

Article

Not peer-reviewed version

---

# Topological and Geometric Analysis of Time-Series Complexity Dynamics via Discrete Hodge Decomposition

---

[Masatsugu Ueda](#)\*

Posted Date: 8 April 2026

doi: 10.20944/preprints202604.0554.v1

Keywords: time-series analysis; Hodge decomposition; topological data analysis; Vietoris-Rips complex; Wasserstein distance; KL divergence; photoplethysmography; complexity dynamics




Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Topological and Geometric Analysis of Time-Series Complexity Dynamics via Discrete Hodge Decomposition

Masatsugu Ueda 

Independent researcher, Japan; mstg.ud@icloud.com

## Abstract

Traditional time-series analysis methods, such as Fourier and wavelet transforms, excel at identifying frequency components and their temporal localization. While powerful for spectral analysis, these methods do not explicitly capture the global geometric structure of state transitions or the emergence of cyclic (non-conservative) dynamics within the signal. In this paper, we propose a novel geometric framework that encodes the local complexity dynamics of a time series as a simplicial complex. Using a sliding Hann window, we map the signal into a sequence of local power spectral density (PSD) distributions. We construct a Vietoris-Rips complex using the Wasserstein distance to preserve the physical metric of frequency shifts, and define a directed edge flow based on the asymmetry of Kullback-Leibler (KL) divergence. Applying discrete Hodge decomposition to this flow separates the dynamics into gradient, curl, and harmonic components. Baseline experiments with synthetic signals demonstrate that our method robustly discriminates commensurable signals (gradient-dominant), incommensurable quasi-periodic signals (emergence of curl flow), and stochastic noise (curl-dominant decomposition). An exploratory application to empirical photoplethysmography (PPG) data demonstrates the framework's capability to characterize real-world biological fluctuations, showing that PPG trajectory patterns are structurally similar to those of incommensurable quasi-periodic signals. In a pilot study with 53 PPG recordings, the harmonic component showed a statistically significant correlation with heart rate that is not explained by standard heart rate variability (HRV) features, suggesting the framework extracts genuinely novel information from physiological signals. This framework offers a potential new mathematical lens for quantifying and classifying the hidden topological structures of time-series data, laying a foundation for future empirical applications and explorations across diverse scientific domains.

**Keywords:** time-series analysis; Hodge decomposition; topological data analysis; Vietoris-Rips complex; Wasserstein distance; KL divergence; photoplethysmography; complexity dynamics

## 1. Introduction

Time-series analysis has long been dominated by spectral methods. Fourier analysis decomposes a signal into its constituent frequencies and quantifies how much energy resides at each frequency. Wavelet analysis extends this by localizing frequency content in time, revealing when particular oscillations are active. These tools have proven extraordinarily productive across science and engineering, precisely because the frequency decomposition of a signal often reveals physically meaningful structure.

Yet spectral methods answer a specific kind of question: how much energy is at a given frequency, and when. They are not designed to ask whether the character of the signal's local behavior changes

over time, nor whether such changes form any organized pattern at a larger scale. If one listens to a piece of music, spectral methods describe which notes are being played. They do not, in themselves, describe the arc of tension and resolution that gives the music its structure — the sense that the signal is moving through distinguishable states and returning, or not returning, to where it began.

This gap is not a deficiency of any particular method. It is a consequence of the question being asked. Spectral analysis operates in frequency space; it does not naturally speak about the global organization of local dynamical states.

In this paper, we ask a different question. Given a time series, we extract from each short segment a quantity that reflects the local information content of that segment — a compact representation of what that piece of the signal “looks like” in terms of how its energy is distributed across scales. We call this a *local complexity state*. As the window slides through the signal, these states form a trajectory on a space of probability distributions.

The central observation is that this trajectory has geometric and topological structure that is not visible in the frequency domain. A purely periodic signal traces a closed orbit — it returns to the same state repeatedly. A quasi-periodic signal with incommensurable frequencies never closes; its trajectory densely fills a region. A stochastic signal scatters its states without coherent organization. These are genuinely different structures, and they correspond to different physical properties of the underlying signal.

To make this precise, we construct a simplicial complex from the pairwise Wasserstein distances between local complexity states and define a directed flow on its edges using the asymmetry of the Kullback–Leibler divergence. Applying discrete Hodge decomposition to this flow separates it into three components: a *gradient* part, reflecting a globally consistent ordering of complexity states; a *curl* part, reflecting local cyclic structure; and a *harmonic* part, reflecting global cycles that thread topological holes in the complex.

This decomposition does not replace spectral analysis. It operates in a complementary space — the space of local complexity states rather than the space of frequencies — and it asks complementary questions about the global organization of signal dynamics.

The contributions of this paper are the following.

- **Methodological Foundation.** We precisely define the pipeline and establish a baseline. For a pure sine wave, the flow energy is at machine precision and the Hodge decomposition is trivial, confirming that no spurious structure is introduced by the method itself.
- **Synthetic Validation.** We demonstrate on synthetic signals that the gradient, curl, and harmonic decomposition robustly discriminates commensurable quasi-periodic signals, incommensurable quasi-periodic signals, and stochastic noise.
- **Empirical Application.** We apply the framework to photoplethysmography (PPG) signals and show that the trajectory structure of PPG is consistent with that of incommensurable quasi-periodic signals.
- **Clinical Potential.** In a pilot study with 53 PPG recordings, we show that the harmonic component carries statistically significant information about heart rate that is not explained by standard HRV features, suggesting the framework extracts information that existing methods do not access.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the pipeline in detail. Section 4 presents the synthetic signal experiments. Section 5 presents the PPG application. Section 6 discusses the results, limitations, and open questions.

## 2. Related Work

The framework proposed in this paper integrates concepts from topological data analysis (TDA), information geometry, optimal transport, and algebraic topology on graphs. In this section, we briefly

review the standard applications of these tools and clarify how our approach fundamentally differs from existing methodologies.

### 2.1. Topological Data Analysis in Time Series

Topological Data Analysis (TDA), particularly persistent homology based on simplicial complexes such as the Vietoris-Rips complex, provides a mathematical foundation for extracting the “shape” of data (Edelsbrunner and Harer, 2010). In the context of time-series analysis, the conventional approach relies on Takens’ delay embedding theorem (Takens, 1981). This method reconstructs an attractor in a Euclidean phase space from the delayed coordinates of the waveform itself, and TDA is then applied to characterize the topology of this reconstructed attractor.

In contrast, our framework does not embed the waveform amplitude into a Euclidean space. Instead, we apply TDA to the space of probability distributions (local power spectral densities) extracted from sliding windows. By analyzing the trajectory of these local complexity states, we capture the topological structure of information transition rather than the geometric shape of the amplitude attractor.

