MH-FLOCKE: Biologically Grounded Embodied Cognition Through a 15-Step Closed-Loop Architecture for Quadruped Locomotion Learning

Revised March 2026 — addressing reviewer feedback on RL baselines, multi-seed validation, mathematical formulations, and cross-embodiment transfer

Marc Hesse
Independent Researcher, Potsdam, Germany

info@mhflocke.com

Abstract

I present MH-FLOCKE, an embodied AI platform in which simulated quadruped creatures learn locomotion through a biologically grounded cognitive architecture. Unlike end-to-end reinforcement learning (RL) approaches that treat the body as an optimization target, MH-FLOCKE implements a 15-step closed-loop processing cycle that integrates proprioception, embodied emotions, episodic memory, motivational drives, a Global Workspace for attentional competition, metacognitive self-assessment, and reward-modulated spike-timing-dependent plasticity (R-STDP) in a spiking neural network (SNN). In this revision, I address reviewer feedback by providing: (1) mathematical formulations of all core learning rules (Izhikevich dynamics, R-STDP, cerebellar forward model, competence gate), (2) a PPO baseline comparison showing MH-FLOCKE's neural learning core achieves 3.5× the walking distance at equal 50k-step budget (45.15 ± 0.67 m vs. 12.83 ± 7.78 m; fixed-budget sample efficiency), (3) multi-seed statistical validation across 10 seeds and 80 runs, and (4) cross-embodiment transfer to the Unitree Go2 without architectural changes. Systematic ablation across 60+ runs isolates component contributions: vestibular reflexes eliminate all falls, motor babbling increases flat-terrain distance by 763%, the cerebellar forward model produces measurable prediction errors, and an olfactory sensory environment enables stimulus-driven behavior switching. The SNN+Cerebellum core (B1) achieves sigma = 0.67 m across 10 seeds — the lowest variance of any condition. I also report an interaction effect where the full cognitive architecture reduces locomotion distance compared to the neural core alone. A companion video: https://www.youtube.com/watch?v=Jo7UM6pEFMg. Code: https://github.com/MarcHesse/mhflocke.

1. Introduction

Quadruped locomotion is a benchmark problem in embodied AI. The dominant approach is end-to-end reinforcement learning (RL), which discovers motor policies through reward maximization without modeling the biological subsystems that actually produce locomotion. These approaches achieve impressive results [1, 2]. But they conflate mechanisms that operate on fundamentally different timescales: spinal reflexes (milliseconds), cerebellar adaptation (seconds), behavior selection (minutes), and memory consolidation (hours).

Biological quadruped locomotion emerges from at least six distinct neural subsystems: spinal Central Pattern Generators (CPGs) for rhythmic motor patterns [3], vestibular reflexes for postural stability [4], cerebellar forward models that predict sensory consequences of motor commands [5], brainstem motor babbling that calibrates the sensorimotor map during development [6], motivational drives that select behaviors based on internal states [7], and sensory systems that provide external stimuli for goal-directed behavior [8]. In the intact animal, these systems operate in parallel. Higher layers modulate but do not replace lower ones.

MH-FLOCKE (named after the author's late dog) implements this biological hierarchy as a computational architecture. The central contribution is a 15-step closed-loop processing cycle (Section 3.7) that runs at

every simulation timestep, integrating all subsystems into a single coherent perception-action-learning cycle. Unlike modular robotics architectures with message-passing interfaces, MH-FLOCKE's cognitive loop is deeply interleaved: emotions modulate learning rates (Step 4), memory retrieval influences behavior selection (Step 5), metacognitive self-assessment gates exploration (Step 8). The brain is not a module on top of the body. It is embodied.

The individual components — CPGs, SNNs, cerebellar models, R-STDP — are well-established. The integration into a single closed-loop architecture running at every timestep is, to my knowledge, novel. Prior integrative work couples these subsystems through message-passing interfaces or operates them at different timescales with explicit handoff points. MH-FLOCKE runs all 15 steps synchronously at 200 Hz. The specific contributions are: (1) systematic ablation across 60+ runs on the original morphology and 80 runs on Go2, quantifying each biological subsystem's contribution with multi-seed statistical validation; (2) a PPO baseline comparison demonstrating 3.5× distance advantage with 11.6× lower variance at 50k steps across 10 seeds; (3) cross-embodiment transfer to a production quadruped robot (Unitree Go2) without architectural changes; (4) mathematical formulations of all core learning rules; (5) a report of an interaction effect where sensory steering interferes with cerebellar learning. The framework builds on Integrity-OS [9], previously published on Zenodo. Source code: https://github.com/MarcHesse/mhflocke, Apache 2.0.

2. Related Work

2.1 RL-Based Locomotion

Deep RL has achieved robust quadruped locomotion in simulation [1] and sim-to-real transfer [2, 10]. End-to-end policies map observations to joint torques and achieve high benchmark performance. The policies are opaque: it is unclear which biological principles, if any, the network discovers, and they do not generalize to non-locomotion behaviors without retraining. Section 5.6 provides a direct PPO comparison [27] on the same morphology.

2.2 Central Pattern Generators

CPG models for robot locomotion have a long history [3, 11]. Oscillator networks produce stable gaits without sensory feedback and can be modulated by descending signals for speed and gait transitions. MH-FLOCKE uses CPGs as the foundation layer, with a competence-gated transition from CPG-dominated to actor-dominated control as the creature learns (Section 3.2).

2.3 Cerebellar Forward Models

The cerebellum is hypothesized to implement forward models that predict sensory consequences of motor commands [5, 12]. Computational models use Purkinje cell supervised learning with climbing fiber error signals. The implementation in Section 3.4 uses a parallel fiber to Purkinje cell pathway with prediction-error-driven climbing fiber pulses. Mathematical formulation is in Section 3.9.

