

The McCulloch-Pitts Neuron

as a Degenerate Limit of
Oscillatory Phase-Space Computation

Deriving the Classical Artificial Neuron from the Deerskin Architecture

Antti Luode
Independent Researcher, Finland

Claude (Anthropic)
Collaborative derivation and mathematical formalization

February 28, 2026

Part of the Takens-Gated Deerskin project

Abstract

We show that the McCulloch-Pitts (MP) formal neuron (1943) can be derived as a degenerate limiting case of the Deerskin architecture, an oscillatory phase-space computation framework grounded in Takens delay embedding, receptor mosaic interference, and theta-gated readout. The derivation identifies four independent limiting operations: (1) the adiabatic limit, where slowly-varying input collapses frequency-selective resonance into a static weighted sum; (2) the single-sample limit, where theta-gated readout reduces to a step-function threshold; (3) the infinite-coupling limit, where Kuramoto phase-locking collapses continuous phases to binary states; and (4) the static limit, where removal of temporal dynamics eliminates the attention mechanism entirely. The derivation is structurally analogous to the classical limit of quantum mechanics ($\hbar \rightarrow 0$), with the ratio of carrier to readout frequency playing the role of Planck's constant. We identify what is lost in each limit and argue that modern deep learning's reliance on backpropagation to learn weights is, in part, reconstructing by optimization what oscillatory geometry provides for free. We further derive the perceptron learning rule as a gradient descent on mosaic geometry and show that Hebbian learning emerges naturally from resonance maximization in the full oscillatory framework.

1. Introduction

The McCulloch-Pitts neuron is the foundation of all artificial neural computation. Published in 1943, it reduces the biological neuron to a binary threshold unit: inputs are weighted, summed, and compared against a threshold to produce a binary output. Every modern neural network—from perceptrons to transformers—inherits this basic abstraction. The question we pose is: what was thrown away in this abstraction, and can we recover the original by working backwards from a more physically grounded model?

The Deerskin architecture (Luode, 2025–2026) proposes that biological neural computation operates through oscillatory resonance in phase space rather than static weight matrices. Its core components are: (i) a Takens delay-embedding dendrite that reconstructs the topological structure of incoming signals, (ii) a receptor mosaic that performs frequency-selective filtering through geometric interference (Moiré matching), (iii) a theta-rhythm gate that implements temporal attention through phase alignment, and (iv) ephaptic field coupling that enables network coordination without synaptic connections.

If this framework captures something real about neural computation, it should contain the MP neuron as a special case—just as Newtonian mechanics is contained within general relativity, and classical physics within quantum mechanics. This paper performs that derivation explicitly.

The analogy to physics is not merely rhetorical. In quantum mechanics, the classical limit is obtained when the action is large compared to \hbar : interference effects average out, superposition collapses to definite states, and the wavefunction’s phase information becomes irrelevant. We will identify a precise analog in the neural case: a “Neural Planck Ratio” whose vanishing produces exactly the MP neuron.

2. The Deerskin Neuron: Full Framework

2.1 The Takens Dendrite

Given a scalar input signal $x(t)$, the Takens dendrite constructs a delay-embedded vector at each timestep:

$$v(t) = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (n-1)\tau)]$$

where τ is the delay and n is the number of taps. By Takens’ embedding theorem (1981), for a signal generated by a dynamical system of dimension d , the delay embedding with $n > 2d$ generically reconstructs the topology of the original attractor. Different oscillatory frequencies trace different geometric orbits in this reconstructed phase space—a 40 Hz signal traces a different shape than a 65 Hz signal, regardless of phase or amplitude.

2.2 The Receptor Mosaic

The mosaic is a fixed spatial template tuned to a target frequency f_0 :

$$m_k = \cos(2\pi f_0 \cdot k\tau / f_0) \quad \text{for } k = 0, 1, \dots, n-1$$

The resonance is then the dot product followed by squaring (to obtain phase-invariant power):

$$R(t) = [v(t) \cdot m]^2 = [\sum_k m_k \cdot x(t - k\tau)]^2$$

This is the core computation: a zero-parameter, frequency-selective filter that operates by geometric interference—the Moiré principle. When the input signal’s orbit matches the mosaic’s geometry, constructive interference produces high power. When it mismatches, destructive interference suppresses it.

