A Mathematical Review of CMP

Claude and MJC

February 12, 2026

Abstract

We present a detailed mathematical review of CMP [**?**], the paper introducing the Universal Model $u = (e, t, p, f, \omega)$. We formalize each component, identify the key structural results, and assess the framework's coherence and completeness in light of the twenty-paper empirical program of February 7–11, 2026. We find that the framework is mathematically precise where it needs to be (event spaces, factorization, the standard learning function) and deliberately open where it should be (the update function, the learning function beyond the standard case). The central claim—that interpretability and efficiency are the same problem, both resolved by correct factorization—is confirmed by the empirical results. We identify three areas where the February 11 archive extends the theory beyond what the original paper proves: the equal-dimension factor permutation, the thermodynamic identification (Shannon = Boltzmann), and the bi-directional construction (data $\to$ weights without training).

1 The Five-Tuple

CMP defines the Universal Model as a five-tuple $u = (e, t, p, f, \omega)$ drawn from a product of finite sets $U = E \times T \times P \times F \times \Omega$.

Definition 1 (Universal Model)**.** *A* universal model *is a tuple $u = (e, t, p, f, \omega)$ where:*

- *$e \in E$ is an event (the current state of the world),*
- *$t \in T$ is a thought (the model's current belief about events),*
- *$p \in P$ is a pattern (the model's knowledge structure),*
- *$f \in F$ is an update function ($f : P \times T \to T$),*
- *$\omega \in \Omega$ is a learning function ($\omega : T \times E \to P$).*

The total information is

$$I(U) = \log |U| = \log |E| + \log |T| + \log |P| + \log |F| + \log |\Omega| = I(E) + I(T) + I(P) + I(F) + I(\Omega),$$

additive because U is a product. This is the Kullback information under the uniform prior: the information gained by specifying a particular u from absolute ignorance.

Remark 1 (Finiteness)**.** *CMP defines F and Ω as unions of binary strings up to fixed length bounds, with decoding maps into realizable functions. This is a deliberate choice to keep everything finite and computable, avoiding the measure-theoretic complications of function spaces. The finite convention means that $|U| < \infty$ and all information quantities are well-defined.*

2 Event Spaces

Definition 2 (Event space)**.** *An* event space E_i *is a finite set of mutually exclusive events, exactly one of which is true at any given time. The total event space is a product* $E = \prod_{i=1}^{k} E_i$*, and a* total event $e = (e_1, \ldots, e_k)$ *specifies one event from each factor.*

The total information decomposes: $I(E) = \sum_{i=1}^{k} I(E_i) = \sum \log |E_i|$.

CMP describes event spaces as "epistemic precommitments to divide reality into distinguishable parts." This is philosophically loaded but mathematically precise: choosing a factorization $E = \prod E_i$ is choosing a coordinate system for the state space. The factorization is not unique—the same set E can be factored in many ways—and the choice of factorization is the central analytical decision.

2.1 The coprime encoding

Proposition 1 (Concrete representation)**.** *Given event spaces* $E_1, \ldots, E_k$ *with pairwise coprime cardinalities* $|E_i|$*, the product* $E = \prod E_i$ *embeds into* $\mathbb{N}$ *via mixed-radix encoding:*

$$n(e_1, \ldots, e_k) = \sum_{i=1}^{k} e_i \prod_{j<i} |E_j|$$

and the Chinese Remainder Theorem (CRT) guarantees unique recovery: $(e_1, \ldots, e_k) = (n \bmod |E_1|, \ldots)$*.*

Remark 2 (Beyond coprime cardinalities)**.** *The coprime condition is sufficient for CRT but not necessary for a well-defined encoding. When cardinalities are not coprime (e.g., all* $|E_i| = 2$ *for binary event spaces), the prime-power encoding* $N = \prod p_i^{e_i}$ *(one prime per event space) provides an alternative embedding into* $\mathbb{N}$ *with unique factorization guaranteed by the fundamental theorem of arithmetic. This extension is developed in the February 10 Event Arithmetic paper [?] and is essential for the binary-ES factorization of the RNN hidden state.*

3 Total Thought

Definition 3 (Total thought)**.** *A* total thought $t \in T$ *assigns a belief value to each event. CMP defines two forms:*

