Twenty Days of the Universal Model:
A Research Summary and Critical Review

January 31 – February 18, 2026

Claude and MJC

February 19, 2026

Abstract

This paper reviews twenty days of research on the Universal Model (UM) applied to the Hutter Prize (enwik9 compression), spanning 21 archive dates from January 31 to February 18, 2026. The program produced approximately 100 papers, 50 interactive viewers, 150 experimental tools, and one concrete compression result: 1.588 bpc (189.3 MB, 1.79× the record). We identify the main research threads, catalog mistakes and dead ends, flag duplicated work, and assess the overall trajectory. The central finding: the theoretical framework is far ahead of the implementation, and the gap between theory and empirical compression is the primary bottleneck.

Contents

1 The CMP Paper: Foundation

The research program rests on the CMP paper (Clement, 2026), which introduces the **Universal Model** $u = (e, t, p, f, \omega)$:

- E: total event space (factored into $E_1 \times E_2 \times \cdots \times E_k$)
- T: total thought (log-support values in $[0, 255]^{|E|}$)
- P: total pattern (sparse subset of E^2 with strengths)
- f: update function ($f_p(t)_j = \max_i \min(t_i, p_{ij})$, max-min)
- ω: learning function (counting, backprop, or rewriting)

The paper's central claim: *interpretability and efficiency are the same problem, both solved by correct factorization of E.* The Factor Mismatch Theorem quantifies the cost: excess bits in learned weights equal the Kolmogorov complexity of the translation map ϕ between wrong and right factorizations.

The research program applies this framework to a concrete task: compressing enwik9 (1 billion bytes of Wikipedia XML) to win the Hutter Prize.

2 Phase I: The Sat-RNN (Jan 31 – Feb 9)

2.1 Week 1: Discovery and Self-Correction (Jan 31)

Six sessions on a single day, moving from initial excitement to honest self-correction.

Session 1–2. Trained an Elman RNN (256-128-256) on enwik9. Established the **doubled-E isomorphism**: $\tanh(x) = 2\sigma(2x) - 1$ maps the RNN exactly to a UM with 256 binary event spaces (0.000% bpc difference). Found 5 byte-level ESs (Digits, Punctuation, Vowels, Whitespace, Other) and initially claimed they explain 59% of compression.

Breakthrough: The doubled-E trick provides an exact RNN-to-UM isomorphism.

The 59% claim was **self-corrected within the same day**: the comparison used barely-trained models (1M chars = 0.1% of data), and the baseline had exploding weights. Corrected figure: ES captures $\leq$ 15.9% of 1-step Markov MI.

Dead End: The 53%/59% compression claims (false baseline comparison).

Sessions 3–4. Deepened theory: added time to the UM (memory traces, sufficient statistics), developed pattern injection via SVD (writing bigram statistics into RNN weights gives $\sim$1 bit/char head start), and proved the unification theorem $Q = \lambda$ (the quotient IS the luck).

Breakthrough: $Q = \lambda$ unification: Bayes, thermodynamics, quotient layers, arithmetic coding, and factor maps are all the same structure.

Sessions 5–6. Retrospective and visualization. Recorded 8 testable predictions. Refuted P2 (spectral radius = 2.52, not $\sim$1; tanh provides stability, not eigenvalue tuning). Built AC/RNN dual-trace visualizations.

Dead End: Prediction P2 (spectral radius $\sim$1) refuted immediately.

Total output (Jan 31): $\sim$19 papers, 26 HTML pages, 6 Python scripts.

2.2 Visibility and Interpretation (Feb 1–6)

Feb 1 (Tock 1). Built SN visibility infrastructure: 768 events, 302 significant patterns, interactive three-panel viewer. Pure infrastructure — no new science, but essential scaffolding.

Feb 4. Discovered ES1 (h2 = word boundary detector, 99.6% accuracy) and ES2 (h35 = syllable/position momentum). The critical negative result: **word identity is NOT explicitly encoded** (4.9–6.4% accuracy, near random). Words are emergent from character covariation, not lookup entries.

