Expert Assessment of the FUM Self-Improvement Engine Stability Analysis Framework (Preliminary)

1. Introduction

1.1 Purpose

This report provides an expert technical assessment of the document "SIE Stability Analysis Framework (Preliminary)" dated 4/2/2025. The objective is to critically evaluate the proposed framework for analyzing and ensuring the stable operation of the FUM Self-Improvement Engine (SIE). This assessment examines the framework's understanding of the problem complexity, the justification for its novel analytical approach, the rigor of the presented mathematical formalism, the validity of the empirical results, the identified limitations, and the feasibility of the planned future work. The evaluation incorporates relevant principles and findings from control theory, reinforcement learning, computational neuroscience, and related fields to provide context and recommendations for subsequent development phases.

1.2 SIE Context and Stability Challenge

The Self-Improvement Engine (SIE) represents a core component of the FUM learning architecture. As detailed in the preliminary framework document, it integrates multiple distinct reward signals (TD_error, novelty, habituation, self_benefit) along with potential external rewards. A key feature is its non-linear modulation of Spike-Timing-Dependent Plasticity (STDP) through a sigmoid function applied to the total reward signal (Rtotal). The resulting synaptic weight update, $\Delta wij \propto \eta \cdot (1+mod\_factor) \cdot Rtotal \cdot eij$, introduces significant complexity.

The central challenge, correctly identified in the preliminary report, is to guarantee stable learning dynamics within this intricate system. This involves preventing undesirable outcomes such as reward hacking (where the system optimizes internal metrics at the expense of external goals), uncontrolled oscillations in network activity or weights, unbounded synaptic weight growth, and ensuring convergence towards desired behavioral objectives. The interaction between multiple reward components, the non-linear plasticity modulation, and the feedback loops inherent in the learning process necessitates a formal stability analysis.

1.3 Scope of Review

This review systematically analyzes the preliminary framework document, covering:

1. **SIE Complexity and Framework Justification:** Assessing the report's characterization of the SIE's complexities and the rationale for requiring a novel

analytical approach beyond standard Reinforcement Learning (RL) methods.

2. **Mathematical Formalism:** Evaluating the proposed system modeling using difference equations and the preliminary Lyapunov stability analysis, including its assumptions and limitations.
3. **Empirical Validation:** Reviewing the simulation setup, results, and the conclusions drawn, particularly regarding the role of synaptic scaling.
4. **Contextualization:** Comparing the proposed methods against state-of-the-art techniques in non-linear control, stochastic approximation theory, multi-objective reinforcement learning (MORL), and the analysis of plastic spiking neural networks (SNNs) with homeostasis.
5. **Limitations and Future Work:** Assessing the acknowledged limitations and the proposed plan for Phase 2 development.
6. **Overall Assessment and Recommendations:** Synthesizing the findings and providing actionable recommendations for the next phase of the SIE stability analysis.

Insights are drawn from established theoretical principles and contemporary research findings, including those presented in supporting documents.[1]

1.4 Preliminary Overall Assessment

The reviewed document presents a commendable initial effort to tackle the complex stability problem of the FUM SIE. It correctly identifies the key challenges and the inadequacy of standard analytical tools, proposing a hybrid approach combining theoretical modeling (Lyapunov analysis) with empirical simulation. The iterative development of the simulator and the empirical identification of synaptic scaling's critical role are positive steps. However, the current framework exhibits significant gaps. The mathematical formalism, particularly the Lyapunov analysis, is preliminary, lacks stochastic rigor, and is disconnected from the crucial empirical finding regarding homeostatic scaling. The simulation environment remains simplified, and the stability assessment lacks quantitative depth. Formal verification methods are notably absent. The framework implicitly addresses a multi-objective learning problem without leveraging appropriate MORL concepts. Consequently, while the direction is sound, substantial refinement in theoretical rigor, simulation realism, and integration of relevant analytical techniques (stochastic approximation, homeostasis modeling, MORL principles, formal verification) is required to achieve robust stability guarantees.

2. Analysis of SIE Complexity and the Need for a Bespoke Framework

2.1 Characterization of SIE Challenges

The preliminary report accurately identifies several key features of the SIE that contribute to its complexity and necessitate a dedicated stability analysis framework (Sections 1 & 2):

- **Multi-Objective Nature:** The SIE combines multiple internal drives (novelty, habituation, self_benefit) with performance-related signals (TD_error, W_EXTERNAL) using a weighted sum to form Rtotal. This inherently creates a multi-objective optimization scenario where objectives may conflict (e.g., exploration driven by novelty versus exploitation potentially favored by TD_error or stability implied by habituation). Balancing these objectives dynamically is a core challenge. Multi-objective optimization problems often lack a single optimal solution, instead featuring a set of trade-off solutions known as the Pareto front.[3] The SIE must navigate this implicit trade-off space.
- **Non-Linear Modulation:** The STDP update rule incorporates a non-linear modulation factor, mod_factor=2σ(Rtotal)−1, leading to an effective update proportional to 2σ(Rtotal)Rtotal. This introduces a state-dependent (via Rtotal) non-linearity that deviates significantly from the linear updates often assumed in simpler RL or plasticity models. Non-linear systems can exhibit complex behaviors like multiple equilibria, bifurcations, and chaos, requiring specialized analytical tools.[4]
- **Coupled Dynamics:** The system exhibits strong coupling. Synaptic weights (W) influence network activity and state representation (e.g., cluster IDs). This activity affects the calculation of reward components (TD error, novelty, etc.), which combine into Rtotal. Rtotal then modulates the plasticity rule (ΔW), closing the loop. Furthermore, the value function (V) used for TD error is also learned and depends on network states influenced by W. Analyzing components in isolation is therefore insufficient.
- **Emergent State Space:** The reliance on dynamically formed cluster IDs as states for TD learning adds another layer of complexity, as the state representation itself evolves during learning.

2.2 Justification for a Novel Analytical Approach

The preliminary report argues convincingly that standard RL stability analyses are insufficient for the SIE (Section 2). This justification is well-founded:

- **Limitations of Standard RL Analysis:** Conventional RL analyses often assume linear dynamics, simpler reward structures (single scalar objective), or known transition models. Model-free RL algorithms, while powerful, often lack inherent stability guarantees and can be highly sensitive to hyperparameters, initialization, and reward shaping, sometimes requiring careful tuning or environment simplification to ensure convergence.[1] Stability is frequently a major challenge, especially in continuous control or complex state spaces.[1]

- **Relevance of Non-Linear Control Theory:** The SIE's characteristics—non-linear update rules and coupled feedback loops—place it firmly in the domain of non-linear dynamical systems. Techniques from non-linear control, such as Lyapunov stability theory, phase plane analysis, and potentially bifurcation analysis, are necessary to understand its behavior.[4] Linearization around an equilibrium point, a common technique, only provides local stability information and may not capture the global dynamics.[4]
- **Multi-Objective Optimization Context:** As noted above, the SIE implicitly operates as a multi-objective system. Standard RL optimizes a single objective. Multi-Objective Reinforcement Learning (MORL) specifically addresses problems with multiple, often conflicting, reward signals.[2] Analyzing the SIE requires considering concepts like Pareto optimality and the potential pitfalls of simple scalarization methods (like the weighted sum used for Rtotal), which may not explore the full Pareto front or can be sensitive to objective scaling.[21]