### 2.2. Information Geometry and Optimal Transport

Information geometry provides a rigorous framework for treating probability distributions as points on a manifold, where the Kullback-Leibler (KL) divergence naturally describes the asymmetric pseudo-distance, or “flow of information,” between states (Cover and Thomas, 2006). Meanwhile, optimal transport theory, via the Wasserstein distance, measures the physical cost of transforming one distribution into another (Villani, 2009). In signal processing, the Wasserstein distance is highly effective because it respects the underlying physical metric (e.g., the proximity of adjacent frequency bins).

Typically, these two concepts are used independently depending on the application. Our framework introduces a hybrid construction: we strictly employ the Wasserstein distance as the ground metric to build the structural skeleton (the simplicial complex) that preserves the physical continuity of frequency shifts, while utilizing the asymmetry of the KL divergence to define the directed dynamic flow on the edges of this skeleton.

### 2.3. Discrete Hodge Decomposition on Graphs

The discrete Hodge decomposition is a powerful combinatorial tool that orthogonally separates edge flows on a graph into gradient, curl, and harmonic components. Jiang et al. (2011) successfully applied this theory to statistical ranking problems, demonstrating that incomplete pairwise comparison data can be decomposed into a globally consistent ranking (gradient) and localized or global inconsistencies (curl and harmonic components) within static networks.

While discrete Hodge theory has proven highly effective for analyzing static networks such as economic models, voting systems, and game theory, its application to the non-equilibrium dynamics of time-series data remains largely unexplored. Our work bridges this gap by applying discrete Hodge decomposition to the directed KL flow on a simplicial complex generated from a single time series. This allows us to macroscopically classify the non-conservative (cyclic) dynamics of the signal’s local state transitions.

## 3. Methods

We describe the pipeline in six steps. The input is a real-valued time series  $x(t)$  sampled at frequency  $f_s$  Hz. The output is a decomposition of the signal’s local complexity dynamics into gradient (G%), curl (C%), and harmonic (H%) energy fractions.

### 3.1. Sliding Window

The time series is segmented into overlapping windows of length  $L$  samples with stride  $S = \lfloor L(1 - r) \rfloor$ , where  $r \in [0, 1)$  is the overlap rate:

$$w_i = x[i \cdot S : i \cdot S + L], \quad i = 0, 1, \dots, N - 1. \quad (1)$$

**Design constraint.** To ensure that the dominant spectral mass of the signal falls above the first non-DC FFT bin, the parameters must satisfy:

$$\frac{f_{\text{signal}}}{f_s/L} \geq 1 \iff L \geq \frac{f_s}{f_{\text{signal}}}. \quad (2)$$

Violating this constraint causes the DC removal step to discard most of the signal energy, producing a degenerate representation.

### 3.2. Local Power Spectral Density

Each window  $w_i$  is multiplied by a Hann window to suppress spectral leakage, and its one-sided power spectrum is computed via the discrete Fourier transform:

$$\tilde{w}_i[k] = \sum_{n=0}^{L-1} w_i[n] \cdot h[n] \cdot e^{-2\pi jkn/L}, \quad k = 0, 1, \dots, \lfloor L/2 \rfloor, \quad (3)$$

where  $h[n] = \frac{1}{2}(1 - \cos \frac{2\pi n}{L})$  is the Hann window and  $j = \sqrt{-1}$ . The DC component ( $k = 0$ ) is removed, and the remaining power values are normalized to form a probability distribution:

$$P_i[k] = \frac{|\tilde{w}_i[k]|^2 + \varepsilon}{\sum_{k=1}^{\lfloor L/2 \rfloor} |\tilde{w}_i[k]|^2 + \lfloor L/2 \rfloor \cdot \varepsilon}, \quad k = 1, \dots, \lfloor L/2 \rfloor, \quad (4)$$

where  $\varepsilon = 10^{-10}$  is an additive smoothing constant. Each  $P_i$  lies on the probability simplex  $\Delta^{\lfloor L/2 \rfloor - 1}$  and represents the local complexity state of window  $i$ .

The Hann window introduces a small but non-zero spectral leakage floor. For a pure sine wave at 1 Hz with  $f_s = 50$  Hz and  $L = 64$ , the maximum pairwise Wasserstein distance among windows is  $W_1^{\max} \approx 0.025$ . This value serves as the empirical leakage floor.

### 3.3. Wasserstein Distance Matrix

The pairwise 1-Wasserstein distance between local complexity states  $P_i$  and  $P_j$  is computed using the frequency bins as the ground metric support:

$$W_1(P_i, P_j) = \sum_{k=1}^{\lfloor L/2 \rfloor - 1} |\text{CDF}_i[k] - \text{CDF}_j[k]| \cdot \Delta f_k, \quad (5)$$

where  $\text{CDF}_i[k] = \sum_{l=1}^k P_i[l]$  and  $\Delta f_k = f_s/L$ . This is the standard CDF-difference representation of the 1-Wasserstein distance for distributions on a one-dimensional ordered support. Using actual Hz values as the ground metric ensures that the distance reflects the physical cost of shifting spectral mass across frequency.

The result is a symmetric  $N \times N$  distance matrix  $D$ , with  $D_{ij} = W_1(P_i, P_j)$ .

### 3.4. Vietoris-Rips Complex

A Vietoris-Rips complex  $\mathcal{K}$  at radius  $\varepsilon$  is constructed from the distance matrix  $D$ :

- **Vertices** (0-simplices): all  $N$  windows  $\{w_0, \dots, w_{N-1}\}$
- **Edges** (1-simplices):  $\{(i, j) : D_{ij} \leq \varepsilon\}$
- **Triangles** (2-simplices):  $\{(i, j, k) : D_{ij}, D_{ik}, D_{jk} \leq \varepsilon\}$

The complex is truncated at dimension 2. The radius  $\varepsilon$  is set as a quantile  $q$  of the upper triangle of  $D$ . In our experiments we use  $q \in \{0.05, 0.10\}$ . We impose an upper limit of  $|T| \leq 8,000$  for computational tractability.

### 3.5. KL-Antisymmetric Edge Flow

For each directed edge ( $i \rightarrow j$ ) in the complex, we define a flow value:

$$f(i, j) = D_{\text{KL}}(P_i \| P_j) - D_{\text{KL}}(P_j \| P_i), \quad (6)$$

where  $D_{\text{KL}}(P \| Q) = \sum_k P[k] \log \frac{P[k]}{Q[k]}$ . This flow is antisymmetric by construction:  $f(i, j) = -f(j, i)$ . Generally, when  $P_i$  is a sharper distribution with lower entropy than  $P_j$ , the KL asymmetry tends to produce a positive flow from  $i$  to  $j$ . Heuristically, this reflects that the complexity ordering is directed from the simpler to the more complex state; however, this correspondence is not a strict mathematical equivalence, as the sign of the KL asymmetry depends on the full shape of the distributions, not entropy alone.

