

1 Microscopic Foundation: Neural Dynamics

We consider a minimal circuit consisting of a single presynaptic neuron connected to a single postsynaptic neuron through a unidirectional synapse of weight w . This section specifies the dynamics governing the postsynaptic membrane potential, the synaptic coupling, and the generation of spikes.

1.1 Membrane Potential Dynamics

The postsynaptic neuron is modeled as a Leaky Integrate-and-Fire (LIF) unit, a standard reduction of the Hodgkin–Huxley formalism that retains sub-threshold integration and thresholding [1]. Its membrane potential $V(t)$ evolves according to the conservation of current across the cell membrane, modeled as a parallel RC circuit with leakage resistance R_m and membrane capacitance C_m [2]:

$$\tau_m \frac{dV(t)}{dt} = -(V(t) - E_L) + R_m I_{\text{syn}}(t) + R_m I_{\text{ext}}(t), \quad (1)$$

where $\tau_m = R_m C_m$ is the membrane time constant, E_L is the resting potential, $I_{\text{syn}}(t)$ is the synaptic current from the presynaptic neuron, and $I_{\text{ext}}(t)$ is an external current, typically modeled as an injected current or Gaussian white noise.

1.2 Spike Generation

A spike is defined as the instant t_{post}^k at which the membrane potential crosses a fixed threshold θ from below:

$$t_{\text{post}}^k : \quad V(t_{\text{post}}^k) = \theta \quad \text{and} \quad \left. \frac{dV}{dt} \right|_{t=t_{\text{post}}^k} > 0. \quad (2)$$

The derivative condition distinguishes the upward threshold crossing from subsequent repolarization. Upon firing, V is reset to $V_{\text{reset}} < \theta$ and the dynamics in (1) are suspended for a refractory period τ_{ref} , after which integration resumes from V_{reset} [2].

The spike trains of both neurons are written as sums of Dirac distributions:

$$\rho_{\text{pre}}(t) = \sum_k \delta(t - t_{\text{pre}}^k), \quad \rho_{\text{post}}(t) = \sum_k \delta(t - t_{\text{post}}^k). \quad (3)$$

The presynaptic spike times $\{t_{\text{pre}}^k\}$ are taken as given (e.g., drawn from a Poisson process), while the postsynaptic times $\{t_{\text{post}}^k\}$ are determined by (1) and (2).

1.3 Synaptic Interaction

The synaptic current is determined by the presynaptic spike train filtered through a postsynaptic current (PSC) kernel and scaled by the synaptic weight w . Under the *current-based* approximation, which treats synaptic currents as independent of the postsynaptic membrane potential, the synaptic current is [1]:

$$I_{\text{syn}}(t) = w \int_{-\infty}^t \alpha(t - s) \rho_{\text{pre}}(s) ds, \quad (4)$$

where $\alpha(t) = \tau_s^{-1} e^{-t/\tau_s} \Theta(t)$ is an exponential PSC kernel with synaptic time constant τ_s and $\Theta(t)$ the Heaviside step function [2].

2 Three-Factor Plasticity Model

We now turn to the evolution of the synaptic weight w . The plasticity rule belongs to the class of *three-factor learning rules* reviewed by Frémaux and Gerstner [3]. In standard Spike-Timing-Dependent Plasticity (STDP), weight changes depend on the correlation of pre- and postsynaptic spike times; three-factor rules gate this local signal with a global neuromodulatory factor representing reward or error feedback.

2.1 Weight-Dependent Scaling and Stability

Before specifying the plasticity dynamics, we define the weight-dependent scaling functions that appear in the learning rule. The weight w represents the efficacy of the synapse from the presynaptic to the postsynaptic neuron and is constrained to the interval $[0, w_{\max}]$, where w_{\max} is a physiological saturation limit.

