Graph Wiring: Eigenstructures for vector datasets and LLM operations

Lorenzo Moriondo
Independent Researcher - tuned.org.uk
ORCID: 0000-0002-8804-2963

22 February 2026

Abstract

Graph wiring builds discrete graph from arbitrary vector spaces leveraging the Laplacians function for search, partitioning, and topological analysis; **this new technique enables faster, semantic-aware exploration of datasets and latent spaces** in LLMs and embedding models. Starting from any vector space, it is based on transposing any matrix into feature space and constructing a wiring graph via nearest-neighbour pairing. I show that the resulting graph Laplacian acts as a discrete Laplace–Beltrami operator and that, minimising its Rayleigh quotient, the core operation is equivalent to minimising the Dirichlet energy of the feature manifold. Under specific metric constraints (conformal gauge), this Dirichlet minimisation is mathematically equivalent to area minimisation, implying that *graph wiring* implementations effectively constructs a discrete minimal surface (or worldsheet) in feature space.

This result establishes a bridge between surface-optimization principles in physical networks, where minimal surfaces map to string-theoretic worldsheets, and spectral graph algorithms. I develop this theoretical foundation, demonstrating how the synthetic λ_τ index acts as a probe of this worldsheet geometry, and show that invariants of the wired graph enable advanced operations on high-dimensional data unavailable to purely geometric methods.

The technique and its performances (no reliance on $O(N^2)$ computations) as an approximating algorithm opens to **next-generation methods and tools to analyse, measure, operate and supervise large high-dimensional collections** of vectors like LLM models' weights. Implementations are presented in `arrowspace` and `surfface` codebases.

1 Introduction

Building forward from the experience with `arrowspace` (the first *graph wiring* implementation) [1] [2], I introduce here *Graph Wiring as a generalisable algorithmic technique*. This is a generic technique applicable to any collection of vectors that makes possible different applications in analysis, operations and supervising of Large Language Models (LLMs) and any related dataset. Sections below define the process in formal and engineering terms, here is a non-comprehensive list of applications made possible or computationally more accessible:

- **Vector database and retrieval**
  - *Lambda-aware RAG similarity.* Improve semantic capture in search for RAGs by blending cosine with λ_τ scores that encode feature–space roughness over the ArrowSpace manifold [2], enhancing tail quality and long-term multi-query stability.

- *Spectral document triage.* Active document selection for RAG using λ_τ as a cheap proxy for "how much an item deviates from learned feature structure," prioritising spectrally informative but semantically relevant items.
- *Spectral drift monitoring.* OOD / drift monitoring for retrieval indices via temporal tracking of taumode distributions and tail-quality metrics derived from CVE experiments, optimising for better tails quality of results.
- *Hallucination guardrails.* Spectral guardrails for hallucination detection: flag queries whose nearest neighbours lie in sparsely populated or high-λ regions of the feature manifold as structurally unsupported context
- *Bundle-aware search.* Multi-vector "query bundle" search by indexing sub-vectors as ArrowSpace rows and using the manifold Laplacian to respect higher-order feature covariation across items.

- **Training data curation for LLMs**
  - *Epiplexity-style data selection.* Data selection for pretraining or continued finetuning by preferring documents that are spectrally representative yet structurally diverse in λ_τ, mirroring epiplexity-style "structural content" but using ArrowSpace as a computationally cheap estimator [11].
  - *Spectral curricula.* Curriculum design where batches are ordered by spectral complexity (e.g., low to high Rayleigh energy) to encourage stable representation learning and reduce catastrophic interference.
  - *Spectral de-duplication.* De-duplication and near-duplicate suppression in embedding space using feature-space Laplacians rather than pairwise cosine thresholds, improving coverage of the long tail of patterns.
  - *Domain mixing in λ_τ-space.* Spectral mixture-of-domains: cluster documents in λ_τ–space and schedule domain-balanced sampling that preserves manifold coverage rather than raw token counts.

- **Online supervision and safety**
  - *Production retrieval telemetry.* Live retrieval quality monitoring in production by tracking ArrowSpace tail-head ratios over time, using the CVE tail metrics as a template score for RAG stability.
  - *Spectral anomaly alarms.* Spectral anomaly alarms for abuse / prompt-injection patterns, by modelling historical query embeddings and flagging future queries whose λ_τ statistics fall outside learned manifolds.
  - *Spectral active learning for safety.* Active-learning loops for LLM moderation heads: triage user messages near spectral decision boundaries for human review to refine safety classifiers efficiently.

- **Structural and mechanistic analysis of embedding models / LLMs**
  - *SAE feature outlier scoring.* Sparse Autoencoders (SAE) decoder feature analysis: treat decoder columns $W_{\text{dec}} \in \mathbb{R}^{d\times|F|}$ as ArrowSpace items, build the feature-space Laplacian over transposed centroids, and use per-feature Rayleigh / λ as a proxy for "how much a feature deviates from the learned manifold of all features" [10].
  - *Cross-layer feature flow spectra.* Cross-layer feature flow graphs: ingest SAE feature similarity graphs as ArrowSpace graphs, apply spectral decomposition to separate smooth (persistent) from rough (transient) feature trajectories across depth [10].

- *F×F feature-circuit graph.* F×F spectral Laplacian as a feature-circuit detector: use ArrowSpace Stage 4 Laplacian-of-Laplacian on SAE dictionaries to find tightly bound feature communities that are candidates for mechanistic circuits.
- *Spectral weight signatures beyond DOCS.* Weight-matrix spectral signatures: complement DOCS by comparing MLP / attention weights via their λ_τ spectra on a shared ArrowSpace manifold [2], distinguishing layers with identical DOCS distributions but different structural geometry [9].
- *Spectrally guided steering targets.* Spectrally guided steering: choose SAE features for activation editing whose λ-trajectories are smooth across layers, yielding stabler multi-layer steering interventions than purely cosine-based selection [10].
- *Epiplexity-oriented feature mining.* Epiplexity-adjacent diagnostics: interpret outlier λ_τ features or circuits as structurally rich candidates for high-epiplexity subspaces, informing where mechanistic effort or additional data will most improve OOD robustness [11].

- **Spectral/topological search supervision**
  - *Tail-aware retrieval benchmarks.* Standardise evaluation of new RAG retrievers by reusing the CVE tail-quality and NDCG metrics as primary scores, explicitly optimising for better tails quality of the results as a proxy for long-term multi-query stability.
  - *Spectral A/B testing of indices.* Spectral A/B testing of retrieval stacks: compare different indices (HNSW, flat cosine, ArrowSpace) by their induced λ_τ distributions and tail-head ratios under identical query sets, rather than topline recall alone.
  - *Spectral cache management.* Topologically aware cache eviction: evict items from vector stores that live in highly redundant (low-λ) regions, preserving rare, high-epiplexity structures that support future OOD queries [11].

1.1 Motivation: Beyond geometric similarity in vector spaces

Nearest-neighbour methods in vector spaces are typically justified by geometric similarity alone (cosine or Euclidean distance), yet physical networks and many engineered systems optimise functionals that are not purely metric [2]: surface area, action, energy dissipation, or boundary-constrained variational quantities. The motivating question is whether one can build a scalable, graph-native surrogate for such physical optimisation principles, while remaining compatible with the data-engineering reality of large N and high ambient dimension, i.e. a representation in which *diffusion energy* and *manifold structure* become first-class signals alongside semantic proximity. In `arrowspace`, this is instantiated by constructing a feature-space manifold $L = \text{Laplacian}(C^\top)$ and using 1D spectral indices (notably λ_τ) to bias retrieval toward candidates that are consistent with the learned wiring, rather than relying on cosine similarity alone.