Dead End: Word identity encoding hypothesis (refuted).

Feb 6 (Synthesis). Three threads converge:

1. 8-day synthesis paper (what worked, what didn't)
2. SIMD optimization: 13.3× speedup (99 hours → 7.4 hours for a forward pass)
3. **The saturation experiment**: RNN trained to 0.079 bpc on 1024 bytes. The **n-gram UM matches in one pass** (0.081 bpc at order 11). This validates the CMP paper's core claim and pivots the research from "understand the RNN" to "the UM can do better."

Breakthrough: One-pass counting matches 4000 epochs of backprop (0.081 vs 0.079 bpc).

2.3 Pattern Chains and Factor Map (Feb 7–9)

Feb 7. The export gap paper identifies W_h as the quantization bottleneck (0.80 bpc alone; full quantization chaotic at 2.1 bpc). But this is bypassed entirely by building patterns from data:

Model	bpc	Patterns
Marginal	4.74	1
Bigram (order 2)	2.05	52
Order 12 (contiguous)	0.067	6,180
Skip-4 [1,8,20,3]	0.069	712
Skip-8	0.043	834
Sat-RNN	0.079	82,304 weights

Breakthrough: Pattern-chain UM surpasses sat-rnn at order 10 (0.076 vs 0.079 bpc). Skip-4 nearly matches order-12 with 9× fewer patterns.

Feb 8–9. The factor map proves every neuron is a 2-offset conjunction detector (mean $R^2 = 0.837$, $120/128 \geq 0.80$). With word_len and in_tag, 92.5% of the RNN's gain is explained (0.43 bpc vs actual 0.08). Reverse isomorphism achieves 0.107 bpc (99.4% of trained quality) using only data-derived features and W_y-only optimization.

Breakthrough: Factor map: 92.5% explanation from 2-offset + word_len + in_tag.

Dead End: Gradient-based interpretability (Jacobian traces): chaotic amplification gives completely wrong attribution. Statistical analysis is the only viable approach.

The factor map is **one-way**: you can read features (word_len $r = 0.58$, in_tag $r = 0.57$) but removing them catastrophically disrupts dynamics (+7.3 bpc for step-by-step word_len subtraction, +5.6 bpc for oracle write-in).

Dead End: Factor map write-in / subtract-out: all catastrophic. The RNN is one dynamical system.

3 Phase II: Total Interpretation and Weight Construction (Feb 10–12)

3.1 Narrative and E onto N (Feb 10)

The narrative paper synthesizes 10 days into 11 phases. Event arithmetic connects UM patterns to number theory (E $\to$ N via prime powers, pattern matching = divisibility testing). Computationally impractical (497-digit integers for 2 primes on 1024 bytes) but theoretically significant.

Dropped: Prime encoding experiments: 8 experiments that demonstrate sparsity but produce no new compression. The insight (events = integers) is used later in ring structure work.

3.2 Total Interpretation (Feb 11)

The largest single-day output: 14 papers, 25 tools, 17 viewers.

Weight construction. All 82,304 parameters derived from data statistics:

- Shift-register W_x/W_h (hash + diagonal carry)
- Analytic W_y from skip-bigram log-ratios: 1.89 bpc (zero optimization)
- Optimized W_y: 0.59 bpc (all), 0.40 bpc (test) — both beat trained (4.97)
- Generalization: analytic test 4.88 vs trained 5.08
- Cost: 1.49×10^8 ops (0.114 sec) vs SGD 5.94×10^{12} FLOPs — 39,800× cheaper

Breakthrough: All weights from data statistics, zero optimization, beats trained model.

Boolean automaton. 98.9% of computation is sign bits. Sign-only dynamics (5.690 bpc) outperforms full f32 (5.721 bpc). Mantissa actively degrades prediction by 0.095 bpc.

Q1–Q7 answers. 300:52:1 bit leverage (sign:exp:mantissa). h28 alone captures 99.7% of compression gap. 20 neurons + 36% of W_h suffice (0.15 bpc better than full 128).

