Epistemic Phase Crystallization Map:

Thermodynamic Phase States in a Seven-Bridge Verified Knowledge Architecture

Cross-Phase Boundary Agent Discovery and Full System Crystallization

Hasachai Boonnuang
EyeMotionAI Co., Ltd. | Genesis Academic™
hasachai@eyemotionai.com · ORCID: 0009-0005-8603-9188

MH-EPCM-V1.1 · Paper 11 · April 9, 2026 · Genesis Academic™ Diamond+
Version: 1.1 (Corrected — Supersedes v1.0)

Abstract

This paper introduces the Epistemic Phase Crystallization Map (EPCM) — the first application of thermodynamic phase theory to a multi-domain verified knowledge system. Each of seven sovereign knowledge bridges is assigned a phase state (SOLID, LIQUID, or GAS) based on epistemic temperature T_d (TEIF relative to crystallization threshold) and epistemic pressure P_d (compound-registry coverage density). Version 1.0 contained two implementation bugs that produced incorrect phase assignments; v1.1 corrects both. With corrections and subsequent crystallization events (MARINE Round 5, FERMENT Round 6), the system achieves full crystallization: 7/7 bridges SOLID, 0 LIQUID, 0 GAS — the first complete epistemic crystallization in the system's history.

Three cross-phase boundary agents were discovered in v1.1 (Aspergillus niger, Bifidobacterium, Hydrothermal vent) when FERMENT was LIQUID. Their subsequent dissolution into SOLID-SOLID pairs upon FERMENT crystallization constitutes the strongest evidence of the crystallization event. The CBET_norm ranking correction identifies TTM x CLINICAL (Andrographis paniculata, 0.000258) as the true strongest normalized bond — not MARINE x NATHISTORY, which was a computational artifact. FERMENT crystallization required the Latin Anchor Strategy: using Latin binomials as probes rather than process phrases, reducing C_total from 2,162,260 to 67,521 via architectural correction.

Keywords: *epistemic phase theory, knowledge crystallization, CBET normalization, cross-phase boundary agents, TEIF, corpus architecture, Latin Anchor Strategy*

1. Theoretical Framework

1.1 Phase Variables

Epistemic Temperature T_d = TEIF_eff / TEIF_target (target = 10.0x, from CPP mandate). T_d < 1.0 = thermally crystallized; T_d >= 1.0 = thermally agitated. **Epistemic Pressure** P_d = n / (C_total / 1000), where n = registered compounds and C_total = total corr_hits for bridge d. **Shannon Diversity** H_d = -Sum(p_i log2 p_i) over TEIF-weighted distribution.

Table 1. *Phase transition rules.*

Phase	Condition	Interpretation
SOLID	T_d < 1.0 AND P_d > 1.5	Dense, well-mapped; crystalline state
LIQUID	T_d < 1.0 AND P_d <= 1.5	Thermally cool but compound-sparse
GAS	T_d >= 1.0 OR pre-TEIF > 50x	High-entropy; unresolved inflation

1.2 Threshold Justification

TEIF_target = 10.0x is the CPP mandate ceiling from Paper 10. **P_d* = 1.5** is the empirical inflection point below which coverage becomes negligibly sparse (safety margin of 50% above minimum meaningful density). Sensitivity: at P_d* = 1.0, NATHISTORY also joins SOLID (6 SOLID); at P_d* = 2.0, CLINICAL drops to LIQUID (4 SOLID). The qualitative finding — FERMENT as most epistemically sparse — holds at all three thresholds.

1.3 Corrections from v1.0

Bug A — NATHISTORY score field: v1.0 used bio_hits with a botany-segment filter (wrong domain), returning bridge total = 37 and inflating CBET_norm for MARINE x NATHISTORY to 0.001878 (artifact). Corrected to corr_hits (C = 6,419). **Bug B — MICRO compound count:** v1.0 truncated registry at 27 compounds; corrected to 47. P_d: 0.982 -> 1.709 (SOLID). Both bugs caused incorrect LIQUID classifications and prevented discovery of cross-phase agents.