The empirical motivation is reinforced by the CVE dense-embedding test campaign [31], where λ-aware spectral search (taumode) delivers a consistent lift over cosine across 18 real vulnerability queries: the average score for head/tail (H/T) quality improves from 0.833 (cosine) to 0.887 (taumode), i.e. a +0.054 absolute gain across all query–rank cells, and taumode wins top-1 on all 18/18 queries in terms of head/tail quality comparison [31]. Critically for retrieval-augmented generation (RAG), the advantage is not confined to the head results, the cumulative score advantage grows approximately linearly to +0.65 by rank 15 (diagram 22 Cumulative Score Advantage), indicating that the spectral information improves tail quality [31]. Tail stability metrics show the same pattern: taumode achieves the best head/tail ratio on 14/18 queries and the lowest tail coefficient of variation on 14/18 queries; suggesting that manifold-aware scoring provides more predictable deep-rank behaviour under multi-query workloads [31]. The second test run on the Dorothea dataset

provides hints for `arrowspace` being effective also as a classifier that can compare with UMAP on clustering while taumode fails in this class of tasks as it has been designed for search and not for classification.

This data supports the paper's further step in the `arrowspace` direction: to introduce a graph-native variational surrogate that is *operationally useful when semantic linking exists* (dense embeddings, manifold structure) and that provides principled failure signals when it does not (high-dimensional sparse domains without coherent neighbourhood structure). In other words, λ_τ should be read as a cheap proxy for deviation from learned structure that improves retrieval and tail stability when the manifold hypothesis is approximately valid, and as a diagnostic that indicates when alternative compressions or operators (e.g. richer spectral coordinates, nonnegative embeddings, or Dirac-type worldsheets) are required. These further operators are being developed in `surfface` [30] that is the extension of `arrowspace` into a more general-purpose setting.

1.2 Graph wiring as manifold reconstruction

The intuition is that this approach can power effective procedures that applies to other classes of problems beside search. For this reason I introduce the more generic concept of *Graph wiring*: the procedure that takes a data matrix $X \in \mathbb{R}^{N\times F}$, transposes to feature-space $X^\top \in \mathbb{R}^{F\times N}$, constructs a sparse neighbourhood graph over features, and uses the resulting Laplacian as a discrete differential operator. The wired graph is then treated as a discrete geometric object: its Laplacian induces a quadratic energy, and its spectrum yields harmonics used for embedding, indexing, and search. This wired graph is used in `arrowspace` by attaching to it the Rayleigh-based score computation (taumode) but this is just a choice of tool for the search problem; what other tools can be attached to a wired graph?

1.3 Contributions: Connecting Laplacian eigenmaps, minimal surfaces, and spectral search

The paper strengthens the core demonstration that (i) wired feature graphs lie in the class of neighbourhood graphs whose Laplacians converge to Laplace–Beltrami operators; (ii) the ArrowSpace-style Rayleigh energies converge to continuum Dirichlet energies [17]; and (iii) with conformal parametrisation assumptions, Dirichlet-energy minimisers correspond to minimal surfaces, justifying a controlled "surface minimisation" interpretation of spectral embeddings. Secondary contributions describe how these constructions integrate into an interface-level computational engine and how λ_τ-style bounded energies support scalable search and dataset operations.

1.4 Outline of the paper

Sections 2–7 develop the formal chain from manifold sampling → graph Laplacians → Dirichlet energies → minimal surfaces, and Section 8 sketches validation and implementation. The concluding sections discuss what is proven vs. conjectured, how to validate assumptions in production pipelines, and open problems in discrete worldsheets and string-inspired graph algorithms.

2 Mathematical Preliminaries

This section includes the mathematical walkthrough for all the layers involved in building the wired graph first and approximating surface minimisation with Dirichlet energy distribution later. The engineering/operative discussion continues from Section 7.

2.1 Manifolds and Riemannian geometry

Let $(\mathcal{M}, g)$ be a smooth, compact d-dimensional Riemannian manifold embedded in $\mathbb{R}^F$ with $d \ll F$, volume form $dV_{\mathcal{M}}$, and a sampling density $p : \mathcal{M} \to (0, \infty)$ such that $\int_{\mathcal{M}} p \, dV_{\mathcal{M}} = 1$. We observe i.i.d. samples $x_1, \ldots, x_n \in \mathcal{M}$ with law $d\mu = p \, dV_{\mathcal{M}}$, which formalises the manifold hypothesis and defines the continuum target for discrete operators.

2.2 Laplace–Beltrami and Dirichlet energy

For $f \in C^\infty(\mathcal{M})$, the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ is (given our sign convention) the divergence of the gradient with respect to g, $\Delta_{\mathcal{M}} f := \mathrm{div}_{\mathcal{M}}(\nabla_{\mathcal{M}} f)$, and the associated Dirichlet energy is

$$\mathcal{E}_{\mathcal{M}}(f) = \frac{1}{2} \int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \, dV_{\mathcal{M}}.$$

Dirichlet energy is the continuum quadratic functional that will arise as the limit of graph Rayleigh/Dirichlet energies under standard manifold-sampling and kernel-graph scalings.

2.3 Graph Laplacians

Given weighted adjacency $W = (w_{ij})$ on n samples, degrees $D = \mathrm{diag}(D_{11}, \ldots, D_{nn})$ with $D_{ii} = \sum_j w_{ij}$, the unnormalised Laplacian is $L = D - W$ and the symmetric normalised Laplacian is $L_{\mathrm{sym}} = D^{-1/2} L D^{-1/2}$. These are discrete diffusion operators and their spectra encode global geometry.

2.4 Rayleigh quotient

For $f \in \mathbb{R}^n \setminus \{0\}$,

$$R_n(f) = \frac{f^\top L f}{f^\top f} = \frac{1}{2} \frac{\sum_{i,j} w_{ij} (f_i - f_j)^2}{\sum_i f_i^2}.$$

Eigenvectors of L are variational minimisers of R_n under orthogonality constraints, mirroring continuum variational principles for $\Delta_{\mathcal{M}}$.

2.5 Minimal surfaces and surface optimisation

For an immersion $X : \Sigma \to \mathbb{R}^D$ of a 2D domain Σ, the Nambu–Goto area functional is

$$\mathcal{S}_{\mathrm{NG}}(X) = \int_\Sigma \sqrt{\det(\partial_\alpha X \cdot \partial_\beta X)} \, d^2\sigma,$$

(and in standard string-theory normalisations one often includes an overall "tension" prefactor, e.g. $(2\pi\alpha')^{-1}$, which we suppress here).[23] The Polyakov functional with auxiliary metric h is

$$\mathcal{S}_{\mathrm{P}}(X, h) = \frac{1}{2} \int_\Sigma \sqrt{\det h} \, h^{\alpha\beta} \, \partial_\alpha X \cdot \partial_\beta X \, d^2\sigma,$$

(where for Lorentzian worldsheets one typically writes $\sqrt{-\det h}$, while the above is the Euclidean-signature form) [23][24]. Classically, the two are equivalent on-shell in the sense that varying $\mathcal{S}_{\mathrm{P}}$ with respect to h yields the constraint that h is Weyl/conformally related to the induced metric $g_{\alpha\beta} = \partial_\alpha X \cdot \partial_\beta X$, i.e. $h_{\alpha\beta} = e^{\phi(\sigma)} g_{\alpha\beta}$, and substituting this back into $\mathcal{S}_{\mathrm{P}}$ recovers $\mathcal{S}_{\mathrm{NG}}$ (up to the suppressed overall prefactor), yielding the Dirichlet-energy viewpoint on minimal surfaces in conformal gauge [24][23].