2.3 Deeper Implications of SIE Complexity

The interaction of the identified complexities leads to further analytical challenges not fully elaborated in the preliminary report:

- **Implicit MORL Dynamics:** The SIE framework tackles a MORL problem implicitly through weighted scalarization (Rtotal) rather than employing explicit MORL algorithms designed to find or approximate the Pareto set.[16] This design choice, while potentially simpler to implement initially, significantly complicates stability analysis. The system's trajectory in the weight/policy space is governed by the gradient of this scalarized reward, which depends heavily on the chosen weights (WTD,Wnovelty, etc.) and the current state. Without mechanisms to explicitly manage objective trade-offs or explore the Pareto front, the system might exhibit oscillations as it shifts focus between conflicting objectives, converge to potentially undesirable points on the Pareto front (e.g., prioritizing an internal drive over external performance), or fail to converge stably, especially if the Pareto front is non-convex.[21] The stability analysis must therefore consider not just boundedness but the nature of convergence within this multi-objective landscape.
- **Compound Non-Linearities:** The specific form of the non-linear modulation, 2σ(Rtotal)Rtotal, introduces a complex gain factor into the learning dynamics. This gain is approximately linear for small ∣Rtotal∣, but saturates in magnitude for large

|Rtotal|. Crucially, the sign of the effective learning term depends only on the sign of Rtotal (since σ(Rtotal) has the same sign as Rtotal), but its magnitude is non-monotonically related to Rtotal. This complex, state-dependent gain interacts with other system non-linearities (e.g., neuron activation functions, TD learning updates) and the coupled dynamics of W and V. Such interactions in non-linear systems can lead to unexpected behaviors like bifurcations, multiple stable states, or sensitivity to initial conditions [4], which are difficult to predict using simplified analyses or bounds that treat components as static. The stability analysis must grapple with the specific nature of this compound non-linearity and its potential to induce different dynamical regimes.

3. Evaluation of the Proposed Mathematical Formalism

3.1 System Modeling via Difference Equations

The preliminary report proposes modeling the SIE-STDP dynamics using a system of non-linear difference equations (Section 3.1). The core equation for weight updates is given as:

Code snippet

```
\Delta W = \left \cdot \Delta t - \lambda W \cdot \Delta t
```

This approach is a standard practice in analyzing discrete-time learning systems, providing a necessary simplification of the underlying continuous-time SNN dynamics. Including a linear weight decay term (λW) is a common and often essential regularization technique to prevent unbounded weight growth. Representing the complex STDP rule, including the non-linear reward modulation, within this framework is appropriate as a first-order model.

However, this model inherently abstracts away significant details. Spike timing precision, the specific dynamics of the chosen neuron model (e.g., LIF), the network's connectivity structure, and the exact mechanism generating eligibility traces (E) are simplified or omitted. Furthermore, the critical mechanism of synaptic scaling, later shown empirically to be vital for stability, is modeled separately and not integrated into this core equation. This separation hinders a unified analysis of the interacting plasticity and homeostatic forces. While simplification is necessary, the current model may omit factors crucial for accurately predicting stability boundaries.

3.2 Lyapunov Stability Analysis

The report presents a preliminary Lyapunov stability analysis (Section 3.2), first considering only the weights (W) and then attempting a coupled analysis including the value function (V).

- **Analysis of L(W)=21 // W // F2:** This is a standard candidate function for analyzing weight dynamics. The derivation of the change ΔL≈ηeff·Rtotal·⟨W,E⟩F−λ‖W‖F2 correctly identifies the balance between the learning term (driven by reward and eligibility) and the decay term. The resulting condition for ΔL≤0, namely λ // W // F2≥ηeff·Rtotal·⟨W,E⟩F, highlights that decay dominates when the weight norm is sufficiently large. The subsequent bounding exercise, yielding an approximate condition // W // F≥λ2ηRmax // E // F, suggests that the system tends towards an equilibrium where weight growth driven by learning (proportional to ηRmax // E // F) is counteracted by decay (proportional to λ // W // F). This provides a basic intuition for how decay prevents unbounded growth *beyond* a certain magnitude related to the system's error signals and parameters.
  However, this analysis suffers from significant limitations acknowledged partially in the report:
  1. It treats ηeff, Rtotal, and E as bounded constants or variables independent of W, ignoring the coupled dynamics. Rtotal and E are functions of the network state, which depends on W.
  2. The analysis is deterministic, ignoring the stochastic nature of rewards, state transitions, and potentially spike timings inherent in the SIE and underlying SNN.
  3. It completely omits synaptic scaling, which the empirical results demonstrate is essential for preventing weight collapse under the chosen parameters. The derived condition based solely on λ is therefore insufficient to explain the observed stable behavior.
  4. The condition derived suggests stability *outside* a ball defined by // W // F, indicating decay dominates for large weights, but doesn't guarantee convergence *to* a bounded region or stability *near* the origin without stronger assumptions on E.
- **Coupled Analysis Attempt L(W,V):** Recognizing the need to include the value function V (since Rtotal depends on the TD error, which involves V) is crucial. The proposed coupled Lyapunov function L(W,V)=21 // W // F2+2c // V // 2 is a plausible starting point. The expression for ΔL includes terms related to both weight changes and TD learning updates.

The report correctly concludes that rigorously analyzing this coupled system is mathematically complex. The TD update term introduces dependencies on future states and rewards, inherent stochasticity, and non-linear interactions between W and V. Finding conditions for $\Delta L \le 0$ that hold universally is challenging. This complexity is typical when applying Lyapunov methods directly to RL algorithms, especially those with non-linear function approximation (here, the SNN implicitly represents the policy and potentially affects state representation) and coupled learning dynamics.[6] Synthesizing and verifying Lyapunov functions for neural network-based controllers or systems is an active research area often requiring sophisticated tools.[29]

3.3 Planned Future Analyses

The report outlines plans for convergence, multi-objective, and gaming analyses (Sections 3.3-3.5).

- **Convergence Analysis:** Identifying fixed points ($\Delta W=0, \Delta V=0$) and analyzing their stability is a standard approach in dynamical systems. However, for the SIE, this involves solving complex non-linear equations in a high-dimensional space, further complicated by stochasticity. The Ordinary Differential Equation (ODE) method from stochastic approximation theory [40] offers a more appropriate framework. This method analyzes the convergence of the stochastic iteration by studying the stability of an associated deterministic ODE. Applying it to the SIE would require deriving this ODE for the coupled system (including scaling) and carefully verifying the necessary conditions on noise properties (likely Markovian, requiring specific extensions of standard theorems [40]) and step sizes.
- **Multi-Objective Analysis:** Planning to use Pareto optimality concepts is highly relevant given the SIE's multi-objective nature. This analysis should aim to characterize the trade-offs the SIE achieves between its different reward components under various parameter settings (e.g., the weights $W_{TD}, W_{novelty}$, etc.). Techniques might involve estimating the attained region of the Pareto front or analyzing how the implicit scalarization R_{total} guides the system.[3] Understanding stability within this MORL context is also crucial.[2]
- **Gaming Analysis:** Identifying parameter regimes prone to reward hacking is essential for robust operation. This requires analyzing scenarios where internal rewards (novelty, self-benefit) might be maximized in ways detrimental to external task performance, potentially by finding attractive but spurious fixed points or limit cycles in the system dynamics.

