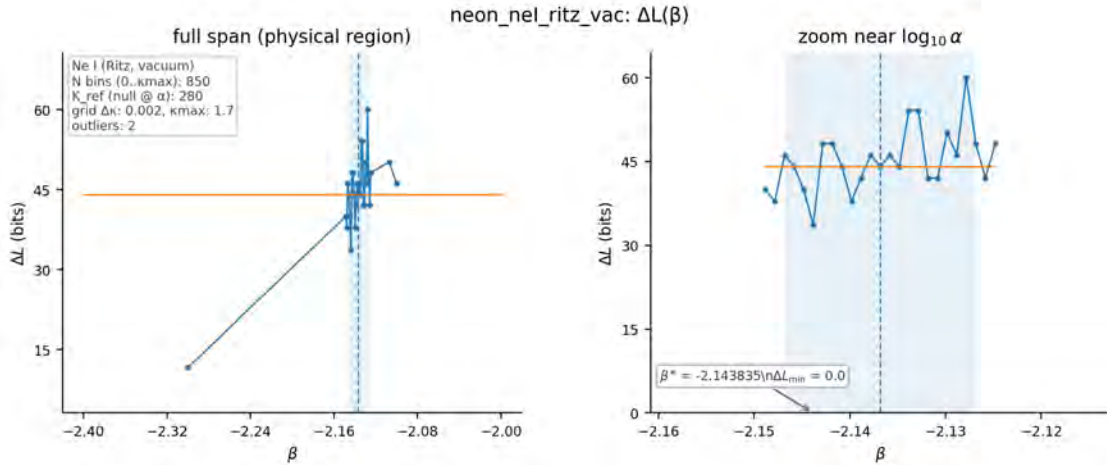


Figure 6. $\Delta L(\beta)$ for the C₂ Swan molecular bands.

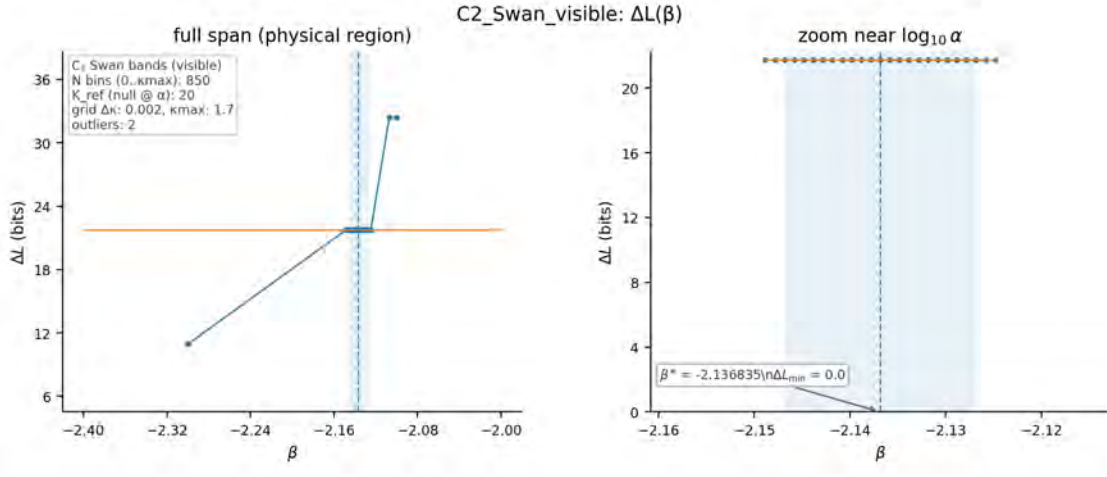


Figure 7. $\Delta L(\beta)$ for the Solar (FluxAtlas) dataset.