3 Laplacian Eigenmaps: The Foundation

3.1 Neighbourhood graphs and kernels

Fix a bandwidth $t = t_n \downarrow 0$ and define a heat kernel weight

$$w_{ij} = k_t(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{4t}\right) \mathbf{1}\{\|x_i - x_j\| \le \varepsilon_n\},$$

with $\varepsilon_n \downarrow 0$ possibly coupled to t_n. Let $L_{n,t}$ be the unnormalised Laplacian associated to W.

3.2 Belkin–Niyogi convergence (operator level)

Theorem 3.1 (Graph Laplacian $\to$ Laplace–Beltrami, informal but standard)**.** *Assume $x_1, \dots, x_n \sim p\, dV_{\mathcal{M}}$ on a smooth compact $(\mathcal{M}, g)$, $t = t_n \to 0$, and scaling such that local neighbourhoods shrink while remaining well-populated (e.g. $nt^{d/2+2} \to \infty$ under heat-kernel constructions). Then for $f \in C^\infty(\mathcal{M})$ and sample points x_i,*

$$\frac{1}{nt^{d/2+1}}(L_{n,t}f)(x_i) \;\longrightarrow\; c_d\, \Delta_{\mathcal{M}} f(x_i),$$

where $c_d > 0$ depends only on d.

Explanation. This theorem states that the wired graph Laplacian is a consistent estimator of a manifold differential operator, which is the mathematical entry point for translating graph energies into geometric energies.

3.3 ε-graphs vs. k-NN graphs (scaling regimes)

Assumption 3.2 (Shrinking neighbourhoods with concentration)**.** Either (i) $\varepsilon_n \to 0$ and $n\varepsilon_n^d \to \infty$ for ε-graphs; or (ii) $k_n \to \infty$ and $k_n/n \to 0$ for k-NN graphs, with $k_n \asymp \log n$ ensuring connectivity with high probability.

Explanation. Assumption 3.2 captures the regime in which k-NN wiring is sparse yet geometrically faithful, allowing the same convergence story as ε-graphs.

3.4 Rayleigh quotients converge to Dirichlet energy

Theorem 3.3 (Rayleigh $\to$ Dirichlet, variational limit (informal))**.** *Under Theorem 3.1 and Assumption 3.2, for sequences $f_n \in \mathbb{R}^n$ representing restrictions of $f \in H^1(\mathcal{M})$ to the sample, the properly rescaled graph energy satisfies*

$$\frac{1}{n^2 t^{d/2+1}}\, f_n^\top L_{n,t} f_n \;\longrightarrow\; C \int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2\, p(x)\, dV_{\mathcal{M}}(x),$$

for a constant $C > 0$ depending on conventions.

Explanation. This is the key quadratic-form convergence: graph smoothness energies converge to Dirichlet energies, hence minimising Rayleigh quotients approximates minimising Dirichlet energy on $\mathcal{M}$.

4 k-NN Graph Construction as Manifold Sampling

4.1 Feature manifold hypothesis

Given $X \in \mathbb{R}^{N \times F}$, define feature-vectors $\phi_j = X_{j,\cdot}^\top \in \mathbb{R}^N$. Assume features $\{\phi_j\}_{j=1}^F$ lie near a low-dimensional manifold $\mathcal{M}_F \subset \mathbb{R}^N$ of intrinsic dimension $d_F \ll N$. This is the "metadata manifold" assumption needed for feature-first wiring.

4.2 Connectivity and geometric fidelity of k-NN wiring

Proposition 4.1 (k-NN graph in the eigenmap regime (informal))**.** *Let $\{\phi_j\}_{j=1}^{F} \sim p_F \, dV_{\mathcal{M}_F}$ on a smooth compact $\mathcal{M}_F$ of dimension d_F. If $k_F \asymp C \log F$ for sufficiently large C, then the undirected symmetrised k_F-NN graph is connected with high probability, and its neighbourhood radius shrinks to 0 while containing k_F samples.*

Explanation. Proposition 4.1 is what makes the ArrowSpace wiring compatible with the eigenmaps convergence results: it justifies using k-NN rather than ε-graphs without leaving the asymptotic regime.

4.3 JL stability (optional but practical)

Lemma 4.2 (Neighbourhood stability under JL projections (informal))**.** *Let $R \in \mathbb{R}^{r \times N}$ be a JL transform with $r = O(\log F/\varepsilon^2)$ and set $\tilde{\phi}_j = R\phi_j$. Then pairwise distances satisfy $\|\tilde{\phi}_i - \tilde{\phi}_j\| = (1 \pm \varepsilon)\|\phi_i - \phi_j\|$ with high probability, implying that k-NN neighbourhoods are preserved up to controlled perturbations.*

Explanation. Lemma 4.2 justifies the engineering step "reduce dimension before wiring" without invalidating the manifold approximation.

5 From Minimal Surfaces to Dirichlet Energy

5.1 Dirichlet energy and harmonic maps

Let $X : \Sigma \to \mathbb{R}^D$ be a smooth map from a 2D domain. The Dirichlet energy is

$$\mathcal{E}_{\mathrm{Dir}}(X) = \frac{1}{2} \int_{\Sigma} \|\nabla_{\Sigma} X\|^2 \, dA.$$

Critical points satisfy the Euler–Lagrange equation $\Delta_{\Sigma} X = 0$, so Dirichlet minimisers are harmonic maps.

5.2 Polyakov $\leftrightarrow$ Nambu–Goto (classical equivalence)

Theorem 5.1 (Polyakov and Nambu–Goto equivalence, classical)**.** *Let $X : \Sigma \to \mathbb{R}^D$ be an immersion and h a Riemannian metric on Σ. If h is chosen conformal to the induced metric $g_{\alpha\beta} = \partial_\alpha X \cdot \partial_\beta X$, then critical points of $\mathcal{S}_{\mathrm{P}}(X, h)$ in X coincide with critical points of $\mathcal{S}_{\mathrm{NG}}(X)$, and the Euler–Lagrange equation is the minimal surface equation (vanishing mean curvature).*

Explanation. Theorem 5.1 is the continuum bridge that turns a quadratic Dirichlet-type energy into an area minimisation statement, under parametrisation constraints.

5.3 Graph Dirichlet energy as discrete Polyakov

For an embedding $Y \in \mathbb{R}^{n \times D}$ of graph vertices into $\mathbb{R}^D$, define the graph Dirichlet energy

$$\mathcal{E}_{\mathrm{graph}}(Y) = \mathrm{Tr}(Y^{\top} L Y) = \frac{1}{2} \sum_{i,j} w_{ij} \|Y_i - Y_j\|^2.$$

This is a quadratic functional directly analogous to $\mathcal{S}_{\mathrm{P}}$: it is a discrete Dirichlet energy induced by the graph Laplacian.