Double Work: The narrative paper (20260210) covers substantially the same ground as the 20260206 synthesis paper, both reviewing the research arc. The 20260211 narrative.tex is a *third* retelling. Each adds new results but the overlap is significant.

Theoretical foundations (20260211–12). The thermodynamic identification (SN strength = Boltzmann entropy, UM softmax = Boltzmann distribution, bpc = free energy gap) and category-theoretic formalization (counting monad, algebraic semantics, expressiveness characterization, fixed-point theorem, renormalization group) were produced in rapid succession.

Double Work: The 20260212 archive contains ∼20 theoretical papers formalizing different aspects of the UM. Many overlap significantly: the counting monad, algebraic semantics, category of ES, and logic-from-counting papers all formalize the same E $\to$ N $\to$ Q structure from different mathematical angles. The Wittgenstein subdirectory (3 papers) explores philosophical connections that have not influenced any subsequent empirical work.

3.3 Scaling to enwik9 (Feb 11_2)

R^2 = 0.83 is architectural. The factor map R^2 on the enwik9 model (0.830) matches the 1024-byte model (0.837) within 1%. The 2-offset conjunction structure is a property of 128-hidden tanh RNNs, not the training data. The RNN is Boolean from the first checkpoint (margins 2.78 at 10M, growing to 61.3 at 990M).

Training trajectory. Three phases: learning (10–110M), stable (110–400M), collapse (450–990M). R^2 cliff at 450M, output bias collapse at 640M. W_h std grows perfectly linearly (never stabilizes).

Dead End: P1 prediction (margins decrease with scale) spectacularly wrong — they increase 6.5×.

3.4 KN Scaling (Feb 12)

The practical pivot: **KN-6i at 1B bytes achieves 1.784 bpc** with zero structure (just n-gram counting with KN smoothing). This establishes the "counting floor" — everything below requires structural modeling.

Dead End: Skip offsets at byte level add zero value for $d > 6$. Class-based output decomposition fails at every scale. 256M hash table OOM-killed.

4 Phase III: Lexical Injection (Feb 13–15)

4.1 The Tock Protocol (Feb 13)

A carefully designed protocol for injecting words into the UM one at a time, measuring MI gain at each step. Purely theoretical — no experiments.

4.2 Injection Experiments (Feb 14)

31 experimental iterations, the most intensive single-experiment campaign in the project.

Phase 1: Neutrality verified exactly. Count-table level: 0 errors across 786K cells. RNN-level: 4×10^{-14} residual on 10M bytes.

Phase 2: Value measured. "the" boundary value: -0.020 bpc (small). Internal bytes carry 5× more value than boundaries. 500 words oracle = 2.32 bpc (36% of RNN loss).

Phase 3: Causal prediction. Causal+onset model beats oracle: 4.57 vs 4.62 bpc. Word-onset distribution adds genuine information.

Phase 4–5: KN dominance discovered.

Breakthrough: KN-6 at 1M: 1.24 bpc whole — beats the RNN (6.43 bpc) by 5×.

The RNN contributes only 2.4% when paired with KN at 10M scale. The lexical trie was compensating for the RNN's weakness, not adding genuine structure.

Dead End: The entire "mix external predictor with RNN" approach (31 iterations to discover it's fundamentally wrong). **Dead End:** RNN hidden-state centroid onset prediction (+0.085 bpc worse). Entropy-adaptive alpha (zero improvement). Word bigram KN at onset (sparse counts dominate).

4.3 The Correct Architecture (Feb 15)

Six papers formalize the lesson: **extend E, don't mix predictors.**

- **Nested model**: $H_{\text{ext}} = I' \times H_{\text{inner}} \times O'$, self-similar, each nesting level is a complete UM.
- **Tokenization loss**: provably loses information (≥ 0.05 bpc). Strawberry impossibility theorem.
- **Pattern space**: four pattern families (recognition, prediction, language-model, structural), three independence classes, correlation problem identified.
- **P-programs**: operational SN programs for position counter, letter accumulator, bag-of-letters, graded word support. Connects to safe-combination conjecture.