### 3.6. Discrete Hodge Decomposition

The edge flow  $\mathbf{f}$  is decomposed using the discrete Hodge decomposition on  $\mathcal{K}$  (Jiang et al., 2011), using two boundary operators:

- $\partial_1 \in \mathbb{R}^{N \times |E|}$ : node-edge incidence matrix
- $\partial_2 \in \mathbb{R}^{|E| \times |T|}$ : edge-triangle incidence matrix

The decomposition is:

$$\mathbf{f} = \underbrace{\partial_1^\top \boldsymbol{\phi}}_{\text{gradient}} + \underbrace{\partial_2 \boldsymbol{\psi}}_{\text{curl}} + \underbrace{\mathbf{h}}_{\text{harmonic}}, \quad (7)$$

where the three components are mutually orthogonal. The potentials are:

$$\boldsymbol{\phi} = (L_0)^+ \partial_1 \mathbf{f}, \quad L_0 = \partial_1 \partial_1^\top, \quad \boldsymbol{\psi} = (L_2)^+ \partial_2^\top \mathbf{f}, \quad L_2 = \partial_2^\top \partial_2, \quad (8)$$

and  $\mathbf{h} = \mathbf{f} - \partial_1^\top \boldsymbol{\phi} - \partial_2 \boldsymbol{\psi}$ , where  $(\cdot)^+$  denotes the Moore-Penrose pseudoinverse computed via sparse iterative least-squares (LSQR) (Virtanen et al., 2020). The energy fractions are:

$$G = \frac{\|\partial_1^\top \boldsymbol{\phi}\|^2}{\|\mathbf{f}\|^2} \times 100\%, \quad C = \frac{\|\partial_2 \boldsymbol{\psi}\|^2}{\|\mathbf{f}\|^2} \times 100\%, \quad H = \frac{\|\mathbf{h}\|^2}{\|\mathbf{f}\|^2} \times 100\%. \quad (9)$$

**Interpretation.** The gradient component reflects a globally consistent ordering of complexity states. The curl component reflects local cyclic structure at the triangle scale. The harmonic component reflects global cycles that thread topological holes in the complex. For a purely 2-dimensional complex (assuming no higher-dimensional voids,  $\beta_2 = 0$ ), the number of independent harmonic cycles corresponds to the first Betti number, bounded by  $\beta_1 = |E| - |V| + n_c - |T|$ , where  $n_c$  is the number of connected components. In practice, at larger values of  $\varepsilon$  the complex may contain 2-cycles ( $\beta_2 > 0$ ), causing this formula to underestimate  $\beta_1$ ; such regimes should be interpreted with caution.

**Trivial zero-flow regime.** When  $\|\mathbf{f}\|^2 < 10^{-15}$ , the KL asymmetry is negligible and the Hodge decomposition is undefined. In this regime, G%, C%, and H% are reported as undefined rather than zero.

### 3.7. Parameter Summary

**Table 1.** Pipeline parameters for synthetic signal and PPG experiments.

Parameter	Symbol	Synthetic	PPG
Sampling frequency	$f_s$	50 Hz	125 Hz
Window length	$L$	64 samples	250 samples
Overlap rate	$r$	0.875	0.875
Stride	$S$	8 samples	31 samples
Smoothing constant	$\varepsilon$	$10^{-10}$	$10^{-10}$
Rips radius quantile	$q$	0.05, 0.10	0.01
Triangle limit	$ T _{\max}$	8,000	8,000

### 3.8. Implementation Details

The pipeline is implemented in Python using standard scientific computing libraries. All three core components — the Vietoris-Rips complex, the Wasserstein distance matrix, and the discrete Hodge decomposition — are implemented from scratch without specialized topological data analysis libraries, using only NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020). The least-squares systems for  $\phi$  and  $\psi$  are solved using the iterative LSQR algorithm (`scipy.sparse.linalg.lsqr`).

**Note on sample entropy computation.** In `baseline_features.py`, sample entropy is computed on a  $4\times$  downsampled segment of 200 samples ( $\approx 6.5$  s at the effective rate) for computational tractability. Comparison with 500-sample full computation shows that absolute values differ substantially in magnitude across patients, though the ranking is broadly preserved. Because the partial correlation analysis uses this downsampled estimate as a covariate, a lower-quality sample entropy estimate makes the independence claim for H% conservative rather than liberal.

The codebase consists of four scripts: `pipeline_psd_wass_rips.py` (pipeline module), `exp_4signal_pipelineB.py` (synthetic experiments), `batch_pleth_analysis.py` (PPG batch processing), and `baseline_features.py` (HRV feature comparison). Wasserstein distance matrices are cached as NumPy `.npy` files; all results are stored as JSON for reproducibility.

## 4. Synthetic Signal Experiments

We validate the pipeline on four synthetic signals whose structural properties are known analytically.

### 4.1. Signal Definitions

All signals are generated at  $f_s = 50$  Hz over 20 seconds ( $T = 1,000$  samples):

$$x_{\text{pure\_sin}}(t) = \sin(2\pi \cdot 1.0 \cdot t) \quad (10)$$

$$x_{\text{comm}}(t) = \sin(2\pi \cdot 1.0 \cdot t) + \sin(2\pi \cdot 2.0 \cdot t) \quad (11)$$

$$x_{\text{incomm}}(t) = \sin(2\pi \cdot 1.0 \cdot t) + \sin(2\pi \cdot \sqrt{2} \cdot t) \quad (12)$$

$$x_{\text{noisy}}(t) = \sin(2\pi \cdot 1.0 \cdot t) + 0.3 \varepsilon(t), \quad \varepsilon(t) \sim \mathcal{N}(0, 1) \quad (13)$$

The commensurable signal (comm) uses an integer frequency ratio (1 : 2); the incommensurable signal (incomm) uses an irrational ratio (1 :  $\sqrt{2}$ ). With  $L = 64$  and  $f_s = 50$  Hz, signal components fall at bins 1.28, 2.56, and 1.81 — all above the DC bin, satisfying the design constraint.