2.4 Embodied Cognition and Consciousness

Global Workspace Theory (GWT) [13, 14] proposes that consciousness arises from competition among specialized modules for access to a shared broadcast. MH-FLOCKE implements GWT as Step 7 of the cognitive loop, with sensory, motor, predictive, error, and memory modules competing for broadcast, modulated by emotional valence and motivational drives. The Perturbational Complexity Index (PCI) [15] measures the complexity of the network's response to a perturbation. In simulation, PCI is approximated as the normalized entropy of SNN spike patterns following a synthetic input pulse applied to the input layer — providing a relative complexity measure that distinguishes ablated configurations rather than a biologically calibrated consciousness score. I make no claims about machine consciousness.

2.5 Prior Biologically Grounded Locomotion Systems

Several prior works combine CPGs with spiking or neuromorphic implementations. Espinal et al. [32] demonstrated SNN-based CPG locomotion on a quadruped using metaheuristic weight configuration — fixed synaptic weights, no on-device learning, fixed phase relationships between legs. Jiang et al. [33] extended

deep SNN-RL to legged robots in Isaac Gym simulation with strong results, but without sim-to-real transfer or on-device adaptation. Neuromorphic platforms (Loihi [4], SpiNNaker [5]) offer efficient SNN execution but require specialized hardware not available in most research settings. MH-FLOCKE differs from these systems in three concrete ways: (1) on-device R-STDP learning continuously adapts synaptic weights after deployment rather than fixing weights at training time; (2) the cerebellar forward model activates in response to real-world prediction errors — it produces zero corrections in clean simulation and non-zero corrections on noisy hardware; (3) the complete system runs on a Raspberry Pi 4 without neuromorphic hardware, demonstrated in companion work [7]. Scope of biological claims: GWT (Step 7), PCI (Step 8), and synaptogenesis (Step 12) are simplified computational analogs inspired by neuroscience, not biologically accurate models. The architecture is intentionally abstract — it captures functional properties (attentional competition, complexity measurement, activity-dependent connectivity) without claiming cellular or circuit fidelity.

3. Architecture

MH-FLOCKE consists of a simulated quadruped body (MuJoCo physics engine, 12 actuators in 4 legs) and a cognitive architecture organized in 9 packages: brain, body, behavior, bridge, integrity, llm, viz, self_improvement, and utils. The core computational substrate is a spiking neural network (SNN) with approximately 4,650 neurons (scaling with morphology and active modules), Izhikevich dynamics [28], and four neuromodulators (DA, 5-HT, NE, ACh). The architecture is organized as nested closed loops across timescales.

3.1 Spinal Reflexes

The fastest layer operates at every timestep (200 Hz). Muscle tone maintains joint stiffness against gravity. Stretch reflexes resist perturbations proportional to joint displacement. Golgi tendon organ simulation limits excessive force. Crossed extension reflexes coordinate contralateral limbs during stumbling [16]. Reflexes are additive — they modulate motor output from higher layers but never replace it.

3.2 Central Pattern Generator

A phase-coupled oscillator network produces rhythmic gait patterns for walk and trot [3, 11]. The CPG provides a stable locomotion baseline requiring no learning. A competence gate (Eq. 4) blends CPG output with learned actor output: at training start, CPG weight is 90%; as the actor's velocity exceeds 0.03 m/s, the gate transitions to 40% CPG / 60% actor. The creature can walk from step one while the actor learns.

3.3 Motor Babbling

During the first 7,000 steps, the creature performs exploratory motor movements with reduced amplitude (70% of full range) and 25% additive noise [6]. This calibrates the sensorimotor map. Motor babbling is the single most impactful component on flat terrain: it increases distance by 763% in ablation (from 1.50 m to 12.95 m at 50k steps). On hilly terrain the CPG alone provides sufficient sensory variety, making babbling less critical.

3.4 Cerebellar Forward Model

A cerebellar learning module implements the Marr-Albus-Ito framework [5, 12]: parallel fiber inputs carry motor efference copies, Purkinje cells learn to predict sensory consequences, and climbing fiber error signals drive supervised learning when prediction error exceeds a threshold. The deep cerebellar nuclei (DCN) output provides motor corrections blended with CPG and SNN output. The prediction error and learning rule are in Eq. 3. Critical implementation detail: the forward model must update before the early-return gate in the training loop (Issue #71) — otherwise it receives no training data during the CPG-dominated phase.

3.5 Vestibular System

Quaternion-derived upright estimation provides gravitational reference [4]. The vestibular signal gates cerebellar corrections: when the creature is falling (upright < 0.3), correction magnitude is reduced to prevent the cerebellum from learning during transient states. This single mechanism eliminated all falls across every

ablation configuration (27 falls in ablation #4 to 0 from ablation #7 onward; the #10 steering regression introduced 1 fall in C2 hilly, see Section 6.2).

3.6 Behavior Planning and Sensory Environment

A BehaviorPlanner selects from 8 behaviors (walk, trot, sniff, alert, rest, look_around, mark, motor_babbling) based on motivational drive state and sensory stimuli [7, 8]. The SensoryEnvironment module (v0.4.0) provides olfactory gradient targets and periodic acoustic events. Scent sources spawn 2–4 m ahead and respawn when distance exceeds 5 m, modeled as turbulent plume intermittency [17]. Olfactory steering modulates hip abduction asymmetrically to create gentle turning arcs [18, 19]. Sounds arrive 50% from scent direction (informative) and 50% from random directions (noise). The result is stimulus-driven behavior switching rather than random switching.