3.4 Deeper Implications of the Mathematical Formalism

The preliminary mathematical formalism reveals critical gaps and misalignments with the system's nature:

- **Stochastic Approximation Framework Discrepancy:** The current Lyapunov analysis applies tools primarily suited for deterministic systems to an inherently stochastic, iterative learning process. The SIE update rule, $W_{n+1}=W_n+\Delta W_n$, where ΔW_n depends on stochastic rewards and state transitions, fits the definition of a stochastic approximation algorithm.[40] Rigorous analysis of such algorithms typically involves the ODE method, which examines convergence in expectation or almost surely by relating the discrete updates to a limiting ODE.[40] This requires verifying specific conditions on the noise process (e.g., Martingale difference or specific Markovian properties) and the step-size sequence (η). The current deterministic $\Delta L \le 0$ approach fails to capture the stochastic dynamics and long-term convergence properties adequately. A fundamental shift towards stochastic approximation theory is necessary for rigorous proofs.
- **Omission of Homeostasis in Formalism:** A major disconnect exists between the mathematical model analyzed in Section 3 (STDP + decay) and the empirically validated system in Section 6 (STDP + decay + scaling). The empirical results clearly show that synaptic scaling is essential for stability under the tested conditions, preventing the weight collapse observed otherwise. Homeostatic mechanisms like synaptic scaling are known to fundamentally alter network dynamics and are often required to stabilize biologically plausible plasticity rules like STDP.[46] By omitting scaling from the core mathematical model, the stability analysis in Section 3 fails to address the mechanisms actually responsible for the observed stability. A meaningful theoretical analysis must integrate the mathematical description of synaptic scaling (e.g., multiplicative adjustments based on activity, or periodic normalization) directly into the system equations *before* applying stability tools. Analyzing stability based solely on weight decay provides limited insight into the behavior of the complete, functioning system.

4. Review of Empirical Validation and Simulation Results

4.1 Experimental Setup and Simulator Evolution

The preliminary report details an empirical validation approach using a custom simulator (simulate_sie_stability.py). The simulator underwent iterative enhancements, starting from a baseline model and progressively incorporating Leaky Integrate-and-Fire (LIF) neuron dynamics, Excitatory/Inhibitory (E/I) balance, timing-based eligibility traces, and finally, synaptic scaling (Section 5, Analysis Log). This iterative development strategy is sound, allowing for the isolation and study of different components' effects. The final validation results presented (Section 6.3) are based on the most complete version of the simulator, using a specific set of parameters including LAMBDA (λ)=0.001 and ETA (η)=0.01. The scenario involves a simplified network structure, random state transitions, and periodic external rewards alongside periodic synaptic scaling.

4.2 Key Empirical Findings

The analysis of simulation data from the enhanced simulator (including synaptic scaling) yielded several key observations (Section 6.3):

- **Observed Stability:** Under the tested parameter configuration, the total reward signal (Rtotal), the modulation factor (mod_factor), the Frobenius norm of the weight matrix (// W // F), and the average state value function (V(state)) were reported as appearing stable over the 10,000 simulation steps.
- **Bounded Weight Norm:** Crucially, the weight norm (// W // F) stabilized at a non-zero value (around 14.2), contrasting sharply with previous simulations *without* scaling where weights collapsed towards zero even with the same weight decay (λ=0.001).
- **Critical Role of Synaptic Scaling:** This directly demonstrates the essential role of the implemented synaptic scaling mechanism in maintaining non-trivial weight structures and preventing runaway depression or collapse in the presence of the STDP rule and moderate weight decay. This aligns strongly with findings in computational neuroscience emphasizing the necessity of homeostatic mechanisms to stabilize potentially unstable Hebbian plasticity.[46]
- **Component Interactions:** A weak positive correlation (0.0426) was noted between the Novelty and Self-Benefit reward components, offering a preliminary glimpse into potential interactions between different objectives within the SIE.

4.3 Critique of Empirical Validation

While the empirical results provide valuable initial insights, particularly regarding synaptic scaling, the validation approach has limitations, many acknowledged in the report's future work section (Section 9):

- **Simulation Simplifications:** The simulator, despite enhancements, employs significant simplifications. Random state transitions, a basic eligibility trace model, periodic rewards, and a specific implementation of synaptic scaling may not fully capture the complexities of the SIE operating within a structured task environment with sparse, correlated inputs and potentially more sophisticated biological mechanisms (e.g., detailed inhibitory plasticity, dynamic clustering). The stability observed might be contingent on these simplifications.
- **Limited Parameter Exploration:** The stability claims are based primarily on a

single parameter set for the full simulator configuration (with scaling). Given the sensitivity to parameters like λ and η observed in earlier sweeps (without scaling), the observed stability might be localized to a small region in the parameter space. Comprehensive parameter sweeps *including* variations in scaling parameters, decay rates, learning rates, and SIE reward component weights are essential to map stability boundaries, as planned for future work.

- **Qualitative Stability Assessment:** The report relies on terms like "appears stable" based on summary statistics (mean, std dev) and likely visual inspection of time series plots. This lacks quantitative rigor. Formal tests for stationarity, convergence criteria for V(state), or analysis of the distribution of fluctuations are needed for a more definitive assessment of stability.
- **Absence of Instability Cases:** Demonstrating parameter regimes where the *full* system (with scaling) *does* become unstable (e.g., exhibits oscillations, runaway weights, or reward hacking) would be highly informative for understanding the limits of stability and the critical parameters governing it. The current results only show one stable operating point.

4.4 Deeper Implications of Empirical Findings

The empirical results, particularly the contrast between simulations with and without scaling, suggest deeper implications for the stability analysis:

- **Synergy of Scaling and Decay:** The findings strongly indicate that linear weight decay (λ) alone, at the tested strength, is insufficient to stabilize the SIE's modulated STDP rule. Synaptic scaling appears necessary to provide the required homeostatic control. This implies that stability arises not simply from a passive decay term counteracting learning, but from an active, dynamic interplay between Hebbian plasticity (modulated STDP), passive decay, and activity-dependent homeostatic regulation (scaling). This dynamic interplay is fundamental to the system's behavior.[47] Consequently, any theoretical analysis aiming for predictive power *must* incorporate models of both decay and scaling and analyze their interaction. Focusing solely on decay, as in the preliminary Lyapunov analysis, misses the core stabilizing mechanism identified empirically.
- **Potential for Unseen Dynamics:** The stability observed over 10,000 steps, assessed via basic statistics, might mask more complex underlying dynamics. Non-linear systems with multiple interacting components, feedback loops, and potential delays (e.g., in reward calculation or eligibility trace propagation) are known to exhibit phenomena like slow oscillations, quasi-periodic behavior, or even chaotic dynamics that might not be apparent in short simulations or through simple averaging.[4] The weak correlation reported between Novelty and

Self-Benefit is only a linear snapshot; non-linear interactions could be far more complex. Longer simulation runs coupled with more sophisticated analysis techniques (e.g., frequency analysis of key variables, phase-space reconstruction, recurrence quantification) are warranted to confidently rule out such hidden dynamics before declaring the system robustly stable.