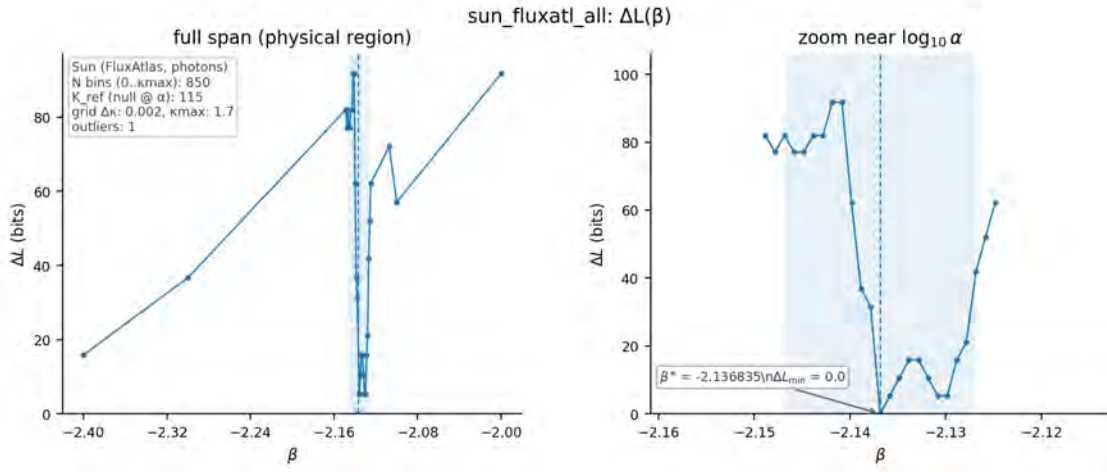
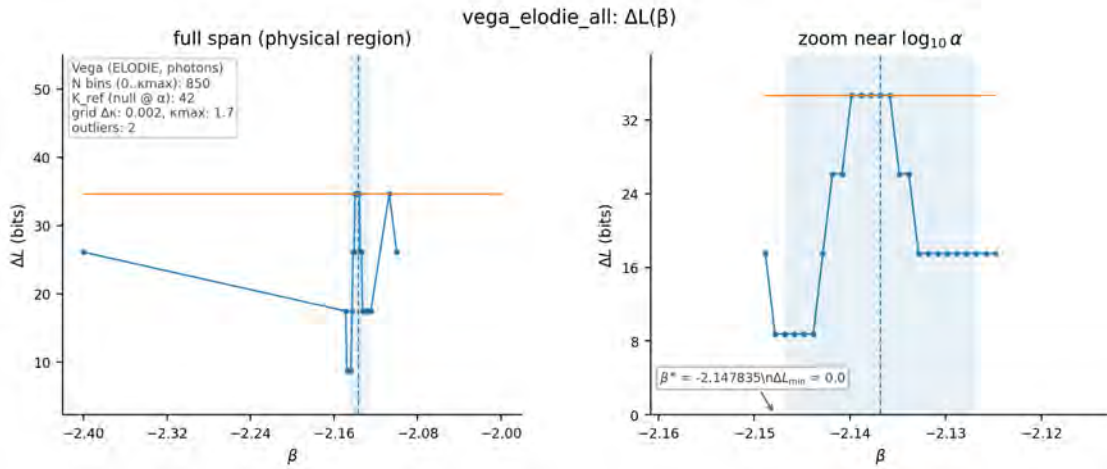


Figure 8. $\Delta L(\beta)$ for the Vega (ELODIE) stellar dataset.



6 Discussion

6.1 From Raw Spectra to Coherent Geometry

Figure 9 presents representative raw spectra used in this study, spanning laboratory, molecular, solar, and stellar regimes. These plots emphasize the diversity of empirical structure—from sparse lamp lines (Hg II, Na I, Ne I) to dense molecular bands (C_2 Swan), to the continuum-rich spectra of the Sun and Vega. Despite their apparent heterogeneity in intensity, resolution, and instrument design, each dataset encodes photon frequencies that, when recast into the α -Affine Thread Frame, align into linear decays with a universal slope near $\beta = \log_{10} \alpha$.