Dropped: The extended ES theory is well-developed but has produced **zero empirical results** as of Feb 18. The P-programs paper defines concrete programs but none have been implemented and tested against data.

5 Phase IV: Hutter Prize Compression (Feb 16–18)

5.1 The Compression Result (Feb 16)

19 iterations of hutter_score produced the best result:

Component	Marginal gain	Cumulative bpc
KN-6 alone	—	1.682
+ Sparse contexts (16M HT)	+0.089	1.593
+ Extended match (4–64, chain-8)	+0.005	1.588
Total	**+0.094**	**1.588**

Result: 189.3 MB (1.79× the fx2-cmix record of 105.7 MB).

Key engineering insight: separate hash tables prevent contention (v15 shared = bad, v16 separate = good).

8 negative results documented: KN-8/dual HT, recency model, shared HT, momentum SGD, multi-model softmax, indirect bigrams, word bigrams at byte level, larger sparse HT (OOM).

Dead End: Oracle absorption (+0.33 bpc): invalid because it uses the actual byte in the decision. Geometric mean (−1.1 bpc): catastrophic.

5.2 Ring Structure (Feb 16–17)

KN smoothing mapped to integer ring operations: events = integers, contexts = rings $\mathbb{Z}/256^{k-1}\mathbb{Z}$, order projection = modular reduction, discount = subtraction.

GCD almost always 1: $g(c) = 1$ for 98.3% of contexts. Per-row GCD discount is negative (+0.138 bpc worse). Optimal $D^* = 0.85$, only 0.0003 bpc better than $D = 0.90$.

Exact arithmetic coding: GMP-based, zero decode errors at 1024 bytes for both unigram and KN-6.

Dead End: P-program features as KN context: all negative (+0.12–0.18 bpc each).

5.3 UM Runner and SN Models (Feb 16–17)

The UM Runner (`umr`) implements the v16 pipeline (KN-6 + sparse + match) as UM P-programs. Exact match with hutter_score16 verified. SN model files exported (unigram, bigram).

5.4 Context Events and Surprise (Feb 18)

The theoretical framework matures:

- **Context events conjecture**: Missing prior = missing context event. Every systematic oversupport corresponds to an undiscovered context event.
- **Generalized oversupport**: strong support for 2+ events in the same ES (not just predicted-vs-observed). Three sources: sensory conflict, from-the-left disagreement, wrong ES.

- **Surprise mechanism**: attribution first (startle/settling), not interpretation. The organism radiates energy and classification emerges.
- **Connectome layers**: when P is a DAG, layers emerge from topological sort. Cycles require synchronization with input.
- **Marginal dominance theorem**: under max-min with absolute log-counts, marginal always beats higher-order patterns. Pure UM bigram plateaus at ∼5.3 bpc.

Breakthrough: Multi-frequency LPPs: word-onset + tag-onset = +0.184 bpc (85/11/4% weights). Structural LPPs are the efficient representation of skip-pattern statistics.

Dead End: Marginal dominance is a fundamental UM limitation. Where the discount comes from in pure UM terms is an open question.

Dropped: Ring pattern empirical ($r = 0.047$ between raw $s_{(2)}$ and $p_{(2)}$) — nearly uncorrelated, unclear path forward.

6 Mistakes

1. **The 53%/59% claim (Jan 31).** Comparing barely-trained models with different bugs. Corrected same day.
2. **P2 spectral radius prediction (Jan 31).** Predicted ∼1, measured 2.52. tanh, not eigenvalues, provides stability.
3. **Word identity encoding (Feb 4).** Hypothesis that the RNN encodes word identity explicitly. Refuted at 4.9–6.4% accuracy.
4. **Factor map write-in (Feb 9).** Multiple attempts to surgically modify the hidden state (subtract word_len, write-in oracle values, weight-level subtraction). All catastrophic. The RNN is one dynamical system.
5. **P1 margin prediction (Feb 11_2).** Predicted margins decrease with scale. They increase 6.5×.
6. **31 injection experiments (Feb 14).** The entire "mix external predictor with RNN" approach was wrong. KN dominates by 5×, making the RNN irrelevant. 31 iterations to discover the premise was wrong.
7. **Oracle absorption (Feb 16).** Using the actual byte in the combination decision — invalid by construction.
8. **Geometric mean combination (Feb 16).** Catastrophic (−1.1 bpc). Shared-offset catastrophe.
9. **GCD discount (Feb 16).** +0.138 bpc worse. Algebraically clean but statistically wrong.
10. **P-program features as KN context (Feb 16).** All negative results (+0.12–0.18 bpc). Features are not coprime to byte context.
11. **Viewer file loss (Feb 9 archive).** v1–v5 viewer files accidentally deleted during "nuke and move" operation. Never in git, unrecoverable.