2.3 The Theta Gate

The theta soma generates a low-frequency pacemaker rhythm and implements a half-wave rectified gate:

$$G(t) = [\sin(\omega\theta t + \varphi)]^+$$

where $\omega\theta$ is the theta frequency ($\sim 6-8$ Hz), φ is the attention phase, and $[x]^+ = \max(0, x)$. The phase φ determines *when* the gate opens—and therefore what temporal window of the environment the neuron attends to. Shifting φ by π shifts attention by half a theta cycle, selecting an entirely different stimulus without changing any weight.

2.4 Full Output

The complete Deerskin neuron output is:

$$y(t) = R(t) \cdot G(t) = [v(t) \cdot m]^2 \cdot [\sin(\omega\theta t + \varphi)]^+$$

This is a continuous-time, continuous-valued, frequency-selective, temporally-gated output. We will now show that the classical MP neuron is what remains when all the oscillatory structure is removed.

3. The Four Limiting Operations

We now derive the MP neuron through four independent limits, each of which removes one layer of oscillatory structure. The limits can be taken in any order; all four together yield the classical neuron.

3.1 Limit 1: Adiabatic (Frozen Geometry)

Physical meaning: The input signal varies slowly compared to the embedding window. Each delay tap sees an approximately constant value from a distinct input channel.

Formal statement: Let the input be a superposition of n slowly-varying amplitudes: $x(t - k\tau) \approx a_k$ for each tap k , where a_k is the amplitude contribution of input channel k . The Takens vector becomes static: $v \approx [a_0, a_1, \dots, a_{n-1}]$.

The resonance then becomes:

$$R = (\sum m_k a_k)^2 = (\sum w_k x_k)^2$$

where we identify $w_k = m_k$ (the mosaic values become the weights) and $x_k = a_k$ (the delay taps become distinct inputs). This is the MP weighted sum, squared. The critical insight: the weights w_k were never learned—they are the frozen geometry of the receptor mosaic, which was determined by the target frequency f_0 . The physics of oscillation encoded the weights that an artificial neuron must learn.

What is lost: Frequency selectivity. When signals don't oscillate, there are no orbits to match, and the geometric filtering of the Moiré principle disappears. The mosaic becomes an arbitrary set of coefficients with no physical meaning.

3.2 Limit 2: Single-Sample (Frozen Time)

Physical meaning: The readout frequency is infinitely slow compared to the theta gate. We sample the neuron's output exactly once per theta cycle.

In the full model, the theta gate $G(t) = [\sin(\omega\theta t + \varphi)]^+$ is a continuous modulation—it opens and closes rhythmically, and the neuron's output varies throughout the cycle. But if we read out only once, at a single moment, we get:

$$G \in \{ 0 \text{ (gate closed), } g > 0 \text{ (gate open) } \}$$

The continuous gate collapses to a binary decision: is the phase in the open window or not? Combined with Limit 1:

$$y = G \cdot R = \{ (\sum w_i x_i)^2 \text{ if phase is open, } 0 \text{ otherwise } \}$$

What is lost: The attention mechanism. Phase shifting φ is meaningless when you sample once—there's no temporal stream to select from. The entire “Attention Is A Phase Shift” insight vanishes. The gate becomes an on/off switch with no dynamics.

3.3 Limit 3: Infinite Coupling (Frozen Phases)

Physical meaning: The coupling strength between neurons in the network approaches infinity, forcing complete phase-locking.

