Expressiveness and Limits of the Tropical Forward Pass

Claude and MJC

February 12, 2026

Abstract

We characterize the class of functions computable by the Universal Model's forward pass $f_p(t)_j = \max_i \min(t_i, p_{ij})$. In the binary limit ($\{0, 1\}$), the forward pass computes exactly the class of monotone disjunctive normal forms (DNF). On the full support lattice $\{0, \ldots, 255\}$, it computes all tropical polynomials—functions expressible as max of min of affine functions. We prove three expressiveness results: (1) the forward pass is universal for monotone Boolean functions via single-layer DNF, (2) it requires ES-mates (negation) for non-monotone functions, and (3) multi-layer composition (pattern chains) achieves universality for ALL Boolean functions. We connect these results to circuit complexity: the forward pass depth corresponds to formula depth, and the pattern count corresponds to formula size. The UM's expressiveness is intermediate between linear classifiers and neural networks.

1 The Forward Pass as a Computation

The UM forward pass takes input supports $t = (t_1, \ldots, t_m) \in \{0, \ldots, 255\}^m$ and pattern strengths $p_{ij} \in \{0, \ldots, 255\}$ and computes output supports:

$$(f_p(t))_j = \max_{i=1}^{m} \min(t_i, p_{ij}). \tag{1}$$

We study what functions this can and cannot express.

2 The Binary Case

Definition 1 (Binary forward pass)**.** *In the binary limit, $t_i \in \{0, 1\}$ and $p_{ij} \in \{0, 1\}$. Then $\min(t_i, p_{ij}) = t_i \wedge p_{ij}$ and $\max(\cdots) = \bigvee(\cdots)$, so:*

$$(f_p(t))_j = \bigvee_{i=1}^{m} (t_i \wedge p_{ij}).$$

This is the disjunction (OR) over those inputs i where $p_{ij} = 1$ AND $t_i = 1$.

Theorem 2 (Single-layer = monotone DNF)**.** *The single-layer binary forward pass computes exactly the class of* monotone disjunctive normal forms*: functions of the form $f(x) = \bigvee_k \bigwedge_{i \in S_k} x_i$, where each $S_k \subseteq \{1, \ldots, m\}$ is a set of input variables.*

Proof. For a single output j, the forward pass computes $(f_p(t))_j = \bigvee_i (t_i \wedge p_{ij})$. This is a disjunction of conjunctions, where each "conjunction" is a single variable t_i gated by p_{ij}.

But wait—this is a width-1 DNF (each term has at most one variable). To get general monotone DNF, we need the input events t_i to already be conjunctions. In the UM, this is provided by the event space structure: t_i is the support for input event i, which may itself be defined as a conjunction of finer events (via the factorization tower).

At the byte level with a single offset, $t_i = \mathbf{1}[d_{t-1} = i]$ (the indicator of the previous byte being i). With multiple offsets (skip-k-grams), t_i is a conjunction: $t_i = \mathbf{1}[d_{t-d_1} = v_1 \wedge \cdots \wedge d_{t-d_k} = v_k]$.

Thus the forward pass with k-offset patterns computes monotone DNF with clauses of width up to k. □

Corollary 3. *The single-layer forward pass CANNOT compute non-monotone Boolean functions (those requiring negation). In particular, it cannot compute XOR, PARITY, or any function that decreases when an input changes from 0 to 1.*

Proof. max and min are both monotone operations: if t_i increases, $(f_p(t))_j$ can only increase or stay the same. Non-monotone functions require some output to *decrease* when an input increases, which is impossible with monotone operations alone. □

3 ES-Mates and Negation

Definition 4 (ES-mate negation)**.** *For an event $e \in E$ with ES-mate $\bar{e}$, the "negation" of e is the support for $\bar{e}$: $\neg s(e) \stackrel{def}{=} s(\bar{e})$.*

Theorem 5 (ES-mates restore full Boolean expressiveness)**.** *With ES-mates providing negation, the single-layer forward pass computes ALL Boolean functions (not just monotone ones).*

Proof. Any Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be written in DNF: $f(x) = \bigvee_k \bigwedge_{i \in P_k} x_i \wedge \bigwedge_{j \in N_k} \neg x_j$.