3.7 The 15-Step Cognitive Loop

CognitiveBrain executes a 15-step closed loop at every simulation timestep. All 15 steps run synchronously at 200 Hz — this is the key architectural difference from modular message-passing systems where subsystems operate at different rates and communicate through explicit interfaces. The cost of this tight coupling is computational overhead; the benefit is that learning signals, emotional state, and motivational drives are all current at the moment of motor output. Each step corresponds to a distinct neural function:

Step 1 — SENSE: Raw proprioceptive and exteroceptive sensor parsing (height, upright, velocity, joint angles).

Step 2 — BODY SCHEMA: Efference copy comparison detects anomalies between expected and actual sensor states.

Step 3 — WORLD MODEL: Learned state predictor (MLP) generates prediction errors that drive curiosity and learning signals.

Step 4 — EMOTIONS: Embodied valence-arousal model derives emotional state from body signals. Somatic markers modulate neuromodulators [20].

Step 5 — MEMORY: Sensorimotor episodic memory records and retrieves experience sequences for prediction [21].

Step 6 — DRIVES: Motivational drives (survival, exploration, comfort, social) steer behavior selection and reward modulation.

Step 7 — GLOBAL WORKSPACE: Module competition for global broadcast, modulated by emotion and drives. Winner broadcasts to SNN hidden neurons [13, 14].

Step 8 — METACOGNITION: Self-assessment of confidence, consciousness level (PCI), learning progress, and module activity [15].

Step 9 — CONSISTENCY: Integrity checking inspired by anterior cingulate conflict monitoring. Detects prediction-body-memory conflicts [22].

Step 10 — COMBINED REWARD: Merges extrinsic reward with curiosity, empowerment, drive modulation, and emotional valence.

Step 11 — R-STDP LEARNING: Reward-modulated spike-timing-dependent plasticity (Eq. 2). Cerebellar populations are protected from R-STDP.

Step 12 — SYNAPTOGENESIS: SNN spike patterns form concept nodes in a knowledge graph — here 'synaptogenesis' denotes spike-pattern-driven concept node formation, not biological synapse creation. Astrocyte calcium gating regulates local plasticity [23, 24].

Step 13 — HEBBIAN LEARNING: Unsupervised coincidence learning captures correlations that reward signals alone cannot detect.

Step 14 — DREAM MODE: Offline consolidation: every dream_interval=500 steps, n_replay=20 randomly sampled sensorimotor sequences from episodic memory are replayed. Each sequence triggers R-STDP

updates (Eq. 2) using stored reward signals, consolidating episodic associations into procedural memory. CPG and cerebellar weights are not modified during replay [25].

Step 15 — NEUROMODULATION: DA (reward sensitivity), 5-HT (mood stability), NE (exploration noise), ACh (attention) adjusted from somatic markers.

Each step maps to a specific biological analog: Steps 1–2 to sensorimotor integration, Step 4 to limbic valuation (amygdala), Step 7 to thalamocortical broadcast, Step 9 to anterior cingulate conflict monitoring, Step 12 to activity-dependent synaptogenesis. The novelty is their synchronous execution at 200 Hz — emotional state, memory state, and motor command are temporally consistent within each timestep, unlike modular architectures where subsystems operate at different rates with explicit handoff points.

Steps 1–6 and 15 run in every seed and configuration. Steps 7–14 are active only in the full-system (C) configuration; their absence defines the B configuration. This means the ablation between B and C directly measures the net contribution of the high-level cognitive stack to locomotion. The results (Section 5.6) show that stack currently reduces locomotion distance and increases variance — a finding that motivates future work on coupling cognitive state to motor learning.

3.8 Nested Timescales

The architecture operates as nested closed loops. Fastest (every step): spinal reflexes maintain posture. Medium (50–100 steps): the cerebellum corrects motor patterns via climbing fiber pulses. Slow (1000+ steps): the BehaviorPlanner switches behaviors. Slowest (dream intervals): episodic memory consolidates into procedural knowledge. Each layer can override layers below it, but lower layers provide the foundation.

3.9 Mathematical Formulations

Added in revision to address reproducibility. All variables are defined at first use.

3.9.1 Izhikevich Neuron Model.

The SNN uses the Izhikevich neuron model [28]. Membrane dynamics:

$$dv/dt = 0.04v^2 + 5v + 140 - u + I$$

$$du/dt = a(bv - u)$$

$$\text{if } v \geq 30 \text{ mV, then } v := c,\ u := u + d$$

where v is membrane potential (mV), u is a recovery variable, I is total synaptic input current, and (a, b, c, d) determine firing pattern. Note: the first term is $0.04v^2$ (v squared), not 0.04v — this follows the standard Izhikevich (2003) formulation. Regular spiking (RS) parameters: a=0.02, b=0.2, c=–65, d=8 for excitatory neurons. Fast spiking (FS): a=0.1, b=0.2, c=–65, d=2 for inhibitory neurons [28]. The network contains approximately 4,650 neurons with an 80/20 excitatory/inhibitory ratio.