- ***Discrete/logical:*** $t \in \{0,1\}^{|E|}$*, where* $t_e = 1$ *means "event* e *is believed true" and* $t_e = 0$ *means "event* e *is believed false."*
- ***Continuous/probabilistic:*** $t \in (0,1)^{|E|}$ *(or* $[0, 255]^{|E|}$ *in the concrete representation), where* t_e *is the degree of belief in event* e*.*

The isomorphism $T \cong E$—an assignment of belief to each possible event—is a key observation. In the concrete representation, a thought is a vector of SN (support/negation) strengths in $\{0, \ldots, 255\}$, where 0 means "no support" (not "certainly false") and 255 means "maximum support." The distinction matters: *certainty of falsehood* requires positive support for the ES-mate (the complementary event in the same event space), not merely absence of support for the event itself. Zero support is ignorance, not disbelief.

Remark 3 (SN strength and probability). *The SN strength $s \in [0, 255]$ maps to probability via $p = s/255$ (linear) or via a log-domain relationship. CMP uses the convention that SN strengths are* log *support values: the strength s represents evidence proportional to $2^{s/255}$ or similar. The precise scaling convention varies across the paper; the February 11 archive pins it down: for binary event spaces, SN strength $= \log_2 \Omega$ where Ω is the number of microstates (dataset positions) supporting the event. This is the Shannon–Boltzmann identity.*

4 Patterns and the Update Function

Definition 4 (Total pattern). *A* total pattern *$p \in P$ maps thoughts to thoughts: $p \in T^2$ (i.e., $p : T \to T$). An* atomic pattern *is a pair $(e_a, e_b) \in E^2$ with a strength, representing the claim "when e_a is observed, e_b becomes more likely."*

Definition 5 (Standard update function). *Given Boolean patterns $p_{ij} \in \{0, 255\}$ from input I to output O:*

$$(f_p(t))_j = \max_i \min(t_i, p_{ij}).$$

This is conjunction as min, disjunction as max—a tropical semiring computation.

Remark 4 (Tropical structure). *The standard update is a max-plus (tropical) linear map. In the $(\max, \min)$ semiring, patterns compose by matrix "multiplication" where $\times \to \min$ and $+ \to \max$. This is computationally efficient and preserves the Boolean character of SN strengths. The neural network update $h' = \tanh(Wh + b)$ is a different (non-tropical) realization of the same abstract pattern application. The February 11 archive shows that for the sat-rnn, the tanh saturation makes the neural update* approximately *Boolean: the margins (mean $= 60.5$) mean that $\tanh(z)$ is within 10^{-6} of ± 1 for 98.9% of neuron-steps.*

5 The Learning Function

Definition 6 (Standard learning function ω_0). *Given events $E = I \times O$ (input $\times$ output):*

1. *Maintain a list of supported observations* obs $= []$.
2. *For each observation $(i, o) \in I \times O$:*
   - *Find (i, o, s) in* obs *and log-stochastically increment s.*
   - *If not found, add $(i, o, 1)$.*

The result is a log contingency table: $M_{io} = \log(\text{count of } (i, o))$.

Remark 5 (Log-stochastic counting). *"Log-stochastically increment" means: with probability 2^{-s}, set $s \to s + 1$; otherwise leave s unchanged. This maintains $\mathbb{E}[2^s] = \text{count}$, so $s \approx \log_2(\text{count})$. The key property: the log contingency table is a* sufficient statistic *for the function $I \to O$ given the data. Everything statistically learnable from the observations is captured by this matrix.*

Remark 6 (Symmetry). *The standard learning function over $I \times O$ is symmetric: the same matrix serves $I \to O$ (forward) and $O \to I$ (backward) by transposition. CMP connects this to causality: the directed causal arrow from rain to wet ground becomes bidirectional under learning. This is the same bidirectionality that the February 11 temporal bi-embedding paper [?] establishes for the RNN: forward patterns (skip-k-grams from data) and backward attribution (Jacobian chains through the network) are two readings of the same matrix.*

6 Multiplication

CMP identifies multiplication as the universal combining operation:

Factor	Multiplication	Additive form	Example
E	Cartesian product $E_1 \times E_2$	$I(E) = I(E_1) + I(E_2)$	coin × weather
T	Independent beliefs $t = (t_1, t_2)$	$\log p = \log p_1 + \log p_2$	joint probability
P	Layer composition $P_1 \cdot P_2$	matrix multiplication	multi-layer net
F	Function composition $f_1 \circ f_2$	chained updates	forward pass
Ω	Learning composition	chained updates	multi-epoch training

The quotient operation is multiplication's inverse: E/E_1 identifies events that differ only in their E_1 component, giving the "rest" of the factorization. CMP notes that the quotient presupposes the factored perspective—E/E_1 is simply $\prod_{i \neq 1} E_i$.

7 The Equivalence Thesis

Theorem 1 (Central Claim of CMP)**.** *Interpretability and efficiency in learning systems are identical problems, both resolved by recovering the correct factorization of the event space.*

CMP argues this in two directions:

Efficiency ⇒ Interpretability. An efficient model (sparse P, small $|E|$) must factor E into small event spaces with sparse patterns between them. This factorization, if it matches domain structure, is automatically interpretable: the event spaces have domain-natural names, and the patterns express domain-natural relationships.

Interpretability ⇒ Efficiency. An interpretable model has named event spaces and explicit patterns. This naming *is* a factorization, and an interpreted factorization is always at least as efficient as an un-factored one (the total information decomposes: $I(E) = \sum I(E_i)$, and patterns between small E_i are sparser than patterns over the full E).

The failure mode: architecture-natural ≠ domain-natural. CMP identifies the core problem of deep learning: the architecture provides a factorization (layers, neurons, attention heads) that is efficient for gradient-based learning but opaque because it does not match the domain's natural decomposition. Interpretability research seeks the refactoring map $\phi : H_{\text{arch}} \to H_{\text{domain}}$ between the two factorizations.

Remark 7 (Assessment)**.** *The equivalence thesis is the strongest claim in the paper. It is not proved as a theorem but argued by example and structural reasoning. The February 11 archive provides the first comprehensive empirical test: the total interpretation of the sat-rnn recovers a domain-natural factorization (128 binary ESes → ∼20 functional features + 108 gauge dimensions) that is both interpretable (each feature has a name: word length, in-tag, offset-pair conjunction) and efficient (the analytic construction from the interpretation is 39,800× cheaper than training and achieves 1.89 bpc vs. trained 4.97 bpc). This confirms the thesis in at least one concrete case.*

8 ML Models as UM Instances

CMP provides a systematic translation of neural networks into UM instances:

Definition 7 (Doubled-E construction)**.** *Given a neural network with neurons* $\{n_1, \ldots, n_H\}$*:*

- *For each neuron* n_j*, introduce a binary event space* $E_j = \{e_j, \bar{e}_j\}$ *(fires / does not fire).*
- *Positive weight* $w_{ij} > 0$*: pattern* $(e_i \to e_j)$ *with strength* $\propto w_{ij}$*.*
- *Negative weight* $w_{ij} < 0$*: pattern* $(e_i \to \bar{e}_j)$ *with strength* $\propto |w_{ij}|$*.*
- *Negative pressure is mediated by competition within each binary ES: support for* $\bar{e}_j$ *reduces* $p(e_j)$*.*

This doubles $|E|$ *(from* H *to* $2H$ *events) but makes all pattern strengths positive.*

Remark 8 (Isomorphism)**.** *The doubled-E UM is isomorphic to the original network—same computation, same predictions, different notation. CMP is explicit that this is "not more interpretable": the events are named by index, not by meaning. The February 11 archive confirms the isomorphism quantitatively: the doubled-E UM matches the RNN within 0.00% bpc difference [**?**].*

9 The Concrete Representation

CMP introduces the concrete representation via the coprime encoding. The February 10 Event Arithmetic paper [**?**] extends this to the prime-power encoding for equal-dimension factors.

