The Column-Matroid Backbone of a Composition-Coherence Fee

A Structural Identity and Its Corollaries

John Komkov
Independent Researcher
https://github.com/jkomkov/bulla

April 2026

Abstract

The Bulla protocol assigns a non-negative integer coherence fee $\mathrm{fee}(G)$ to each composition G of tool specifications, measuring semantic dimensions whose consistency is structurally unverifiable from observable schemas alone. We report that $\mathrm{fee}(G)$ equals the corank of the observable column set in the linear matroid on the full coboundary's columns—a one-line proof from a column-compression identity discovered by a post-hoc audit of a pre-registered spectral experiment. This structural identity yields a closed-form pairwise formula:

$$\mathrm{fee}(A, B) = U(A) + U(B) + |\mathrm{shared\_dims}(A, B)|$$

where $U(X)$ is a per-server invariant (Lean-verified) and $|\mathrm{shared\_dims}|$ counts bilateral semantic dimensions (proved under DFD; verified on 703 real-schema compositions). The backbone theorem also proves the additive decomposition as matroid direct-sum additivity and closes the rank$\leftrightarrow H^1$ formal target via two lines of rank-nullity (Lean-verified, zero `sorry`). An empirical density law localizes the connecting homomorphism of the associated long exact sequence strictly in the interior of its attainable range on all 240 non-trivial compositions tested. The discovery is an instance of *Type D productive falsification*: the pre-registered prediction held, but for a stronger and simpler reason than anticipated. A companion signed-incidence note [6] supplies the field-independence and endpoint-coupling boundaries used throughout: the fee agrees across coefficient fields as a numerical invariant, and pairwise same-dimension blocks are bounded by rank 2 without implying any broader typed-repair characterization.

1 Introduction

The Bulla protocol assigns a non-negative integer coherence fee $\mathrm{fee}(G)$ to each composition G of two or more tool specifications. The fee measures the codimension of an observable coboundary inside the full internal-state coboundary of a cellular sheaf constructed from the composition graph [1]. The protocol note [2] wrote $\mathrm{fee}(G) = \dim H^1(G; \mathcal{F}/\mathcal{O})$, a cohomological framing suggestive enough to motivate the construction but not proven in the strict sense.

This paper reports a structural identity that closes the gap and produces a closed-form pairwise formula for the fee.

The experiment. A pre-registered Cauchy eigenvalue interlacing test on the 703-composition real-schema corpus (§1.1) passed all 703 compositions with zero violations, under the hypothesis that Bulla's coboundary pair instantiates the spectral theory of cellular sheaves [3].

The audit. A post-sweep audit discovered that, on every composition, $\delta_{\mathrm{obs}} = \delta_{\mathrm{full}}[\,:\,, \mathcal{O}\,]$ exactly over $\mathbb{Q}$: the observable coboundary is the principal column submatrix of the full coboundary.

The interlacing holds as a classical submatrix theorem, independent of [3]. We call this *Type D productive falsification*: the pre-registered prediction held, but for a stronger and simpler structural reason than anticipated. The backbone theorem emerged not from searching for a matroid identity but from a standing audit protocol—inspect the construction that produced the data, not just the measurement—applied to a test whose prediction *passed*.[1]

The consequences. The column-compression identity makes fee(G) the corank of a representable matroid. Four corollaries carry the paper's argument:

1. The additive decomposition theorem [4] is matroid direct-sum additivity (Corollary 2.8).
2. The rank$\leftrightarrow$ H^1 identity is proven and Lean-verified (Corollary 2.9).
3. The unilateral fee is a stable per-server invariant, Lean-verified (Corollary 2.13).
4. The bilateral fee counts shared semantic dimensions exactly, yielding the closed-form pairwise formula (Corollary 2.14, Theorem 2.15).

Claim discipline. What is established here is the backbone identity and the corollaries that genuinely follow from it. What failed productively was the older attempt to motivate these consequences through a heavier sheaf-spectral route than the code audit warranted. What remains open is the correct dimension-aware coupled-disclosure objective; the signed-incidence note [6] rules out the separable partition story, but this paper does not claim a replacement characterization.
Further corollaries—T-concentration as uniform-matroid proximity, the repair filtration as a matroid flag, a flats-interaction reformulation of the Unilateral-Bilateral Decomposition, and a row-corank reframing of the submodularity question—appear in [5].