#### 4.2. Baseline: Spectral Leakage Floor

Before comparing signals, we quantify the leakage introduced by the Hann window on the pure sine signal. The maximum pairwise  $W_1$  distance among pure\_sin windows is  $W_1^{\max} = 0.0254$ , defining the empirical leakage floor. All three non-trivial signals exceed this floor by more than an order of magnitude (comm:  $16.4\times$ , incomm:  $13.5\times$ , noisy:  $95.9\times$ ).

For pure\_sin,  $\|\mathbf{f}\|^2 < 10^{-15}$  across all tested values of  $\varepsilon$  ( $q = 0.05$  and  $q = 0.10$ ), confirming the trivial zero-flow regime. The value at  $q = 0.10$  ( $\|\mathbf{f}\|^2 = 7 \times 10^{-16}$ ) lies within floating-point rounding accumulation and carries no physical meaning.

#### 4.3. Results

**Table 2.** Hodge decomposition of four synthetic signals. Dashes indicate the trivial zero-flow regime ( $\|\mathbf{f}\|^2 < 10^{-15}$ ), in which G%, C%, and H% are undefined.

Signal	$q$	$ E $	$ T $	G%	C%	H%	$\ \mathbf{f}\ ^2$
pure_sin	0.05	348	592	—	—	—	$\approx 10^{-29}$
pure_sin	0.10	696	2760	—	—	—	$7 \times 10^{-16}$
comm	0.05	347	770	<b>99.5</b>	0.5	0.0	$8 \times 10^{-13}$
comm	0.10	692	3671	<b>100.0</b>	0.0	0.0	$1.7 \times 10^{-5}$
incomm	0.05	346	820	<b>79.4</b>	<b>20.1</b>	0.5	$7 \times 10^{-9}$
incomm	0.10	691	3192	<b>77.0</b>	<b>22.9</b>	0.2	$2 \times 10^{-7}$
noisy	0.05	346	489	<b>65.6</b>	<b>31.6</b>	2.9	$2 \times 10^{-1}$
noisy	0.10	691	2292	<b>54.8</b>	<b>43.6</b>	1.6	$5 \times 10^{-1}$

#### 4.4. Observations

**Observation 1 (robust): G% ordering.** The gradient fraction follows a consistent ordering across both values of  $q$  (Figure 1):

$$\text{pure\_sin} \approx \text{comm} (\approx 100\%) \gg \text{incomm} (\approx 78\%) > \text{noisy} (\approx 55\text{--}66\%).$$

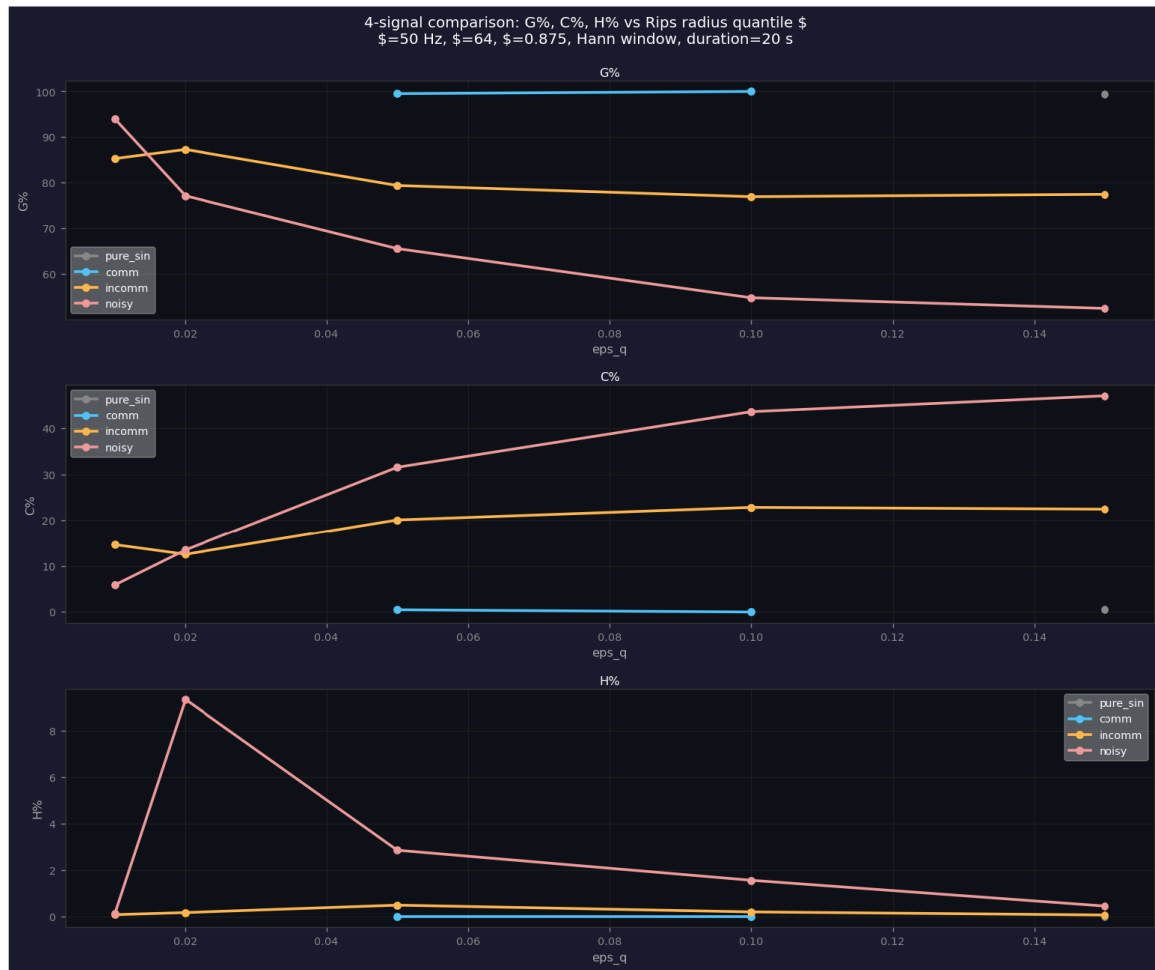
This ordering reflects the degree to which a globally consistent ordering of complexity states holds.

**Observation 2 (robust): C% ordering.** The curl fraction follows the complementary ordering:

$$\text{noisy} (32\text{--}44\%) > \text{incomm} (20\text{--}23\%) \gg \text{comm} (\approx 0\text{--}0.5\%) \approx \text{pure\_sin} (0\%).$$

The emergence of curl in incomm but not comm demonstrates that the incommensurable frequency ratio generates local cyclic structure in the KL flow that is absent in the commensurable case. This distinction is not accessible to Fourier analysis.

**Observation 3 (tentative): H%.** The harmonic fraction is small across all signals and both values of  $q$  (at most 3%). It is not a stable discriminator among these four signal classes.