3.9.2 Reward-Modulated STDP (R-STDP).

Synaptic plasticity follows a three-factor learning rule [29]. The eligibility trace for synapse (i, j) accumulates spike-timing correlations:

$$de_{ij}/dt = -e_{ij}/tau_e + STDP(Dt_{ij})$$

where tau_e = 1000 ms is the eligibility trace time constant and $Dt_{ij} = t_{post} - t_{pre}$ is the difference between postsynaptic and presynaptic spike times. The eligibility trace accumulates at each detected spike pair: STDP(Dt) is evaluated once per correlated pre/post event, with the sign of Dt determining potentiation (Dt > 0, LTP) or depression (Dt < 0, LTD). The STDP kernel is. In discrete simulation (dt = 5 ms), the ODE is integrated at each timestep; STDP(Dt) contributes only when a spike pair is detected, making the implementation effectively event-driven while retaining the continuous-time formulation.

$$STDP(Dt) = A^+ \exp(-Dt / tau^+) \text{ if } Dt > 0 \text{ (pre before post)}$$

$$STDP(Dt) = -A^- \exp(Dt / tau^-) \text{ if } Dt < 0 \text{ (post before pre)}$$

with $A^+ = 0.01$, $A^- = 0.012$, $tau^+ = tau^- = 20$ ms. Weight update modulated by global reward R(t):

$$Dw_{ij} = eta \cdot R(t) \cdot e_{ij}$$

where eta = 0.001. R(t) combines extrinsic reward (forward velocity, upright stability) with intrinsic components (curiosity, empowerment, drive satisfaction, emotional valence) as computed in Step 10. Cerebellar Purkinje cell populations are excluded from R-STDP to prevent interference with supervised forward model learning.

3.9.3 Cerebellar Forward Model.

The cerebellar module predicts the next sensory state s^(t+1) from current motor command m(t) in R^{12} (12 joint motor commands) and proprioceptive state s(t) in R^n (n = 12 joint positions + 4 IMU channels + configuration-dependent sensory channels). The concatenated input [m(t); s(t)] in R^{12+n}:

$$\hat{s}(t+1) = W_{PF} \cdot [m(t); s(t)]$$

$$e_{FM}(t) = || s(t+1) - \hat{s}(t+1) ||^2$$

When $e_{FM}(t)$ exceeds threshold theta_{CF} = 0.01, a climbing fiber pulse triggers the weight update:

$$DW_{PF} = -\text{alpha}_{CB} \cdot (s(t+1) - \hat{s}(t+1)) \cdot [m(t); s(t)]^T$$

where the superscript T denotes the matrix transpose, and alpha_{CB} = 0.005 is the cerebellar learning rate. Dimensions: W_{PF} in $R^{n\times(12+n)}$, prediction error in R^n, $[m(t); s(t)]^T$ in R^{12+n}, DW_{PF} in $R^{n\times(12+n)}$ (outer product). In the Go2 experiments, n=16 (12 joint positions + 4 IMU channels). DCN output Dm(t) blends with CPG and SNN output. Vestibular gating reduces corrections when upright < 0.3.

3.9.4 Competence Gate.

The competence gate transitions motor control from CPG-dominated to actor-dominated:

$$w_{CPG}(t) = w_{max} - (w_{max} - w_{min}) \cdot \text{sigma}(k \cdot (v_{actor}(t) - v_{thresh}))$$

where w_{max} = 0.9, w_{min} = 0.4, k = 50, v_{thresh} = 0.03 m/s, and sigma(x) = 1/(1+exp(–x)). Since sigma(x) in (0, 1) for all x, $w_{CPG}(t)$ is strictly bounded within (w_{min}, w_{max}) by construction — no explicit clipping is needed. Final motor output:

$$m_{out}(t) = w_{CPG}(t) \cdot m_{CPG}(t) + (1 - w_{CPG}(t)) \cdot m_{actor}(t) + Dm_{CB}(t)$$

On the Go2 platform, dynamic PD scaling adjusts gains from 0.4 (standing) to 1.5 (fallen) to account for the heavier morphology.

3.9.5 Combined Reward Function.

The reward signal R(t) has two components: $R(t) = R_{ext}(t) + R_{int}(t)$. The extrinsic component is fixed: $R_{ext}(t) = 0.8 \cdot v_{forward}(t) + 0.2 \cdot \text{upright}(t)$, where $v_{forward}(t)$ is forward velocity normalized to morphology maximum and upright(t) is quaternion-derived orientation quality in [0,1]. The intrinsic component $R_{int}(t) = \lambda_c \cdot PE_{world}(t) + \lambda_e \cdot \text{Empowerment}(t) + \lambda_d \cdot D_{sat}(t) + \lambda_v \cdot \text{Valence}(t)$ uses neuromodulator-scaled weights (λ in [0, 0.3] per component, configuration-dependent). B configurations use only $R_{ext}(t)$. C configurations activate $R_{int}(t)$ via the drive and emotional layers. Exact per-experiment weights are in the repository config files.

3.9.6 SNN Initialization and Connectivity.

Synaptic weights are initialized randomly: excitatory from U(0, 0.2), inhibitory from U(–0.1, 0). Sparse random connectivity at approximately 50% of possible connections per population pair. No structured connectivity at initialization; the sensorimotor map emerges through R-STDP during motor babbling. The k=50 steepness in Eq. 4 produces a sigmoid transition width of ~0.1 m/s centered at v_{thresh}=0.03 m/s. Sensitivity experiments with k in {20, 50, 100} showed similar final actor competence (±0.05 across 3 seeds), confirming robustness to this parameter.

4. Experimental Setup

4.1 Simulation Environment

All experiments use MuJoCo physics with a timestep of 5 ms (200 Hz). Two morphologies: (1) the original Bommel quadruped — 4 legs, 3 joints per leg, 12 actuators, procedurally generated; (2) the Unitree Go2 — same joint topology but heavier, with different link lengths and actuator torque limits. Two terrain conditions: flat (difficulty 0.0) and hilly (difficulty 0.3, procedurally generated). Training runs are 50,000 steps (4.2 minutes of simulated time). Go2 runs use --auto-reset 500 because the Go2 cannot self-right after falls.