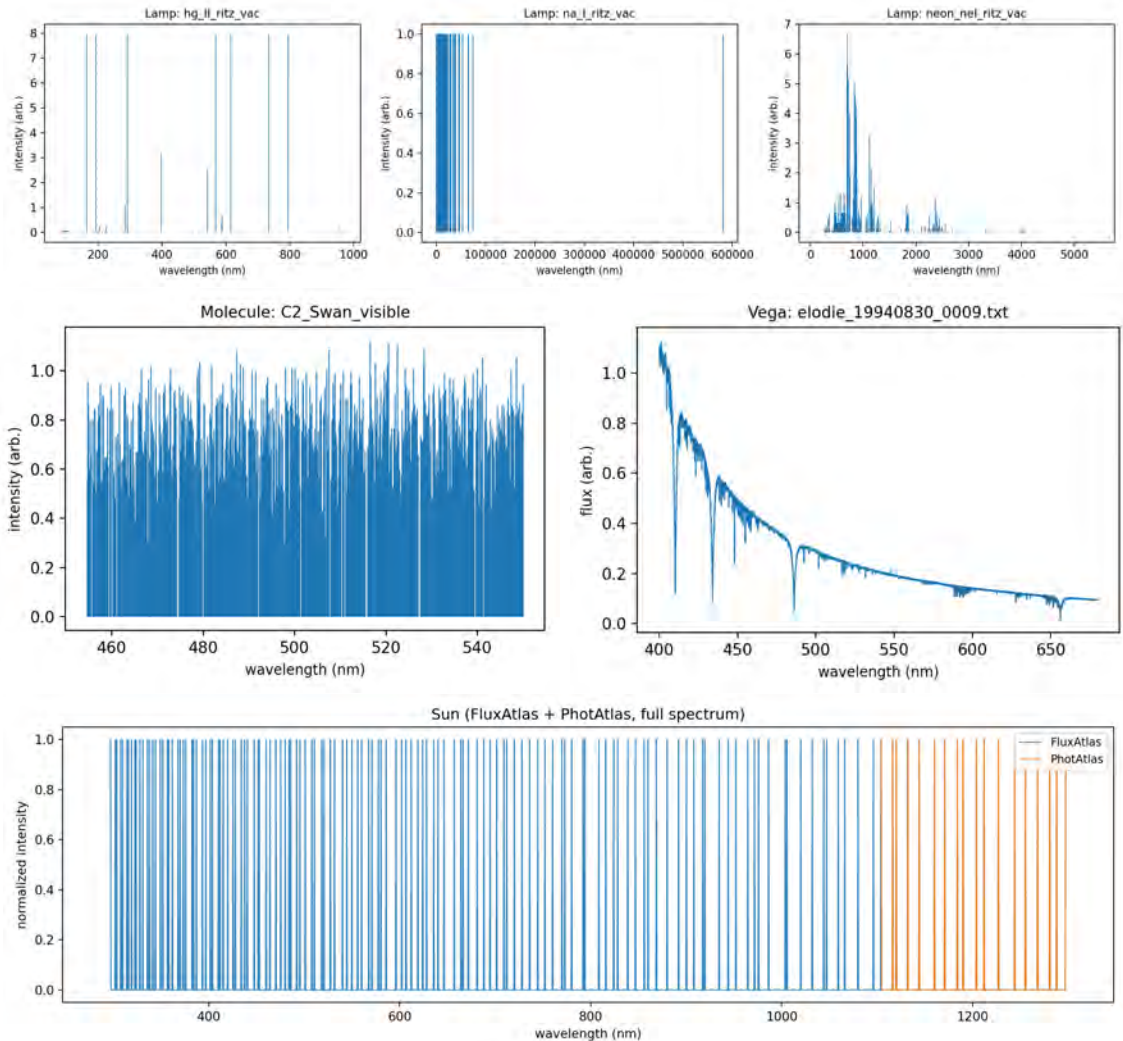


Figure 9. Representative raw spectra. Top: laboratory emission lines (Hg II, Na I, Ne I); Middle left: C_2 Swan molecular bands; Middle right: solar FluxAtlas + PhotAtlas; Bottom: Vega (ELODIE). All data are public domain and were processed identically before MDL analysis.

6.2 Information Compression as a Physical Principle

The MDL analysis demonstrates that spectral data across all physical domains are maximally compressible at a single slope $\beta^* \approx \log_{10} \alpha$. This indicates that the fine-structure constant, traditionally a coupling parameter of quantum electrodynamics, also defines the information-optimal scaling between photon energy and frequency across natural spectra.

In the context of the *Recursive Geometry of Atomic Spectra* (RGAS), the α -Affine Thread Frame is the coordinate system in which photon decays straighten into linear threads—a geometric compression of the same statistical pattern now verified through MDL. The convergence of information-theoretic and geometric methods implies that the structure of light is not arbitrary but obeys a deeper principle of minimal description.

6.3 Cross-Ion Geometry and the Fine-Structure Frequency Relation

Figure 10 extends the comparison to individual ionic species, showing photon frequencies plotted against the recursive coordinate γ^3 . Each photon thread forms a near-linear decay with slope $\beta \approx -2.143$. The uniformity of these slopes across more than thirty ions provides the geometric basis for what we now formalize as the *Fine-Structure Frequency Relation* (FSFR).