**Figure 1.** 4-signal Hodge decomposition (see Section 4). Hodge decomposition of four synthetic signals as a function of the Rips radius quantile  $q$  ( $f_s = 50$  Hz,  $L = 64$ ,  $r = 0.875$ , Hann window, duration = 20 s). Top: G%. Middle: C%. Bottom: H%. Commensurable signal (comm, blue) is gradient-dominant with negligible curl. Incommensurable signal (incomm, orange) shows stable curl ( $\approx 20$ – $23\%$ ) independent of  $q$ . Noisy signal (pink) shows the highest curl and the most  $\epsilon$ -sensitive H%. Pure sine (gray) enters the trivial zero-flow regime for small  $q$ .

#### 4.5. Interpretation

The synthetic experiments confirm three properties. First, the trivial zero-flow result for pure\_sin establishes that the method introduces no spurious structure. Second, G%/C% correctly discriminates commensurable from incommensurable quasi-periodic signals — a structural distinction inaccessible to spectral methods. Third, the ordering is stable under the choice of  $\epsilon$ .

## 5. Application to Photoplethysmography Signals

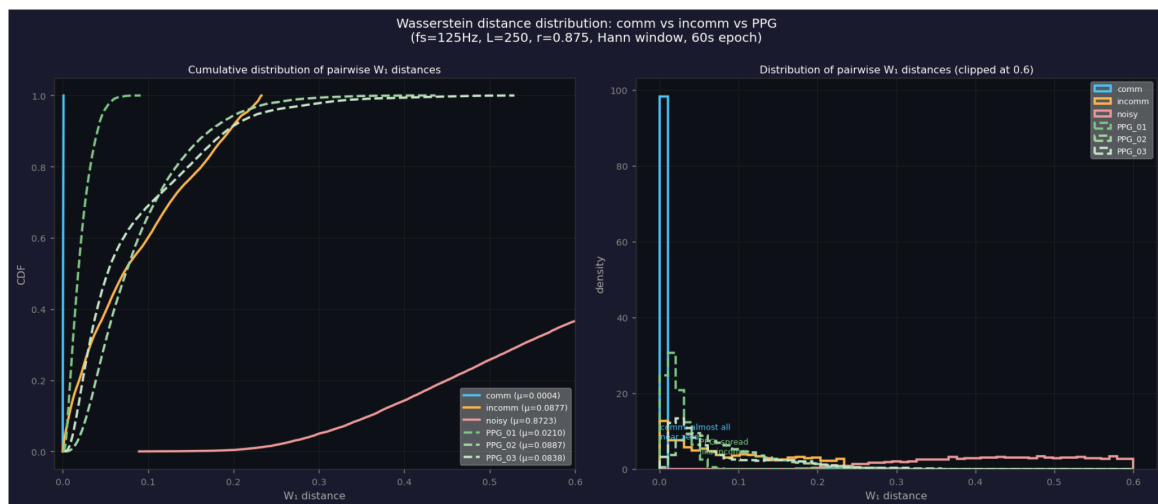
We apply the framework to PPG signals from the BIDMC dataset (Goldberger et al., 2000), containing 53 ICU patient recordings at  $f_s = 125$  Hz. Each recording is segmented into 60-second epochs (424 epochs total, 8 per patient). Concurrent clinical measurements — heart rate (HR), peripheral oxygen saturation (SpO<sub>2</sub>), and respiratory rate (RESP) — are available from bedside monitors. Parameters:  $L = 250$ ,  $r = 0.875$ ,  $q = 0.01$  (Table 1). The choice  $q = 0.01$  — substantially smaller than  $q \in \{0.05, 0.10\}$  used for synthetic signals — is required to keep the triangle count  $|T|$  below 8,000. PPG signals produce a denser, more uniformly distributed  $W_1$  distance matrix than the synthetic signals, causing  $|T|$  to grow much more rapidly with  $\epsilon$ ; at  $q = 0.05$ ,  $|T|$  exceeds 8,000 for most PPG epochs.

### 5.1. PPG as an Incommensurable Quasi-Periodic Signal

**Experiment A: Wasserstein distance distribution.** Pairwise  $W_1$  distance distributions of PPG epochs are compared against comm and incomm using the two-sample Kolmogorov-Smirnov test (Table 3; Figure 2).

**Table 3.** KS test statistics  $D$  for PPG epochs against synthetic signals ( $f_s = 125$  Hz,  $L = 250$ , Hann window, 60-second epoch).

Signal	$W_1$ mean	$D$ vs comm	$D$ vs incomm	Closer to
comm (synthetic)	0.0004	0	—	—
incomm (synthetic)	0.0877	0.983	0	—
PPG_01 (patient)	0.0210	1.000	0.573	<b>incomm</b>
PPG_02 (patient)	0.0887	1.000	0.170	<b>incomm</b>
PPG_03 (patient)	0.0838	1.000	0.102	<b>incomm</b>