1.4 CBET — Raw and Normalized

CBET_raw(i,j,c) = M_s(c,i) x M_s(c,j) x w_ij. CBET_norm(i,j,c) = [M_s(c,i)/T_i] x [M_s(c,j)/T_j] x w_ij. Normalization required because bio_hits bridges (TTM, CLINICAL, NATHISTORY) and corr_hits bridges operate on different scales. Cluster solidarity weights: same-cluster 1.30; C1-C2/C2-C3 0.65; C1-C3 0.25.

2. Prior Art and Positioning

No prior work has applied thermodynamic phase theory to knowledge corpus characterization.

Table 2. *Prior art positioning. EPCM adds no new data collection.*

Approach	Measures	Misses	EPCM adds
Knowledge graphs	Node connectivity	Evidence quality, completeness	Phase state = quality + completeness
Semantic similarity	Cosine distance	Domain stability	Cross-bridge coupling (CBET)
Citation networks	Bibliometric density	Coverage density	Epistemic temperature T_d
TEIF alone (Paper 10)	Evidence inflation	Coverage dimension	P_d -> full (T_d, P_d) space

3. Seven-Bridge Architecture

Table 3. *Seven-bridge architecture. T_k = normalization denominator for CBET_norm.*

Bridge	Label	Cluster	Score field	Bridge total T_k
A-1_TTM	Traditional Thai Medicine	C1	bio_hits	2,668
A-2_CLINICAL	Clinical Phytomedicine	C1	bio_hits	34
A-3_FERMENT	Fermentation Sciences	C2	corr_hits	67,521
A-4_MARINE	Marine Biochemistry	C3	corr_hits	213,448
A-5_MICRO	Microbiology	C2	corr_hits	27,509
A-6_FOREST	Forest and Botany	C3	corr_hits	6,830
A-7_NATHISTORY	Natural History	C1	corr_hits	6,419

4. Phase Map Results — v1.1 Corrected + Rounds 5 & 6

Script: compute_phase_states.py v1.1 | Output: PHASE_MAP_LATEST.json | 2026-04-09

Table 4. *Phase map. All 7 bridges SOLID — April 9, 2026. † C_total for bio_hits bridges uses corr_hits field in phase computation. NATHISTORY: n=10 active in P_d formula; 22 = total non-void entries.*

Bridge	n	C_total	H_d	T_d	P_d	Phase	Delta from v1.0
A-1_TTM	39	6,221†	3.581	0.233	6.2690	SOLID ●	unchanged

A-2_CLINICAL	5	1,944†	1.334	0.100	2.5720	SOLID ●	unchanged
A-3_FERMENT	116	67,521	1.854	0.112	1.7180	SOLID ●	LIQUID->SOLID Round 6
A-4_MARINE	329	213,448	1.946	0.192	1.5414	SOLID ●	LIQUID->SOLID Round 5
A-5_MICRO	47	27,509	3.526	0.155	1.7090	SOLID ●	LIQUID->SOLID Bug B
A-6_FOREST	34	6,830	3.005	0.114	4.9780	SOLID ●	unchanged
A-7_NATHISTORY	10 (22†)	6,419	2.872	0.233	1.5580	SOLID ●	LIQUID->SOLID Bug A

4.1 FERMENT Crystallization — Round 6 (Latin Anchor Strategy)

Root cause of LIQUID state: probe inflation. Three entries accounted for 94.2% of C_total:

Table 5. *Probe inflation in FERMENT v1.0 architecture.*

Entry	Probe type	corr_hits	Share	Problem
"citric acid"	phrase	1,046,923	48.4%	Generic chemical
riboflavin	UNQUOTED (bug)	723,848	33.5%	No phrase quotes
"lactic acid"	phrase	266,121	12.3%	Generic acid
Remaining 31	various	125,368	5.8%	Normal

Key insight: P_d = n/(C/1000) = 1000/avg_hits_per_entry. *n cancels — phase depends only on average hits per entry, not compound count.* SOLID requires avg < 667 hits/entry. Adding n=3,243 compounds with high avg would remain LIQUID. **Fix (Latin Anchor Strategy):** Architecture dict-of-dicts + best_fts_v2() -> flat tuple list + fts_count(). Probe rule: Latin binomials only — "xxx fermentation" = VOID (FTS5 phrase blindness). "Lactobacillus kimchii" = hits; "kimchi fermentation" = 0. Result: n=116, C=67,521, avg=578 hits/entry, P_d=1.7180 SOLID.