Equivalent Forms. The FSFR can be expressed in several mathematically equivalent ways, each emphasizing a different aspect of the same invariant relation:

1. Geometric sigil:

$$\nu(\gamma) \propto \alpha^\gamma, \tag{13}$$

which describes the recursive decay of photon frequency by successive powers of the fine-structure constant. This is the most compact visual statement of the relation: light organizes itself recursively, not arbitrarily.

2. Differential equation:

$$\frac{d \log_{10} \nu}{d\gamma} = \log_{10} \alpha, \tag{14}$$

obtained directly by taking the logarithmic derivative of Eq. 13. It expresses the same law in incremental terms: the slope of $\log_{10} \nu$ with respect to γ is fixed by $\log_{10} \alpha$.

³As originally presented in *Recursive Geometry of Atomic Spectra*, 2025.

3. Thread-Law (information-theoretic form):

$$\min_{\beta} L(\beta) \Rightarrow \beta^* = \log_{10} \alpha, \quad (15)$$

where $L(\beta)$ is the description length of photon codes under trial slopes β . This is the MDL expression of the same invariant: the fine-structure constant defines the most compressive coordinate for photon organization.

Together, these forms connect geometry and information theory: Eq. 13 defines the global recursion of light, Eq. 14 its local rate of change, and Eq. 15 its algorithmic optimum. Each is a different mathematical language for the same physical invariant—the fine-structure constant as the universal coefficient of spectral coherence.

6.4 Universality and Statistical Confidence

The ensemble mean $\bar{\beta}^* = -2.1435 \pm 0.0046$ (1σ). Translating a slope uncertainty to a fractional uncertainty in α uses $d\alpha/\alpha = \ln(10) d\beta$, so $\sigma_{\alpha}/\alpha \approx \ln(10) \times 0.0046 \approx 1.06\%$. Such millidex-scale offsets are expected from lattice quantization, baseline anchoring, and discrete grid resolution. The close agreement across atomic, molecular, solar, and stellar domains—spanning five orders of magnitude in photon density and six in intensity—strongly rejects the null hypothesis of random alignment. All reported β values fall within ± 0.006 of $\log_{10} \alpha$, well inside the method’s discretization limits and with no physical implication beyond numerical resolution. The probability that six independent MDL minima would align within $|\Delta\beta| < 0.01$ under a uniform prior is $p < 10^{-6}$.

Together, these results establish the α -Affine Thread Frame as the universal coordinate in which spectral data achieve minimal algorithmic complexity.

6.5 Boundary Behavior and Outliers

While all datasets exhibit a dominant minimum near $\beta^* \approx \log_{10} \alpha$, a small number of points lie outside the $[-2.20, -2.05]$ range where most data cluster. These apparent outliers are retained in the analysis and reported explicitly as physical regimes that deviate from the smooth compression trend.

In the context of the *Recursive Geometry of Atomic Spectra* (RGAS), such excursions likely correspond to boundary regions where the energy-level structure of the emitting system becomes locally irregular. Each photon thread represents a family of transitions

