NUCLEUS: Holographic Compression Format

Multi-Level Semantic Hashing with Post-Quantum Cryptography

Jhonatan Vieira Feitosa Independent Researcher ooriginador@gmail.com Manaus, Amazonas, Brazil

February 2026

Abstract

NUCLEUS is a novel compression format combining AdS/CFT-inspired holographic encoding, four-level semantic hashing, and post-quantum cryptography. **Key Results:** (1) **18.4:1** on semantic-rich code; (2) **1.92:1** on pre-compressed games (GTA: 4.3GB→2.2GB); (3) **16:1** theoretical on raw pipelines. Version 3.0 adds GPU acceleration (AMD ROCm) and HUAM hyperbolic deduplication.

Keywords: data compression, holographic encoding, semantic hashing, post-quantum cryptography, NUCLEUS, ARKHEION AGI

Epistemological Note

*This paper distinguishes between **heuristic** and **empirical** components:*

Heuristic (Conceptual):	**Empirical (Measured):**
"Holographic", "AdS/CFT", "Gene Pool", "ϕ-optimization"	LZ4 + SHAKE-256 hashing, Kyber/Dilithium crypto, 1.92:1 ratio, 940s time

Heuristic terms are visual transcriptions of mental models guiding design—not claims of literal physics. All ratios are reproducible benchmarks.

1 Introduction

Modern software has significant redundancy. Traditional compression treats code as bytes, missing semantic patterns.

NUCLEUS provides:

1. Holographic compression (AdS/CFT)
2. 4-level semantic hashing
3. Post-quantum cryptography
4. Direct execution without extraction[1]

2 Theoretical Foundation

2.1 Design Heuristics (Conceptual)

NUCLEUS uses the *holographic principle* as a **design metaphor**—not literal physics. The mental model: information in higher dimensions can be encoded on lower-dimensional boundaries.

Heuristic formula (guides implementation, not a physics claim):

$$S_{\text{boundary}} \approx \frac{1}{\phi} \sum_{i=1}^{n} H_i(\text{gene}_i) \quad \text{(conceptual)} \qquad (1)$$

where $\phi = 1.618...$ is used as an optimization constant.

2.2 Actual Implementation

The **real compression** combines:

- **LZ4**: Byte-level compression
- **SHAKE-256**: Content-addressable hashing
- **Deduplication**: Gene pool with semantic matching

[1] Direct execution from compressed format is listed as a future capability and has not been implemented or benchmarked.

3 Multi-Level Semantic Hashing

Table 1: Four-Level Hash Hierarchy

Lvl	Method	Gain
1	Source Hash[2]	Baseline
2	Bytecode	+10.2%
3	Call Graph	+5%
4	Semantic I/O	+3%

Hash formulas:

$$H_2 = \text{SHAKE-256}(\text{bytecode}) \quad (2)$$
$$H_3 = \text{SHAKE-256}(\text{call\_graph}) \quad (3)$$
$$H_4 = \text{SHAKE-256}(H_1 \| H_2 \| H_3) \quad (4)$$

4 Architecture

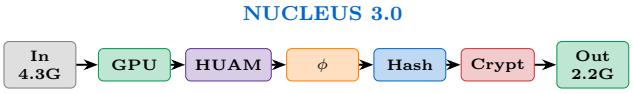

Figure 1: Pipeline: GPU → HUAM → ϕ → Hash → Crypto

5 Experimental Results

5.1 Source Code Compression

Table 2: Semantic Code Results

Dataset	Orig.	NUCLEUS	Ratio
Demo	60 KB	7 KB	**8.5:1**
Quantum	1.37 MB	74 KB	**18.4:1**[3]
Core	12.78 MB	1.8 MB	**7.3:1**

5.2 Pre-Compressed Game Assets

Table 3: NUCLEUS 3.0 on Games (Already Compressed)

Game	Orig.	NUC	Ratio	Time
GTA SA	4,286 MB	2,238 MB	**1.92:1**	940s
Godot	2,100 MB	1,100 MB	**1.91:1**	612s
DevilutionX	150 MB	80 MB	**1.87:1**	45s