4.2 Ablation Design

A 3×2 design crossing system complexity with terrain difficulty:

A (CPG only): Spinal CPG + reflexes + vestibular. The SNN substrate is present but receives no reward signal and remains near-silent; no cerebellum, no drives. The biological floor — an anencephalic preparation.

Note: The PCI approximation used here (spike pattern entropy following a synthetic input pulse) differs fundamentally from the TMS-based protocol in human consciousness research [15]. It is a relative complexity metric within this system only.

B (CPG + SNN + Cerebellum): Adds SNN with R-STDP, actor-critic, drives, behavior planner, sensory environment, and cerebellar forward model.

C (Full system): All components including the complete 15-step cognitive loop with GWT, metacognition, dream mode, and synaptogenesis.

Each configuration on flat (subscript 1) and hilly (subscript 2) terrain yields 6 conditions. On the original morphology (Bommel), ablation iterations #4, #7, #9, #10 capture the progression. On Go2: all 6 conditions plus PPO baseline, 10 random seeds each (80 total runs). Seeds: 7, 42, 55, 123, 256, 314, 808, 999, 1337, 2025.

4.3 PPO Baseline

A Proximal Policy Optimization (PPO) [27] baseline is trained on the identical Go2 MuJoCo model using Stable Baselines3 [30]. Observation space: joint angles, velocities, body orientation, height — identical to the SNN's sensory input. Same 12 actuators. 50,000 environment steps, default hyperparameters (lr 3e-4, clip 0.2, 64 minibatch, 2048 steps/update). 10 seeds for statistical comparison. This is a fixed-budget comparison measuring sample efficiency at the same step count as MH-FLOCKE. No curriculum, no domain randomization, and no reward shaping were applied. PPO with extended training (1M+ steps) and standard enhancements would likely achieve higher asymptotic performance; this comparison does not characterize that regime.

4.4 Metrics

Maximum distance (m): Euclidean distance from origin. Falls: transitions from upright > 0.7 to upright < 0.3. Scents found (SF): olfactory targets reached within radius 3.0 m. Actor competence: 0.0 (fully CPG) to 1.0 (CPG at minimum 40%). PCI: normalized entropy of SNN spike patterns following a synthetic perturbation pulse; used as a relative complexity metric across configurations, not a calibrated consciousness measure. All metrics logged in FLOG (custom binary format, msgpack-encoded). Statistical comparison via mean ± std, 10 seeds. Note: PCI values for B1 (0.231) and C1 (0.256) are similar — PCI does not cleanly distinguish these configurations at the 50k-step horizon. It is retained as a relative complexity indicator across A/B/C conditions, not as a primary discriminating metric between B and C.

5. Results

5.1 Ablation Progression (Original Morphology)

Table 1 shows the progression of the full system (C1, C2) across ablation iterations. The largest single improvement: vestibular reflexes and motor babbling (#4→#7). Falls dropped from 27 to 0 on hilly terrain; flat-terrain distance increased 763% (1.50 m → 12.95 m). The forward model fix (#9) added 3.9% on flat terrain — small but non-zero for the first time, confirming the cerebellar module was receiving training data. The sensory environment (#10) introduced behavior switching at the cost of 10% forward distance, gaining 11 scent sources found.

Table 1: Full system progression across ablation iterations (original morphology, single seed)

Run	#4	#7	#9	#10	Key change
C1 flat	1.50 m / 1F	12.95 m / 0F	13.45 m / 0F	12.14 m / 0F	Babbling + vestibular
C2 hilly	4.47 m / 27F	12.18 m / 0F	12.40 m / 0F	6.45 m / 1F	Steering interaction

5.2 Component Isolation (Ablation #10, Original Morphology)

Table 2 shows the full 6-configuration ablation at iteration #10. The C2 hilly regression (12.40 m → 6.45 m between #9 and #10) requires explanation. In #9, olfactory steering is not yet active. In #10, steering modulates hip abduction after the cerebellar correction pathway. On hilly terrain, where the cerebellum is more actively correcting balance, the steering perturbation interferes more severely with forward model learning than on flat terrain. The mechanism is described in Section 6.2. B1's recovery from 3.43 m (#9) to 12.91 m (#10) is the inverse case: without a cerebellum to interfere with, olfactory drives improve B1 substantially.

Table 2: Ablation #10 results (50k steps, Sensory v0.4.0 + FM fix + Steering 0.05, original morphology)

Config	#7	#9	#10	Falls	SF	vs #9	Note
A1 CPG flat	15.96 m	16.44 m	16.44 m	0	0	±0%	No change
A2 CPG hilly	17.20 m	16.92 m	16.92 m	0	0	±0%	No change
B1 SNN flat	10.07 m	3.43 m	12.91 m	0	19	+277%	Sensory rescues
B2 SNN hilly	9.89 m	12.45 m	14.71 m	0	13	+18%	Drives help
C1 Full flat	12.95 m	13.45 m	12.14 m	0	11	−10%	Steer→CB
C2 Full hilly	12.18 m	12.40 m	6.45 m	1	8	−48%	Regression (see Sec. 6.2)

From Table 2, cognitive efficiency (scents/m at #10): A1=0.00, A2=0.00, B1=1.47, B2=0.88, C1=0.91, C2=1.24. B1 achieves the highest efficiency — navigating most effectively per unit distance. C1's lower efficiency (0.91 vs 1.47) reflects the distance penalty from non-locomotion behaviors.