Organization. §2 states the backbone theorem and its four corollaries. §3 treats the matroid direct-sum proof of Theorem 1. §4 reports an empirical density law. §5 updates the three original Bulla open problems.

1.1 The 703-Composition Real-Schema Corpus

All empirical claims are made on a corpus of MCP (Model Context Protocol) tool-composition instances derived from 57 server manifests scraped from the public `oslook/mcp-servers-schemas` repository on 2026-04-07. The corpus is filtered to servers with `MIN_SCHEMA_FIELDS` ≥ 3 (38 of 57), giving a pairwise closure of $\binom{38}{2} = 703$ compositions.

The Column-Compression Identity (Theorem 2.4) and the Backbone Theorem (Theorem 2.6) are structural and hold for any Bulla composition. The density-law measurement (§4) is corpus-specific.

2 The Column-Matroid Backbone

2.1 Setup

Let $G = (T, E, D)$ be a Bulla composition of tools T, directed edges E between tools, and semantic dimensions D attached to edges. For each tool $t \in T$, write $t.\text{int}$ for the internal-state fields and $t.\text{obs} \subseteq t.\text{int}$ for the observable subset.

Definition 2.1 (Coboundary construction)**.** The Bulla coboundary construction produces two $\mathbb{Q}$-linear maps sharing a common codomain:

$$\delta_{\text{full}}(G) : C^0_{\text{full}}(G) \longrightarrow C^1(G), \tag{1}$$

$$\delta_{\text{obs}}(G) : C^0_{\text{obs}}(G) \longrightarrow C^1(G), \tag{2}$$

[1] Type D is the most dangerous productive-falsification species to miss: the prediction-verification pattern looks like ordinary confirmatory success. The taxonomy and the commit-before-measure discipline that produced the backbone are described in [5].

where $C^0_{\text{full}} = \bigoplus_t \mathbb{Q}[t.\text{int}]$, $C^0_{\text{obs}} = \bigoplus_t \mathbb{Q}[t.\text{obs}]$, and $C^1 = \bigoplus_{(e,d)} \mathbb{Q}$ is the edge-dimension cochain space. The matrix entry at row (e, d) and column (t, f) is -1 if $t = e.\text{src}$ and $f = d.\text{from}$; $+1$ if $t = e.\text{dst}$ and $f = d.\text{to}$; and 0 otherwise, restricted to $f \in t.\text{int}$ (resp. $t.\text{obs}$).

Definition 2.2 (Coherence fee)**.** $\text{fee}(G) := \text{rank}_{\mathbb{Q}}(\delta_{\text{full}}(G)) - \text{rank}_{\mathbb{Q}}(\delta_{\text{obs}}(G)) \geq 0$.

Remark 2.3 (Shared codomain)*.* Both coboundaries map into the same $C^1(G)$: the edge-dimension basis is constructed identically regardless of vertex-stalk scope. This fact is load-bearing for Corollary 2.9.

2.2 The Column-Compression Identity

Theorem 2.4 (Column-Compression Identity)**.** *For every Bulla composition G, there is a column index set $\mathcal{O}(G) \subseteq \{1, \ldots, n_{\text{full}}\}$ such that $\delta_{\text{obs}}(G) = \delta_{\text{full}}(G)[\,:\,, \mathcal{O}(G)\,]$ as an equality of $\mathbb{Q}$-valued matrices.*

Proof. Since $t.\text{obs} \subseteq t.\text{int}$ per tool, every observable basis element appears in the full basis. Let $\mathcal{O}(G)$ be the full-basis indices of observable pairs. Every ± 1 entry in δ_{obs} appears at the same row and corresponding column in δ_{full}; columns not in $\mathcal{O}$ (hidden fields) are absent from δ_{obs}. □

Empirical verification. 703/703 compositions, zero entry mismatches on the 240 non-trivial cases.