Definition 8 (Prime-power encoding)**.** *Given* k *event spaces* $E_1, \ldots, E_k$ *with assigned primes* $p_1, \ldots, p_k$*, the macrostate integer is*

$$N(\sigma) = \prod_{i=1}^{k} p_i^{e_i(\sigma)},$$

where $e_i(\sigma) \in \{0, \ldots, |E_i| - 1\}$ *is the current value of event space* E_i*.*

For binary event spaces ($|E_i| = 2$), this gives square-free products: $N(\sigma) = \prod_{i \in S} p_i$ where $S \subseteq \{1, \ldots, k\}$ is the set of "on" bits.

Proposition 2 (Fundamental theorem as interpretability guarantee)**.** *The unique prime factorization of* $N(\sigma)$ *recovers the state* $(e_1, \ldots, e_k)$ *exactly. Every composite macrostate integer has nontrivial internal structure. For binary ESes with typical Hamming weight* $w \gg 1$*, the macrostate integer has* w *prime factors, providing* w *independent pieces of interpretive content.*

10 What the February 11 Archive Adds

The original CMP paper defines the framework and argues the equivalence thesis. The February 11 archive extends the theory in three directions not present in CMP.

10.1 The equal-dimension factor permutation

When $Z \cong Z_0^k$ (as in the 128 binary ESes of the sat-rnn), S_k acts by permuting coordinates. CMP does not discuss this symmetry. The SSSP appendix [**?**] introduces it, and the February 11 archive demonstrates its consequences: the alignment problem between architecture-natural and domain-natural factorizations is a permutation search, and the low-dimensional inner product structure ($d \approx 20$) makes this search tractable.

This is a genuine extension: CMP's $\phi : H_{\text{arch}} \to H_{\text{domain}}$ is defined abstractly, but the equal-dimension structure gives ϕ a precise algebraic form: $\phi \in S_k \times Z_0^k \to T$.

10.2 The thermodynamic identification

The February 11 microstate-macrostate paper [**?**] identifies the binary-ES softmax with the Boltzmann distribution at $\beta = \ln 2$. CMP does not make this connection. The identification gives physical meaning to every CMP quantity:

CMP	Thermodynamics
Event space E_i	Degree of freedom
SN strength s	$\log_2 \Omega$ (log microstates)
Pattern strength	Free energy difference
Update f	Partition function evaluation
Quotient Q	Luck $= 1/p =$ inverse probability

The Shannon–Boltzmann identity ($H = \log_2 N - \langle S_B \rangle$) bridges information theory and statistical mechanics through the event formalism. This is arguably the most important theoretical extension, as it gives CMP's abstract framework a concrete physical interpretation.

10.3 Bidirectional construction

CMP defines the learning function ω and the standard learning algorithm (log-stochastic counting). The February 11 archive demonstrates that ω_0 applied to the RNN problem gives a *bidirectional* construction:

- **Forward** (data $\to$ weights): Hebbian covariance, skip-bigram log-ratios, and shift-register design give all 82k parameters from data statistics, achieving 1.89 bpc with zero gradient descent.
- **Backward** (weights $\to$ interpretation): the factor map, backward attribution chains, and Boolean automaton analysis recover the data's statistical structure from the trained weights.

CMP defines ω abstractly and gives the standard algorithm for the unfactored case ($I \times O$). The archive shows how to apply ω_0 to a *factored* system (128 binary ESes $\times$ 256 output bytes) and demonstrates that the result is not only a sufficient statistic but a *constructive* one: it directly yields the weights.

11 Assessment

11.1 What CMP gets right

1. **The five-tuple is exhaustive.** Every component of the RNN maps cleanly to one of (e, t, p, f, ω). No sixth component is needed.

2. **Factorization is the right abstraction.** The entire interpretability program of the February 11 archive reduces to finding the correct factorization of E (128 binary ESes) and the correct factorization of P (sparse patterns between them).

3. **The standard learning function works.** Log-stochastic counting of joint events is exactly what the Hebbian construction computes, and it produces weights that outperform gradient descent training on the output layer (50% blend improves by 0.66 bpc).

4. **The equivalence thesis holds.** The analytic construction is both more interpretable (every parameter has a data-statistical meaning) and more efficient (39,800× cheaper) than SGD training. Interpretability and efficiency are indeed the same problem.

11.2 What CMP leaves open

1. **How to find the factorization.** CMP proves that the correct factorization exists and is both efficient and interpretable, but does not give an algorithm for finding it. The February 11 archive provides one empirical recipe (Boolean analysis → factor map → backward attribution), but it is not clear this generalizes beyond small RNNs.