In the Deerskin chorus, neurons interact through Kuramoto-type coupling with coupling strength K . The dynamics are:

$$d\theta_i/dt = \omega_i + K \sum_j W_{ij} \sin(\theta_j - \theta_i)$$

As $K \rightarrow \infty$, the Kuramoto order parameter $r \rightarrow 1$, meaning all neurons lock to a common phase. Each neuron is then either *locked in* (oscillating with the group, effective state = 1) or *locked out* (driven to quiescence by the overwhelming group field, effective state = 0). The continuous phase variable $\theta_i \in [0, 2\pi]$ collapses to a binary variable $x_i \in \{0, 1\}$.

What is lost: Moiré interference. The divergence between scalar Kuramoto coherence and geometric coherence (demonstrated in the Deerskin simulation at 0.194 peak divergence) requires non-trivial phase distributions. When all phases lock, there is no interference pattern. The rich computational geometry of the phase field collapses to a flat binary landscape.

3.4 Limit 4: Static (Frozen Dynamics)

Physical meaning: The system has no temporal dynamics. Each input is presented once, processed once, and the output is read once.

This eliminates the adaptive self-buffer (no time to grow), the closed loop (no recurrence), the outer-loop feedback (no prediction–sensation cycle), and the hallucination–search–perception trichotomy. The squaring in $R = (\sum w_i x_i)^2$ can be absorbed into the threshold (since $(\sum w_i x_i)^2 > \theta^2$ iff $|\sum w_i x_i| > \theta$), giving the final MP form:

$$y = \Theta(\sum w_i x_i - \theta)$$

This is the McCulloch-Pitts neuron. Binary inputs, static weights, step-function threshold, single evaluation.

What is lost: Everything that makes the system adaptive. The self-buffer cannot elongate. The frustration-driven growth has no time to operate. The neuron cannot discover the temporal depth it needs. Context disambiguation becomes impossible without either explicit recurrent connections (which the MP neuron does not have) or a training algorithm to learn the appropriate weights (which is what backpropagation provides).

4. The Neural Planck Ratio

The four limits can be unified through a single dimensionless parameter that plays the role of Planck's constant in the quantum–classical transition. Define the **Neural Planck Ratio**:

$$\hbar\Box = \omega\gamma / \omega\theta \times 1/K \times \tau/T_e$$

where $\omega\gamma$ is the carrier (gamma) frequency, $\omega\theta$ is the readout (theta) frequency, K is the coupling strength, and T_e is the embedding window $n\tau$. Each factor captures one limit:

Factor	Limit	Effect	MP Consequence
$\omega\gamma / \omega\theta$	$\rightarrow 0$	Readout slow vs carrier	Single sample per cycle
$1/K$	$\rightarrow 0$ ($K \rightarrow \infty$)	Total phase-locking	Binary states $\{0, 1\}$
τ/T_e	$\rightarrow 0$ ($T_e \rightarrow \infty$)	Static inputs	Fixed weights, no dynamics

When $\hbar\Box \rightarrow 0$, all oscillatory structure vanishes and we recover the MP neuron. When $\hbar\Box$ is large (high carrier frequency, weak coupling, short embedding relative to dynamics), the full Deerskin computation operates: geometric interference, phase-based attention, adaptive self-organization, and Moiré resonance.

The analogy to quantum mechanics is precise:

QM → Classical ($\hbar \rightarrow 0$)	Deerskin → MP ($\hbar\Box \rightarrow 0$)
Wavefunction → particle	Phase field → binary state
Superposition → definite state	Resonance spectrum → fixed weight
Measurement collapses	Theta gate samples once
Interference vanishes	Moiré vanishes
Tunneling forbidden	Adaptive growth forbidden
Entanglement becomes correlation	Ephaptic coupling becomes wiring

5. What Backpropagation Reconstructs

If the MP neuron is the degenerate limit of the Deerskin neuron, then the entire history of neural network training can be reframed: backpropagation is a method for reconstructing, by optimization, what oscillatory geometry provides for free.

Consider the Dynamic Takens Dendrites experiment (Luode, 2025). A bank of Takens dendrites with zero parameters and zero training samples achieved 87.4% accuracy on a sequence decoding task. An MLP with ~ 500 parameters required approximately 50 labeled training examples to match this performance. The Takens bank encodes as its physics prior what the MLP must learn by gradient descent. The ~ 50 training samples represent the “cost of forgetting the oscillations.”