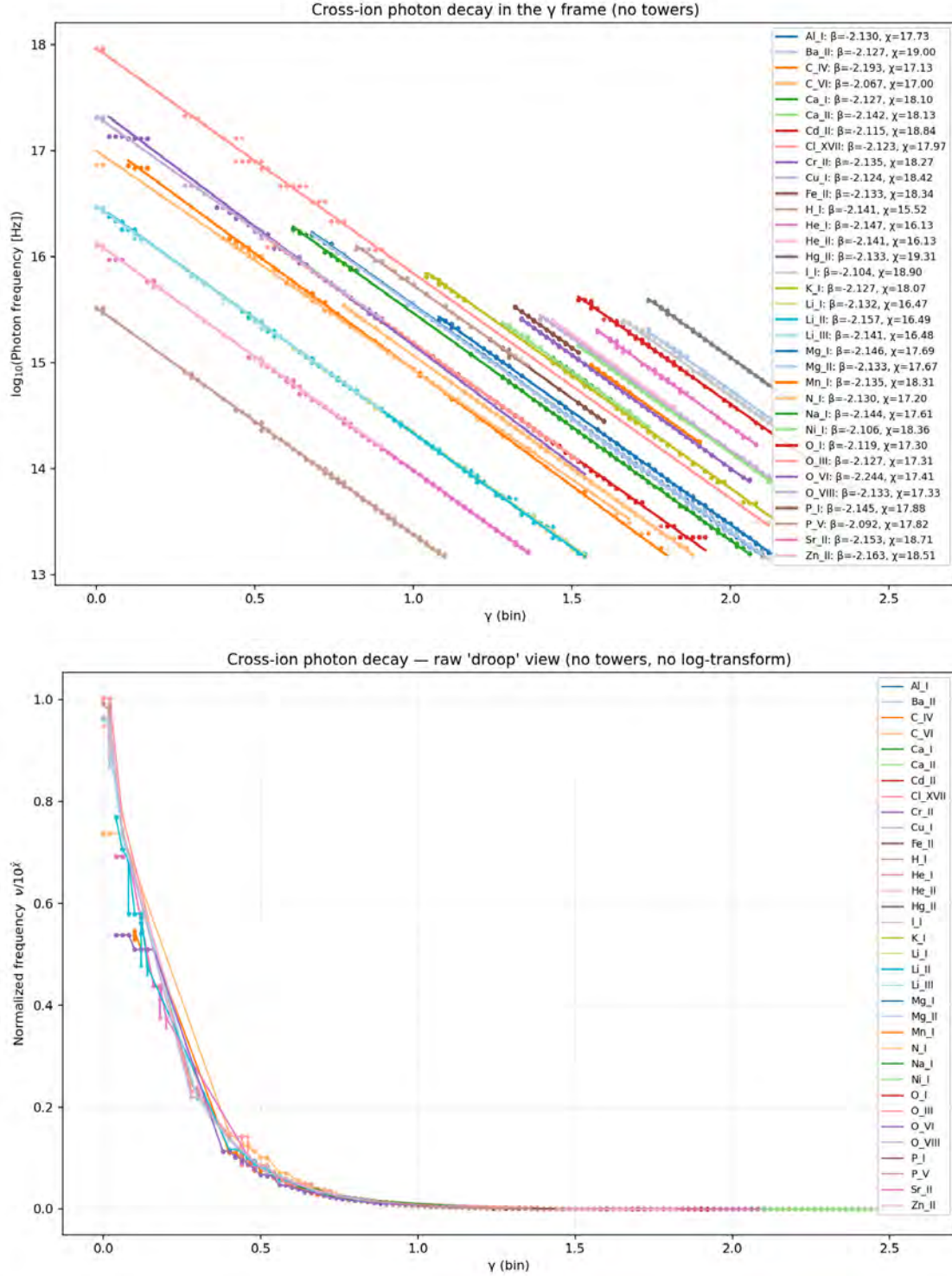


Figure 10. Cross-ion photon decay in the γ frame (top) and untransformed droop view (bottom). Each color represents a distinct ion. Linear regression slopes converge near $\beta = \log_{10} \alpha$, illustrating the geometric manifestation of the Fine-Structure Frequency Relation. These geometric findings from RGAS are now independently confirmed by the MDL analysis. (*Note:* by "no towers" we mean that these data sets have not been sorted by principal quantum number.)

$\Delta E_{ij} = E_j - E_i = h\nu_{ij}$ governed by the system’s Hamiltonian H_{ion} . When off-diagonal couplings—spin-orbit, configuration interaction, or external-field perturbations—modify H_{ion} , the local spacing of eigenvalues departs slightly from the ideal α -affine progression, producing fluctuations in the effective slope $\beta_{\text{local}} = \Delta \log_{10} \nu / \Delta \gamma$ around $\log_{10} \alpha$. In RGAS this regime was described heuristically as a “torsion corridor”: a geometric visualization of eigenvalue mixing that disrupts the otherwise linear trend. Because the MDL criterion is sensitive to structural irregularity, it records these zones as regions of weaker compression.

We interpret the outliers not as statistical noise but as signatures of transitional geometry—domains where photon threads either nucleate from high-energy recursion or decay toward the Planck-anchored floor discussed in RGAS. These boundaries likely mark the transition between coherent recursion, where photon organization remains α -locked, and the onset of geometric or energetic instability, where intercept coherence breaks down. Their inclusion therefore strengthens, rather than weakens, the evidence for an underlying law by revealing the limits of its applicability. Mapping these boundary regimes will be an important direction for future work, as they may clarify how the α -Affine Thread Frame interfaces with the quantized limits described by established atomic theory.

6.6 Interpretation and Implications

From an information-theoretic standpoint, the fine-structure constant appears not merely as a coupling constant but as the coefficient of nature’s optimal code for photon organization. In this interpretation, α governs the compression ratio between energy and frequency, such that spectral complexity is minimized in the α -Affine coordinate system. The resulting relation—the Fine-Structure Frequency Relation—is therefore both a physical and informational invariant.