2.3 The Column Matroid and the Backbone Theorem

Definition 2.5 (Column matroid)**.** $M(G)$ is the linear matroid on ground set $E(G) := \{1, \ldots, n_{\text{full}}\}$ with rank function $\text{rank}_{M(G)}(S) := \dim_{\mathbb{Q}} \text{span}_{\mathbb{Q}}\{\delta_{\text{full}}(G)[\,:\,, j\,] : j \in S\}$.

Theorem 2.6 (Column-Matroid Backbone)**.** *For every Bulla composition G,*

$$\text{fee}(G) \;=\; \text{corank}_{M(G)}(\mathcal{O}(G)) \;:=\; \text{rank}_{M(G)}(E(G)) - \text{rank}_{M(G)}(\mathcal{O}(G)).$$

Proof. $\text{rank}_{M(G)}(E(G)) = \text{rank}_{\mathbb{Q}}(\delta_{\text{full}})$. $\text{rank}_{M(G)}(\mathcal{O}) = \text{rank}_{\mathbb{Q}}(\delta_{\text{full}}[\,:\,, \mathcal{O}\,]) = \text{rank}_{\mathbb{Q}}(\delta_{\text{obs}})$ by Theorem 2.4. Subtract. □

2.4 Corollary 1: Additive Decomposition as Matroid Direct Sum

Definition 2.7 (Dimension-Field Disjointness)**.** A composition G satisfies *DFD* if no (t, f) pair is referenced by two distinct dimension names in D.

Corollary 2.8. *Under DFD, $M(G) = \bigoplus_d M_d(G)$ as a matroid direct sum, and* $\text{fee}(G) = \sum_d \text{fee}_d(G)$*.*

Proof. Under DFD, each column of δ_{full} has nonzero entries only in rows of the single dimension referencing its (t, f) pair, so δ_{full} is block-diagonal with one block per dimension. Let E_d be dimension d's columns; the E_d partition $E(G)$ and $M_d := M(G)|_{E_d}$. Direct-sum rank additivity gives $\text{rank}_{M(G)}(S) = \sum_d \text{rank}_{M_d}(S \cap E_d)$. Applying Theorem 2.6 per dimension:

$$\text{fee}(G) = \sum_d \big[\text{rank}_{M_d}(E_d) - \text{rank}_{M_d}(\mathcal{O} \cap E_d)\big] = \sum_d \text{fee}_d(G). \qquad \square$$

2.5 Corollary 6: Cohomological Gap (Problem 3 Closure)

Corollary 2.9. *For every Bulla composition G,*

$$\mathrm{fee}(G) \;=\; \dim_{\mathbb{Q}} H^1(\delta_{\mathrm{obs}}) - \dim_{\mathbb{Q}} H^1(\delta_{\mathrm{full}})$$

where $H^1(\delta) := \mathrm{coker}(\delta) = C^1/\,\mathrm{im}(\delta)$.

Proof (rank-nullity). The chain complex $C^0 \to C^1 \to 0$ has no C^2, so $H^1(\delta) = \mathrm{coker}(\delta)$. By rank-nullity, $\dim \mathrm{coker}(\delta) = \dim C^1 - \mathrm{rank}(\delta)$. Both coboundaries share C^1 (Remark 2.3), so:

$$\dim H^1(\delta_{\mathrm{obs}}) - \dim H^1(\delta_{\mathrm{full}}) = \big[\dim C^1 - \mathrm{rank}(\delta_{\mathrm{obs}})\big] - \big[\dim C^1 - \mathrm{rank}(\delta_{\mathrm{full}})\big] = \mathrm{fee}(G). \quad \square$$

Proof (long exact sequence). The short exact sequence $0 \to \mathcal{O} \to \mathcal{F} \to \mathcal{F}/\mathcal{O} \to 0$ induces a LES. Since $C^2 = 0$ and the relative complex has trivial degree-1 term ($C^1_{\mathrm{full}} = C^1_{\mathrm{obs}}$), we get $H^1(\mathcal{F}/\mathcal{O}) = 0$ and $\mathrm{fee}(G) = \dim \mathrm{im}(\partial)$ where $\partial : H^0(\mathcal{F}/\mathcal{O}) \to H^1(\mathcal{O})$ is the connecting homomorphism. $\square$

Lean verification. Formally verified in Lean 4 / Mathlib v4.28.0 (Aristotle UUID `044a00b1`). Three theorems: `fee_h1_additive`, `fee_eq_h1_difference`, `range_obs_le_range_full`. No `sorry`.

