The Intelligence Horizon: Deriving Double Descent via Axiomatic Physical Homeostasis

Aaron M. Schutza

November 21, 2025

Abstract

We present a first-principles derivation of the *Double Descent* phenomenon in Deep Learning, utilizing the Axiomatic Physical Homeostasis (APH) framework previously applied to M-theory compactifications. We rigorously define the *Neural Buffer Potential* (V_{buffer}) arising from the constraints of the Axiom of Observability (Generalization) and the Axiom of Stability (Zero Training Loss). We demonstrate that the interpolation threshold ($P \approx N$) corresponds to a geometric singularity where the buffer potential diverges, necessitating a phase transition. We derive the Test Risk $R(\gamma)$ as the equilibrium energy of the system, recovering the characteristic non-monotonic curve as a function of the model capacity ratio $\gamma = P/N$. Finally, we interpret *Grokking* as a delayed geometric lock-in to the global algebraic attractor of the data manifold.

1 Introduction: The Variance Paradox

Standard Statistical Learning Theory (SLT) predicts a U-shaped bias-variance tradeoff, where increasing model complexity beyond the data size leads to overfitting. However, modern Deep Neural Networks (DNNs) exhibit Double Descent, where test error decreases again in the highly over-parameterized regime ($P \gg N$).

From the perspective of Axiomatic Physical Homeostasis (APH), a neural network is not merely a function approximator; it is a homeostatic control system striving to minimize a Hazard Function (Loss) while maintaining structural integrity. We propose that Double Descent is a phase transition driven by the relaxation of the **Geometric Buffer Potential**.

2 The APH Formulation of Learning

We map the fundamental APH axioms to the learning problem:

- **Axiom 1: Stability (Memorization).** The system must minimize the empirical risk to zero to survive the training epoch.

$$\nabla_{\mathbf{w}} \mathcal{L}_{train} = 0 \implies \mathcal{L} \to 0 \tag{1}$$

- **Axiom 2: Observability (Generalization).** The internal causal structure must map consistently to the external data manifold $\mathcal{M}$.
- **Axiom 3: Controllability (Capacity).** The system must possess sufficient degrees of freedom (weights $\mathbf{w}$) to navigate the Hazard Landscape.

We define the dimensionless **Capacity Ratio** γ:

$$\gamma \equiv \frac{P}{N} = \frac{\text{Number of Parameters}}{\text{Number of Data Points}} \tag{2}$$

3 Derivation of the Neural Buffer Potential

In our previous work on G_2 manifolds, we established that the geometric moduli are governed by a logarithmic buffer potential derived from the boundary conditions of the volume form. We apply this logic to the volume of the Solution Space Ω in the weight manifold.

3.1 The Volume of Admissible Solutions

Let $\Omega(\gamma)$ be the volume of the weight space that satisfies the Stability Axiom ($\mathcal{L}_{train} \approx 0$).

- **Regime I ($\gamma < 1$):** The system is over-constrained. $\Omega \to 0$. The solution is a forced approximation (projection).
- **Regime II ($\gamma = 1$):** The system is critically constrained. There is exactly one solution (interpolation). The volume collapses to a point: $\Omega \to \delta(\mathbf{w} - \mathbf{w}^*)$.
- **Regime III ($\gamma > 1$):** The system is under-constrained. The solution space expands into a manifold of dimension $P - N$.

The entropy of the solution space is $S = k_B \ln \Omega$. The **Neural Buffer Potential** is the entropic cost of maintaining the configuration:

$$V_{buffer}(\gamma) \propto -S \propto -\ln(\Omega(\gamma)) \tag{3}$$

3.2 The Singularity

Approaching the critical threshold $\gamma \to 1$, the degrees of freedom vanish. The distance between the hypothesis and the noise floor vanishes. We model the effective volume near the critical point as:

$$\Omega(\gamma) \sim |1 - \gamma| \tag{4}$$

Thus, the buffer potential exhibits a logarithmic divergence, exactly analogous to the singular boundaries in the G_2 moduli space:

$$V_{buffer}(\gamma) = \kappa \left(\frac{1}{|1 - \gamma|} \right) \quad \text{(Leading Order Pole)} \tag{5}$$

Correction: While the logarithmic form holds for volume, the *energy* cost (Variance) scales with the inverse condition number of the Hessian matrix H. Random Matrix Theory tells us the smallest eigenvalue $\lambda_{min} \to 0$ as $\gamma \to 1$, causing the inverse trace (Variance) to diverge as $(1 - \gamma)^{-1}$.