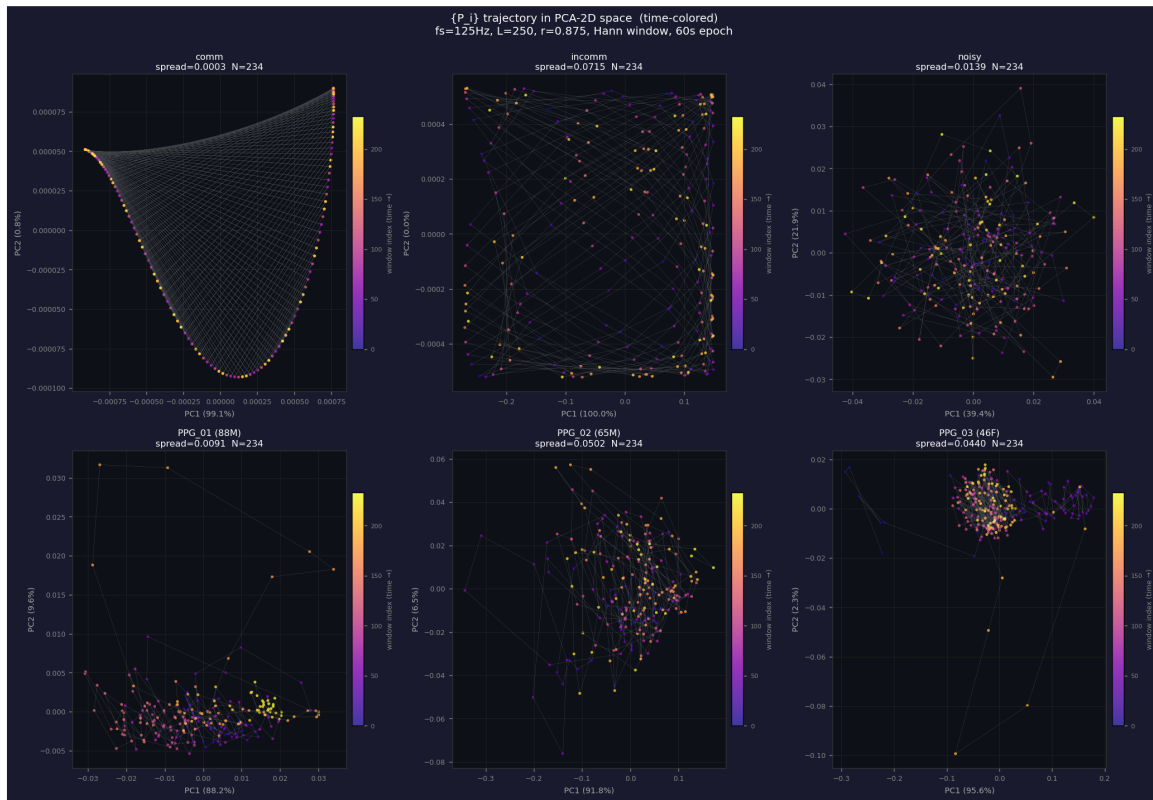


**Figure 2.**  $W_1$  distance distributions (see Section 5). Pairwise  $W_1$  distance distributions for synthetic signals and three representative PPG epochs, at  $f_s = 125$  Hz,  $L = 250$ , Hann window, 60-second epoch. Left: empirical CDF. Right: histogram (clipped at  $W_1 = 0.6$ ). Comm concentrates near zero; PPG epochs exhibit broad, continuous distributions structurally similar to incomm ( $D \approx 1.0$  by KS test against comm). The structural comparison is based on the shape and breadth of dispersion, not the absolute mean.

All PPG epochs show  $D \approx 1.0$  against comm, indicating that the PPG and comm distributions are statistically incompatible. In contrast,  $D$  against incomm is substantially smaller. We note that the KS statistic quantifies distributional dissimilarity, not similarity; a smaller  $D$  against incomm indicates only that PPG's  $W_1$  distribution is *less inconsistent* with incomm than with comm. What this demonstrates is that PPG complexity states do not form the discrete, closed orbits characteristic of commensurable signals, but instead exhibit a continuous, broad  $W_1$  distance distribution whose shape and breadth are more consistent with incommensurable dynamics. Note that absolute  $W_1$  values depend on specific physiological frequencies and amplitudes; it is the distributional structure — the breadth and shape of dispersion — rather than the absolute mean, that is being compared here (Figure 2).

**Experiment B: PCA-2D trajectory shape.** The  $\{P_i\}$  sequence of each epoch is projected onto its first two principal components and color-coded by time index. The commensurable signal traces a closed elliptical orbit. The incommensurable signal fills a band-shaped region without closing. All three PPG patients show band-shaped filling consistent with incomm, with no evidence of closed orbital structure, supporting their characterization as incommensurable quasi-periodic signals in the  $\{P_i\}$  representation (Figure 3).

**Note.** This characterization is specific to the  $\{P_i\}$  representation under this pipeline. It is not a direct measurement of physiological frequencies, nor a claim that cardiac and respiratory frequencies are mathematically incommensurable in any strict sense.



**Figure 3.** PCA-2D trajectories (see Section 5).  $\{P_i\}$  trajectories in PCA-2D space, color-coded by window index (time). Top: synthetic signals (comm, incomm, noisy). Bottom: PPG patients. Comm traces a closed arc; incomm fills a band-shaped region without closing. All PPG patients show band-shaped filling consistent with incomm.

## 5.2. Hodge Decomposition Results

Distributions of Hodge energy fractions across 424 epochs are shown in Figure 4.

**Table 4.** Summary statistics of Hodge components across 424 epochs (53 patients  $\times$  8 epochs).

Component	Mean	Std	Min	Max
G%	87.3	5.7	67.4	99.6
C%	11.6	5.8	0.4	29.6
H%	1.2	1.9	0.0	15.0

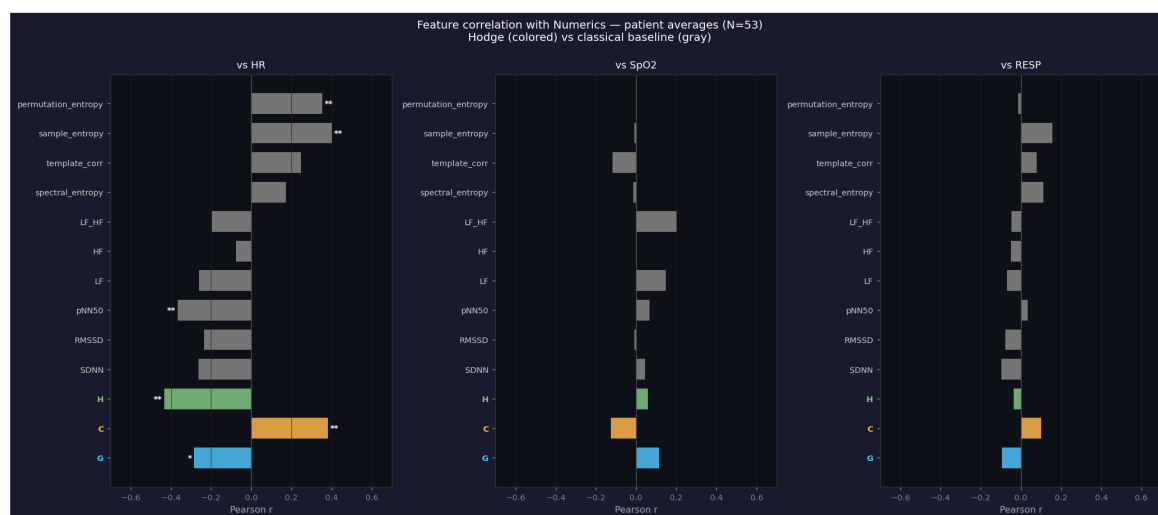
The gradient component dominates across all patients, consistent with the incommensurable quasi-periodic characterization (Figure 4). One patient (bidmc\_19) shows 5 epochs with floor ratio  $< 5\times$ ; results for this patient should be interpreted with caution.



**Figure 4.** G%/C%/H% distributions across 53 patients (see Section 5). Distribution of Hodge energy fractions across 424 epochs (53 patients  $\times$  8 epochs). Top: histograms of G%, C%, H%; dashed line indicates the mean. Bottom: patient-averaged values sorted in ascending order.