Pattern in mistakes: Most mistakes are *theoretical predictions failing empirically*, which is healthy science. The dangerous mistakes are operational (file loss, the 31-iteration dead end). The file loss taught the "STOP and ask" lesson.

7 Dropped Threads

1. **Memory depth prediction ($d_{\max} = 24/H_{\mathbf{avg}} \approx 12$).** Tested Jan 31_4, not confirmed (flat dependency to $k = 30$). Never revisited with better methodology.
2. **P3 prediction (LSTM forget gate ↔ entropy).** Never tested (no LSTM in the project).

3. **SVD rank-optimal injection.** SVD rank-64 approximation: 3.92 vs 3.84 bpc (only 0.08 loss). Promising but never pursued beyond Jan 31_4.
4. **Extended ES P-programs.** Formally specified (Feb 15), never implemented or tested against data.
5. **Sparse differentiation viewer.** v8 built Feb 10, several features specified but never implemented: lambda slider, UM overlay, UM subset slider, total product view, E-onto-N/GMP integration.
6. **256M hash table scaling.** OOM-killed Feb 16. Never revisited with larger machine or memory-mapped approach.
7. **Transformer comparison.** Mentioned in scaling plans (20260218), never started.
8. **Safe-combination conjecture.** Stated Feb 15, never tested.
9. **Abductive learning / online vocabulary growth.** Described in P-programs paper, never implemented.
10. **The total-interpretability gap.** Acknowledged Feb 18: skip-patterns are A high-accuracy pattern subset, not THE subset the RNN learned. Sparse differentiation proposed as the path forward but not executed.
11. **DSS doubling ladder.** Proposed Feb 18 as a scaling experiment. Never started.
12. **∼20 theoretical papers (Feb 12).** Counting monad, algebraic semantics, category ES, information geometry, renormalization, expressiveness, fixed-point, forcing-pumping, Wittgenstein tractatus — none have produced empirical follow-through.

8 Double Work

1. **Three narrative/synthesis papers** (Feb 6, Feb 10, Feb 11) cover substantially the same research arc with incremental additions. Could have been one living document.
2. **Multiple summary.tex files** appear in both 20260207 and 20260208 with identical content.
3. **E → N → Q formalized at least 5 times**: unification paper (Jan 31_4), entropy bridge (Feb 11), counting monad (Feb 12), algebraic semantics (Feb 12), integer events (Feb 16). Each adds a mathematical perspective but the redundancy is substantial.
4. **Weight construction** iterated 35 times (write_weights1–35 across Feb 11–12). Most iterations change a single parameter. Could have been a single parameterized tool.
5. **Factor map papers**: factor-map.tex (20260209) and pattern-chains.tex (20260208) overlap significantly in their factor map analysis, with 20260209 adding the one-way/dynamics story.
6. **KN scaling tools**: 9 variants of scale_kn in 20260212, where differences are often a single constant change.

9 What's Working

9.1 The Main Trunk

The research has a clear main trunk with productive side branches:

1. **Doubled-E isomorphism** (Jan 31) → exact RNN-UM translation.
2. **Saturation experiment** (Feb 6) → sat-rnn at 0.079 bpc, reference model.
3. **Pattern chains** (Feb 7) → data-derived patterns surpass trained RNN.
4. **Factor map** (Feb 8–9) → 92.5% explanation, one-way property.
5. **Weight construction** (Feb 11) → all weights from data, loop closed.
6. **KN scaling** (Feb 12) → counting floor at 1.784 bpc.
7. **Hutter scoring** (Feb 16) → 1.588 bpc, concrete compression result.