5.4 The core approximation theorem (Dirichlet ⇒ surface)

Theorem 5.2 (Graph wiring approximates surface minimisation (structured statement))**.** *Assume:*

1. *(**Manifold regime**) Feature vectors $\{\phi_j\}_{j=1}^{F}$ sample a smooth compact manifold $\mathcal{M}_F$ and wiring satisfies Assumption 3.2.*
2. *(**Operator convergence**) The wired Laplacian L_F converges (after rescaling) to $\Delta_{\mathcal{M}_F}$ in the sense of Theorem 3.1.*
3. *(**Conformal parametrisation**) The continuum limit of the embedding is taken in a conformal gauge so that the Polyakov and Nambu–Goto variational problems coincide as in Theorem 5.1.*

Then minimisers of $\mathcal{E}_{\text{graph}}$ over admissible embeddings Y converge (in the manifold limit $F \to \infty$) to minimisers of the continuum Dirichlet energy on $\mathcal{M}_F$, which, under the conformal parametrisation condition, correspond to minimal surfaces in the target space.

Explanation. Theorem 5.2 is the strengthened "spine": (i) graph wiring gives a consistent Laplace–Beltrami surrogate; (ii) the graph quadratic form converges to Dirichlet energy; and (iii) Dirichlet minimisation corresponds to area minimisation in conformal gauge, hence spectral minimisation approximates surface minimisation.

Proof sketch of Theorem 5.2

The proof proceeds in three steps.

Step 1: Γ-convergence of the graph Dirichlet energy. By Assumption 1 (manifold regime) and the standard ε_n-scaling condition $\varepsilon_n \to 0$, $\varepsilon_n \gg (\log n/n)^{1/d}$, the graph Dirichlet energy

$$\mathcal{E}^{(F)}_{\text{graph}}(Y) = \operatorname{Tr}(Y^\top L_F Y) = \tfrac{1}{2}\sum_{i,j} w_{ij}\|Y_i - Y_j\|^2$$

(suitably renormalised by $n^2\varepsilon_n^{d+2}$ where $d = \dim \mathcal{M}_F$) Γ-converges, as $F \to \infty$, to the continuum Dirichlet energy

$$\mathcal{E}_{\mathcal{M}_F}(y) = \tfrac{1}{2}\int_{\mathcal{M}_F} \|\nabla_{\mathcal{M}_F} y\|^2 \, dV_{\mathcal{M}_F}$$

in the $TL^2(\mathcal{M}_F)$ topology (transport-L^2 induced by optimal transport between empirical and limit measures). This is the main result of Trillos–Slepcev [7] and the related consistency analysis of [8]; see also [**?**] for a self-contained treatment in the unweighted geometric-graph setting. The Γ-convergence comprises two inequalities:

(i) *(liminf)* For any sequence $Y^{(F)}$ with TL^2-limit y, $\liminf_{F\to\infty} \mathcal{E}^{(F)}_{\text{graph}}(Y^{(F)}) \geq \mathcal{E}_{\mathcal{M}_F}(y)$.

(ii) *(recovery sequence)* For every $y \in H^1(\mathcal{M}_F, \mathbb{R}^D)$ there exists $Y^{(F)} \to y$ in TL^2 with $\lim_{F\to\infty} \mathcal{E}^{(F)}_{\text{graph}}(Y^{(F)})$ $\mathcal{E}_{\mathcal{M}_F}(y)$.

Step 2: Convergence of minimisers. Restrict to the admissible set $\mathcal{A}_F := \{Y : Y$ is a conformal embedd and its continuum counterpart $\mathcal{A} := \{y : y$ is a conformal immersion, $h = e^\phi g,\ g = \partial_\alpha y \cdot \partial_\beta y\}$. Under Assumption 1 (compact manifold, bounded density) and the normalisation constraint $\int_{\mathcal{M}_F} |y|^2 \, dV =$ 1 (to ensure coercivity), any sequence of minimisers $Y_*^{(F)} \in \mathcal{A}_F$ of $\mathcal{E}^{(F)}_{\text{graph}}$ is precompact in TL^2. By the fundamental theorem of Γ-convergence [**?**] (see also [**?**]), every cluster point y_* of $\{Y_*^{(F)}\}$ minimises $\mathcal{E}_{\mathcal{M}_F}$ over $\mathcal{A}$.

Step 3: Dirichlet minimisers in conformal gauge are minimal surfaces. By Assumption 3 and the classical equivalence (Theorem 5.1 [23]), any conformal immersion $y \in \mathcal{A}$ satisfying $\delta\mathcal{E}_{\mathcal{M}_F}(y) = 0$ is harmonic, and a harmonic conformal immersion $\Sigma \to \mathbb{R}^D$ parametrises a minimal surface [5][17]: its mean curvature vector vanishes identically. Hence y_* parametrises a minimal surface in $\mathbb{R}^D$.

Conclusion. Combining Steps 1–3: minimisers of $\mathcal{E}_{\text{graph}}$ over $\mathcal{A}_F$ converge (in TL^2, as $F \to \infty$) to a conformal harmonic immersion that parametrises a minimal surface. This is the content of Theorem 5.2. □

Remark (where additional work is needed). The argument is complete given Assumption 3. If conformal gauge must be *derived* rather than imposed, one would need either a reparametrisation theorem (existence of isothermal coordinates on the discrete surface, convergent as $F \to \infty$) or a penalisation argument enforcing conformality in the limit. This is the remaining open step that elevates the result from a structured assumption to a fully unconditional theorem.

5.5 Branching and trifurcations (mostly conjectural)

Meng et al. show that surface minimisation induces topology changes (bifurcation → trifurcation) at a critical thickness parameter χ; in graph wiring, the analogous objects are local high-degree motifs, curvature proxies, or junction-like anomalies in the wired manifold. A precise mapping between χ and graph-theoretic observables (degree distributions, local energy concentrations, discrete curvature) remains open, but the worldsheet interpretation suggests that branching corresponds to concentration of harmonic energy and to singularity-like regions of the induced discrete geometry.

6 ArrowSpace as Discrete Worldsheet Construction

6.1 From data to a *feature-space* Laplacian (algorithmic pipeline)

Given a data matrix $X \in \mathbb{R}^{N\times F}$, we treat the *features* as the sampled objects: for each feature index $j \in \{1, \dots, F\}$ define the feature-vector (a "feature point cloud") $\phi_j := X_{j,\cdot}^\top \in \mathbb{R}^N$. We then build a symmetrised k-NN graph G_F over $\{\phi_j\}_{j=1}^F$ (using a metric induced by the representation, e.g. cosine or ℓ_2), assign kernel weights w_{ij} (e.g. Gaussian / heat-kernel weights), and form a (possibly normalised) graph Laplacian L_F. This mirrors the standard manifold-learning construction where a Laplace–Beltrami operator on an underlying manifold is approximated by a kernel graph Laplacian built from samples.[4][6]

Manifold invariant (feature-first). In ArrowSpace, the primary "manifold" object is the feature-space wiring induced by the Laplacian built on feature representatives (e.g. centroids in feature space), i.e. one targets a Laplacian of the form $L_F \approx$ Laplacian(Transpose(Centroids)) rather than a Laplacian over items. This design choice is what makes subsequent Rayleigh-type energies interpret "roughness" relative to learned *feature co-variation* rather than item adjacency.[**?**][?]