5. Contextualization with State-of-the-Art Stability Analysis Techniques

To properly evaluate the preliminary SIE stability framework, it is essential to compare its methodology against established and emerging techniques used for analyzing stability in related complex systems, such as non-linear control systems, reinforcement learning agents, and plastic spiking neural networks.

5.1 Comparison with Lyapunov-Based Methods

The framework's use of Lyapunov functions aligns with a cornerstone of stability analysis in control theory [4] and its increasing application in machine learning and RL.[6] The core idea of finding an "energy-like" function whose value decreases along system trajectories provides a powerful tool for proving stability.[10]

However, the preliminary analysis in the report represents only a basic application. State-of-the-art Lyapunov-based approaches often incorporate more advanced techniques:

- **Neural Lyapunov Functions:** Given the complexity of the SIE dynamics (non-linear, high-dimensional), finding analytical Lyapunov functions may be intractable. A modern approach involves parameterizing the Lyapunov function itself as a neural network and learning its parameters concurrently with the system or controller, subject to satisfying the Lyapunov conditions.[6] This offers greater flexibility in capturing complex stability landscapes.
- **Formal Verification:** Deriving analytical conditions often involves approximations or restrictive assumptions. To provide rigorous guarantees, especially for neural network components, formal verification methods are employed. These use tools like Sum-of-Squares (SOS) optimization, Satisfiability Modulo Theories (SMT) solvers, Mixed-Integer Linear Programming (MILP), or interval bound propagation methods (like α,β-CROWN) to computationally *prove* that the Lyapunov conditions (e.g., $V(x)>0$ for x偃=0, $\dot{V}(x)\le -\epsilon V(x)$) hold over a specified region of the state space.[30] The SIE framework currently lacks any verification component.
- **Region of Attraction (ROA) Estimation/Maximization:** A key goal in control is often not just proving stability, but determining the set of initial states from which

the system converges to the equilibrium – the Region of Attraction (ROA). Advanced Lyapunov techniques focus on finding the largest possible certified ROA, often by optimizing the Lyapunov function itself or using level sets.[30] The SIE analysis only touches upon boundedness implicitly.

- **Stochastic Lyapunov Theory:** For systems with inherent randomness, like RL agents, deterministic Lyapunov stability ($\dot{V}<0$) is often too strong or inapplicable. Stochastic Lyapunov theory provides analogous concepts, such as stability in probability, mean-square stability, or conditions based on the expected change in the Lyapunov function (e.g., $E[\Delta L]\leq 0$).[6] These are more appropriate for analyzing the SIE's stochastic dynamics.

5.2 Comparison with Stochastic Approximation Theory

As argued previously, the SIE's learning process is fundamentally a stochastic approximation (SA) algorithm.[40] This field provides the most relevant theoretical tools for analyzing its convergence and stability.

The preliminary report's analysis does not leverage key SA concepts:

- **The ODE Method:** This is the central technique in SA theory for analyzing asymptotic behavior. It links the discrete, stochastic iteration $x_{n+1}=x_n+a_n(h(x_n)+M_{n+1})$ to a deterministic Ordinary Differential Equation (ODE) $\dot{x}=h(x)$. The stability properties of the ODE's equilibria (often analyzed using standard Lyapunov methods *for the ODE*) determine the convergence behavior of the stochastic algorithm.[40] The SIE analysis needs to identify the appropriate $h(x)$ for the coupled (W, V, scaling) system and analyze the corresponding ODE.
- **Handling Noise:** SA theory requires careful consideration of the noise term (M_{n+1}). Classical results often assume Martingale difference noise. The noise in the SIE (stemming from rewards, state transitions, eligibility traces) is likely state-dependent and correlated over time (Markovian). Analyzing SA with Markovian noise requires more advanced results, such as extensions of the Borkar-Meyn theorem, which impose specific conditions on the noise process and its interaction with the system dynamics.[40]
- **Step-Size Conditions:** Convergence proofs in SA typically rely on specific conditions for the step-size sequence a_n (analogous to $\eta\cdot\Delta t$ in the SIE model), such as $\sum a_n=\infty$ and $\sum a_n^2<\infty$. The SIE appears to use a constant base learning rate η. Constant step-size SA generally does not guarantee convergence to a single point but rather convergence to a residual error region or a stationary distribution around the optimum. Analysis requires different techniques, focusing on boundedness of moments or using averaging principles.[41]

5.3 Comparison with Multi-Objective Reinforcement Learning (MORL)

The SIE's structure positions it as an implicit MORL system. While the preliminary report acknowledges the multiple reward components, it doesn't analyze the system through the lens of MORL theory.

Key MORL concepts absent from the current analysis include:

- **Pareto Optimality:** Explicitly analyzing the trade-offs between the SIE's objectives (TD_error, novelty, etc.) is crucial. Does the system converge to solutions that are Pareto optimal, meaning no single objective can be improved without degrading another?[3] Where on the Pareto front do solutions lie under different parameter settings?.[15]
- **Limitations of Scalarization:** The SIE uses a weighted sum (Rtotal) to combine objectives. This is a simple scalarization technique. It is well-known that linear scalarization can only find solutions on the *convex hull* of the Pareto front and may fail to find solutions in non-convex regions. It can also lead to uneven coverage of the front and be sensitive to the relative scaling of objectives.[16] More advanced MORL methods use techniques like decomposition [16], hypervolume maximization [22], or learning sets of policies [21] to better handle complex Pareto fronts. The potential limitations of the SIE's simple scalarization need investigation.
- **Stability in MORL:** While a developing area, research exists on convergence and stability guarantees for specific MORL algorithms or settings.[2] These might offer relevant analytical approaches, particularly those considering policy gradient methods or actor-critic structures which bear some resemblance to the SIE's interaction between weight updates and value function learning.

5.4 Comparison with SNN Plasticity and Homeostasis Research

The SIE operates within a spiking neural network context, incorporating STDP and homeostatic mechanisms.

- **Homeostasis Necessity:** The empirical finding that synaptic scaling is crucial for stability aligns perfectly with a large body of computational neuroscience literature. Simple Hebbian or STDP rules are often inherently unstable, leading to runaway potentiation or depression. Homeostatic mechanisms, such as synaptic scaling (multiplicatively adjusting weights to maintain a target firing rate) or intrinsic plasticity (adjusting neuronal excitability), are considered essential for maintaining stable yet plastic network function.[46] The SIE framework correctly identifies this empirically, but the theoretical analysis needs to catch up by modeling these mechanisms.