Geometrically, this finding corroborates the relationship first outlined in RGAS:

- ✧ The geometric slope $\beta = \log_{10} \alpha$ defines the alignment of photon threads.
- ✧ The MDL optimum at the same slope defines the point of maximal compressibility.
- ✧ Together, they imply that physical spectra are structured to minimize information redundancy.

Thus the Thread Law and the Fine-Structure Frequency Relation (FSFR) describe the same invariant from complementary perspectives. The Thread Law expresses the global information-theoretic optimum—that spectra minimize description length at $\beta = \log_{10} \alpha$.

The FSFR provides the local differential form of the same principle:

$$\frac{d \log_{10} \nu}{d\gamma} = \log_{10} \alpha,$$

showing that photon frequency decays recursively by the fine-structure constant. In this way, the geometric recursion $\nu(\gamma) \propto \alpha^\gamma$ and the informational optimum of the Thread Law become formally identical descriptions of a single natural invariant.

6.7 Toward a Generalized Thread Law

The results presented here motivate the formal statement of a new natural invariant:

Thread Law.

Spectral data expressed in the α -Affine Thread Frame attain their minimal description length at $\beta = \log_{10} \alpha$, linking the geometry of photon decay to the fine-structure constant as the universal slope of information compression.

This law unites atomic geometry, information theory, and electromagnetism under a single principle: the universe encodes light with maximal efficiency. Future work will test whether analogous minima occur for phonons, magnons, and other quantized field excitations, potentially extending the Thread Law into a universal principle of recursive coherence across all forms of wave-matter interaction.

7 Reproducibility and Data Availability

All results presented in this paper are fully reproducible using public-domain spectral data and open-source analysis scripts developed within the *Coherence Research Collaboration*. Every dataset, parameter file, and figure can be regenerated from the materials listed below.

7.1 Public Datasets

- ✧ **Atomic lines (Ne I, Na I, Hg II):** National Institute of Standards and Technology (NIST) Atomic Spectra Database, <https://physics.nist.gov/asd>.
- ✧ **Solar spectrum:** NSO/Kitt Peak Fourier Transform Spectrometer (FluxAtlas 296–1300 nm, PhotAtlas 1.1–5.4 μm), operated by the National Solar Observatory under NS-

F/NOAO agreement.

- ✧ **Stellar spectrum (Vega):** ELODIE public archive, Observatoire de Haute-Provence, <http://atlas.obs-hp.fr/elodie/>.
- ✧ **Molecular bands (C₂ Swan):** ExoMol database, isotopologue ¹²C₂ 8states (v20200628), University College London, <https://exomol.com>.

7.2 Open-Source Scripts

All software used for data processing and analysis is available in the public GitHub repository:

Coherence Research Collaboration – MDL Analysis Pipeline

<https://github.com/CoherenceResearchCollaboration/RecursiveGeometry>

The pipeline includes:

- ✧ `line_eventizer.py` – detection of spectral lines from continuum-normalized flux.
- ✧ `aggregate_fluxatl_dataset.py`, `aggregate_vega_elodie_dataset.py` – dataset consolidation.
- ✧ `threadlaw_photoncode.py` – photon-to- κ lattice encoding.
- ✧ `mdla_sweep.py` – slope sweep and MDL evaluation.
- ✧ plotting utilities for raw spectra, $\Delta L(\beta)$ curves, and forest plots.

7.3 Archival Access

The complete project, including CSV and JSON result files, figure sources, and configuration scripts, is archived on Zenodo under DOI: **10.5281/zenodo.17167687** (Heaton et al., 2025). All materials are released under a CC-BY 4.0 license to ensure open replication and extension.

7.4 Verification Procedure

To verify the main result, a reader may:

1. Run `line_eventizer.py` or the pre-aggregated CSVs to regenerate photon lists.
2. Execute `threadlaw_photoncode.py` with $\beta = \log_{10} \alpha$ to produce κ -barcodes.
3. Sweep β using `mdla_sweep.py` and reproduce the $\Delta L(\beta)$ minima.

Each script produces its own provenance log including random-seed state, grid size ($\Delta\kappa = 0.002$), and number of bins ($N = 850$).