6.2 Rayleigh energies and the synthetic λ_τ index

For an item vector $x \in \mathbb{R}^F$ (a coordinate over features), define the Rayleigh energy

$$R(x) \;=\; \frac{x^\top L_F x}{x^\top x}.$$

Interpreting L_F as a discrete Laplace-type operator, $x^\top L_F x$ is a graph Dirichlet energy measuring how "non-smooth" x is on the wired feature manifold, in the same sense that Dirichlet energies are the quadratic forms associated to Laplacians in the continuum.[8]

To obtain a stable bounded scalar coordinate, ArrowSpace introduces the synthetic index

$$\lambda_\tau(x) = \frac{R(x)}{R(x) + \tau}, \qquad \tau > 0,$$

which maps unbounded energies to $[0, 1)$ and can be used for λ-aware search, stratification, and tail-aware retrieval heuristics.[**?**][**?**]

6.3 Worldsheet interpretation: Dirichlet ⇒ area (where valid)

The conceptual bridge to surface optimisation is variational: in the continuum, Dirichlet energy of a map and area coincide under conformal parametrisation, so energy-minimising (harmonic) conformal immersions parametrise minimal surfaces.[23][5][17] This motivates interpreting graph Dirichlet energies (quadratic forms of Laplacians) as discrete Polyakov-type energies, and thus as computational proxies for surface-minimising behaviour when the discretisation regime is appropriate.[8][4]

Proof program (generic graph wiring ≈ surface minimisation). A defensible route to a theorem is to separate what is *known* from what is *additional*: (i) show (after the right normalisation and scaling) that the graph Dirichlet energy $\frac{1}{2}\sum_{i,j} w_{ij}(u_i - u_j)^2$ Γ-converges to a continuum Dirichlet energy on a manifold sampled by $\{\phi_j\}$; (ii) use the fundamental theorem of Γ-convergence to obtain convergence of (approximate) minimisers; (iii) impose (or prove) a conformal-gauge condition so that continuum Dirichlet minimisers correspond to area-minimising (minimal-surface) solutions.[8][6][23] Steps (i)–(ii) are by now standard in the variational analysis of graph-based learning, while step (iii) is the "string-inspired" ingredient that typically requires extra structure or a parametrisation constraint.[8][5]

6.4 Coarsening as renormalisation (centroids, Nyström, multiscale Laplacians)

For large F, one can reduce the feature graph by centroiding or landmarking in feature space, and build a smaller Laplacian $L_{F'}$ that approximates the leading spectrum (and thus the dominant quadratic-form geometry) of L_F. Such Nyström-style approximations are widely used in scalable spectral methods and come with perturbation-style guarantees under structural assumptions on the Laplacian / affinity matrix.[**?**][8] Under this view, coarsening is a computational analogue of coarse-graining: it aims to preserve the low-frequency ("smooth worldsheet") geometry while trading away high-frequency detail that is often noisy at finite sample size [6][**?**].

6.5 Convergence and approximation (informal finite-sample rate)

A representative bias–variance trade-off in kernel graph Laplacian analyses takes the form

$$\left|R_n(f_n) - \mathcal{E}_{\mathcal{M}}(f)\right| = O\left(\varepsilon_n^2 + \frac{\log n}{n\varepsilon_n^d}\right) \quad \text{with high probability,}$$

where ε_n is the neighbourhood/kernel scale and d is the intrinsic dimension of the manifold being sampled. This encapsulates the practical fragility of Laplacian methods at large intrinsic dimension (shrinking neighbourhoods reduces bias, but overly sparse neighbourhoods increase variance), and it is one reason to motivate feature coarsening, projections, and careful kernel tuning in deployments [4][8].

7 Experimental Validation

As mentioned above, `arrowspace` has been tested for capabilities, speed and accuracy on a dense text embeddings representation (CVE dataset) and a sparse high-dimensional representation (Dorothea dataset). In both cases the wiring of the graphs on these datasets was successful and brought meaningful results in terms of search (CVE) and clustering (Dorothea). Code and results are available at [31].

As a initial approach: on synthetic manifolds with known Laplace–Beltrami spectra, validate convergence by tracking eigenvalue error and Rayleigh energy convergence as n grows. On real embedding datasets (e.g. vision/text features), validate intrinsic dimension estimates, spectral stability under subsampling, and downstream stability of λ_τ-aware retrieval and clustering.

7.1 Applications: Spectral Search and Dataset Operations

λ-aware vector search can be treated as a *two-stage retrieval operator* that first applies a spectral prior and only then spends semantic compute on reranking. Concretely, each candidate embedding x is assigned an energy $E(x) = \frac{x^\top Lx}{x^\top x}$ (a Rayleigh-quotient style smoothness/roughness score on the learned feature-space graph), and the system restricts retrieval to a narrow energy band $E(x) \in [E_-, E_+]$ before applying cosine similarity or cross-encoder scoring within that band. This yields a practical blend of geometry (semantic proximity) and diffusion energy (alignment to the learned manifold): low-energy items behave like "typical" manifold-consistent points, while high-energy items behave like "high-curvature" or rare-structure points that are often useful for novelty, counterexamples, or out-of-distribution screening. In operational terms, λ-banding acts as a cheap, stable prefilter that reduces candidate set entropy and makes downstream reranking both faster and more tail-reliable under multi-query workloads.

Dataset fingerprinting extends the same spectral object L into a *versionable invariant* for corpora and indexes. A robust fingerprint can be defined by (i) a truncated Laplacian spectrum $\{\lambda_1, \ldots, \lambda_k\}$ (or several local spectra over subgraphs), and (ii) summary statistics of the λ_τ distribution induced by the library's 1D linearisation (e.g., quantiles, tail-mass, or mixture parameters). Truncation makes the fingerprint resilient to small-scale noise while remaining sensitive to structural changes in the learned manifold; pairing eigenvalues with λ_τ summaries captures both "global shape" (connectivity/expansion encoded by low-frequency modes) and "population placement" (how items occupy energy bands). This supports dataset operations such as deduplication across snapshots, provenance tracking for derived indexes, regression testing of pipeline changes (e.g., clustering parameters or graph wiring), and fast compatibility checks when merging shards, since mismatched spectral fingerprints indicate incompatible manifold geometry even if simple vector statistics (mean norm, PCA variance) look similar.

Drift detection can then be formalized as monitoring *spectral perturbations* as proxies for manifold deformation. At the global level, changes in the truncated spectrum (e.g., $\|\lambda^{(t)}_{1:k} - \lambda^{(t-1)}_{1:k}\|_2$) and in the spectral gap (commonly $\lambda_2 - \lambda_1$ for the unnormalized Laplacian, or the analogous gap under the chosen normalization) provide early warning that connectivity, mixing time, or cluster separability has changed, even before downstream task metrics degrade. At the local level, computing gaps and low-rank spectra on sliding subgraphs (e.g., neighborhoods around frequently retrieved features or high-usage centroids) enables targeted alarms: a collapsing local gap suggests emerging bridges, topology changes, or new modes that may destabilize retrieval tails, while a widening gap suggests fragmentation or over-regularization. Coupling these signals with λ_τ distribution shift (especially in the tails) yields an actionable drift monitor: if spectral invariants move but semantic reranking performance remains stable, the system can schedule re-indexing; if both move, it can trigger active learning, re-clustering, or re-anchoring to restore manifold consistency.