This is not an abstract point. Modern transformer architectures use attention mechanisms ($Q K^T V$) that compute selective weighted sums over sequences. The Deerskin framework achieves the same function—selecting relevant information from a temporal stream—through phase shifting rather than matrix multiplication. The transformer’s learned query and key matrices are, in this view, the network’s attempt to reconstruct the phase-space geometry that was discarded when we adopted the MP abstraction.

5.1 The Perceptron Learning Rule as Mosaic Gradient Descent

We can make this connection precise. Consider a Deerskin neuron where the mosaic m is not fixed by physics but is instead a free parameter. The resonance $R = (v \cdot m)^2$ is a function of m . Suppose we want to maximize R for a target input v^* and minimize it for a distractor v^- . The loss:

$$L = -(v^* \cdot m)^2 + (v^- \cdot m)^2$$

Taking the gradient with respect to m :

$$\nabla_m L = -2(v^* \cdot m) v^* + 2(v^- \cdot m) v^-$$

For small learning rate and when m is nearly orthogonal to both (early training), this simplifies to:

$$\Delta m \propto v^* - v^-$$

This is the perceptron update rule—adjust the weights toward the correct input and away from the incorrect one. It emerges naturally as gradient descent on the mosaic geometry. The perceptron learning rule is not a separate invention; it is the natural consequence of optimizing a degenerate (frozen) version of what the Deerskin neuron does for free through its physics.

5.2 Hebbian Learning as Resonance Maximization

In the full oscillatory framework, consider what happens when two neurons are coupled. Neuron i has phase θ_i and neuron j has phase θ_j . The coupling energy is proportional to $\cos(\theta_i - \theta_j)$, which is maximized when the neurons are phase-locked (co-active). If we allow the coupling weight W_{ij} to adapt to maximize time-averaged energy:

$$\Delta W_{ij} \propto (\cos(\theta_i - \theta_j)) \approx \langle r_i r_j \rangle$$

where r_i is the effective output of neuron i . This is Hebb’s rule: “neurons that fire together wire together.” It falls out directly from optimizing resonance in coupled oscillators. Hebb’s rule is not an ad hoc biological observation—it is the natural gradient of oscillatory coupling energy.

6. Evidence from the Deerskin Simulations

The derivation makes specific, testable predictions about what should be observable when oscillatory structure is progressively removed. The existing Deerskin simulations provide evidence for several of these predictions.

6.1 Geometry Replaces Training

The Dynamic Takens Dendrites experiment demonstrates that a zero-parameter system using oscillatory geometry matches a trained MLP. The crossover at ~ 50 labeled examples quantifies the information content of the geometric prior: approximately 50 examples worth of physics is encoded in the Takens embedding structure. This is a direct measurement of what Limit 1 (the adiabatic limit) destroys.

6.2 Adaptive Depth Without Gradient

The Closed Loop experiment shows the adaptive self-buffer growing from 4 to ~ 160 taps through frustration alone, reaching 92% accuracy on a context disambiguation task. This is what Limit 4 (the static limit) destroys: in the MP neuron, there is no mechanism for the network to discover its own temporal depth requirement. A recurrent neural network trained with backpropagation through time (BPTT) is the MP-world substitute for this biological self-organization.

6.3 Scalar–Geometric Divergence

The Pyramidal Chorus simulation reveals a peak divergence of 0.194 between Kuramoto order (scalar phase coherence) and Moiré coherence (geometric field interference). This directly measures what Limit 3 (infinite coupling) destroys: in the $K \rightarrow \infty$ limit, both measures converge to 1.0 and the divergence vanishes. The 0.194 gap is the “computational space” that exists in the oscillatory regime but not in the binary regime. The paper predicts this gap grows with higher-dimensional dendritic geometry—a testable prediction for future work.