The amplitudes of potentiation and depression are modulated by soft-bound functions [1]:

$$A_+(w) = \eta_+(w_{\max} - w), \quad A_-(w) = \eta_- w, \quad (5)$$

where η_+ and η_- are learning rates. These linear dependencies cause the rate of weight change to diminish as w approaches either boundary. This prevents unbounded growth of the weight and biases the dynamics toward the interior of $[0, w_{\max}]$. However, because the full weight update (defined below in (10)) involves a signed modulation signal, the soft bounds alone do not strictly guarantee $w \in [0, w_{\max}]$; in practice, a hard clipping step $w \leftarrow \max(0, \min(w, w_{\max}))$ may be applied after each update.

2.2 Local Dynamics: The Eligibility Trace

A central feature of this model is that coincident pre- and postsynaptic spikes create a temporary memory called the *eligibility trace* $E(t)$ [3]. This trace allows the synapse to bridge the temporal gap between millisecond-scale neural activity and delayed reward signals. It evolves as:

$$\tau_e \frac{dE(t)}{dt} = -E(t) + S(t), \quad (6)$$

where τ_e is a decay time constant, typically on the order of hundreds of milliseconds to seconds for reinforcement learning tasks [1].

The driving term $S(t)$ captures the instantaneous STDP induction. To define it, we introduce filtered spike-history variables for each neuron:

$$\tau_+ \frac{dx_{\text{pre}}(t)}{dt} = -x_{\text{pre}}(t) + \rho_{\text{pre}}(t), \quad (7)$$

$$\tau_- \frac{dy_{\text{post}}(t)}{dt} = -y_{\text{post}}(t) + \rho_{\text{post}}(t), \quad (8)$$

where τ_+ and τ_- set the widths of the potentiation and depression windows, respectively. Experimental measurements place these values in the range 20–40 ms [4]. The STDP induction

term then combines Long-Term Potentiation (LTP) and Long-Term Depression (LTD):

$$S(t) = \underbrace{A_+(w) x_{\text{pre}}(t) \rho_{\text{post}}(t)}_{\text{LTP}} - \underbrace{A_-(w) y_{\text{post}}(t) \rho_{\text{pre}}(t)}_{\text{LTD}}, \quad (9)$$

with $A_+(w)$ and $A_-(w)$ as defined in (5).

2.3 Global Dynamics: Neuromodulated Update

The weight evolves under the product of the eligibility trace and a global neuromodulatory signal $M(t)$:

$$\frac{dw(t)}{dt} = M(t) E(t). \quad (10)$$

To construct $M(t)$, we first define smooth estimates of the instantaneous firing rates by low-pass filtering the spike trains with a rate time constant τ_r :

$$\tau_r \frac{dr_{\text{pre}}(t)}{dt} = -r_{\text{pre}}(t) + \rho_{\text{pre}}(t), \quad (11)$$

$$\tau_r \frac{dr_{\text{post}}(t)}{dt} = -r_{\text{post}}(t) + \rho_{\text{post}}(t). \quad (12)$$

Unlike a rectangular sliding-window estimator, these exponentially weighted averages are smooth functions of time and consistent in form with the other filtered quantities in the model.

As a simplified stand-in for a more general objective, we define the instantaneous reward as a penalty on the deviation of the postsynaptic rate from half the presynaptic rate:

$$R(t) = -\left(r_{\text{post}}(t) - \frac{1}{2} r_{\text{pre}}(t)\right)^2. \quad (13)$$

The modulation signal $M(t)$ is then a Reward Prediction Error (RPE), computed as the difference between $R(t)$ and a slowly adapting baseline $\bar{R}(t)$ [3]:

$$M(t) = R(t) - \bar{R}(t), \quad (14)$$

where $\bar{R}(t)$ tracks the running average of the reward:

$$\tau_{\bar{R}} \frac{d\bar{R}(t)}{dt} = -\bar{R}(t) + R(t). \quad (15)$$

The baseline enables bidirectional regulation: performance better than expected yields $M > 0$ (reinforcing the current eligibility trace), while performance worse than expected yields $M < 0$ (weakening it).

