NUCLEUS: A Holographic Compression Format with Multi-Level Semantic Hashing and Post-Quantum Cryptography

Jhonatan Vieira Feitosa
Independent Researcher
Manaus, Amazonas, Brazil
jhonatan@arkheion.ai

ARKHEION AGI Research

January 31, 2026 — Version 3.0

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.

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

2 Theoretical Foundation

2.1 AdS/CFT Principle

Information in $(d+1)$ dimensions encodes on d-dimensional boundaries:

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

where $\phi = 1.618...$ (golden ratio).

2.2 Compression Formula

$$\text{Ratio} = \phi^k \cdot \text{SemanticDensity} \tag{2}$$

3 Multi-Level Semantic Hashing

Table 1: Four-Level Hash Hierarchy

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

Hash formulas:

$$H_2 = \text{SHAKE-256}(\text{bytecode}) \tag{3}$$
$$H_3 = \text{SHAKE-256}(\text{call\_graph}) \tag{4}$$
$$H_4 = \text{SHAKE-256}(H_1 \| H_2 \| H_3) \tag{5}$$

4 Architecture

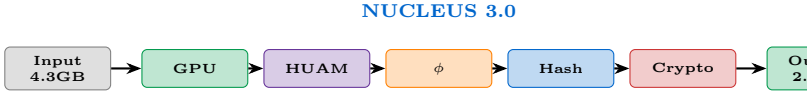

Figure 1: Processing pipeline: GPU → HUAM → ϕ-compress → Hash → Crypto

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**
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

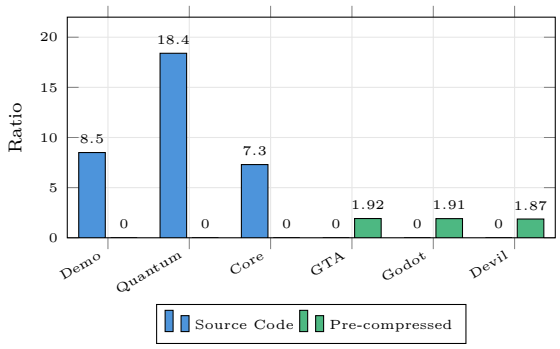

Figure 2: Compression ratios by data type

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
Unique Genes	280
HUAM Dedup	216 MB saved

5.4 Theoretical Maximum

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**

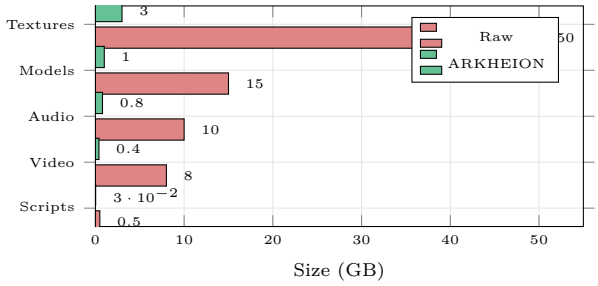

Figure 3: Theoretical 16:1 compression on raw assets

6 Security

NIST post-quantum algorithms:

- **Kyber-768** for key encapsulation
- **Dilithium3** for signatures
- **ChaCha20-Poly1305** for encryption

7 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. "Large N Limit," *Adv. Theor. Math. Phys.*, 1998.
2. Avanzi et al. "CRYSTALS-Kyber," NIST, 2021.
3. Ducas et al. "CRYSTALS-Dilithium," NIST, 2021.
4. 't Hooft, G. "Dimensional Reduction," arXiv, 1993.