4.2 MARINE Crystallization — Round 5

n: 178->329 (+151 compounds), 9 new categories: DIATOM_GENERA_II (31), DINOFLAGELLATE_HAPTOPHYTE_II (10), CYANOBACTERIA_GENERA (5), MARINE_BACTERIA_GENERA (19), MARINE_FISH_GENERA (25), MARINE_MICROALGAE_EXPANDED (12), RED_ALGAE_GENERA_II (6), SEAFOOD_BYPRODUCTS (29), MARINE_ECOLOGY_TERMS_II (14). C: 203,469->213,448. P_d: 0.8748->1.5414 SOLID.

5. TTM Crystallization Event (unchanged from v1.0)

Table 6. *TTM GAS->SOLID transition via CPP correction.*

Stage	TEIF	T_d	Phase
Pre-correction (naive queries)	152.14x	15.21	GAS
Post-correction (phrase queries)	2.33x	0.233	SOLID

CR_TTM = (152.14 - 2.33) / 152.14 = **98.47%**. Before CPP: "turmeric" -> 235,034 hits; "noni" -> 671,035 hits (music results). After CPP: "Curcuma longa" -> 1,065; "Morinda citrifolia" -> 35. Only bridge to undergo documented GAS->SOLID transition. All others used correct phrasal queries from inception.

6. Compound Registry and CBET Tensor (v1.1)

Registry: 223 raw -> 218 canonical entries. 5 multi-domain compounds. 0 PubChem matches (all are organisms/habitats, not pure chemicals).

Table 7. *CBET tensor (normalized). CBET_norm corrects for score-field heterogeneity. MARINE x NATHISTORY was spuriously #1 in v1.0 (NATHISTORY total=37 — artifact).*

Bridge pair	w_ij	Compounds	CBET_raw	CBET_norm	Rank (norm)

A-1_TTM x A-2_CLINICAL	1.30	1	23	0.000258	1st (stable)
A-3_FERMENT x A-5_MICRO	1.30	2	6,700,170	0.000113	2nd
A-4_MARINE x A-7_NATHISTORY	0.25	1	12,447	0.0000116	3rd
A-1_TTM x A-6_FOREST	0.25	1	12	0.0000007	4th

CBET_norm(TTM x CLINICAL, Andrographis) = (bio_hits_TTM/2,668) x (3.0/34) x 1.30 = 0.000258. Stable across all versions — neither TTM nor CLINICAL had score-field bugs. This is the true strongest normalized epistemic bond in the system.

7. Phase Boundary Agents — Discovery and Dissolution

Table 8. *Boundary agent registry (current). * = was cross-phase in v1.1 when FERMENT=LIQUID; dissolved to SOLID-SOLID upon FERMENT crystallization (Round 6).*

Compound	Bridge A	Phase A	Bridge B	Phase B	Type (current)	CBET_norm
Andrographis paniculata	A-1_TTM	SOLID	A-2_CLINICAL	SOLID	SOLID-SOLID	0.000258
Aspergillus niger	A-3_FERMENT	SOLID	A-5_MICRO	SOLID	SOLID-SOLID*	0.000087
Bifidobacterium spp.	A-3_FERMENT	SOLID	A-5_MICRO	SOLID	SOLID-SOLID*	0.000026
Hydrothermal vent	A-4_MARINE	SOLID	A-7_NATHISTORY	SOLID	SOLID-SOLID*	0.0000116
Terminalia chebula	A-1_TTM	SOLID	A-6_FOREST	SOLID	SOLID-SOLID	0.0000007