- **Three-Factor Learning Rules:** The reward modulation (mod_factor) applied to STDP makes the SIE's learning rule a form of "three-factor" plasticity. In these rules, synaptic change depends on pre-synaptic activity, post-synaptic activity (captured by STDP), and a third, often global, modulatory signal (here, Rtotal).[58] This third factor, often linked to neuromodulators like dopamine in biology, typically conveys information about reward, novelty, or context, gating or scaling the plasticity induced by local activity.[60] Analyzing reward-modulated STDP (R-STDP) involves understanding the dynamics of eligibility traces (which bridge the gap between local activity and delayed rewards) and how the reward signal interacts with STDP's learning window.[62] Stability analyses of R-STDP consider these interactions.[61] The SIE analysis could benefit significantly by drawing upon the specific theoretical tools and insights developed for R-STDP.

5.5 Comparative Summary Table

The following table summarizes the comparison between the preliminary SIE framework's approach and relevant state-of-the-art techniques.

Approach Name	Core Principle	Applicability to SIE	Strengths	Weaknesses/Challenges for SIE Analysis	Key References
SIE Framework (Prelim.)	Simplified difference eq. model; Preliminary deterministic Lyapunov analysis ($\Delta L \le 0$); Empirical simulation.	Direct application attempt.	Identifies problem; Combines theory/simulation; Iterative simulator development; Empirically finds scaling importance.	Lacks rigor (stochasticity, homeostasis ignored in theory); Simulation simplified; No verification; MORL aspect ignored; Convergence goal unclear.	*(User Query Document)*
Classical Lyapunov Theory	Find energy function V(x) s.t. $\dot{V}(x)<0$	Used preliminarily for L(W) and	Powerful, well-established for	Difficult for high-dim, non-linear	[4]

	along trajectories for deterministic systems.	L(W,V). Relevant for analyzing the limiting ODE in SA.	deterministic systems. Provides stability certificates.	systems; Finding V(x) is hard; Doesn't directly handle stochasticity; ROA analysis complex.	
Neural Lyapunov Synthesis	Parameterize V(x) as NN; Learn V(x) concurrently with controller/system to satisfy Lyapunov conditions.	Potential future direction if analytical V(x) intractable for SIE's limiting ODE or full system.	Handles complex dynamics; Automates V(x) finding; Can be integrated with learning.	Non-convex optimization; Requires verification; Training stability challenges.	28
Formal Verification Methods	Use SOS, SMT, MILP, Bound Propagation to rigorously prove properties (e.g., Lyapunov conditions) over a region.	Necessary future step to provide guarantees for SIE stability conditions, especially if used in critical systems.	Provides formal guarantees (soundness); Can handle NN components; Can certify ROAs.	Computationally expensive; Scalability challenges for large systems/NNs; Requires precise system model/bounds.	30
Stochastic Approx. Theory	Analyze convergence/stability of xn+1=xn+an(h(xn)+Mn+1) via ODE method, noise analysis, step sizes.	**Highly Relevant:** SIE update is SA. ODE method needed for rigorous convergence/stability analysis.	Directly addresses iterative, stochastic nature of learning; Provides tools for convergence analysis	Requires deriving limiting ODE; Needs verification of noise conditions (Markovian); Step-size analysis	40

			(mean, a.s.); Handles noise.	(constant η) requires specific techniques.	
Multi-Objective RL (MORL)	Learn policies balancing multiple conflicting objectives; Analyze Pareto optimality and trade-offs.	**Highly Relevant:** SIE implicitly solves a MORL problem. Analysis should incorporate MORL concepts.	Explicitly handles objective trade-offs; Provides framework for analyzing performance across objectives; Methods for finding Pareto set.	SIE uses simple scalarization (potential issues); Stability analysis in MORL is less mature; Requires defining objectives clearly.	3
SNN Homeostasis Analysis	Model and analyze mechanisms (scaling, intrinsic plasticity) that stabilize neural activity/plasticity.	**Essential:** Empirical results show scaling is critical. Theory *must* incorporate homeostasis models.	Explains biological stability; Provides models for scaling/other mechanisms; Analyzes interaction with plasticity (STDP).	Integrating homeostasis into coupled SIE analysis adds complexity; Precise biological mechanisms vs. simplified models.	46
Reward-Modulated STDP Analysis	Analyze three-factor rules involving pre/post activity and a modulatory signal (e.g., reward).	**Relevant:** SIE uses R-STDP variant. Analysis can leverage specific R-STDP literature.	Addresses interaction of local plasticity and global signals; Considers eligibility traces, reward delays; Analyzes stability/lear	Theoretical analysis can be complex; Specific models vary; Interaction with homeostasis needs consideration.	60

			ning.		

5.6 Implications of Contextualization

This comparison highlights several key points:

- **Convergence Goal Mismatch:** The preliminary Lyapunov analysis seems aimed at proving convergence to an equilibrium or ensuring strict boundedness ($\Delta L \leq 0$). However, the underlying algorithm is a constant step-size stochastic approximation, which typically converges to a *region* around the optimum, not a single point.[41] The analysis framework needs to align with the expected convergence behavior. Is the goal stability within a bounded operating region, or convergence to a specific state? The tools used should match the goal (e.g., SA analysis of mean-square boundedness vs. Lyapunov stability of an ODE equilibrium).
- **Significant Verification Gap:** There is a major gap between the analytical conditions derived in the preliminary report (which are incomplete and based on strong assumptions) and the level of assurance needed for a complex system like FUM. Modern approaches in safety-critical control and RL heavily rely on formal *verification* tools to computationally prove that stability conditions hold over relevant operating regions.[30] The current SIE framework lacks any plan for such verification, relying solely on analytical attempts and limited simulation. Bridging this gap is crucial for establishing trust in the SIE's stability.

6. Overall Assessment and Recommendations for Phase 2

6.1 Synthesis of Strengths

The preliminary "SIE Stability Analysis Framework" represents a valuable initial step towards addressing a critical and complex challenge. Its strengths include:

- **Problem Recognition:** Clearly articulates the complex stability issues arising from the SIE's multi-objective nature, non-linear plasticity modulation, and coupled dynamics.
- **Need for Novelty:** Correctly identifies that standard RL or linear system analysis techniques are insufficient, justifying the need for a bespoke framework.
- **Hybrid Approach:** Adopts a sensible strategy combining theoretical analysis (Lyapunov methods) with empirical validation (simulation).
- **Iterative Simulation:** Demonstrates good practice by progressively enhancing the simulator to incorporate more biologically relevant features (LIF neurons, E/I balance, synaptic scaling).
- **Key Empirical Finding:** Successfully identifies the crucial role of synaptic scaling

in stabilizing the weight dynamics, providing a critical insight for future theoretical work.