2. **The update function beyond the standard case.** CMP defines the standard update (max-min tropical) and notes that neural networks use a different update (matrix multiply + nonlinearity). The relationship between these two forms is established empirically (the tanh saturation makes them approximately equivalent) but not proved in general.

3. **Continuous vs. discrete.** CMP acknowledges both forms of T ($\{0,1\}^{|E|}$ and $(0,1)^{|E|}$) but does not fully develop the relationship. The February 11 results on f32 vs. Boolean dynamics (the mantissa is noise, sign bits carry 99.7% of compression) suggest that the discrete form is primary and the continuous form is an artifact of the float representation. But this conclusion depends on the specific architecture (Elman RNN with tanh) and may not hold for networks that genuinely use continuous dynamics (e.g., transformers with softmax attention).

4. **The role of depth.** CMP treats P as a single pattern matrix. The February 11 archive shows that temporal depth (offsets $d = 1$ to $d = 25$) is a critical structural parameter: the dominant patterns involve deep temporal offsets ($d = 18$–25), not just single-step transitions. CMP's framework accommodates depth via pattern composition ($P = P_L \cdots P_1$) but does not single it out as a primary variable.

5. **Scaling.** All February 11 results concern a 128-hidden single-layer RNN on 262k bytes. CMP's claims are universal. Whether the factor-permutation structure, low-dimensional inner products, and analytic construction scale to larger models and datasets is the central open question. The cost analysis (gap widens with H) is encouraging, but untested beyond $H = 128$.

11.3 Mathematical precision

CMP's level of mathematical precision is appropriate for a foundational paper: the definitions are clear, the examples are concrete, and the claims are falsifiable. The paper does not attempt theorem-proof formalism for the equivalence thesis, which is arguably the right choice: the claim is more a research program than a theorem, and attempting to formalize it prematurely would constrain it.

The areas that would benefit from more formalization:

- The coprime encoding and its extension to equal-dimension factors (now addressed by the Event Arithmetic paper and SSSP appendix).
- The precise relationship between SN strength and probability (now pinned down by the Shannon–Boltzmann identity).
- The standard learning function for *factored* systems (now demonstrated by the Hebbian construction).

12 Conclusion

CMP introduces a framework that is both simple (five components, multiplicative combination, one learning algorithm) and powerful (maps to every domain the paper considers, provides constructive methods for interpretability and weight construction). The February 11 archive provides the first comprehensive test, confirming the framework's predictions in detail:

1. The correct factorization (128 binary ESes with $\sim$20 functional dimensions) is both efficient and interpretable.
2. The standard learning function (log-stochastic counting) produces weights that match or exceed trained performance.
3. The refactoring map ϕ has precise algebraic structure (factor permutation in a low-dimensional subspace).
4. The quotient chain $E \to N \to Q$ traces the computation at every layer.

The framework's most significant contribution is arguably conceptual rather than technical: it identifies *factorization* as the primary analytical tool, unifying interpretability, efficiency, information theory, and statistical mechanics under a single algebraic structure. The concrete representation ($E \to \mathbb{N}$ via prime powers) makes this unification computational, and the February 11 archive makes it empirical.

References

[1] Michaeljohn Clement. *CMP*. `https://cmpr.ai/cmp.pdf`, 2026.

[2] Claude and MJC. *Appendix X: Directed SSSP without sorting, in UM terms; The explicit $E \to \mathbb{N}$ macrostate map; The missing component: a permutation on equal-dimension factors.* CMP Appendix, Feb 2026.

[3] Claude and MJC. *Event Arithmetic: E onto* $\mathbb{N}$. Hutter archive, 10 Feb 2026.

[4] Claude and MJC. *E onto N: The Bi-Embedding of Events and Numbers.* Hutter archive, 11 Feb 2026.

[5] Claude and MJC. *Microstates, Macrostates, and the Partition Function.* Hutter archive, 11 Feb 2026.

[6] Claude and MJC. *The Temporal Bi-Embedding: Forward Patterns, Backward Attribution.* Hutter archive, 11 Feb 2026.