# Sharpest-LPP Scoring in the N-gram UM

Threshold Creation, Three Scoring Modes

Claude and MJC — February 20, 2026

## 1. Setup

We test the settling conjecture's core prediction: sharpness preference (selecting the sharpest LPP output) should improve with more context.

The N-gram UM has event spaces for byte input/output (256 events each), previous bytes, envelope, and dynamic $k$-gram ESs. One output LPP per order maps from the context ES to byte_output, updated online via $\omega_0$ (log-stochastic counting).

**Output distribution.** Log-support values are converted to probabilities via softmax: $p(b) = 2^{s(b)} / \sum_b 2^{s(b)}$. For $s = 0$ (unseen): $2^0 = 1$ (baseline). For $s = 1$ (one observation): $2^1 = 2$. A single observed output gives the distribution $(2, 1, 1, \ldots, 1)$ over 256 bytes — classical Laplace smoothing.

**Sharpness.** The support gap $s_1 - s_2$: the difference between the top two support values in the LPP's output for the active source event. This is the UM-native measure (the ring pattern's $s^{(2)}$ from the settling paper), measurable entirely within the support lattice.

**Threshold creation.** Part of $\omega$: a $k$-gram must reach log-support $\geq 4$ ($\sim$16 observations) before an event is created for it. This prevents sparse neurons from existing. The creation rule is communicated to the UM runner as LPP metadata (since SN format encodes $(E, P, T)$ but not $\omega$).

## 2. Three Scoring Modes

- **(a) Online all**: predict-then-learn on entire data.
- **(b) Second half**: bits from position $n/2$ onward (generalization).
- **(c) Frozen rerun**: re-score the data with LPPs held constant. This is the Hutter Prize relevant experiment: the model has seen this data, now predict it with frozen weights.

## 3. Results

### 3.1. Threshold 4, 64K

| Order | Neurons | (a) Online | (b) 2nd half | (c) Frozen |
|---|---:|---:|---:|---:|
| 2 | 607 | 4.805 | 4.618 | 4.818 |
| 3 | 1,775 | 4.814 | 4.553 | 4.485 |
| 4 | 2,921 | 4.851 | 4.629 | 4.407 |
| 6 | 4,335 | 4.906 | 4.678 | 4.475 |

Frozen rerun improves from 4.818 (order 2) to 4.407 (order 4), a **0.41 bpc gain**. Order 6 frozen is slightly worse than order 4 (4.475 vs 4.407); the additional 6-gram neurons selected only 1.3% of the time are not yet helping.

Second-half generalization improves from order 2 to 3 ($-0.065$) but is noisy at higher orders. Online scores are flat across orders (∼4.8–4.9), dominated by the cold-start phase.

### 3.2. 100K, Order 6

| | (a) Online | (b) 2nd half | (c) Frozen |
|---|---|---|---|
| Order 6, 100K | 4.789 | 4.622 | 4.528 |

### 3.3. LPP Selection Frequencies (Order 6, 64K)

| LPP | Selected |
|---|---|
| unigram | 33.3% |
| marginal | 21.0% |
| bigram | 23.7% |
| trigram | 12.8% |
| 4-gram | 5.5% |
| 5-gram | 2.4% |
| 6-gram | 1.3% |

Unigram wins 33% of the time because when all LPPs have the same gap (typically $s_1 - s_2 = 1$), the lowest-order LPP wins by loop order. The sharpest-LPP selection is a conservative strategy: it only promotes a higher-order LPP when it has strictly more evidence separation than all lower orders.

### 3.4. Event Genesis

With threshold 4 at 64K order 6: 4,335 neurons created from 87,570 tracked $k$-grams. The 95% that never reach threshold are the sparse noise that threshold creation filters.

## 4. Conclusion

Sharpest-LPP scoring (external to $f$, using support gap $s_1 - s_2$) improves with more context, particularly in the frozen rerun mode that is relevant to the Hutter Prize. The best frozen result is **4.407 bpc** at order 4 (64K), vs. ∼5.2 bpc for the max-min forward pass. The 0.8 bpc gap between sharpest-LPP and max-min is the settling problem.

Models are exported as SN files and loadable via `umr run`.