7.2 Summary

The chain of reasoning that this paper is attempting to prove is: neighbourhood graphs approximate Laplace–Beltrami operators; graph Rayleigh quotients approximate Dirichlet energies; and, in conformal gauge, Dirichlet minimisers correspond to minimal surfaces.

The first link is supported by the Laplacian Eigenmaps line of work, where a weighted graph Laplacian built from local neighbourhood relations is treated as a discretization of the manifold Laplace–Beltrami operator, with convergence results available under sampling and bandwidth conditions (and with careful attention to normalization and density effects).

The second link is essentially variational: the Rayleigh quotient of a (graph) Laplacian is proportional to a Dirichlet form measuring the squared gradient of a function over the underlying discrete structure, so minimizing a Rayleigh quotient is equivalent to minimizing a Dirichlet energy subject to constraints. The third link imports the string-theoretic viewpoint that, after choosing conformal gauge, the worldsheet action reduces to a Dirichlet-type energy in the embedding coordinates, and stationary points correspond to minimal-area surfaces (with the caveat that the gauge fixing and target geometry conditions matter for what is rigorously implied).

The conjectural portion is that the full `arrowspace` pipeline (approximate neighbours, coarsening, heterogeneous data) *remains in the asymptotic regime sufficiently well that minimal-surface reasoning predicts detailed physical motifs* (e.g. Meng-type trifurcations) at finite sample sizes. This is nontrivial because each engineering choice perturbs the operator: approximate k-NN and sparsification alter graph connectivity and effective kernel bandwidth; centroid coarsening replaces the raw sample measure with a compressed measure; and heterogeneous feature distributions induce non-uniform sampling that can bias naive Laplacians unless normalization or density-correction is applied. In this view, "minimal surface" is best treated as a *controlled approximation*: the discrete Dirichlet functional $x^\top Lx$ acts as a computable surrogate for surface area only insofar as the learned Laplacian L retains the correct low-frequency geometry of the underlying manifold, and only insofar as the relevant modes are stable under the pipeline's approximations. The practical claim is therefore operational rather than metaphysical: even if the continuum analogy is imperfect, the induced spectral quantities (low-end spectrum, local spectral gaps, and energy-band structure) can still serve as reliable diagnostics of when the learned manifold is being deformed, when new branches are forming, and when topological motifs are likely to appear.

In production, assumptions are validated via spectral stability under subsampling, intrinsic dimension diagnostics, k/bandwidth sensitivity, and task-level regressions of performance against these geometric stability measures. Subsampling tests probe whether truncated spectra and spectral gaps persist under random deletions, which is a necessary condition for treating the observed eigenstructure as a property of the data manifold rather than an artifact of finite sampling. Sensitivity sweeps over k and kernel bandwidth assess whether the operator is in a regime where neighbourhood graphs remain locally consistent (too small: disconnected graphs and unstable gaps; too large: oversmoothing and loss of curvature information), while intrinsic-dimension diagnostics help detect when the assumed manifold model is breaking down due to heavy heterogeneity or near-random structure. Finally, performance is regressed against these stability measures to ensure that improvements (e.g. from λ-aware candidate banding or drift monitors) correlate with improved geometric consistency rather than accidental dataset-specific effects. An implementation of this advanced concept that includes the approximate equivalence beteen Dirichlet and surface minimisation has been developed and can be tested in the `surfface` repository [30].

8 Discussion

8.0.1 What is Proven vs What is Conjectured

The rigorous part of the story concerns the chain from sampled manifolds to graph Laplacians and then to Dirichlet energies: under standard assumptions on sampling density, neighbourhood scale, and kernel choice, we know that graph Laplacians converge (in operator and spectral senses) to the Laplace–Beltrami operator, and that discrete Rayleigh quotients converge to continuum Dirichlet energies. What is also proven, in classical differential geometry and string theory, is the equivalence between minimising Dirichlet energy and minimising area for isometric immersions in conformal gauge, which justifies treating harmonic maps as minimal surfaces [16][17][18]. What remains conjectural is that the full ArrowSpace graph-wiring pipeline—including approximate neighbours, coarsening, feature-first wiring, and the synthetic λ_τ index—lands in the regime where these asymptotic results tightly approximate Meng-type surface minimisation on realistic, finite, noisy datasets, and that the induced "worldsheet" geometry captures all morphologically relevant degrees of freedom. `surfface` [30] is a candidate to bridge these last step to make the concept started by `arrowspace` into an approximate algorithm for surface minimisation, this is promising as `arrowspace` demonstrated to be very effective on improving quality on tail/head ratios for vector search [31], in particular on dense, highly semantically linked datasets. If this conjecture is proved correct for certain values for parameters with which the pipeline is computed, the class of applications described in this paper would be among the first applications of string-theoretic principles to software engineering.

8.0.2 How to Validate Assumptions in Production Pipelines

In practice, a production pipeline cannot directly inspect the underlying manifold, so we validate assumptions indirectly via a battery of stability and scaling tests that approximate the theoretical limits. One typically monitors spectral stability under subsampling (e.g. checking that leading eigenvalues and eigenvectors of the feature Laplacian change slowly as more data arrive), estimates intrinsic dimension and compares it to feature count, and probes locality by varying k and bandwidth over reasonable ranges while tracking downstream metrics such as retrieval accuracy or clustering quality. Additional checks include monitoring how λ_τ distributions evolve under time-varying data, and whether geometric quantities (e.g. approximate curvature or motif counts) exhibit convergence as more samples are added; together, these diagnostics provide empirical evidence that the graph wiring is operating within the regime where the manifold and minimal-surface approximations are meaningful rather than pathological. Some examples for this kind of tests are available in this subpackage

8.0.3 Relation to Diffusion Maps, Spectral Clustering, and TDA

Diffusion maps [19], spectral clustering [20], and many constructions in topological data analysis (TDA)[21] [22] can all be understood as different ways of extracting structure from the same underlying Laplace-type operators and neighbourhood graphs. Diffusion maps emphasise long-time heat-kernel behaviour and diffusion distances [19], spectral clustering focuses on low-lying eigenvectors as relaxed solutions to cut problems [20], and TDA often builds simplicial complexes from neighbourhood relations and studies their homology [21] [22]. Graph wiring and ArrowSpace sit naturally in this ecosystem: they adopt the same Laplacian backbone, but interpret its quadratic form as a discrete worldsheet energy and use spectral quantities primarily as proxies for minimal surfaces and branching structures, providing a complementary geometric lens that is particularly tuned to surface optimisation and physical-network morphology rather than solely to clustering or topological invariants.

9 Conclusion

Graph Wiring provides a computable bridge from large-scale vector data to discrete geometric operators whose energies converge to Dirichlet functionals, and under conformal parametrisation this supports a principled approximation to surface minimisation. The immediate open problems are to tighten finite-sample bounds tailored to k-NN wiring with ANN error, formalise discrete Polyakov/Nambu–Goto correspondences on sparse graphs, and derive explicit connections between physical thickness parameters (such as χ in Meng et al.) and observable graph invariants in wired feature manifolds.