Remark 2.10 (Upper bound). Since ∂ is linear, $\mathrm{fee}(G) \le n_{\mathrm{full}} - n_{\mathrm{obs}} = |\mathcal{O}^c|$. The gap measures *column redundancy*—hidden-column directions already implied by observed relations. Section 4 measures this.

Remark 2.11 (Notation retirement). The informal fee $= \dim H^1(G; \mathcal{F}/\mathcal{O})$ is retired: under the literal relative-complex reading, $H^1(\mathcal{F}/\mathcal{O}) = 0$. The honest content is $\dim H^1(\delta_{\mathrm{obs}}) - \dim H^1(\delta_{\mathrm{full}})$.

2.6 Corollaries 8a and 8b: the Closed-Form Pairwise Fee

For a pairwise composition $G = A + B$, define the *unilateral fee* $U(X) := \mathrm{fee}(X, \varnothing)$ (composition of X with a synthetic null probe) and the *bilateral correction* $B(A, B) := \mathrm{fee}(A, B) - U(A) - U(B)$.

Definition 2.12 (Convention-Hidden Partition)**.** A composition G satisfies the *convention-hidden partition* (CHP) if every field f that appears as d.from or d.to for some dimension d satisfies $f \notin t$.obs on the tool t that carries it. Equivalently: inferred convention fields are never observable.

CHP holds by construction in the current Bulla classifier, which removes inferred convention fields from the observable schema. It is stated as an explicit hypothesis so that the theorem's scope is clear: a future system that observes some convention fields would require a different analysis.

Corollary 2.13 (Unilateral stability)**.** $\mathrm{fee}_A(A + B) = U(A)$ *and* $\mathrm{fee}_B(A + B) = U(B)$, *where* fee_X *denotes the fee contribution of X-only rows. The composition partner does not affect the unilateral row-class fee.*

Verification. 703/703. Lean-verified (UUID `01333bb4`, 2 theorems, 0 `sorry`).

Corollary 2.14 (Bilateral fee counts shared dimensions)**.** *Under DFD and CHP,* $B(A, B) = |\mathrm{shared\_dims}(A, B)|$, *where* shared_dims *counts semantic dimensions with bilateral rows.*

Proof. By Corollary 2.8, $B(A,B) = \sum_d B_d$ where $B_d := \mathrm{fee}_d - U_d(A) - U_d(B)$. If d is not shared, all dimension-d rows are unilateral, $\mathrm{fee}_d = U_d(A) + U_d(B)$, and $B_d = 0$.

If d is shared, let p tools in A and q tools in B carry dimension d. By CHP, both d.from and d.to are hidden columns. Under DFD the $p+q$ hidden columns appear only in dimension-d rows. The unilateral A-rows (differences of A-columns) have rank $p-1$; the unilateral B-rows rank $q-1$; one bilateral row $-e_{(a_i,f_A)} + e_{(b_j,f_B)}$ is independent of both subspaces (it has nonzero projection onto both column groups). Equivalently, the dimension-d rows form the signed incidence matrix of a connected bipartite graph on $p+q$ vertices, which has rank $p+q-1$ over any field. So $\mathrm{rank}_d^{\mathrm{full}} = p+q-1$. In δ_{obs} all $p+q$ columns are absent, giving $\mathrm{rank}_d^{\mathrm{obs}} = 0$. Therefore

$$B_d = (p+q-1) - (p-1) - (q-1) = 1.$$

□

Empirical verification. 703/703 (all satisfy both DFD and CHP), Pearson $r = 1.000$ on the 25 nonzero-B cases. Pre-registered for cross-registry replication.