6.2 Synthesis of Weaknesses

Despite these strengths, the framework in its current state exhibits significant weaknesses that must be addressed:

- **Lack of Theoretical Rigor:** The mathematical formalism, particularly the Lyapunov analysis, is preliminary and lacks the necessary rigor.
  - It largely ignores the stochastic nature of the learning process, attempting to apply deterministic stability concepts directly (addressed in Section 3.4).
  - It fails to integrate the empirically critical mechanism of synaptic scaling into the core theoretical model, creating a disconnect between theory and observation (addressed in Section 3.4).
  - Formal proofs of stability or convergence are absent, and the derived analytical conditions rely on strong, potentially unrealistic assumptions (acknowledged in Section 8).
- **Simulation Limitations:** The empirical validation, while useful, relies on a simplified simulation environment.
  - Key aspects of realistic operation (structured tasks, sparse activity, detailed plasticity rules, dynamic clustering) are missing (Section 9).
  - Parameter space exploration is insufficient, with stability demonstrated only for a single configuration of the full model (Section 9).
  - The assessment of stability is qualitative ("appears stable") rather than based on rigorous quantitative measures (addressed in Section 4.3).
  - Potential for complex, hidden dynamics (e.g., slow oscillations) may be missed by current analysis methods (addressed in Section 4.4).
- **Methodological Gaps:**
  - **Verification:** There is no strategy outlined for formally verifying any derived stability conditions using state-of-the-art tools (addressed in Section 5.6).
  - **MORL:** The implicit multi-objective nature of the SIE is not explicitly analyzed using appropriate MORL concepts or tools (addressed in Section 2.3).
  - **Convergence:** The analysis does not clearly define the target convergence property (point vs. region) and uses methods potentially mismatched with the constant step-size nature of the algorithm (addressed in Section 5.6).

6.3 Evaluation of Future Work and Integration Plans

The future work outlined in Section 9 and the integration plans in Section 7 are generally appropriate but require refinement based on the weaknesses identified above.

- **Parameter Sweeps (Future Work 1):** Absolutely essential, but must be performed with the *full* simulator (including scaling) and should systematically aim to map stability boundaries, not just confirm stability at default parameters. Variations should include scaling parameters, decay rates, learning rates, and relative weights of SIE components.
- **Enhance Simulator (Future Work 2):** Necessary for increasing confidence, but adds complexity. Prioritization should be given to features most likely to impact stability dynamics, such as more realistic eligibility trace models (e.g., capturing finer temporal details relevant to STDP [68]) and potentially sparsity constraints.
- **Refine Theory (Future Work 3 / Phase 2):** This is the most critical area. The plan to "revisit Lyapunov analysis attempting to incorporate scaling" is necessary but insufficient. A more fundamental shift in the theoretical framework is required, as detailed in the recommendations below.
- **Test Specific Scenarios (Future Work 4):** A valuable addition for targeted validation, particularly for exploring reward hacking potential and robustness to specific perturbations.
- **FUM Integration (Section 7):** The plan to integrate stability conditions or adaptive tuning is logical but premature. The feasibility depends entirely on deriving *robust, verifiable* stability conditions in Phase 2. The computational cost of mechanisms like synaptic scaling needs careful consideration at scale. Derived stability conditions should ideally be independent of network size, but this requires theoretical justification and empirical validation.

6.4 Actionable Recommendations for Phase 2

To address the identified weaknesses and build a robust stability framework, Phase 2 should prioritize the following:

1. **Adopt a Stochastic Approximation (SA) Framework:** Re-cast the theoretical analysis within the formal framework of stochastic approximation.
    - Derive the associated Ordinary Differential Equation (ODE) that describes the expected evolution of the coupled system state (including weights W, value function parameters V, and any state variables related to synaptic scaling).
    - Analyze the stability of this limiting ODE, identifying its equilibrium points and using appropriate tools (e.g., Lyapunov functions *for the ODE*) to assess their stability properties.
    - Explicitly address the nature of the noise (Markovian) and the implications of using a constant learning rate (η), leveraging relevant SA theorems and techniques.[40]

2. **Integrate Homeostasis into the Theoretical Model:** The mathematical model used for analysis *must* incorporate the dynamics of synaptic scaling (and any other relevant homeostatic mechanisms present in FUM). This could involve adding differential/difference equations governing scaling factors or modifying the weight update equation directly. Stability analysis should then be performed on this integrated system.[46]
3. **Employ Appropriate Lyapunov or Alternative Techniques:** If Lyapunov analysis is pursued (e.g., for the limiting ODE or directly on the stochastic system), utilize methods suitable for the context:
   - Stochastic Lyapunov functions focusing on stability in expectation or mean-square.[6]
   - Consider learning-based approaches (Neural Lyapunov Functions) if analytical functions prove intractable.[28]
   - Alternatively, explore other stability analysis techniques from non-linear systems theory if Lyapunov methods prove too difficult.[4]
4. **Enhance Simulation Analysis Rigor:** Move beyond qualitative stability assessments.
   - Implement quantitative metrics for convergence (e.g., time to reach a tolerance band, stationarity tests on key variables) and boundedness.
   - Conduct longer simulations specifically designed to detect oscillations or complex dynamics (e.g., using spectral analysis, return maps, or Lyapunov exponents if feasible).
   - Systematically map stability boundaries via extensive parameter sweeps using the enhanced simulator (including scaling and other relevant features). Visualize stability regions in parameter space.
5. **Explicitly Address the Multi-Objective Nature:** Analyze the SIE from a MORL perspective.
   - Characterize the implicit Pareto trade-offs being made by the weighted sum Rtotal.
   - Investigate potential issues arising from the simple scalarization, especially under conflicting objectives or non-convex Pareto fronts.[16]
   - Assess whether the system reliably converges to desirable regions of the objective space across different conditions.
6. **Develop a Formal Verification Strategy:** Outline a plan for how key stability properties or conditions, once theoretically derived or empirically identified, can be formally verified.
   - Evaluate the feasibility of applying tools like MILP, SMT solvers, or certified bound propagation methods to relevant parts of the SIE model (potentially simplified or abstracted).[30]

    - Focus verification efforts on critical parameter ranges or operating conditions identified through theoretical analysis and simulation sweeps.
7. **Leverage R-STDP Literature:** Incorporate theoretical insights specifically from the analysis of reward-modulated STDP, paying attention to the role of eligibility traces, reward timing, and known stability conditions for such three-factor rules.[60]

7. Conclusion

7.1 Summary of Findings

The "SIE Stability Analysis Framework (Preliminary)" document represents a necessary and well-motivated effort to understand and ensure the stability of a critical, complex component of the FUM architecture. The framework correctly identifies the unique challenges posed by the SIE's multi-objective reward structure, non-linear STDP modulation, and coupled dynamics, rightly concluding that standard analytical tools are insufficient. The proposed hybrid approach, combining preliminary theoretical modeling with iterative empirical simulation, is a reasonable starting point. A key contribution is the empirical identification of synaptic scaling as an essential mechanism for stabilizing weight dynamics in the simulated system.