9.0.1 Implications for AI Operations and Data Engineering

For AI operations and data engineering, treating feature-space graphs as discrete worldsheets equipped with a surface-like energy has several concrete implications. First, it offers a principled way to define λ-aware search, dataset partitioning, and drift detection that are sensitive not only to pairwise similarity but also to the global geometry of feature manifolds, potentially improving retrieval robustness and monitoring. Second, it provides a bridge between continuous physical reasoning (via Dirichlet and area minimisation) and large-scale vector infrastructure, allowing domain experts to encode physical priors—such as smoothness, curvature constraints, or expected branching motifs—directly into the graph wiring and indexing scheme. Finally, by grounding heuristic graph operations in a partially rigorous geometric framework, it opens a path to more auditable and scientifically interpretable ML systems, where changes in performance can be traced to shifts in manifold geometry rather than opaque black-box behaviour.

9.0.2 Summary of Rigorous Contributions

The rigorous core of graph wiring can be summarised as a sequence of controlled approximations: high-dimensional feature data are assumed to lie on a smooth low-dimensional manifold; k-NN graphs with appropriate scaling reconstruct its local geometry; graph Laplacians converge to the Laplace–Beltrami operator; and discrete Rayleigh quotients converge to continuum Dirichlet energies. Combined with the classical equivalence between Dirichlet energy minimisers and minimal surfaces under conformal parametrisation, this yields a mathematically defensible claim that spectral embeddings derived from the wired feature graph approximate minimal surfaces in feature space. The introduction of the bounded λ_τ index provides a stable scalar proxy for these energies, and the complexity analysis shows that the entire construction is computationally tractable at the scales relevant for modern data and physical-network workloads.

9.0.3 Open Problems and Next Steps

Several key questions remain open at both theoretical and practical levels. On the theory side, we lack sharp finite-sample error bounds that tie specific graph-wiring choices (e.g. choice of k, kernel, coarsening strategy) to quantitative approximations of surface area in physically relevant regimes, and we do not yet have a full discrete analogue of the Polyakov/Nambu–Goto equivalence. On the applied side, we need systematic benchmarks comparing ArrowSpace-based minimal-surface approximations to classical PDE or variational solvers across domains such as Meng-type networks, fluid interfaces, and orbit design, as well as ablation studies that map where the approximation starts to break. Next steps include deriving tighter probabilistic guarantees for λ_τ-based search, developing adaptive wiring schemes that respond to local curvature estimates, and integrating richer physical constraints (e.g. anisotropy, inhomogeneous media) into the graph energy itself.

9.0.4 Graph Wiring as a General Technique

Beyond ArrowSpace and Meng-type morphologies, graph wiring should be viewed as a general recipe for turning high-dimensional data into a discrete geometric object amenable to physical and engineering reasoning. The recipe is agnostic to the origin of features: whether they come from language models, vision encoders, scientific simulations, or sensor arrays, the same steps—feature-first wiring, Laplacian construction, spectral analysis, and energy-based indexing—can be applied to induce a worldsheet-like structure. In this sense, graph wiring is not a one-off approximation for a particular physics problem but a reusable pattern for constructing manifolds, surfaces, and approximate energies on top of arbitrary representation spaces; thereby giving practitioners a unifying language in which to talk about "shape", "curvature", and "tension" in their data in ML practitioners terms. An implementation of this concept as a general technique is being developed and can be tested in the `arrowspace` [1] `surfface` repository [30].

9.0.5 Open Problems in Discrete Worldsheets and String–Inspired Graph Algorithms

Interpreting feature graphs as discrete worldsheets immediately raises a host of string-inspired questions that remain almost entirely unexplored in the discrete, data-driven setting. We would like to understand how to define and compute discrete analogues of mean curvature, extrinsic curvature, and stress–energy tensors on sparse graphs; how to formulate and solve discrete versions of Polyakov or Nambu–Goto actions with constraints; and whether renormalisation or coarse-graining ideas from random-surface theory can guide multiscale graph coarsening and spectral approximations. Another open direction is to relate worldsheet deformations to changes in downstream performance or robustness, effectively turning string-inspired deformations into actionable levers in graph algorithms and ML pipelines. Progress on these fronts would elevate graph wiring from a useful approximation to a richer mathematical framework for discrete surface-based computation.

9.0.6 `arrowspace` and Research on Metamaterials

For metamaterials and other engineered physical media, `arrowspace` offers a concrete computational tool for exploring and optimising families of microstructures viewed as points on a high-dimensional manifold. By wiring feature representations of candidate unit cells or network morphologies into a worldsheet and using λ_τ and spectral embeddings as proxies for effective mechanical or transport properties, researchers can search for designs that approximate minimal surfaces or other extremal geometries known to have desirable behaviour (e.g. high stiffness-to-weight ratios, tailored band gaps). In this context, the same spectral and worldsheet operations used for vector search and clustering become mechanisms for navigating design spaces, interpolating between existing metamaterials, and detecting emergent motifs such as trifurcations or gyroid-like patterns, thereby linking the abstract mathematics of graph wiring directly to experimental and industrial agendas in structured matter.

Substituting a Dirichlet/Laplacian worldsheet with a Dirac worldsheet is a plausible way to turn `arrowspace`/`surfface` from "geometry-aware indexing" into a library that can also *simulate* (classes of) metamaterials, because many metamaterial phenomena are more naturally expressed in terms of an effective *Hamiltonian* (often Dirac-like) than in terms of an energy-minimizing surface functional. In the current formulation, the discrete Dirichlet energy $x^\top Lx$ encodes smoothness on a learned manifold graph and is well-aligned with diffusion, clustering, and minimal-surface analogies. A Dirac formulation instead starts from an operator D whose square yields a Laplacian-type operator (schematically $D^2 \sim \Delta$ up to curvature/connection terms), and whose *spectrum and edge modes* encode topological and dispersive properties that are central in photonic/phononic/mechanical metamaterials and topological media; Dirac equations are explicitly used as effective models in metamaterial/topological settings, and have been demonstrated in photonic metamaterial simulations of Dirac physics [29]. This shift changes what λ "means": eigenvalues of L

emphasize diffusion/mixing and smoothness, while eigenvalues of D (and the structure of its near-zero modes) emphasize band structure, chirality, and boundary-localized states—often the actual engineering targets.

9.0.7 Graph Wiring and Quantum Computing

If an equivalence between graph wiring and surface minimisation were proved in a form strong enough or if it will be found to hold under the discretisations used in realistic pipelines, it would open several quantum-computing applications: (i) surface-code layout and lattice-surgery compilation could be recast as an explicit discrete surface-optimisation problem, where patch merge/split schedules and braiding-free interactions are chosen by minimising an induced "worldsheet cost" subject to hardware locality constraints, potentially yielding lower space–time overheads for fault-tolerant circuits [25][26]. (ii) more generally, any workflow that already treats computation as geometry on a 2D/3D lattice—such as topological-code scheduling and distillation-block design—could use the surface-minimisation objective as a unifying optimisation primitive for routing, deformation, and coarse-graining of logical patches across time [26][25]. (iii) because many quantum graph/linear-algebra routines work directly with graph Laplacians, a proved surface-minimisation interpretation could also motivate new Laplacian-native quantum subroutines (or improved preconditioning/graph reductions) for graph-based QML tasks such as quantum spectral clustering [27]. (iv) finally, device-level qubit placement and swap-minimising compilation on connectivity-limited hardware can be approached with spectral graph methods; a surface-minimisation equivalence would provide an additional physically meaningful objective for these spectral placements, linking embedding/layout choices to an explicit discrete surface energy rather than purely heuristic criteria [28].