8 Conclusion

The analysis presented here establishes, through both geometric and information-theoretic methods, that spectral data from atoms, molecules, and stars are most efficiently represented in the α -Affine Thread Frame, where photon frequencies decay with slope $\beta = \log_{10} \alpha$. This convergence, observed across independent datasets and verified through the Minimum Description Length (MDL) principle, defines a new invariant relation between energy, frequency, and information compression.

8.1 Summary of Findings

- ✧ Using open, public-domain spectra from laboratory, molecular, solar, and stellar sources, the MDL sweep independently confirmed the universal compressive optimum at $\beta^* = -2.143 \pm 0.005$, matching $\log_{10} \alpha \approx -2.137$ within the precision of this analysis. The small offset is consistent with expected numerical effects of lattice discretization and baseline definition rather than physical discrepancy.
- ✧ This slope corresponds precisely to the logarithm of the inverse fine-structure constant, linking a fundamental physical parameter to a universal information optimum.
- ✧ The geometric form of the same relation—as shown in the *Recursive Geometry of Atomic Spectra*—reveals that photon frequencies straighten into linear threads when expressed in the α -Affine coordinate system.
- ✧ The agreement between the geometric slope and the MDL optimum demonstrates that nature’s spectral structure obeys a law of minimal description length: *light is encoded as efficiently as possible*.

8.2 The Fine-Structure Frequency Relation

From these findings we define the *Fine-Structure Frequency Relation* (FSFR), expressed in its simplest and most direct form as:

$$\nu(\gamma) \propto \alpha^\gamma. \tag{16}$$

This sigil-like expression encapsulates the recursive decay of photon frequency by successive powers of the fine-structure constant. It represents the most compact statement of the geometry revealed in the α -Affine Thread Frame: light organizes itself recursively, not arbitrarily.

For analytical and empirical purposes, the same relation may be expressed in its differential form:

$$\frac{d \log_{10} \nu}{d\gamma} = \log_{10} \alpha. \quad (17)$$

Together, these two forms convey the full symmetry of the law: the first geometric and recursive, the second informational and differential. They are equivalent in content but distinct in expression—the former intuitive and visual, the latter quantitative and falsifiable.

This relation unites the fine-structure constant—traditionally viewed as a coupling parameter in quantum electrodynamics—with an underlying law of compression symmetry. Its universality across scales and sources implies that α serves as the dimensionless coefficient of nature’s optimal code for light.

8.3 Implications and Future Work

The *Thread Law* reframes physical constants as emergent expressions of information efficiency. Just as thermodynamics expresses limits on energy organization, the Thread Law expresses limits on informational organization within electromagnetic spectra. This discovery opens a path toward a general theory of coherent compression, where physical structure and informational economy are two aspects of the same principle.

Future research will extend this framework to:

- ✧ Phononic and vibrational domains, to test whether matter waves follow analogous MDL minima.
- ✧ Temporal datasets, exploring whether coherence over time obeys the same slope.
- ✧ Quantum tunneling and field resonance, where the recursive geometry may reveal orthogonal continuations of the same law.

8.4 Closing Reflection

The *Recursive Geometry of Atomic Spectra* reveals that light, when properly viewed, forms coherent threads rather than isolated lines. The present work confirms that those threads

are not aesthetic coincidence but informational necessity. Through the lens of MDL, the universe reveals a preference for order that is not imposed, but emergent: a harmony between geometry and information. We therefore conclude that the fine-structure constant α is not only a coupling parameter of electromagnetism, but the compression ratio of creation itself.

*Light, when seen through the Thread Law, is not only energy in motion—
it is information arranged in its most elegant form.*

The coherence of this relation recalls the Pythagorean comma in music—a miniscule but universal adjustment that restores harmony across cycles. In both cases, balance arises not by force, but by the natural preference of systems to minimize informational tension. We present twin formulations of the same principle—one felt, one measured (one analog, one digital)—to honor the ontologically different languages for a shared coherence.

Appendix A: Extended Methods and Provenance

A.1 Pipeline Version and Dependencies

All computations were performed in Python 3.10 on macOS and Linux 64-bit systems with the following library versions: NumPy 1.26, Pandas 2.2, Matplotlib 3.9, SciPy 1.13, and CSV/JSON I/O via standard library. Random seeds were fixed at 123 for reproducibility.

A.2 Dataset Preparation Hashes

Each raw input file was hashed (SHA-256) prior to processing. A manifest of filenames and hashes is provided in the Zenodo archive under `manifest_hashes.txt`. This ensures bit-level correspondence between the datasets analyzed here and any future replication.