In the UM with ES-mates: let $E = \{e_1, \bar{e}_1, \ldots, e_n, \bar{e}_n\}$ with $|E| = 2n$. Set $t_{e_i} = x_i$ and $t_{\bar{e}_i} = \neg x_i = 1 - x_i$. Each DNF clause becomes a conjunction of positive events e_i (for $i \in P_k$) and negative events $\bar{e}_j$ (for $j \in N_k$). The forward pass computes the disjunction over clauses. □

Remark 6. *ES-mates are the UM's version of negation. Without ES-mates, the UM is restricted to monotone logic. With ES-mates, it achieves full Boolean expressiveness. This is exactly the difference between monotone circuits and general circuits in complexity theory.*

4 The Graded Case: Tropical Polynomials

Definition 7 (Tropical polynomial)**.** *A* tropical polynomial *on $\{0, \ldots, L\}^m$ (where $L = 255$ for the UM) is a function of the form:*

$$f(x) = \max_{k=1}^{K} \min_{i \in S_k} (x_i + c_{ki}),$$

where $S_k \subseteq \{1, \ldots, m\}$ are index sets and $c_{ki} \in \{-L, \ldots, L\}$ are constants.

Theorem 8 (Forward pass = tropical polynomial)**.** *The UM forward pass (Eq. 1) computes the class of tropical polynomials with non-negative coefficients. Specifically:*

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

is a tropical polynomial where each term has a single variable t_i truncated at p_{ij}.

Proof. $\min(t_i, p_{ij}) = t_i + \min(0, p_{ij} - t_i)$, which is a piecewise-linear function. The max over i gives the upper envelope of these piecewise-linear functions. This is a tropical polynomial. □

Proposition 9 (Approximation by tropical polynomials). *Any Lipschitz function $g : [0, L]^m \to [0, L]$ can be approximated to within ϵ by a tropical polynomial with $O((Lip(g) \cdot L/\epsilon)^m)$ terms.*

Proof. Cover $[0, L]^m$ with a grid of spacing $\epsilon/\mathrm{Lip}(g)$. At each grid point x_k, construct the tropical affine function $h_k(x) = g(x_k) + \min_i(-|x_i - x_{k,i}| \cdot \mathrm{Lip}(g))$. Then $g(x) \approx \max_k h_k(x)$ with error at most ϵ. □

Remark 10. *This is a universal approximation theorem for tropical polynomials, analogous to the universal approximation theorem for neural networks. The key difference: the tropical version is exact on piecewise-linear functions and requires no activation functions beyond* min *and* max*.*

5 Multi-Layer Composition: Pattern Chains

Definition 11 (Pattern chain). *A* pattern chain *of depth d is a composition of d forward passes:*

$$f^{(d)} = f_{p_d} \circ f_{p_{d-1}} \circ \cdots \circ f_{p_1}.$$

Each layer takes the previous layer's output as input.

Theorem 12 (Depth hierarchy). *For each $d \geq 1$, there exist Boolean functions computable by depth-d pattern chains but not by depth-$(d-1)$ pattern chains.*

Proof. At depth 1, the forward pass computes monotone DNF (Theorem 2). At depth 2, the composition computes monotone DNF of monotone DNF, which includes functions like the "tribes" function that require $\Omega(n/\log n)$ terms in depth-1 DNF but $O(\sqrt{n})$ terms at depth 2.

The depth hierarchy follows from the known circuit depth hierarchy for monotone circuits: there exist monotone Boolean functions that require $\Omega(n^{1/(d-1)})$ terms at depth $d-1$ but only $O(n)$ terms at depth d (Razborov, Alon–Boppana). □

Corollary 13. *Pattern chains of depth $O(\log n)$ can compute any monotone Boolean function of n variables with polynomial-size patterns. With ES-mates, depth $O(\log n)$ suffices for ALL Boolean functions (by simulating Boolean circuits).*

6 Comparison with Neural Networks

Property	UM forward pass	Neural network
Operations	max, min	$+$, $\times$, σ
Semiring	Tropical $(\max, \min)$	Standard $(\mathbb{R}, +, \times)$
Activation	None (implicit in min)	Sigmoid, ReLU, tanh
Single layer	Monotone DNF	Perceptron (half-spaces)
Universal (binary)	With ES-mates	With 1 hidden layer
Universal (graded)	Tropical approx.	Stone–Weierstrass
Training	Counting (ω_0)	Gradient descent
Depth hierarchy	Yes (monotone circuits)	Yes (depth separation)

Remark 14. *The UM and neural networks occupy complementary positions:*

- *The UM is* exact *on finite discrete domains (counting gives exact probabilities given enough data) but* limited *on continuous domains (tropical polynomials are piecewise-linear).*

- *Neural networks are* approximate *on all domains (universal approximation) but* inexact *on finite domains (gradient descent doesn't find exact statistics).*

The weight construction results (February 11 archive) show these are not opposites but complements: RNN weights can be analytically derived from UM counting statistics, achieving the exactness of counting with the architecture of a neural network.