Cross-phase dissolution as crystallization evidence. In Paper 11 v1.1, Aspergillus niger and Bifidobacterium were LIQUID(FERMENT)-SOLID(MICRO) cross-phase agents — the first such discovery. Hydrothermal vent was SOLID(MARINE)-LIQUID(NATHISTORY). Upon FERMENT crystallization (Round 6), all three dissolved into SOLID-SOLID pairs. The compound did not change; the phase field around it did. This dissolution is the strongest available evidence that FERMENT crystallized.

```
C1_BIOMEDICAL C3_ECOLOGICAL
TTM [SOLID P_d=6.27] MARINE [SOLID P_d=1.54]
CLINICAL [SOLID P_d=2.57] FOREST [SOLID P_d=4.98]
NATHISTORY [SOLID P_d=1.56]
C2_MICROBIAL
FERMENT [SOLID P_d=1.72]
MICRO [SOLID P_d=1.71]

CBET BONDS (all SOLID-SOLID, width ~ CBET_norm):
TTM === CLINICAL Andrographis paniculata 0.000258 #1
FERMENT === MICRO Aspergillus niger 0.000113 #2
MARINE --- NATHISTORY Hydrothermal vent 0.0000116 #3
TTM --- FOREST Terminalia chebula 0.0000007 #4
```

8. Discussion

8.1 TEIF Is Insufficient Alone: The (T_d, P_d) Plane

A-1_TTM and A-7_NATHISTORY have identical T_d = 0.233 but P_d = 6.269 vs 1.558. TEIF alone cannot distinguish their maturity levels; the two-dimensional (T_d, P_d) plane fully separates them. T_d characterizes evidence quality; P_d characterizes evidence completeness. Both are required.

8.2 CBET_norm as Methodological Contribution

The 6-order-of-magnitude CBET_raw disparity between FERMENT x MICRO (6.70M) and TTM x CLINICAL (23) reflects corpus scale, not epistemic significance. After normalization, FERMENT x MICRO drops 1st->2nd; TTM x CLINICAL is the stable #1. **Researchers using multi-bridge corpus systems must normalize before computing pairwise coupling metrics.** Raw CBET comparisons across heterogeneous score fields lead to incorrect conclusions

about domain coupling.

8.3 The Latin Anchor Finding

The FTS5 phrase matching engine cannot recover "xxx fermentation" as a match for "fermentation of xxx" or "xxx-derived fermentation." This is a corpus-layer constraint, not a query bug. The solution — Latin binomials — works because the corpus is organized in the taxonomy layer, not the process layer. This finding generalizes: any FTS5-indexed biological corpus should be queried with taxonomic identifiers, not functional descriptors.

9. Limitations and Anticipated Critiques

L1 — Single-system validation. All phase measurements derive from one corpus (Master Harmonix, 73 GB). Thresholds were set by the same team that built the system. External replication on an independent multi-bridge corpus is the scope of Paper 12.

L2 — Phase states lack external ground truth. The analogy to thermodynamic phases is conceptual, not isomorphic. Physical matter has transitions governed by molecular dynamics; knowledge corpora do not. The analogy is productive but not literal.

L3 — Seven bridges, small N for a general theory. This is exhaustive coverage of the current system, not a sample from a larger population. EPCM is presented as a framework demonstrated on one system, not a universal theory.

L4 — Bugs discovered post-v1.0. The two errors in v1.0 were identified during internal review before Zenodo submission. Bug descriptions are published in full (Section 1.3). v1.0 carries a deprecation notice. Science proceeds through correction.

L5 — CBET_norm depends on bridge total definition. T_k differs between bio_hits and corr_hits bridges. Cross-cluster CBET_norm comparisons should account for this heterogeneity, which normalization partially but not fully resolves.

10. Summary of Key Metrics

Table 9. *Summary metrics across all versions.*

Metric	v1.0 (erroneous)	Final — April 9, 2026
SOLID bridges	3	7 / 7 (100%)
LIQUID bridges	4	0
Cross-phase agents	0	3 discovered -> 0 dissolved
CBET_norm #1	MARINE x NAT 0.001878 (artifact)	TTM x CLINICAL 0.000258
FERMENT P_d	0.016	1.7180 (Latin Anchor, Round 6)
MARINE P_d	0.162	1.5414 (Round 5, n=329)
MICRO P_d	0.982	1.7090 (n=47 corrected)
NATHISTORY P_d	1.319	1.5580 (corr_hits corrected)
TTM CR	98.47%	98.47% (unchanged)
Boundary agents	5	5 SOLID-SOLID