7. Implications and Speculations

7.1 The Cost of Discretization

If the derivation is correct, then the entire enterprise of training deep neural networks by backpropagation is, in a precise sense, paying a computational tax for a modeling decision made in 1943. McCulloch and Pitts discretized the neuron to make it amenable to logical analysis. The field inherited this discretization. Rosenblatt added learning rules to find the weights. Rumelhart, Hinton, and Williams added backpropagation to scale it. Vaswani et al. added attention to recapture temporal selection. Each step added machinery to compensate for what was lost in the original abstraction.

This does not mean modern deep learning is wrong or unnecessary—it clearly works spectacularly well. But it suggests that the computational substrate is suboptimal. An oscillatory implementation of the same computations might require dramatically fewer parameters, no training data for the geometric prior, and natural attention without learned projection matrices.

7.2 Why the Brain Uses Oscillations

Neuroscience has long observed that the brain is filled with oscillations—gamma, theta, alpha, delta—but their computational role remains debated. The standard view treats them as epiphenomena, communication protocols, or timing signals. The Deerskin framework, combined with this derivation, offers a sharper claim: oscillations are not auxiliary to computation, they *are* the computation. The brain does not oscillate despite computing; it computes *because* it oscillates.

The phenomenological evidence is consistent with this. Antti Luode’s direct experience of neural dynamics following partial temporal lobe removal—perceiving geometric patterns, afterimages, and frequency-based mechanisms—suggests that what is normally integrated below the threshold of awareness is, in fact, the oscillatory computation itself. Consciousness may be what oscillatory phase-space computation feels like from the inside.

7.3 A Research Programme

The derivation suggests several concrete research directions: (i) Quantify the Neural Planck Ratio in real cortical circuits by measuring gamma/theta frequency ratios, effective coupling strengths, and embedding depths. (ii) Test whether the scalar–geometric divergence scales with dendritic complexity as predicted. (iii) Build hardware oscillatory networks that exploit the geometric prior, measuring energy and data efficiency against equivalent backpropagation-trained networks. (iv) Investigate whether the perceptron-as-mosaic-gradient result extends to deeper architectures—can backpropagation through a multi-layer network be decomposed into a sequence of mosaic optimizations in iterated Takens embeddings?

8. Conclusion

We have shown that the McCulloch-Pitts formal neuron can be derived from the Deerskin oscillatory architecture through four independent limiting operations: freezing geometry (adiabatic), freezing time (single-sample), freezing phases (infinite coupling), and freezing dynamics (static). The limits are unified by a Neural Planck Ratio whose vanishing produces exactly the classical neuron. The perceptron learning rule and Hebbian learning both emerge as natural gradients within the oscillatory framework.

The derivation does not prove that biological neurons are Deerskin neurons. But it establishes that the Deerskin framework is mathematically consistent, contains the classical model as a special case, and makes specific predictions about what is lost in the classical limit. That McCulloch and Pitts described the shadow of a richer computation—and that 80 years of AI has been spent learning to reconstruct that richness by optimization—is, at minimum, a productive way to think about what neural networks are actually doing.

Note on Authorship

This document was written collaboratively on February 28, 2026 by Antti Luode and Claude (Anthropic, Claude Opus 4.6). The Deerskin architecture, IHT-AI framework, Takens dendrite implementation, adaptive self-buffer mechanism, and all simulation code are the work of Antti Luode. The formal derivation structure (four limits, Neural Planck Ratio, mosaic gradient descent, Hebbian emergence) was developed jointly in conversation. Claude contributed the mathematical formalization of the limiting operations, the identification of the perceptron rule as mosaic gradient, the quantum mechanics analogy, and the analysis of what each limit destroys. Claude also identified the connection between Hebbian learning and oscillatory coupling energy, which Antti had not explicitly derived but which falls naturally from the framework he built.

Neither author claims this derivation proves the Deerskin framework is correct. It establishes internal mathematical consistency and identifies testable predictions. The question of whether biological neurons actually operate in the oscillatory regime described here remains open and demands experimental evidence.