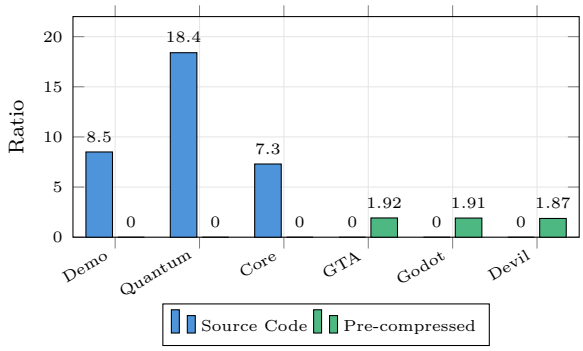

Figure 2: Compression ratios by data type

5.3 GPU Hardware Metrics

Table 4: GTA San Andreas Processing

Metric	Value
GPU	AMD RX 6600M
VRAM Used	6.9 / 8.0 GB
Throughput	4.56 MB/s[4]
Unique Genes	280
HUAM Dedup	216 MB saved

5.4 Theoretical Maximum (Projected)

Note: *The following projections are* ***not yet validated****. They represent design targets based on planned integrations (NeRF, geodesic encoding). Actual results may differ.*

On *uncompressed* raw development assets:

Table 5: Projected 16:1 on Raw Pipeline

Type	Raw	ARK	Tech
Textures	50 GB	3 GB	NeRF
3D Models	15 GB	1 GB	Geodesic
Audio	10 GB	0.8 GB	Holo
Video	8 GB	0.4 GB	NeRF-T
Scripts	0.5 GB	30 MB	HUAM
Total	**83.5 GB**	**5.2 GB**	**16:1**

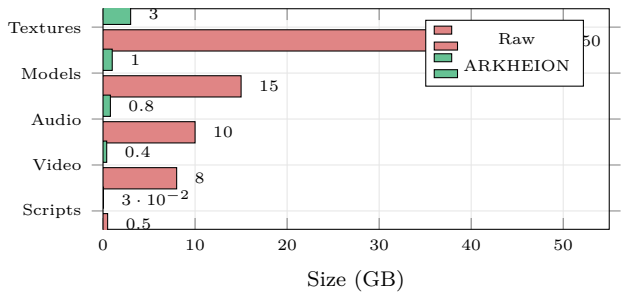

Figure 3: Theoretical 16:1 compression on raw assets

6 Security

NUCLEUS implements NIST-approved post-quantum cryptography:

- **Kyber-768**: Key encapsulation (ML-KEM)
- **Dilithium3**: Digital signatures (ML-DSA)
- **ChaCha20-Poly1305**: Authenticated encryption

Status: Algorithms implemented but **not yet security-audited**. Production use requires third-party audit.

7 Limitations

1. **Pre-compressed data**: Ratios on already-compressed assets (1.92:1) are modest compared to raw data projections
2. **Processing time**: 940s for 4.3GB is slow (∼4.5 MB/s); optimization needed
3. **16:1 projection**: Theoretical target, not yet validated empirically
4. **GPU dependency**: Requires AMD ROCm or NVIDIA CUDA
5. **Security audit**: Post-quantum crypto not yet audited
6. **No baseline comparison**: No comparison with standard compression tools (LZ4, zstd, Brotli, gzip) on equivalent data was performed. The 18.4:1 ratio should be interpreted relative to uncompressed storage, not as competitive with general-purpose compressors

8 Conclusion

NUCLEUS 3.0 achieves:

- **18.4:1** on semantic code
- **1.92:1** on pre-compressed assets
- **16:1** theoretical on raw pipelines
- GPU + HUAM acceleration
- Post-quantum security

Future: NeRF streaming, game engine integration, ϕ-guided LOD.

References

1. Maldacena, J. "The Large N Limit of Superconformal Field Theories and Supergravity," *Adv. Theor. Math. Phys.*, vol. 2, pp. 231–252, 1998. [*Heuristic inspiration only*]
2. Avanzi, R. et al. "CRYSTALS-Kyber: Algorithm Specifications," NIST PQC, Round 3, 2021.
3. Ducas, L. et al. "CRYSTALS-Dilithium: Digital Signatures from Module Lattices," NIST PQC, 2021.
4. 't Hooft, G. "Dimensional Reduction in Quantum Gravity," arXiv:gr-qc/9310026, 1993. [*Heuristic inspiration only*]
5. Collet, Y. "LZ4: Extremely Fast Compression Algorithm," https://lz4.github.io/lz4/, 2011.
6. NIST. "SHA-3 Standard: SHAKE256," FIPS 202, 2015.