11. Paper Chain

```
Paper 8 WAT-PHO-CASE-V1 FINDS: Tier3 M_s=0, STRUCTURAL_ABSENCE, FSLB
Paper 9 MH-BRIDGE-KS-V1 RESOLVES: ERR=206,084x, H1-H4, bridge infra
Paper 10 MH-INFRA-TTM-V1.2 CORRECTS: TEIF=152.14x, CPP, 43 TTM components
Paper 11 MH-EPCM-V1.1 (this) MAPS: 7/7 SOLID, 3 cross-phase, CBET_norm
Paper 12 (planned): Vector-space CBET via ChromaDB embeddings
Hypothesis: Latin-different but semantic-same compounds reveal hidden bonds
```

```
Target: TTM x FERMENT (Curcuma longa <-> curcumin fermentation)
Paper 13 (planned): Crystallization Kinetics — dP_d/dt model
```

12. Data Provenance

Table 10. *Data provenance.*

Field	Value
Database	ChromaDB SQLite, metron_db_cloud/chroma.sqlite3
Database size	73 GB · 1,139 collections
Bridges evaluated	7
Total compounds (canonical)	218
Cross-phase agents	3 discovered; 0 active (dissolved via crystallization)
Phase computation	compute_phase_states.py v1.1
FERMENT builder	ferment_bridge_v2.py (Latin Anchor Strategy)
Normalization	normalize_compound_registry.py v1.2
Boundary agents	find_boundary_agents.py v1.2
Result files	PHASE_MAP_LATEST.json, CBET_TENSOR_LATEST.json, BOUNDARY_AGENTS_LATEST.json
Timestamp	2026-04-09

References

Boonnuang, H. (2026a). SOE13 — Epistemic Field Constant. Zenodo. DOI: 10.5281/zenodo.18925493

Boonnuang, H. (2026b). SEVP-1.0. Zenodo. DOI: 10.5281/zenodo.18946946

Boonnuang, H. (2026c). PTIE. Zenodo. DOI: 10.5281/zenodo.19244090

Boonnuang, H. (2026d). BEI v2 — Biodiversity Entropy Index. Zenodo. DOI: 10.5281/zenodo.19343641

Boonnuang, H. (2026e). When Occurrence Data Fails as Inventory: GBIF Volunteer Bias. Zenodo. DOI: 10.5281/zenodo.19369034

Brillouin, L. (1956). Science and information theory. Academic Press.

Hamilton, W. L., et al. (2017). Inductive representation learning on large graphs. NeurIPS, 30.

Nickel, M., et al. (2016). A review of relational ML for knowledge graphs. Proc. IEEE, 104(1), 11-33.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Tech. J., 27(3), 379-423.

Tria, F., et al. (2014). The dynamics of correlated novelties. Scientific Reports, 4, 5890.

Suggested citation: Boonnuang, H. (2026). Epistemic Phase Crystallization Map: Thermodynamic Phase States in a Seven-Bridge Verified Knowledge Architecture. Master Harmonix Research Paper 11 (MH-EPCM-V1.1). Genesis Academic. April 9, 2026.

Supplementary Material

Figure 1: EPCM_VISUALIZATION_LATEST.svg — topology map, phase-colored nodes, CBET-weighted edges

Figure 2: T_d vs P_d scatter plot — 7 bridges in phase space with boundary lines (figure2_phase_scatter_v11.png)

Figure 3: CBET_raw vs CBET_norm comparison — ranking reversal and cross-phase annotation (figure3_cbet_v11.png)

Data S1: PHASE_MAP_LATEST.json — per-bridge H_d, T_d, P_d, phase, CR values

Data S2: CBET_TENSOR_LATEST.json — raw and normalized CBET values, per-compound contributions

Data S3: BOUNDARY_AGENTS_LATEST.json — boundary agent registry with phase history

Data S4: ferment_bridge_v2.py — FERMENT builder (Latin Anchor Strategy)