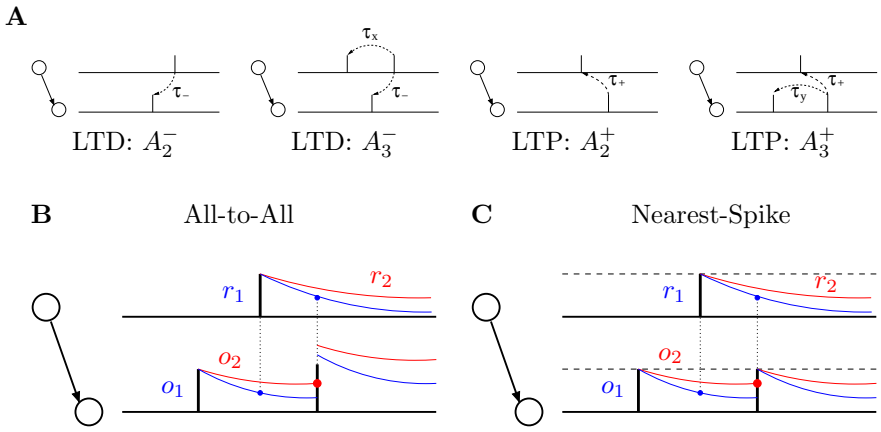

Figure 1: Schematic description of the triplet learning rules. **A**. Schematic description of the two terms contributing to long-term depression (LTD) controlled by A_2^- and A_3^- and the two long-term potentiation (LTP) terms controlled by A_2^+ and A_3^+. A presynaptic spike after a postsynaptic one (post $\rightarrow$ pre) induces LTD if the temporal difference is not much larger than τ_- (pair term, A_2^-). The presence of a previous presynaptic spike gives a further contribution (2-pre-1-post triplet term, A_3^-) if the interval between the two presynaptic spikes is not much larger than τ_x. Similarly, the triplet term for LTP depends on 1 presynaptic spike, but 2 postsynaptic spikes. The presynaptic spike must occur *before* the second postsynaptic one with a temporal difference not much larger τ_+. **B**. Time course of detectors of pre- and postsynaptic events r_1, r_2, o_1 and o_2. The presynaptic variables r_1 and r_2 are increased by a fixed amount upon arrival of a presynaptic spike. Analogously, postsynaptic variables are updated upon postsynaptic firing. With All-to-All interactions, each postsynaptic spike interacts with all previous postsynaptic spikes and vice versa, i.e. the internal variables r_1, r_2, o_1 and o_2 accumulate over several postsynaptic spike timings. The red and blue dots denote the values of those internal variables "read" by the triplet model whenever a spike occurs; e.g. the value of the postsynaptic variable o_1 is "read out" at the moment of presynaptic spike arrival leading to synaptic depression proportional to the momentary value of o_1 (blue dot). Similarly, the value of the presynaptic variable r_1 and the postsynaptic variable o_2 are "read out" at the moment of the second postsynaptic spike and determine the amplitude of synaptic potentiation. **C**. Same as in **B**, but with Nearest-Spike interactions: the extension of the spike interaction is restricted to the last spike; no accumulation occurs.