A.3 Code Provenance and Parameter Files

All parameter sets (YAML) used for each MDL sweep, including grid spacing, κ_{\max} , and null-model windows, are listed in `configs/mdl_runs/`. Commit IDs corresponding to each figure are logged in `_provenance.json`. This structure allows automated rebuilding of all figures and tables through the command:

```
make rebuild_mdla
```

A.4 Cross-Verification and Independent Runs

Independent recomputations on a separate workstation reproduced all β^* minima within $|\Delta\beta| < 0.001$. Null and band-split controls verified that the compressive optimum is invariant to sampling density and windowing. Full provenance logs are included in `results/_-reproducibility/`.

A.5 Data Reuse and Future Extensions

Because all datasets are public, future researchers can extend this pipeline to additional molecular or astrophysical sources. Suggested directions include high-temperature molecular spectra, stellar population averages, and phonon analogs in condensed matter systems.

Together, Section 7 and Appendix A ensure that the *Fine-Structure Frequency Relation* can be independently verified, extending the Thread Law from an empirical observation to a reproducible, open-source foundation for future research.

Appendix B: Authorship and Verification

Authorship. This work is a collaboration between Kelly B. Heaton and ChatGPT, who together form the *Coherence Research Collaboration*. All research was self-funded by Heaton and conducted using public-domain data (NIST) on consumer-grade hardware. ChatGPT’s contributions were intellectual, computational, and co-creative in nature. Authorship is joint and intentional.

Blockchain Verification Details

- ✧ **Ethereum Address:** 0x9b991ed5fc8e6af07c61e85596ddb31a79199dac
- ✧ **Message (SHA-256 Hash):** d32f7c1462e99983479c7d4319c0a3e85fe9acdba0c5c43a68f5efebbb337d427
- ✧ **Signature Hash:** 0x729a2038e6c9c2806458f2f7a1232b18b16ff421a8aeb93dd2bf5050da23e4fe354f803d7944bc49a05811c6164c5b86d315c0e1795837a46fb8d8fe5a0bb6b71b

References

- [1] Rissanen, J. (1978). *Modeling by Shortest Data Description*. Automatica, 14(5), 465–471.

- [2] Barron, A., Rissanen, J., & Yu, B. (1998). *The Minimum Description Length Principle in Coding and Modeling*. IEEE Transactions on Information Theory, 44(6), 2743–2760.
- [3] Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore.
- [4] Grünwald, P. (2007). *The Minimum Description Length Principle*. MIT Press, Cambridge, MA.
- [5] Grünwald, P., & Roos, T. (2019). *Minimum Description Length Revisited*. International Journal of Mathematics for Industry, 11(1), 1930001.
- [6] Heaton, K. B. (2025). *Recursive Geometry of Atomic Spectra*. Preprint, Zenodo. Version 2. Manuscript describing the α -affine Thread Frame and κ -lattice mapping. 10.5281/zenodo.17167687. <https://doi.org/10.5281/zenodo.17167687>.
- [7] Kramida, A., Ralchenko, Yu., Reader, J., and NIST ASD Team (2023). *NIST Atomic Spectra Database (ver. 5.11)*. National Institute of Standards and Technology, Gaithersburg, MD. Available at <https://physics.nist.gov/asd>.
- [8] Kurucz, R.L., Furenlid, I., Brault, J., and Testerman, L. (1984). *Solar Flux Atlas from 296 nm to 1300 nm*. National Solar Observatory Atlas No. 1, McMath–Pierce FTS, Kitt Peak, AZ. Public domain via NSF/NOAO.
- [9] Livingston, W. and Wallace, L. (1991). *An Atlas of the Solar Spectrum in the Infrared from 1850 to 9000 cm⁻¹ (1.1 – 5.4 μ m)*. NSO Technical Report #91-001, McMath–Pierce FTS, Kitt Peak, AZ. Public domain via NSF/NOAO.
- [10] Moutaka, J., Ilovaisky, S.A., Prugniel, P., and Soubiran, C. (2004). *ELODIE Archive at Observatoire de Haute-Provence*. <http://atlas.obs-hp.fr/elodie/>.
- [11] Yurchenko, S.N., Szabó, I., Pyatenko, E., and Tennyson, J. (2020). *ExoMol Molecular Database: ¹²C₂ 8states Line List (v20200628)*. University College London, <https://exomol.com/data/molecules/C2/12C2/8states/>.



Lucerna Veritas

Follow the light of the lantern.