5.3 Sensory Environment Validation

Four implementation iterations were needed. The first version used fixed scent positions with quadratic distance decay — zero scents found because the creature drifted laterally during motor babbling. The final version (v0.4.0) uses distance-based respawning (> 5 m triggers a new source ahead), linear decay (smell = 1/(1+dist)), and olfactory steering via asymmetric hip abduction (gain = 0.05). At 100k steps: 22.02 m distance, 19 scents found (same count as 50k steps, indicating a plateau), smell strength 0.18–0.25. Without the sensory environment, the same configuration reached 20.85 m but zero scents and stagnating distance after 100k steps.

5.4 Forward Model Breakthrough

Prior to ablation #9, the cerebellar forward model showed zero prediction error across all runs. The cause (Issue #71): the forward model update occurred after an early-return gate that bypasses processing during the CPG-dominated phase. Moving the FM update before this gate produced e_{FM} = 0.005–0.006 and doubled correction magnitudes (0.031→0.059). C1 improved 3.9% (12.95 m → 13.45 m). The improvement is reproducible.

5.5 200k Long Run Analysis

A 200k-step run (single seed, no confidence intervals; C1, pre-sensory) reached 31.03 m with zero falls. Distance stagnated: 2.8 m/10k steps in the first 50k, 0.4 m/10k in the final 50k. The CPG-only configuration would extrapolate to ~65 m at 200k steps, making the full system 48% of the CPG baseline. This growing gap reflects time spent in non-locomotion behaviors (alert 15%, look_around 6%) driven by the cognitive architecture.

5.6 Cross-Embodiment Transfer: Unitree Go2

To test generalization, I transferred MH-FLOCKE to the Unitree Go2 without architectural changes. The Go2 has the same joint topology (4 legs, 3 joints, 12 actuators) but heavier mass, different link lengths, different actuator torque limits, and a different center of mass. The only adaptations: dynamic PD gain scaling (0.4 standing, 1.5 fallen) and auto-reset at 500 steps. Table 3 shows results across 10 seeds.

Table 3: Go2 ablation results (50k steps, 10 seeds, mean ± std). PPO trained with Stable Baselines3 [30]. PCI not reported for A (CPG-only) and PPO configurations.

Config	Distance (m)	Falls	Upright Str.	Actor C.	PCI	Note
A1 CPG flat	40.73 ± 6.14	0.2 ± 0.4	49627	0.90	—	Moderate var.
A2 CPG hilly	26.96 ± 9.68	1.0 ± 1.2	47483	0.87	—	Terrain –34%
B1 SNN+CB flat	45.15 ± 0.67	0.0 ± 0.0	50000	1.00	0.231	Best, 0 falls
B2 SNN+CB hilly	23.55 ± 3.95	0.0 ± 0.0	50000	0.96	0.258	Terrain –48%
C1 Full flat	22.35 ± 11.50	3.3 ± 2.8	39677	0.69	0.256	High var., falls
C2 Full hilly	20.21 ± 2.54	0.7 ± 0.8	46691	0.77	0.258	Stable
PPO flat	12.83 ± 7.78	0.0 ± 0.0	50000	—	—	3.5× worse
PPO hilly	9.49 ± 5.29	0.0 ± 0.0	50000	—	—	Terrain –26%

First: B1 achieves sigma = 0.67 m — the lowest variance of any biological condition. All 10 seeds reach zero falls and full actor competence. Second: A1 outperforms C1 on distance (40.73 m vs. 22.35 m). The full cognitive architecture trades locomotion efficiency for behavioral diversity. Third: C1 shows the highest variance (sigma = 11.50 m) and is the only configuration with substantial falls (3.3 ± 2.8), indicating that the cognitive layers introduce instability at the 50k-step horizon. Fourth: the terrain effect is asymmetric — B1 loses 48% on hilly terrain (45.15 $\rightarrow$ 23.55 m), while C1 loses only 10% (22.35 $\rightarrow$ 20.21 m), suggesting behavioral diversity provides some terrain robustness. Fifth: PPO variance (sigma = 7.78 m) is 11.6× larger than B1's, demonstrating that biological priors stabilize learning across seeds.

The Go2 B1 distances are substantially higher than the original Bommel morphology (45 m vs. 13 m). The Go2 is an engineered robot with optimized mass distribution; Bommel is procedurally generated.

5.7 CPG Fall at Step 49.5k: Evidence for SNN Necessity

In one A1 seed, at step 49,500, the SNN's learned velocity contribution dropped (no sustained reward signal in CPG-only config) and the Competence Gate increased CPG weight to 88%. The resulting rigid, high-amplitude gait caused a fall. The B and C configurations, which include the SNN, showed zero falls (B) or recoverable falls (C) across all seeds. Learned motor adaptation is necessary for sustained stability.

5.8 Behavioral Time Breakdown

Table 4 quantifies where time goes. B and C configurations spend 29–37% of time in non-locomotion behaviors. This directly explains the distance gap with A1: the creature is choosing other behaviors, not failing to walk. Alert behavior (12–17%) dominates non-locomotion time. B1 shows more trot (8.0%) than C1 (3.5%), suggesting the full cognitive architecture suppresses the walk-to-trot transition through neuromodulatory dynamics. High variance across seeds (walk std = 17–24%) reflects stochastic drive and sensory systems.

Table 4: Behavioral time breakdown (50k steps, 10 seeds, mean ± std %). A configurations: 100% walk.