However, the current framework requires substantial development to provide robust stability guarantees. The theoretical analysis, particularly the Lyapunov treatment, lacks the necessary rigor for the stochastic, non-linear, adaptive nature of the SIE. It notably fails to incorporate the crucial homeostatic mechanisms identified empirically and does not leverage appropriate tools from stochastic approximation theory. The empirical validation, while insightful, is based on simplified simulations and limited parameter exploration, with stability assessed qualitatively. Furthermore, the framework lacks strategies for formal verification and does not explicitly address the implications of the SIE's implicit multi-objective nature.

7.2 Path Forward

The path forward for Phase 2 must focus on bridging the significant gaps between the preliminary theory, the complex empirical behavior revealed by simulation, and the state-of-the-art in stability analysis for complex learning systems. Success hinges on adopting a more rigorous theoretical foundation grounded in stochastic approximation theory, explicitly modeling the interplay between plasticity and homeostasis (synaptic scaling), conducting extensive and quantitatively analyzed simulations to map stability boundaries, and developing a strategy for formal verification. Addressing the multi-objective aspects explicitly will also be crucial for understanding convergence towards desired outcomes. By systematically tackling these areas, the FUM project can move towards a more comprehensive and trustworthy understanding of SIE stability.

Works cited

1. Neural Lyapunov and Optimal Control - arXiv, accessed April 11, 2025, https://arxiv.org/html/2305.15244v4
2. Lexicographic Multi-Objective Reinforcement Learning - IJCAI, accessed April 11, 2025, https://www.ijcai.org/proceedings/2022/0476.pdf
3. Multi-objective optimization - Wikipedia, accessed April 11, 2025, https://en.wikipedia.org/wiki/Multi-objective_optimization
4. Nonlinear control - Wikipedia, accessed April 11, 2025, https://en.wikipedia.org/wiki/Nonlinear_control
5. Nonlinear Systems Analysis - SIAM Publications Library, accessed April 11, 2025, https://epubs.siam.org/doi/book/10.1137/1.9780898719185
6. openreview.net, accessed April 11, 2025, https://openreview.net/pdf/8864e380755262c5e55267e7f10e9bff5e4d817f.pdf
7. Actor-Critic Reinforcement Learning for Control with Stability Guarantee, accessed April 11, 2025, http://ras.papercept.net/images/temp/IROS/files/2895.pdf
8. Reinforcement Learning for Control with Stability Guarantee - OpenReview, accessed April 11, 2025, https://openreview.net/forum?id=vBNTeQ7dPP
9. On the Stability and Convergence of Robust Adversarial Reinforcement Learning: A Case Study on Linear Quadratic Systems, accessed April 11, 2025, https://proceedings.neurips.cc/paper/2020/file/fb2e203234df6dee15934e448ee88971-Paper.pdf
10. Stability Theory for Nonlinear Systems, accessed April 11, 2025, http://www.diag.uniroma1.it/oriolo/amr/material/stability.pdf
11. Introduction to Nonlinear Control: Stability, Control Design, and Estimation - Amazon.com, accessed April 11, 2025, https://www.amazon.com/Introduction-Nonlinear-Control-Stability-Estimation/dp/0691240485
12. Nonlinear Control Systems – Notes and Study Guides - Fiveable, accessed April 11, 2025, https://library.fiveable.me/nonlinear-control-systems
13. Analysis and Control of Nonlinear Systems, accessed April 11, 2025, https://mas-e.engineering.berkeley.edu/course/analysis-and-control-of-nonlinear-systems/
14. Stability and Feedback Control of Nonlinear Systems - Scientific & Academic Publishing, accessed April 11, 2025, http://article.sapub.org/10.5923.j.control.20201001.02.html
15. Distributional Pareto-Optimal Multi-Objective Reinforcement Learning - NeurIPS, accessed April 11, 2025, https://proceedings.neurips.cc/paper_files/paper/2023/file/32285dd184dbfc33cb2d1f0db53c23c5-Paper-Conference.pdf
16. Pareto Set Learning for Multi-Objective Reinforcement Learning - arXiv, accessed April 11, 2025, https://arxiv.org/html/2501.06773v1

17. NeurIPS Poster Distributional Pareto-Optimal Multi-Objective Reinforcement Learning, accessed April 11, 2025, https://neurips.cc/virtual/2023/poster/70393
18. Enabling Pareto-Stationarity Exploration in Multi-Objective Reinforcement Learning: A Weighted-Chebyshev Multi-Objective Actor-Critic Approach | OpenReview, accessed April 11, 2025, https://openreview.net/forum?id=BPQMd2gTYI
19. PA2D-MORL: Pareto Ascent Directional Decomposition Based Multi-Objective Reinforcement Learning | Proceedings of the AAAI Conference on Artificial Intelligence, accessed April 11, 2025, https://ojs.aaai.org/index.php/AAAI/article/view/29148
20. [2501.06773] Pareto Set Learning for Multi-Objective Reinforcement Learning - arXiv, accessed April 11, 2025, https://arxiv.org/abs/2501.06773
21. Multi-Objective Reinforcement Learning using Sets of Pareto Dominating Policies, accessed April 11, 2025, https://www.jmlr.org/papers/volume15/vanmoffaert14a/vanmoffaert14a.pdf
22. MORL-Prompt: An Empirical Analysis of Multi-Objective Reinforcement Learning for Discrete Prompt Optimization - ACL Anthology, accessed April 11, 2025, https://aclanthology.org/2024.findings-emnlp.577.pdf
23. Safe and Balanced: A Framework for Constrained Multi-Objective Reinforcement Learning, accessed April 11, 2025, https://arxiv.org/html/2405.16390v1
24. Gradient Shaping for Multi-Constraint Safe Reinforcement Learning, accessed April 11, 2025, https://proceedings.mlr.press/v242/yao24a/yao24a.pdf
25. A Distributional View on Multi-Objective Policy Optimization - Proceedings of Machine Learning Research, accessed April 11, 2025, http://proceedings.mlr.press/v119/abdolmaleki20a/abdolmaleki20a.pdf
26. A Distributional View on Multi-Objective Policy Optimization - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2005.07513
27. (PDF) Personalized robotic control via constrained multi-objective reinforcement learning, accessed April 11, 2025, https://www.researchgate.net/publication/375254025_Personalized_robotic_control_via_constrained_multi-objective_reinforcement_learning
28. Lyapunov-Regularized Reinforcement Learning for Power System Transient Stability - NSF Public Access Repository, accessed April 11, 2025, https://par.nsf.gov/servlets/purl/10289980
29. Reviews: Learning Stable Deep Dynamics Models, accessed April 11, 2025, https://proceedings.neurips.cc/paper_files/paper/2019/file/0a4bbceda17a6253386bc9eb45240e25-Reviews.html
30. Lyapunov-stable Neural Control for State and Output Feedback: A Novel Formulation for Efficient Synthesis and Verification - arXiv, accessed April 11, 2025, https://arxiv.org/html/2404.07956v1
31. Lyapunov Neural Network with Region of Attraction Search - arXiv, accessed April 11, 2025, https://arxiv.org/html/2403.10621v1
32. Learning and Verifying Maximal Taylor-Neural Lyapunov functions - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2408.17246
33. Lyapunov-stable neural-network control - Robotics, accessed April 11, 2025,