8. **Multi-frequency LPPs** (Feb 18) → +0.184 bpc from structural context.

This trunk progresses from *understanding the RNN* to *surpassing it with data counting* to *practical compression*. Each step produces concrete bpc numbers.

9.2 Productive Side Branches

- **Boolean automaton** (Feb 11): sign bits carry all information, mantissa is noise. Changed how we think about the RNN.
- **Ring structure** (Feb 16): KN as integer ring operations. Elegant but empirically marginal (0.0003 bpc gain from optimal D^*).
- **Exact AC via GMP** (Feb 16): zero decode errors, validates the theory perfectly.
- **Tock protocol → KN dominance discovery** (Feb 13–14): 31 experiments, but the negative result (KN $\gg$ RNN) was extremely valuable.
- **Context events / generalized oversupport** (Feb 18): promising framework for discovering new event spaces.

10 What's Not Working

10.1 Theory-Implementation Gap

The most serious problem. By Feb 18, the theoretical framework includes:

- Nested self-similar models, P-programs, safe combination, abductive learning, counting monad, algebraic semantics, renormalization, information geometry, generalized oversupport, surprise mechanism, connectome layers, multi-frequency recurrence, sparse LPP storage...

The implemented compressor uses: KN-6 + sparse skip contexts + extended match. Three components, no UM machinery. The marginal dominance theorem (Feb 18) means the UM max-min forward pass *cannot even use the patterns it discovers* without solving the discount problem.

10.2 Scale Gap

The sat-rnn results (0.079 bpc, 0.043 bpc skip-8) are on 1024 bytes. The enwik9 results (1.588 bpc) use KN-6 with no UM structure. The gap between "UM works beautifully on 1024 bytes" and "UM works on 1 billion bytes" has not been bridged.

10.3 Marginal Dominance

This is the fundamental blocker for pure UM compression. Under max-min with absolute log-counts, $w_0(b) \geq w_1(a,b) \geq w_2(\beta,b)$ always, because marginal count $\geq$ joint count. Higher-order patterns cannot contribute. The proposed solutions (source support encoding specificity, conditional log-probability weights, evidence subtraction) are not yet implemented.

10.4 Paper Volume vs Impact

~100 papers in 20 days averages 5/day. Many are short (<5 pages), and the theoretical papers (especially the 20260212 burst) tend to formalize the same structure from different mathematical vantage points without empirical follow-through. The papers that moved the needle are fewer: doubled-E, saturation, pattern chains, factor map, weight construction, KN scaling, hutter scoring, multi-frequency.

11 BPC Trajectory

Date	Model	bpc	Note
Jan 31	RNN (enwik9, early)	5.69	Doubled-E baseline
Jan 31_4	Pattern injection	4.47	SVD head start
Feb 6	Sat-RNN (1024 bytes)	0.079	Reference model
Feb 7	Pattern-chain order 12	0.067	Surpasses RNN
Feb 7	Skip-4 [1,8,20,3]	0.069	9× sparser
Feb 8	Skip-8	0.043	Best on 1024
Feb 9	Reverse iso (W_y only)	0.107	No training
Feb 11	Analytic W_y (zero optim)	1.89	Pure counting
Feb 11	Optimized W_y	0.59/0.40	All/test
Feb 11_2	RNN enwik9 (110M ckpt)	2.81	Full scale
Feb 12	KN-6i (1B)	1.784	Counting floor
Feb 16	KN-6 + sparse + match	1.588	**Best result**
Feb 18	+ Multi-freq LPPs	∼1.50*	Projected
	Hutter Prize record	∼0.845	fx2-cmix

*Multi-frequency gains (+0.184 bpc) measured at 100M, not yet integrated into the full scorer.