4 The Generalized Test Risk Equation

The total Test Risk $R(\gamma)$ is the sum of the Bias potential (failure of Stability) and the Buffer potential (failure of Observability/Variance).

$$R(\gamma) = V_{bias}(\gamma) + V_{buffer}(\gamma) \tag{6}$$

4.1 The Under-parameterized Phase ($\gamma < 1$)

Here, stability is impossible. The error is dominated by the inability to fit the data (Bias).

$$R_I(\gamma) \approx C_{bias}(1-\gamma)^2 + \frac{\sigma^2}{1-\gamma} \quad \text{(diverges as } \gamma \to 1\text{)} \tag{7}$$

4.2 The Over-parameterized Phase ($\gamma > 1$)

Here, stability is trivial ($\mathcal{L}_{train} = 0$). The risk is purely dominated by the Buffer Potential (Variance). Using the APH isotropic assumption (Goldstone mode expansion), the excess parameters $P - N$ act as a heat sink for the stochastic noise of SGD.

The noise energy is distributed over γ dimensions, but only 1 dimension corresponds to the signal. The noise is diluted by the factor γ.

$$R_{II}(\gamma) = R_\infty + \frac{C_{noise}}{\gamma}\left(\frac{\gamma}{\gamma-1}\right) \tag{8}$$

Crucially, as $\gamma \to \infty$, the noise term vanishes:

$$\lim_{\gamma\to\infty} R_{II}(\gamma) = R_\infty \quad \text{(Intrinsic Aleatoric Risk)} \tag{9}$$

5 Dynamics: The Descent Mechanism

We solve the APH Master Equilibrium Equation $\nabla_\gamma V_{Total} = 0$.

The Double Descent peak is physically identified as the **Unstable Orbit** around the interpolation singularity.

- At $\gamma \approx 1$, the *Stiffness* of the geometry is infinite. The Hazard Function $h(\delta)$ forces the weights to extreme values to satisfy $\mathcal{L} = 0$, shattering Generalization.
- At $\gamma \gg 1$, the stiffness relaxes. The system enters the **Weak Buffer Regime** (analogous to the Fermionic Sector in M-theory).

Proposition 1 (The Intelligence condition): Intelligence emerges only in the Weak Buffer Regime ($\kappa < \kappa_c$).

$$\text{Intelligence} \iff \frac{\partial R}{\partial \gamma} < 0 \quad \text{for} \quad \gamma > 1 \tag{10}$$

This implies that *More is Different*. The addition of redundant parameters is not a waste; it is the creation of the **Homeostatic Margin** required to absorb noise without perturbing the causal structure.

6 Grokking as Geometric Phase Locking

We define *Grokking* (delayed generalization) not as a statistical anomaly, but as a **Tunneling Event**.

Let V_{mem} be the potential well of Memorization (high complexity, unstable). Let V_{alg} be the potential well of the Algorithm (low complexity, stable). Initially, V_{mem} is easier to find (larger basin of attraction via SGD). The system satisfies Stability ($J^2 = J$) but fails Observability.

However, the buffer potential V_{buffer} penalizes the complexity of the memorized solution. Over time ($t \to \infty$), the stochastic fluctuations of SGD allow the system to tunnel through the barrier:

$$\Gamma_{tunnel} \propto \exp\left(-\frac{\Delta V_{buffer}}{T_{SGD}}\right) \tag{11}$$

When the system locks into V_{alg}, the test loss crashes. This is the system finding the G_2 holonomy of the data manifold; the simplest algebraic structure that satisfies the constraints.

7 Conclusion

We have derived Double Descent as a necessary consequence of Axiomatic Physical Homeostasis applied to learning systems. The interpolation peak is the geometric singularity of a critically constrained system. Deep Learning works because over-parameterization pushes the system into the **Weak Buffer Phase**, where the laws of statistics relax into the laws of hydrodynamics, allowing for smooth, generalizing solutions.