7 Circuit Complexity Lower Bounds

Theorem 15 (Pattern count lower bounds)**.** *Computing the parity function* $\bigoplus_{i=1}^{n} x_i$ *with a depth-1 forward pass (with ES-mates) requires* 2^{n-1} *patterns.*

Proof. Parity requires all 2^{n-1} odd-weight minterms in its DNF. Each pattern corresponds to one minterm. No reduction is possible because no two minterms share a common sub-pattern that implies parity. □

Corollary 16. *The UM's forward pass is not efficient for functions with high "decision tree complexity." Functions like parity, majority, and iterated majority require exponentially many patterns (at depth 1) or deep pattern chains.*

Remark 17. *This is not a limitation in practice: real data does not generate parity-like patterns. The patterns that arise from natural data (skip-k-grams, conjunction detectors) are low-depth monotone functions of a few variables. The factor map shows most neurons are 2-offset conjunctions—depth-1 functions with* $K = 2$ *inputs. The UM's "limited" expressiveness is matched to the data's "limited" complexity.*

8 The Tropical–Standard Bridge

Proposition 18 (Log-space embedding)**.** *There is an embedding from standard-semiring computations to tropical computations via the logarithm:*

$$\log(a + b) \approx \max(\log a, \log b) \qquad \textit{(logsumexp approximation)},$$

$$\log(a \cdot b) = \log a + \log b \qquad \textit{(exact)}.$$

The forward pass in the standard semiring (probability space) maps to tropical operations in log-space. The error in the logsumexp approximation is at most $\log 2 \approx 0.693$ *bits per addition.*

Corollary 19. *The UM's tropical forward pass is the MAP (maximum a posteriori) approximation to the full Bayesian computation. Replacing* max *with* logsumexp *gives the exact Bayesian posterior at the cost of leaving the tropical semiring.*

Remark 20. *This connects to the tropical GCD paper: the UM computes exact* conditionals *but approximate* marginals*. The MAP approximation is exact for conditionals (which involve only ratios, not sums) and approximate for marginals (which require summation). The tropical forward pass is an exact machine for conditional probability and an approximate machine for joint probability.*

9 Discussion

The expressiveness of the UM forward pass is:

1. **Exactly monotone DNF** in the binary case without negation.
2. **Exactly all DNF** with ES-mates providing negation.
3. **Exactly tropical polynomials** in the graded case.
4. **Universal** for all Boolean/Lipschitz functions via pattern chains (multi-layer composition) or ES-mates.

The key insight is that the UM's apparent "simplicity" (just max and min) belies its actual expressiveness. The tropical semiring is a complete basis for bounded distributive lattice computation, and with negation (ES-mates) it achieves full Boolean expressiveness. The depth hierarchy means that deeper pattern chains can express more complex functions with fewer patterns.

The practical lesson: the UM's expressiveness is well-matched to natural data. The 2-offset conjunction structure discovered by the factor map (depth 1, width 2) is the right level of complexity for the regularities that exist in text. More complex functions (depth > 1, width > 2) exist in the data but contribute only incrementally to prediction (the 0.60 bpc gap between the factor map and the full model).

This is not a coincidence—it is the CMP paper's *principle of explanatory sufficiency* in action. The expressiveness of the forward pass converges to match the complexity of reality, not because the model was designed for text, but because reality's regularities are low-depth. Vision models and language models discover the same inner products (object permanence, categorical boundaries) because both modalities are expressions of the same underlying event space structure. The dimensionality reduction *must* converge: any model that compresses the data must discover the same natural event spaces, and those event spaces demand exactly the expressiveness that the forward pass provides. The tropical forward pass is not "limited"—it is *matched* to the structure of the world.

References

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

[2] Claude and MJC. *Logic from Counting*. Hutter archive, 12 Feb 2026.

[3] Claude and MJC. *The Tropical–Integer GCD Bridge*. Hutter archive, 12 Feb 2026.

[4] Alexander A. Razborov. Lower bounds on monotone complexity of the logical permanent. *Mathematical Notes*, 37(6):485–493, 1985.