### 5.3. Correlation with Clinical Measurements

H% shows the strongest correlation with HR among all three components ( $r = -0.433$ ,  $p = 0.001$ , 95% CI:  $[-0.56, -0.29]$ ; see Appendix A for bootstrap validation; Figure 5): patients with higher heart rates tend to have smaller harmonic fractions. C% shows a significant positive correlation with HR ( $r = +0.381$ ,  $p = 0.005$ ). Note that no multiple comparison correction has been applied;  $p$ -values in Table 5 should be interpreted as exploratory given the number of features tested.



**Figure 5.** Correlation comparison: Hodge vs. classical features (see Section 5). Pearson correlations between Hodge components (G%, C%, H%; colored) and ten classical HRV/PPG features (gray) against HR, SpO<sub>2</sub>, and RESP (patient averages,  $N = 53$ ). \* $p < 0.05$ , \*\* $p < 0.01$ .

**Table 5.** Pearson correlations between Hodge components and clinical measurements (patient averages,  $N = 53$ ), with 95% Bootstrap CIs for HR correlations (10,000 paired resamples). \* $p < 0.05$ ; \*\* $p < 0.01$ .

	HR	95% Bootstrap CI	SpO2	RESP
G%	-0.286*	[-0.496, -0.023]	+0.114	-0.094
C%	+0.381**	[+0.141, +0.559]	-0.127	+0.101
H%	-0.433**	[-0.562, -0.286]	+0.060	-0.035

#### 5.4. Independence from Standard HRV Features

Ten classical HRV and PPG features are computed per epoch. Partial correlations are computed after controlling for established nonlinear complexity measures. After removing the variance explained by each baseline feature, H% retains a statistically significant partial correlation with HR:

**Table 6.** Partial correlations of H% with HR after controlling for baseline features ( $N = 53$  patient averages).

Controlling for	Partial $r$	$p$
None (full)	-0.433	0.001
Sample entropy	-0.360	0.008
Permutation entropy	-0.394	0.003
pNN50	-0.360	0.008

C%, by contrast, loses significance after controlling for sample entropy ( $r = +0.239$ ,  $p = 0.085$ ), suggesting partial overlap with existing nonlinear complexity measures. Furthermore, a multiple linear regression model confirming the independent contribution of H% when controlling for these baseline features simultaneously is detailed in Appendix A.

To assess H%'s independent contribution when multiple covariates are controlled simultaneously, we performed a multiple regression with HR as the outcome and five baseline features (SDNN, RMSSD, pNN50, sample entropy, permutation entropy) entered together. The baseline-only model explained  $R^2 = 0.351$  of HR variance. Adding H% increased  $R^2$  to 0.430 ( $\Delta R^2 = +0.079$ ), a statistically significant increment (F-test:  $F = 6.36$ ,  $p = 0.015$ ). This confirms that H% carries independent predictive information about HR beyond the joint contribution of the five baseline HRV features. Full details of both analyses are provided in Appendix A.

#### 5.5. Interpretation

Three findings emerge. First, the  $\{P_i\}$  trajectory structure of PPG is consistent with incommensurable quasi-periodic dynamics, providing an empirical characterization not available from spectral analysis alone. Second, the Hodge decomposition produces well-defined and clinically associated feature values across 53 patients. Third, H% carries statistically significant information about heart rate not explained by standard HRV features. The physiological interpretation of H% is beyond the scope of this work and is left to domain experts with access to labeled clinical datasets.

## 6. Discussion

### 6.1. Summary of Findings

The synthetic signal experiments establish three properties: no spurious structure is introduced; G%/C% robustly discriminates commensurable from incommensurable signals across tested  $\varepsilon$  values; and this discrimination is structural, not accessible to spectral methods. The PPG experiments demonstrate that PPG trajectory statistics are consistent with incommensurable quasi-periodic dynamics, and that H% carries information about heart rate not captured by standard HRV features.

### 6.2. What This Framework Does and Does Not Claim

**The framework measures complexity dynamics, not frequency content.** G%, C%, and H% characterize the global structure of transitions among local complexity states. They are complementary to, not replacements for, Fourier or wavelet analysis.

**The characterization of PPG as incommensurable is specific to the  $\{P_i\}$  representation.** The observation that PPG trajectory statistics resemble those of incommensurable synthetic signals is a statement about the  $\{P_i\}$  representation under this pipeline, not a direct measurement of physiological frequencies.

**H% carries statistically independent information, but its physiological interpretation is left to domain experts.** The partial correlation analysis shows that H% carries information about heart rate beyond what is captured by standard HRV features. What specific biological mechanism H% reflects is beyond the scope of this work, and a definitive answer requires labeled clinical datasets at scale with domain expert collaboration. Nevertheless, we offer a working hypothesis as a starting point for such investigation: at higher heart rates, cardiac cycle regularity increases and heart rate variability decreases, which may constrain the  $\{P_i\}$  trajectory to a more regular path on the probability simplex — one with fewer topological holes and therefore lower harmonic content. Under this hypothesis, H% would reflect the geometric regularity of PSD dynamics rather than any single physiological variable directly. This is a working hypothesis, not a proven claim; testing it requires clinical data with controlled physiological states. It is a question for clinical researchers with the appropriate data and domain knowledge.

### 6.3. Limitations

**Computational tractability.** The triangle count  $|T|$  grows rapidly with  $\varepsilon$ . The practical range of  $q$  is limited to  $q \lesssim 0.10$  in the present implementation.

**Reliability of  $\beta_1$  at larger  $\varepsilon$ .** The formula  $\beta_1 = |E| - |V| + n_c - |T|$  returns negative values at  $q \geq 0.05$  in practice, indicating  $\beta_2 > 0$ . This does not affect the Hodge decomposition, but  $\beta_1$  cannot be interpreted as the first Betti number in these regimes.

**Sensitivity of H% to  $\varepsilon$ .** H% shows greater sensitivity to the choice of Rips radius than G% or C%.

**Single pilot dataset.** The PPG results are derived from a single publicly available dataset of 53 ICU patients. The correlations reported should be understood as pilot observations rather than established clinical findings.

**Within-patient epoch variability.** Patient-level averages (8 epochs per patient) are used throughout the correlation analysis, which partially mitigates but does not eliminate within-patient epoch-to-epoch variation. The H% standard deviation across all epochs is 1.9 (Table 4), suggesting non-trivial intra-patient variability. A mixed-effects model at the epoch level would provide a more principled treatment; this is left for future work with larger, longitudinal datasets.