9.1 Disclaimer

The mathematical walkthrough in Section 2 has been reconstructed by LLMs from code developed by the author to set in formally recognisable steps the backbone of the computation and it does not reflect a formal understanding but it is instead the stepwise progression of the intuition pathway that led to the definition of Graph Wiring as an algorithm for surface minimisation. In case of errors or discrepancies always refer to the latest version of the code and documentation at [1] and [30].

10 Acknowledgements

This work made use of LLM tools to assist with drafting, editing, and code/documentation iteration; all technical claims, experiments, and final text remain my responsibility. I thank Enverge Lab for providing the computation cluster on which the experiments reported in this work were executed. I am also grateful to my GitHub Sponsors for their support, which materially enables me to continue this research as an independent researcher. If you would like to support ongoing work on ArrowSpace and related research, please consider sponsoring me at my sponsors page. Your sponsorship helps fund compute time, open-source maintenance, and the time required to develop, test, and communicate these ideas in a transparent and reproducible way.

References

[1] L. Moriondo, *ArrowSpace-rs: Spectral Vector Search with Lambda-Tau Indexing.* Rust codebase for spectral indexing, 2025. https://github.com/Mec-iS/arrowspace-rs

[2] L. Moriondo. ArrowSpace: introducing Spectral Indexing for vector search. *Journal of Open Source Software*, 10, 2025. doi:10.21105/joss.09002. URL: https://joss.theoj.org/papers/10.21105/joss.09002.pdf.

[3] B. Osting. Consistency of Dirichlet partitions. *SIAM Journal on Mathematical Analysis*, 49(6):4251–4279, 2017. URL: https://arxiv.org/pdf/1708.05472.pdf.

[4] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In *Proceedings of COLT 2005*, 2005. URL: https://www.aoml.noaa.gov/ftp/pub/phod/thacker/SSAmanifoldPapers/BelkinNiyogi2005.pdf.

[5] J. Jost. *Two-Dimensional Geometric Variational Problems.* John Wiley & Sons, Chichester, 1991.

[6] M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems (NIPS)*, 2008. URL: https://www.aoml.noaa.gov/ftp/pub/phod/thacker/SSAmanifoldPapers/BelkinNiyogi2008.pdf.

[7] N. García Trillos and D. Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 2016. arXiv:1403.6355. URL: https://www.math.cmu.edu/users/slepcev/cont_TV_PC_new.pdf.

[8] N. García Trillos and D. Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018. arXiv:1508.01928. URL: https://www.math.cmu.edu/users/slepcev/spectral_gamma10.pdf.

[9] Z. Min and X. Wang. DOCS: Quantifying Weight Similarity for Deeper Insights into Large Language Models. *arXiv preprint*, arXiv:2501.16650, 2025. URL: https://arxiv.org/abs/2501.16650.

[10] Balcells et al. Analyze Feature Flow to Enhance Interpretation and Steering in Language Models. *arXiv preprint*, arXiv:2502.03032, 2025. URL: https://arxiv.org/abs/2502.03032.

[11] M. Finzi, S. Qiu, Y. Jiang, P. Izmailov, J. Z. Kolter, and A. G. Wilson. From Entropy to Epiplexity: Rethinking Information for Computationally Bounded Intelligence. *arXiv preprint*, arXiv:2601.03220, 2026. URL: https://arxiv.org/abs/2601.03220.

[12] J. Hoffmann et al. Training Compute-Optimal Large Language Models. *arXiv preprint*, arXiv:2203.15556, 2022. URL: https://arxiv.org/abs/2203.15556.

[13] Sridhar Mahadevan, *CMPSCI 791BB: Advanced ML - Spectral Graph Theory.* University of Massachusetts, Amherst, 2006. https://people.cs.umass.edu/~mahadeva/cs791bb/lectures-s2006/lec4.pdf

[14] Guangliang Chen, *Math 253: Rayleigh Quotient Lecture Notes.* San Jose State University, 2020. https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec4RayleighQuotient.pdf

[15] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela, *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.* arXiv preprint arXiv:2005.11401, 2020. Accepted at NeurIPS 2020.

[16] J. Jost. *Two-Dimensional Geometric Variational Problems*. John Wiley & Sons, 1991. Chapter on Dirichlet energy and area for harmonic maps and minimal surfaces in conformal gauge.

[17] C. De Lellis. Geometric variational problems. Lecture notes, Universität Freiburg, 2018. URL: http://home.mathematik.uni-freiburg.de/analysis/GeomAnalysisWS1819/GV18.pdf.

[18] B. Loustau. Minimal surfaces and quasi-Fuchsian structures. Lecture notes, 2015. URL: https://brice.loustau.eu/ressources/Asheville.pdf.

[19] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. doi:10.1016/j.acha.2006.04.006. URL: https://www.math.pku.edu.cn/teachers/yaoy/Fall2011/Lafon06.pdf.

[20] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi:10.1007/s11222-007-9033-z. URL: http://www.cs.cornell.edu/courses/cs6241/2019sp/readings/vonLuxburg-2007-spectral.pdf.

[21] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010. doi:10.1090/mbk/069.

[22] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009. doi:10.1090/S0273-0979-09-01249-X. URL: https://www.math.kth.se/math/GRU/2013.2014/SF2704/Papers/Topologyanddata.pdf.

[23] D. Tong. *String Theory*. Lecture notes, University of Cambridge, 2009. URL: https://www.damtp.cam.ac.uk/user/tong/string/string1.pdf.

[24] UCSB Physics (PHYS 230A). *Week 3: Polyakov action* (lecture notes). URL: https://web.physics.ucsb.edu/~phys230A/lectures/wk3_stactionlc.pdf.

[25] C. Horsman, A. G. Fowler, S. Devitt, and R. Van Meter. Surface code quantum computing by lattice surgery. *arXiv preprint* arXiv:1111.4022, 2012. URL: https://arxiv.org/abs/1111.4022.

[26] D. Litinski. A game of surface codes: large-scale quantum computing with lattice surgery. *Quantum*, 3:128, 2019. URL: https://quantum-journal.org/papers/q-2019-03-05-128/.

[27] I. Kerenidis, J. Landman, A. Prakash, and A. Szilágyi. Quantum spectral clustering. *arXiv preprint* arXiv:2007.00280, 2020. URL: https://arxiv.org/abs/2007.00280.

[28] M. J. Baum, C. M. Dawson, J. Hines, and R. B. Thompson. Using spectral graph theory to map qubits onto connectivity-limited devices. *arXiv preprint* arXiv:1910.11489, 2019. URL: https://arxiv.org/abs/1910.11489.

[29] A. Di Gaspare, S. G. Zamharir, C. Knox, A. Yagmur, S. Sasaki, M. Salih, L. Li, E. H. Linfield, J. Freeman, and M. S. Vitiello, "Second and third harmonic generation in topological insulator-based van der Waals metamaterials," *Light: Science & Applications*, vol. 14, no. 1, p. 337, 2025. DOI: 10.1038/s41377-025-01847-5.

[30] L. Moriondo (@Mec-iS) `surfface-rs`: Approximating Meng-Type Surface Minimisation in Rust. *GitHub Repository* (2026). https://github.com/tuned-org-uk/surfface-rs

[31] `arrowspace-rs`: Extensive test campaign on CVE and Dorothea dataset for searching and classification. *Blog Post* (2026). https://www.tuned.org.uk/posts/016_arrowspace_performance_results