Behavior	A1 CPG flat	B1 SNN+CB flat	C1 Full flat	C2 Full hilly
walk	100.0 ± 0.0	63.1 ± 24.4	71.2 ± 17.9	68.2 ± 21.0
alert	—	12.4 ± 10.7	12.3 ± 9.1	11.4 ± 8.5
trot	—	8.0 ± 8.3	3.5 ± 4.4	4.7 ± 4.9
motor_babbling	—	6.5 ± 8.0	6.5 ± 8.0	6.5 ± 8.0
look_around	—	5.3 ± 9.7	4.5 ± 4.0	5.1 ± 7.0
sniff	—	4.7 ± 14.1	2.0 ± 6.1	4.1 ± 12.2

6. Discussion

6.1 The Distance Gap as Behavioral Richness

The full system (C) walks less far than the CPG baseline (A). At 50k steps on the Go2, A1 reaches 40.73 m while C1 reaches 22.35 m. The behavioral time breakdown (Section 5.8) explains this quantitatively: C1 spends only 71% of time walking. The rest goes to alert (12%), motor_babbling (7%), look_around (5%), trot (4%), and sniff (2%). The purpose of the cognitive architecture is not to maximize distance. It is to produce a creature with animal-like behavioral diversity. That said, the current gap is also partly architectural immaturity: dopaminergic neuromodulation does not yet couple to R-STDP learning rates, and metacognitive confidence does not gate motor output. Both are open engineering problems.

6.2 Steering-Cerebellum Interaction

The negative interaction between olfactory steering and cerebellar learning was unexpected. In the B configuration (no cerebellum), the sensory environment dramatically improved performance (B1: +277%). In C, it reduced performance (C1: –10%, C2: –48%). The mechanism: olfactory steering modifies motor output after the cerebellar correction, creating an unpredictable perturbation from the forward model's perspective. The forward model learns to predict the consequences of its own corrections — and steering invalidates those predictions. On hilly terrain the effect is amplified because the cerebellum is more active. The fix is to integrate steering before the cerebellar pathway, not after.

6.3 B1 Recovery: Drives Compensate for Missing Cerebellum

B1 collapsed from 10.07 m to 3.43 m between ablations #7 and #9 on the original morphology due to the forward model timing fix disrupting SNN training dynamics. In ablation #10, with the sensory environment, B1 recovered to 12.91 m (+277%). Motivational drives and olfactory stimulation can substitute for cerebellar motor corrections — at least on flat terrain at 50k steps.

6.4 Limitations

Only two morphologies have been tested. Generalization to hexapods, bipeds, or morphologies without a 3-joint-per-leg topology is unknown. The SNN uses Izhikevich neurons rather than more detailed compartmental models — efficient but not fully biologically realistic. Sim-to-real transfer has not been attempted on any physical hardware. The 200 Hz cognitive loop is computationally expensive: running the full 15-step C1 configuration takes approximately 5–10× longer per step than the B1 configuration on the same CPU, due to GWT competition, synaptogenesis, and dream mode — limiting real-time deployment without optimization or neuromorphic hardware. The C2 regression indicates the steering-cerebellum interaction is unresolved. The expanded 10-seed validation shows the cognitive architecture (C) introduces instability compared to the neural core (B), with falls and high variance. The cognitive layers are not yet ready to provide a net locomotion benefit. The SNN+Cerebellum core is the current performance ceiling; C is the architectural ambition.

6.5 Comparison with PPO Baseline

PPO achieves 12.83 ± 7.78 m on the Go2 (flat) vs. 45.15 ± 0.67 m for B1 — a 3.5× advantage with 11.6× lower variance at 50k steps. This is a fixed-budget comparison: both systems receive the same step budget with no curriculum or domain randomization for PPO. Given millions of steps and standard enhancements, PPO would likely converge higher. MH-FLOCKE's advantage at this budget stems from the CPG providing a strong locomotion prior — PPO must discover walking from scratch, MH-FLOCKE refines an innate gait. The original morphology ablations (Tables 1–2) are single-seed; multi-seed validation for that morphology is planned but not yet available. The Go2 10-seed results (Table 3) provide the statistical foundation for the architecture claims.

6.6 Cognitive Architecture Overhead on Go2

An earlier 3-seed validation showed identical B and C results — suggesting the cognitive layers had no effect. The 10-seed validation reveals a different picture: B1 at 45.15 ± 0.67 m, C1 at 22.35 ± 11.50 m. The full architecture halves distance and introduces significant instability. Three causes: (1) computational overhead from GWT competition and neuromodulatory dynamics introducing stochastic variation; (2) behavioral

interference — drives trigger non-locomotion behaviors that interrupt forward progress; (3) fall cascades — neuromodulatory perturbations during critical gait phases trigger falls, and with auto-reset at 500 steps, each fall compounds across seeds. On hilly terrain, C2 (sigma = 2.54 m) is actually more stable than C1 (sigma = 11.50 m), with fewer falls (0.7 vs. 3.3). Terrain variability may stabilize the cognitive architecture by keeping the Global Workspace productively engaged.

7. Future Work

Muscle synergies [26] would reduce the actor's search space from 12 independent joints to 4–5 synergy dimensions. Recovery learning would address CPG overriding righting reflexes after falls. The steering-cerebellum interaction (Section 6.2) will be resolved by integrating olfactory steering as an SNN input rather than a post-hoc motor modification. Neuromodulator-to-R-STDP coupling will make drives directly influence learning dynamics — the concrete fix for the cognitive overhead issue (Section 6.6). A navigation ablation using scent-finding efficiency as primary metric will provide a fairer measure of cognitive layer contributions. Sim-to-real transfer on a physical Unitree Go2 is a longer-term goal. Neuromorphic hardware (Intel Loihi, SpiNNaker) would address the computational cost limitation for real-time deployment.