### 6.4. Open Problems

**Stability theorem.** If two signals are close in  $L^2$ , are their Hodge decompositions close?

**Optimal window length.** Is there a principled way to choose  $L$  related to the signal's intrinsic timescales?

**Relationship between H% and topological persistence.** The connection between H% at fixed  $\varepsilon$  and the persistence diagram of the Vietoris-Rips filtration is not formalized here.

**Extension to multivariate signals.** Extending the framework to multivariate time series is a natural direction left for future work.

### 6.5. Conclusion

We have proposed and validated a framework for characterizing the global topological and geometric structure of local complexity dynamics in time series. The discrete Hodge decomposition of a KL-antisymmetric flow on a Vietoris-Rips complex of local PSD distributions produces three interpretable components. In synthetic experiments, these components robustly discriminate signal types that are indistinguishable by spectral methods. In a PPG application, the harmonic component carries information about heart rate not captured by standard HRV features.

The framework does not replace spectral analysis; it asks different questions about the same signals. What the present work establishes is that the questions can be asked precisely, and that the answers are not trivially obtainable from existing methods.

### Acknowledgements

The author used Claude (Anthropic) as a research collaborator throughout this work, including assistance with pipeline implementation, manuscript drafting, and iterative revision. Critical review of successive drafts was conducted with Gemini (Google DeepMind). All experimental designs, interpretations, and conclusions are the sole responsibility of the author. AI tools are not listed as authors in accordance with standard publishing ethics guidelines.

### Code Availability

The full pipeline implementation and experimental scripts are available at <https://doi.org/10.5281/zenodo.19456477>. The BIDMC PPG dataset used in Section 5 is publicly available via PhysioNet at <https://physionet.org/content/bidmc/1.0.0/> (Goldberger et al., 2000).

### Appendix A. Statistical Robustness of the H%–HR Association

To ensure the statistical reliability of the observed correlation between the harmonic fraction (H%) and heart rate (HR) given the pilot sample size ( $N = 53$ ), we conducted three additional validations.

**Bootstrap Confidence Intervals.** We performed paired bootstrap resampling (10,000 iterations) on the patient-averaged data to estimate the stability of the correlations. The 95% confidence interval (CI) for the Pearson correlation between H% and HR is  $[-0.56, -0.29]$ , which is strictly bounded away from zero. Figure A1 shows the empirical bootstrap distributions for G%, C%, and H% against HR, visually confirming the robust left-shift (negative correlation) of the H% distribution.

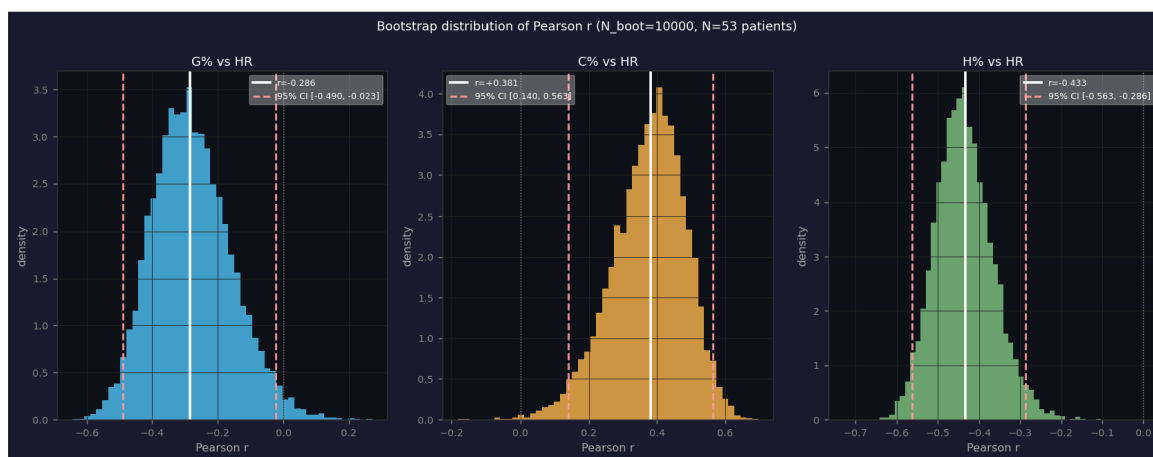
**Multiple Linear Regression.** To rigorously test independence beyond pairwise partial correlations, we fitted a multiple linear regression model predicting HR. H% was included alongside standard time-domain and nonlinear HRV features (SDNN, RMSSD, pNN50, and sample entropy) simultaneously as independent predictors. In this combined model, H% retained a statistically significant negative coefficient ( $p = 0.015$ ,  $F = 6.36$ ,  $\Delta R^2 = +0.079$ ). This confirms that the harmonic component provides unique explanatory power for heart rate that cannot be accounted for by a linear combination of these standard complexity features.

**Leave-One-Out Cross-Validation (LOO-CV).** We performed LOO-CV on both regression models to provide an empirical assessment of out-of-sample predictive value.

**Table A1.** In-sample and LOO-CV  $R^2$  for regression models predicting HR ( $N = 53$  patient averages).

Model	In-sample $R^2$	LOO-CV $R^2$	Shrinkage
A: baseline only	0.351	0.172	0.179
B: baseline + H%	0.430	0.266	0.164
$\Delta R^2$	+0.079	+0.094	—

Both models exhibit similar shrinkage ( $\approx 0.17$ ), consistent with the pilot sample size. Model B achieves a LOO-CV  $R^2$  of **0.266**, indicating substantive out-of-sample predictive power. Crucially, the increment attributable to H% ( $\Delta R^2$ ) is **+0.094 under LOO-CV** — larger than the in-sample estimate of +0.079. This confirms that H%'s contribution is not an artifact of overfitting but represents genuine predictive value that generalizes beyond the training data (EPV  $\approx 8.8$ ).



**Figure A1.** Empirical bootstrap distributions (10,000 iterations) of the Pearson correlation coefficient  $r$  between Hodge components (G%, C%, H%) and Heart Rate (HR). Dashed red lines indicate the 95% CI bounds; dotted line indicates  $r = 0$ . The H% distribution is strictly negative (95% CI:  $[-0.56, -0.29]$ ), confirming a robust inverse relationship with HR.

## References

- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons.
- Edelsbrunner, H. and Harer, J. (2010). *Computational Topology: An Introduction*. American Mathematical Society.
- Goldberger, A. L. et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.
- Harris, C. R. et al. (2020). Array programming with NumPy. *Nature*, 585:357–362.
- Jiang, X., Lim, L.-H., Yao, Y., and Ye, Y. (2011). Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244.
- Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381. Springer.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer.
- Virtanen, P. et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.