https://www.roboticsproceedings.org/rss17/p063.pdf
34. Lyapunov-stable Neural Control for State and Output Feedback: A Novel Formulation - arXiv, accessed April 11, 2025, https://arxiv.org/abs/2404.07956
35. [2109.14152] Lyapunov-stable neural-network control - arXiv, accessed April 11, 2025, https://arxiv.org/abs/2109.14152
36. Certified Training with Branch-and-Bound: A Case Study on Lyapunov-stable Neural Control, accessed April 11, 2025, https://openreview.net/forum?id=8ctju6iFcn
37. Neural Lyapunov Control for Discrete-Time Systems, accessed April 11, 2025, https://proceedings.neurips.cc/paper_files/paper/2023/file/08bf1773e94763b6cc366ee7c6582f27-Paper-Conference.pdf
38. Verified-Intelligence/Lyapunov_Stable_NN_Controllers: Lyapunov-stable Neural Control for State and Output Feedback - GitHub, accessed April 11, 2025, https://github.com/Verified-Intelligence/Lyapunov_Stable_NN_Controllers
39. Fossil 2.0: Formal Certificate Synthesis for the Verification and Control of Dynamical Models, accessed April 11, 2025, https://arxiv.org/html/2311.09793v2
40. www.jmlr.org, accessed April 11, 2025, http://www.jmlr.org/papers/volume26/24-0100/24-0100.pdf
41. The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning | SIAM Journal on Control and Optimization, accessed April 11, 2025, https://epubs.siam.org/doi/10.1137/S0363012997331639
42. Stochastic approximation algorithms: Overview and recent trends - Indian Academy of Sciences, accessed April 11, 2025, https://www.ias.ac.in/article/fulltext/sadh/024/04-05/0425-0452
43. A Tutorial Introduction to Reinforcement Learning - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2304.00803
44. Stability and Convergence of Distributed Stochastic Approximations with large Unbounded Stochastic Information Delays - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2305.07091
45. Stochastic Approximation: from Statistical Origin to Big-Data, Multidisciplinary Applications, accessed April 11, 2025, https://tzelai.ckirby.su.domains/pubs/2021_StochApprox.pdf
46. Supervised learning of spatial features with STDP and homeostasis using Spiking Neural Networks on SpiNNaker - arXiv, accessed April 11, 2025, https://arxiv.org/html/2312.02659v2
47. Unsupervised Heart-rate Estimation in Wearables With Liquid States and A Probabilistic Readout - eScholarship, accessed April 11, 2025, https://escholarship.org/content/qt81z8g7ft/qt81z8g7ft_noSplash_956d49a7b3c09a114dee16511bf7e989.pdf
48. The interplay between homeostatic synaptic scaling and ... - eLife, accessed April 11, 2025, https://elifesciences.org/reviewed-preprints/88376/reviews
49. Biologically Plausible Models of Homeostasis and STDP: Stability and Learning in Spiking Neural Networks - UC Irvine, accessed April 11, 2025, https://sites.socsci.uci.edu/~jkrichma/Carlson2013-IJCNN2013.pdf
50. Homeostatic Plasticity and STDP: Keeping a Neuron's Cool in a Fluctuating World

- PMC, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC3059670/
51. Biologically plausible models of homeostasis and STDP: Stability and learning in spiking neural networks | Request PDF - ResearchGate, accessed April 11, 2025, https://www.researchgate.net/publication/261315653_Biologically_plausible_models_of_homeostasis_and_STDP_Stability_and_learning_in_spiking_neural_networks
52. Safe Deep Reinforcement Learning Control with Self-Learned Neural Lyapunov Functions and State Constraints - ONERA, accessed April 11, 2025, https://w3.onera.fr/copernic/sites/default/files/Safe_Deep_Reinforcement_Learning_Control_with_Self-Learned_Neural_Lyapunov_Functions_and_State_Constraints.pdf
53. Convergence Rate of Stochastic Approximation Algorithms in the Degenerate Case | SIAM Journal on Control and Optimization, accessed April 11, 2025, https://epubs.siam.org/doi/10.1137/S0363012995281730
54. Reinforcement Learning-Based Control of Nonlinear Systems Using Lyapunov Stability Concept and Fuzzy Reward Scheme - King's Research Portal, accessed April 11, 2025, https://kclpure.kcl.ac.uk/portal/files/136810333/crlnlf_final.pdf
55. Multi-Objective Deep Reinforcement Learning for Personalized Dose Optimization Based on Multi-Indicator Experience Replay - ResearchGate, accessed April 11, 2025, https://www.researchgate.net/publication/366664602_Multi-Objective_Deep_Reinforcement_Learning_for_Personalized_Dose_Optimization_Based_on_Multi-Indicator_Experience_Replay
56. Deep Unsupervised Learning Using Spike-Timing-Dependent Plasticity - arXiv, accessed April 11, 2025, https://arxiv.org/html/2307.04054v2
57. Purely STDP-based assembly dynamics: Stability, learning, overlaps, drift and aging - PMC, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10124856/
58. Incorporating structural plasticity into self-organization recurrent networks for sequence learning - Frontiers, accessed April 11, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1224752/full
59. STDP-Driven Rewiring in Spiking Neural Networks under Stimulus-Induced and Spontaneous Activity - MDPI, accessed April 11, 2025, https://www.mdpi.com/2313-7673/8/3/320
60. Three-Factor Learning in Spiking Neural Networks: An Overview of Methods and Trends from a Machine Learning Perspective - arXiv, accessed April 11, 2025, https://arxiv.org/html/2504.05341v1
61. Three-Factor Learning in Spiking Neural Networks: An Overview of Methods and Trends from a Machine Learning Perspective - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2504.05341
62. Three Factor Learning with Lava, accessed April 11, 2025, https://lava-nc.org/lava/notebooks/in_depth/three_factor_learning/tutorial01_Reward_Modulated_STDP.html
63. 19.5 Summary | Neuronal Dynamics online book, accessed April 11, 2025,

https://neuronaldynamics.epfl.ch/online/Ch19.S5.html
64. Functional Requirements for Reward-Modulated Spike-Timing-Dependent Plasticity - PMC, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6634722/
65. Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules - PMC, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC4717313/
66. (PDF) Theoretical Analysis of Learning with Reward-Modulated Spike-Timing-Dependent Plasticity. - ResearchGate, accessed April 11, 2025, https://www.researchgate.net/publication/221617962_Theoretical_Analysis_of_Learning_with_Reward-Modulated_Spike-Timing-Dependent_Plasticity
67. A Learning Theory for Reward-Modulated Spike-Timing-Dependent Plasticity with Application to Biofeedback | PLOS Computational Biology, accessed April 11, 2025, https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000180
68. Stochastic Models of Neural Synaptic Plasticity - arXiv, accessed April 11, 2025, https://arxiv.org/pdf/2010.08195