Corollaries 2.13 and 2.14 together yield:

Theorem 2.15 (Closed-form pairwise formula)**.** *Under DFD and CHP, for every pairwise composition* (A, B)*,*

$$\boxed{\mathrm{fee}(A,B) \;=\; U(A) + U(B) + |\mathrm{shared\_dims}(A,B)|}$$

where $U(X)$ *is a per-server invariant and* $|\mathrm{shared\_dims}|$ *counts the semantic dimensions with bilateral rows.*

Proof. Immediate from Corollary 2.13 ($\mathrm{fee}_A = U(A)$, $\mathrm{fee}_B = U(B)$) and Corollary 2.14 ($B(A,B) = |\mathrm{shared\_dims}|$). □

Theorem 2.15 decomposes the fee into local invariants (unilateral, depending only on each server's schema) and a pairwise interaction term (bilateral, counting shared dimensions). The formula requires no matrix computation—only schema inspection.

Example 2.16. `github + notion`: U(`github`) $= 10$, U(`notion`) $= 1$, two shared dimensions (`date_format_match`, `sort_direction_match`), so fee $= 10 + 1 + 2 = 13$. Verified.

3 Theorem 1 under DFD

The additive decomposition theorem of [4] states that, under DFD, the coherence fee splits by named semantic dimension: $\mathrm{fee}(G) = \sum_d \mathrm{fee}_d(G)$. Corollary 2.8 gives this as a one-line consequence of matroid direct-sum additivity.

What Corollary 2.8 adds beyond the original linear-algebraic proof sketch is a *name* for the structure. The statement "$M(G)$ decomposes as a matroid direct sum under DFD" carries through to other matroid operations (intersection, union, truncation, minors) without reproving each. It also sharpens the open question of whether DFD is necessary: the question becomes "characterize the matroid-direct-sum compositions that violate DFD," which has a precise formulation in terms of column support overlap in non-block-diagonal representations.

Empirical verification. The per-dimension fee identity $\mathrm{fee}(G) = \sum_d \mathrm{fee}_d(G)$ was verified on 1596 compositions (the full pairwise closure of the 57-manifest registry, not restricted to the real-schema subset). Zero violations. On the real-schema subset, 45 of 45 compositions with nonzero per-dimension fees satisfy the identity exactly.

DFD as hypothesis. DFD holds on all 703 real-schema compositions, so the hypothesis is not binding on the current corpus. Whether it is necessary in general—whether matroid direct sum can hold without DFD—remains open.

4 Observation 1: the Interior-Regime Density Law

Corollary 2.9 gives $\text{fee}(G) = \dim \text{im}(\partial)$ where $\partial : H^0(\mathcal{F}/\mathcal{O}) \to H^1(\mathcal{O})$. The normalized ratio $\rho(G) := \text{fee}(G)/(n_{\text{full}} - n_{\text{obs}})$ measures how surjective ∂ is.

Observation 4.1 (Interior regime)**.** On every non-trivial composition in the 703 corpus ($n_{\text{full}} - n_{\text{obs}} > 0$), the ratio $\rho(G)$ lies strictly in $(0, 1)$. Neither extremum is realized.

Table 1: Distribution of $\rho(G)$ on 240 non-trivial compositions.

mean	median	stdev	min	max
0.603	0.667	0.175	0.250	0.933

68.3% of compositions land in $\rho \in [0.50, 0.75)$—tight enough to be called a *density law*: the ratio of rank-contributing hidden columns to total hidden columns is concentrated around a single value across structurally distinct compositions.

In applied terms: on real MCP compositions, the cohomological obstruction is always present, never saturated, and its magnitude is regularized to roughly 60% of its attainable maximum.

Pre-registered replication. A second MCP registry snapshot will be procured and the same ρ measurement run. Four locked branches: *O1-pass* (mean $\in [0.50, 0.70]$, modal mass $> 50\%$, strict interior); *O1-narrow* (interior preserved, concentration shifts); *O1-broken* (some $\rho \in \{0, 1\}$); *O1-trivial* (no non-trivial compositions).

5 Three Open Problems as Backbone Reductions

The Bulla research program posed three structural open problems. The backbone theorem (Theorem 2.6) is the protagonist of all three updates: each problem is either closed or reduced to a sharper question by the matroid framing.

5.1 Problem 3 (Rank$\leftrightarrow H^1$): Closed

Corollary 2.9 proves $\text{fee}(G) = \dim H^1(\delta_{\text{obs}}) - \dim H^1(\delta_{\text{full}})$ as an identity of $\mathbb{N}$-valued invariants. The Lean 4 formalization (Aristotle UUID `044a00b1`) closes the problem in three theorems and zero `sorry`, reducing what was projected as a multi-sprint cellular-cohomology formalization to an application of `Submodule.finrank_quotient_add_finrank` (rank-nullity for cokernels).