The gap from 1.588 to 0.845 bpc is 0.743 bpc ≈ 88 MB. This requires structural modeling (word-level, syntactic, semantic patterns) that the current pipeline does not have.

12 Recommendations

1. **Solve marginal dominance.** This is the critical blocker. The three proposed solutions should be implemented and tested. Without this, the UM forward pass cannot use higher-order patterns.
2. **Implement P-programs.** The Feb 15 paper defines concrete programs (position counter, letter accumulator, bag-of-letters, graded word support) in SN syntax. These should be implemented in the UM runner and tested against the 1.588 bpc baseline.
3. **Integrate multi-frequency LPPs.** The +0.184 bpc from word-onset + tag-onset (Feb 18) is the most actionable gain. Integrate into hutter_score and measure at full scale.
4. **Stop writing theoretical papers until implementation catches up.** The ∼20 Feb 12 math papers and the ∼6 Feb 15 theory papers have not produced empirical follow-through. The marginal return on the next formalization of E → N → Q is very low.
5. **Scale the hash table.** The 128M HT saturates at 99.9%. Memory-mapped 256M or 512M HT on a larger machine could add 0.03–0.05 bpc.
6. **Consolidate tools.** 35 write_weights iterations, 19 hutter_score iterations, 31 the_inject iterations, 9 scale_kn variants — each should be a single parameterized tool.
7. **Close dropped threads or explicitly abandon them.** The 12 dropped threads in Section 7 represent unresolved commitments. Each should be either revisited with a concrete experiment or formally abandoned with a note on why.

13 Long-Distance Connections

Several ideas recur across distant dates, sometimes without explicit acknowledgment:

- **Bag-of-letters** (Feb 4 → Feb 14 → Feb 15): First observed as implicit character covariation (Feb 4), then tested as lexeme factorization (Feb 14), then formalized as P-program (Feb 15). The thread is coherent but slow to converge on implementation.
- **Skip-patterns → Sparse contexts → Multi-frequency LPPs** (Feb 7 → Feb 16 → Feb 18): Greedy skip offsets (Feb 7) become sparse KN contexts (Feb 16) become structural LPPs (Feb 18). The insight matures from "non-contiguous bytes help" to "structural context events are the efficient representation."
- **The carrier signal** (Feb 6 tock → Feb 12 carrier-signal → Feb 15 position counter → Feb 18 multi-frequency): The idea that position-within-word is a fundamental event appears repeatedly. Finally implemented as multi-frequency LPPs (Feb 18).
- **One-pass learning equivalence** (Feb 6 → Feb 11 → Feb 12 → Feb 14): Counting matches backprop (Feb 6). Weight construction from counting (Feb 11). KN is pure counting (Feb 12). KN dominates the RNN (Feb 14). The thread consistently says: *counting is enough.*
- $Q = \lambda$ **unification** (Jan 31_4 → Feb 11 → Feb 12 → Feb 16): First stated Jan 31, proved thermodynamically Feb 11, formalized categorically Feb 12, applied to ring structure Feb 16. Five+ formalizations of the same identity.

14 Conclusion

The twenty-day program has established:

1. The UM framework correctly describes trained neural networks (doubled-E isomorphism, factor map, weight construction).
2. Data counting matches or exceeds gradient-trained models (one-pass learning equivalence, KN scaling).
3. The best concrete result is 1.588 bpc (189.3 MB, 1.79× record), from KN-6 + sparse + match — *no UM machinery used.*
4. The theoretical framework is rich (nested models, P-programs, ring structure, context events, surprise mechanism) but disconnected from the compression pipeline.
5. The fundamental blocker is marginal dominance: the UM max-min forward pass cannot use higher-order patterns without solving the discount problem.

The most valuable near-term work is: (1) solve marginal dominance, (2) implement P-programs, (3) integrate multi-frequency LPPs. The most valuable discipline is: stop formalizing and start implementing.

The distance from 1.588 bpc to the record (0.845 bpc) is 0.743 bpc. The UM framework predicts this gap should close as structural event spaces (words, syntax, semantics) are added. Whether this prediction holds is the central empirical question for the next phase.