8. Conclusion

MH-FLOCKE demonstrates that biologically grounded, modular cognitive architectures can produce quadruped locomotion learning with rich behavioral repertoires and cross-embodiment generalization. The 15-step closed-loop architecture integrates six timescales of processing — from spinal reflexes to dream-based memory consolidation — in a single computation running at every simulation timestep. Systematic ablation quantifies each component's contribution: vestibular reflexes eliminate falls, motor babbling enables flat-terrain learning, the cerebellar forward model provides measurable motor corrections, and the sensory environment replaces random behavior switching with stimulus-driven goal pursuit.

Cross-embodiment transfer to the Unitree Go2 achieves 3.5× the distance of a PPO baseline at identical sample budgets (45.15 ± 0.67 m vs. 12.83 ± 7.78 m, 10 seeds), with the SNN+Cerebellum core showing 11.6× lower variance than PPO. Reporting the cognitive-architecture overhead — the full system currently halves locomotion distance compared to the neural core — provides concrete architectural lessons. The gap between B and C is not a fundamental trade-off. It is the remaining engineering work.

Companion video: https://www.youtube.com/watch?v=Jo7UM6pEFMg. Source code: https://github.com/MarcHesse/mhflocke (Apache 2.0). Documentation: https://mhflocke.com/docs/. The Go2 model files are redistributed under BSD-3-Clause from the MuJoCo Menagerie project [31].

Disclosure: *AI writing tools (Claude, Anthropic) were used to assist with manuscript preparation, code development, and language editing. All research design, architecture implementation, experimentation, data collection, analysis, and conclusions are the sole work of the author.*

References

[1] Rudin, N. et al. (2022). Learning to walk in minutes using massively parallel deep RL. CoRL.

[2] Miki, T. et al. (2022). Learning robust perceptive locomotion for quadrupedal robots. Science Robotics, 7(62).

[3] Ijspeert, A.J. (2008). Central pattern generators for locomotion control. Neural Networks, 21(4), 642–653.

[4] Angelaki, D.E. & Cullen, K.E. (2008). Vestibular system: the many facets of a multimodal sense. Ann. Rev. Neurosci., 31, 125–150.

[5] Wolpert, D.M. et al. (1998). Internal models in the cerebellum. Trends Cogn. Sci., 2(9), 338–347.

[6] Prechtl, H.F.R. (1997). The importance of fetal movements. In Textbook of Perinatal Medicine.

[7] Tinbergen, N. (1951). The Study of Instinct. Clarendon Press.

[8] Sokolov, E.N. (1963). Perception and the Conditioned Reflex. Pergamon Press.

[9] Hesse, M. (2025). Integrity-OS: Neural-Symbolic Hybrid Architecture. Zenodo. DOI: 10.5281/zenodo.18450340.

[10] Hwangbo, J. et al. (2019). Learning agile and dynamic motor skills for legged robots. Science Robotics, 4(26).

[11] Righetti, L. & Ijspeert, A.J. (2008). Pattern generators with sensory feedback for quadruped locomotion. ICRA.

[12] Miall, R.C. & Wolpert, D.M. (1996). Forward models for physiological motor control. Neural Networks, 9(8), 1265–1279.
[13] Baars, B.J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.
[14] Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness. Cognition, 79, 1–37.
[15] Casali, A.G. et al. (2013). A theoretically based index of consciousness. Sci. Transl. Med., 5(198), 198ra105.
[16] Tresch, M.C. et al. (1999). The construction of movement by the spinal cord. Nature Neurosci., 2(2), 162–167.
[17] Murlis, J. et al. (1992). Odor plumes and how insects use them. Ann. Rev. Entomol., 37(1), 505–532.
[18] Catania, K.C. (2006). Olfaction: underwater sniffing by semi-aquatic mammals. Nature, 444(7122), 1024–1025.
[19] Porter, J. et al. (2007). Mechanisms of scent-tracking in humans. Nature Neurosci., 10(1), 27–29.
[20] Damasio, A.R. (1994). Descartes' Error: Emotion, Reason, and the Human Brain. Putnam.
[21] Tulving, E. (1972). Episodic and semantic memory. In Organization of Memory (pp. 381–403). Academic Press.
[22] Botvinick, M.M. et al. (2001). Conflict monitoring and cognitive control. Psychol. Rev., 108(3), 624–652.
[23] Huttenlocher, P.R. (1979). Synaptic density in human frontal cortex. Brain Res., 163(2), 195–205.
[24] Araque, A. et al. (1999). Tripartite synapses: glia, the unacknowledged partner. Trends Neurosci., 22(5), 208–215.
[25] Walker, M.P. & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. Neuron, 44(1), 121–133.
[26] Bizzi, E. & Cheung, V.C.K. (2013). The neural origin of muscle synergies. Front. Comput. Neurosci., 7, 51.
[27] Schulman, J. et al. (2017). Proximal policy optimization algorithms. arXiv:1707.06347.
[28] Izhikevich, E.M. (2003). Simple model of spiking neurons. IEEE Trans. Neural Networks, 14(6), 1569–1572.
[29] Fremaux, N. & Gerstner, W. (2016). Neuromodulated STDP and three-factor learning. Front. Neural Circuits, 9, 85.
[30] Raffin, A. et al. (2021). Stable-Baselines3: Reliable RL implementations. JMLR, 22(268), 1–8.
[31] Zakka, K. et al. (2022). MuJoCo Menagerie: A collection of physics-based simulation models. GitHub: google-deepmind/mujoco_menagerie.
[32] A. Espinal et al. (2016). Quadrupedal robot locomotion: A biologically inspired approach and its hardware implementation. Comput. Intell. Neurosci.
[33] X. Jiang et al. (2023). Fully spiking neural network for legged robots. arXiv:2310.05022.