The Rolv Sparse Memory Threshold (RSMT): A Universal Rule for Memory-Efficient Sparse Computation

Author: Rolv Eitrem Heggenhougen Affiliation: Independent Researcher / Inventor

Abstract

The Rolv Sparse Memory Threshold (RSMT) defines the exact density at which sparse matrix storage becomes more memory-efficient than dense storage. Modern AI workloads are constrained primarily by VRAM, making this threshold essential for scaling models and selecting optimal representations. RSMT provides a simple, mathematically derived, datatype-dependent rule that applies across all hardware platforms and sparse formats. This work formalizes a decision boundary that has long been missing in the field of sparse computation.

1. Introduction

Sparse computation is central to modern AI, yet the field has lacked a clear, quantitative rule for determining when sparse storage becomes more memory-efficient than dense. Engineers have relied on intuition, heuristics, or framework defaults, despite the fact that VRAM — not compute — is the dominant bottleneck in large-scale models.

I created the Rolv Sparse Memory Threshold (RSMT) to formalize this break-even point. RSMT provides a universal, hardware-agnostic rule that determines exactly when sparse storage uses less memory than dense. It is simple, deterministic, and applies to any datatype or framework. This threshold should have existed years ago; here, it is defined explicitly.

2. The RSMT Formula

Let:

- b = bytes per stored value
- i = bytes per index

The RSMT density threshold d is:

$$d = \frac{b}{b + i}$$

If the actual density of non-zero elements is below this threshold, sparse storage is more memory-efficient than dense.

3. Derivation

Dense storage requires:

$$D = m \times n \times b$$

Sparse storage (CSR approximation) requires:

$$S \approx \text{nnz} \times (b + i)$$

Substituting $\text{nnz} = d \cdot m \cdot n$ and solving for $S < D$:

$$d < \frac{b}{b+i}$$

This inequality defines the Rolv Sparse Memory Threshold.

4. Practical Thresholds

Table 1. Practical RSMT Thresholds for Common Datatypes

Value Type	Index Type	b (bytes)	i (bytes)	RSMT d	Interpretation
float32	int64	4	8	0.333	Use sparse if density < 33%
float16	int64	2	8	0.20	Use sparse if density < 20%
float32	int32	4	4	0.50	Use sparse if density < 50%

These thresholds make the tradeoff explicit and predictable.

5. Why RSMT Matters

Sparse compute is only useful when you know when to use it. RSMT provides a deterministic, mathematically correct rule for selecting between sparse and dense representations. It eliminates guesswork, scales across all hardware, and ensures that memory-critical workloads can take full advantage of sparsity.

6. Conclusion

RSMT is a foundational rule for modern AI memory economics. It is simple, universal, and long overdue. By formalizing the threshold, this work provides the field with a clear, reliable decision boundary that improves efficiency and enables larger, more capable models.