Table 1: Summary of notation.

Symbol	Description	Equation
$V(t)$	Postsynaptic membrane potential	(1)
E_L	Resting (leak) potential	(1)
θ	Spike threshold	(2)
V_{reset}	Reset potential after spike	§1.2
$\rho_{\text{pre}}(t), \rho_{\text{post}}(t)$	Spike trains (sums of Dirac deltas)	(3)
$\alpha(t)$	Postsynaptic current kernel	(4)
$I_{\text{syn}}(t)$	Total synaptic current	(4)
$I_{\text{ext}}(t)$	External / noise current	(1)
w	Synaptic weight	(10)
$A_+(w), A_-(w)$	Weight-dependent LTP/LTD amplitudes	(5)
$x_{\text{pre}}(t)$	Presynaptic spike-history trace	(7)
$y_{\text{post}}(t)$	Postsynaptic spike-history trace	(8)
$S(t)$	STDP induction term	(9)
$E(t)$	Eligibility trace	(6)
$r_{\text{pre}}(t), r_{\text{post}}(t)$	Exponentially filtered firing rates	(11)–(12)
$R(t)$	Instantaneous reward signal	(13)
$\bar{R}(t)$	Reward baseline (running average)	(15)
$M(t)$	Neuromodulatory signal (RPE)	(14)

2.4 Summary of Symbols

2.5 Parameter Values

References

- [1] Wulfram Gerstner, ed. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. S.l.: Cambridge Univ. Press, 2014. ISBN: 978-1-107-63519-7 978-1-107-06083-8. DOI: [10.1017/CB09781107447615](https://doi.org/10.1017/CB09781107447615).
- [2] Peter Dayan and Laurence F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Computational Neuroscience. Cambridge, Mass.: MIT Press, 2001. ISBN: 978-0-262-04199-7 978-0-262-54185-5.
- [3] Nicolas Frémaux and Wulfram Gerstner. “Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules”. In: *Frontiers in Neural Circuits* 9 (Jan. 2016), p. 85. ISSN: 1662-5110. DOI: [10.3389/fncir.2015.00085](https://doi.org/10.3389/fncir.2015.00085). (Visited on 12/12/2025).
- [4] Guo-qiang Bi and Mu-ming Poo. “Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type”. In: *The Journal of Neuroscience* 18.24 (Dec. 1998), pp. 10464–10472. ISSN: 0270-6474, 1529-2401. DOI: [10.1523/JNEUROSCI.18-24-10464.1998](https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998). (Visited on 02/18/2026).
- [5] A Destexhe et al. “Kinetic Models of Synaptic Transmission”. In: ().

Table 2: Model parameters and representative values. Ranges are drawn from the cited experimental and modeling literature.

Parameter	Description	Typical value	Source
τ_m	Membrane time constant	10–20 ms	[2]
R_m	Membrane resistance	10–100 M Ω	[2]
C_m	Membrane capacitance	τ_m/R_m	—
E_L	Resting potential	–70 mV	[2]
θ	Spike threshold	–55 mV	[2]
V_{reset}	Reset potential	–70 mV	[2]
τ_{ref}	Absolute refractory period	2–5 ms	[1]
τ_s	Synaptic time constant (PSC)	2–10 ms	[5]
w_{max}	Maximum synaptic weight	model-dependent	—
η_+	LTP learning rate	model-dependent	—
η_-	LTD learning rate	model-dependent	—
τ_+	Potentiation window width	20–40 ms	[4]
τ_-	Depression window width	20–40 ms	[4]
τ_e	Eligibility trace decay	0.1–1.0 s	[1]
τ_r	Rate-estimation time constant	50–200 ms	—
$\tau_{\bar{R}}$	Reward baseline time constant	1–10 s	[3]