The key structural insight: the Problem 3 target was hard only under the assumption that the proof required sheaf-theoretic machinery. The column-compression identity (Theorem 2.4) shows the two coboundaries are related by a column inclusion, not a generic morphism. Under column inclusion, rank-nullity suffices.

5.2 Problem 1 (Termination): Structural Half Closed

The repair loop produces an increasing sequence $\mathcal{O}_0 \subseteq \mathcal{O}_1 \subseteq \cdots \subseteq E(G)$ of observable column sets. Under the backbone, this is a flag in $M(G)$ with rank jumps in $\{0, 1\}$ at each step (a matroid flag property).

The structural half of Problem 1—that the repair loop terminates in at most $\text{fee}(G)$ rank-increasing steps—was Lean-verified (UUID `ff78c99c`, 3 theorems). The backbone adds a run-time assertion: after each confirmed probe, check whether the new column lies in $\text{cl}_{M(G)}(\mathcal{O}_i)$. If it does, the probe was rank-redundant and a `ProbeSoundnessViolation` is raised.

The practical half—bounding the number of oracle calls needed to produce a rank-increasing step—remains open and depends on oracle-quality assumptions outside the matroid structure.

5.3 Problem 2 (Submodularity): Sharpened

The Bulla boundary fee $\mathrm{bf}(P)$ was tested for submodularity on the partition lattice and found 4061 violations across the corpus. Under the backbone, $\mathrm{bf}(P)$ is a difference of row-matroid corank functions:

$$\mathrm{bf}(P) = \mathrm{corank}_{M_{\mathrm{row}}^{\mathrm{full}}}(\mathrm{intra}(P)) - \mathrm{corank}_{M_{\mathrm{row}}^{\mathrm{obs}}}(\mathrm{intra}(P)).$$

Each corank is individually supermodular, but a difference of supermodular functions is neither sub- nor supermodular in general. The 4061 violations are structurally explained: they measure the direction in which the difference fails for this specific pair of row matroids.

The sharpened research question is: under what conditions on the weak map $M_{\mathrm{row}}^{\mathrm{obs}} \to M_{\mathrm{row}}^{\mathrm{full}}$ is the corank difference sub- or supermodular? This is a bona fide matroid-theoretic question with real structure, connected to the weak-maps literature (Lovász, Kung).

A Lean Verification Summary

Twelve theorems verified in Lean 4 against Mathlib, zero `sorry`:

File	UUID	Thms	Content
BullaCorollary6	`044a00b1`	3	rank↔ H^1 (Problem 3)
TerminationStructural	`ff78c99c`	3	Repair loop bound (Problem 1)
UBD1	`edb6c604`	4	Range incl. + DFD + $B \geq 0$
Corollary8a	`01333bb4`	2	Unilateral stability
Total		**12**	**0 sorry**

B Code and Proof Artifacts

The column-submatrix audit, the $\rho(G)$ measurement script, and all four Lean files are available at `https://github.com/jkomkov/bulla`. The SHA-256 anchor chain of the Column-Matroid Backbone memo (v1–v5) and all pre-registration anchors are recorded in the repository's technical claim ledger.

References

[1] J. Komkov, "The Coherence Fee: Measuring Hidden Convention Risk in Hierarchical Tool Compositions," 2026.

[2] J. Komkov, "Bulla Protocol Note," 2026.

[3] J. Hansen and R. Ghrist, "Toward a spectral theory of cellular sheaves," *J. Appl. and Comput. Topology*, vol. 3, no. 4, pp. 315–358, 2019.

[4] J. Komkov, "Additive decomposition of the coherence fee under Dimension-Field Disjointness," 2026.

[5] J. Komkov, "The Column-Matroid Backbone of the Bulla Coherence Fee," internal research memo v5, 2026.

[6] J. Komkov, "Signed-Incidence Structure in Compositional Verification: A Structural Note," `papers/signed-incidence/paper/signed-incidence.pdf`, 2026.