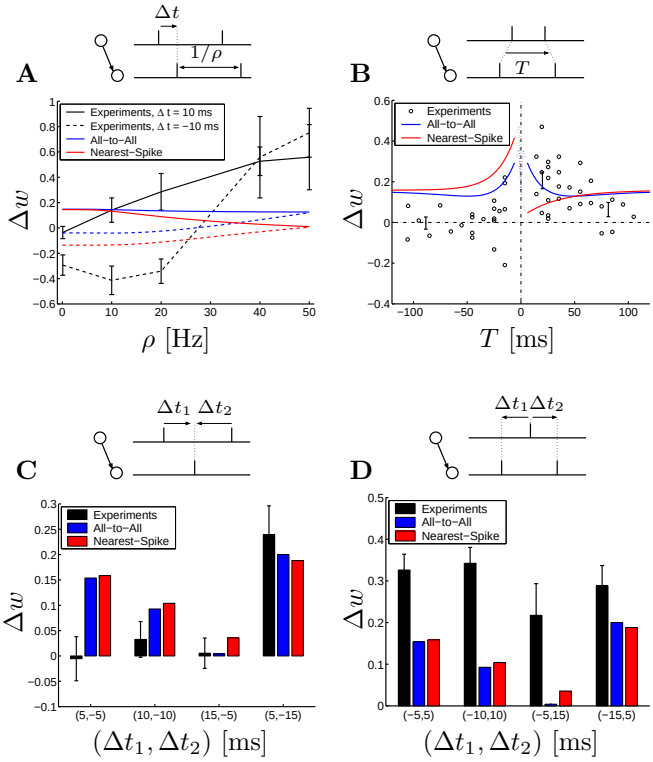

Figure 2: Failure of pair-based STDP learning rules. In all four subgraphs, black lines or symbols denote experimental data, blue lines correspond to the All-to-All pair model and the red lines to the Nearest-Spike pair model (see text for details). **A**. Weight change in a pairing protocol as a function of the frequency ρ (*solid lines*: $\Delta t = +10$ ms, *dashed lines*: $\Delta t = -10$ ms) . Black lines and data points (with errors) redrawn from Sjöström (2001).The experimental data are neither reproduced at high nor at low values of the repetition frequency ρ. **B**. Quadruplet protocol. Black circles are redrawn from Wang et al. (2005). **C**. Triplet protocol for the *pre-post-pre* case and **D** for the *post-pre-post* case. Black dots in **B** and black bars (and standard errors) in **C** and **D** from Wang et al. (2005). The asymmetry of the experimental results (no potentiation for $(\Delta t_1, \Delta t_2) = (5 \text{ ms}, -5 \text{ ms})$ in **C** but strong potentiation for $(-5 \text{ ms}, 5 \text{ ms})$ in **D** is not captured by the pair-based models.

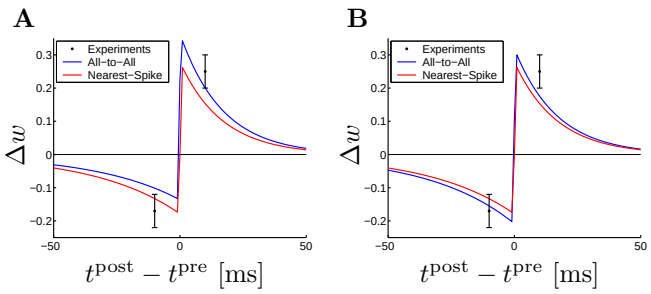

Figure 3: The triplet learning rule can reproduce the STDP learning window. Weight change induced by a repetition of 60 pairs of pre- and postsynaptic spike with a delay of Δt at a repetition frequency of 1 Hz. **A**. Weight change as a function of the time difference between post- and presynaptic spike timing for the full triplet model and **B** for the minimal triplet model. The parameters taken for the triplet models are those which correspond to the hippocampal culture data. See table **??**. Experimental data points and standard errors redrawn from Wang et al. (2005).

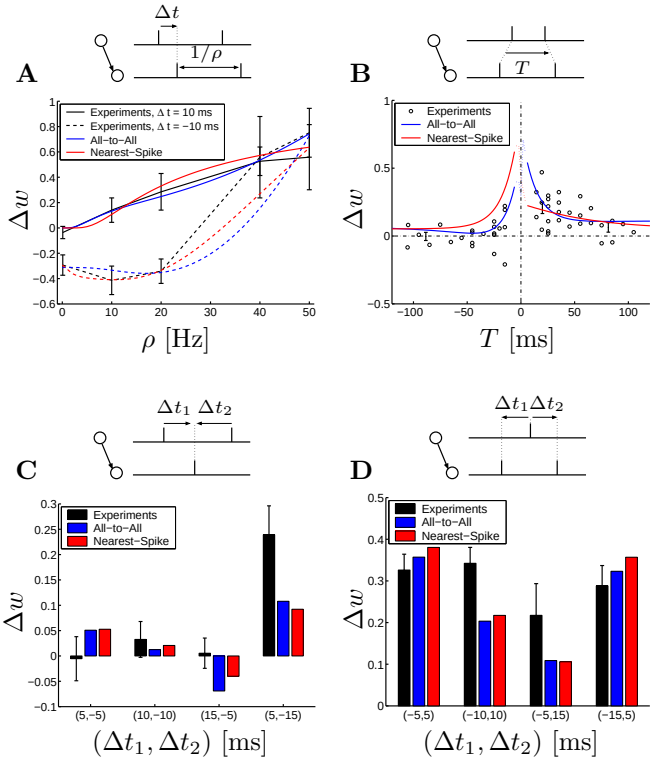

Figure 4: The full triplet learning rule succeeds to reproduce the pairing experiment and most of the triplet and quadruplet experiments. In all four subgraphs, black lines or circle denote experimental data, blue lines correspond to the All-to-All pair model and the red lines to the Nearest-Spike pair model. **A**. Weight change in a pairing protocol as a function of the frequency ρ (*solid lines*: $\Delta t = +10$ ms, *dashed lines*: $\Delta t = -10$ ms) . Black lines and data points (with errors) redrawn from Sjöström (2001). **B**. Quadruplet protocol. Black circles are redrawn from Wang et al. (2005). **C**. Triplet protocol for the *pre-post-pre* case and **D** for the *post-pre-post* case. Black dots in **B** and black bars (and standard errors) in **C** and **D** from Wang et al. (2005). The triplet-based models succeed to reproduce the asymmetry in triplets protocols (no potentiation for $(\Delta t_1, \Delta t_2) = (5 \text{ ms}, -5 \text{ ms})$ in **C** and strong potentiation for $(-5 \text{ ms}, 5 \text{ ms})$) in **D**: for those triplets the model results (with All-to-All interactions) are within 1.1 σ (standard error of experimental data) whereas the results of the pair-based models are off by more than 4 σ.

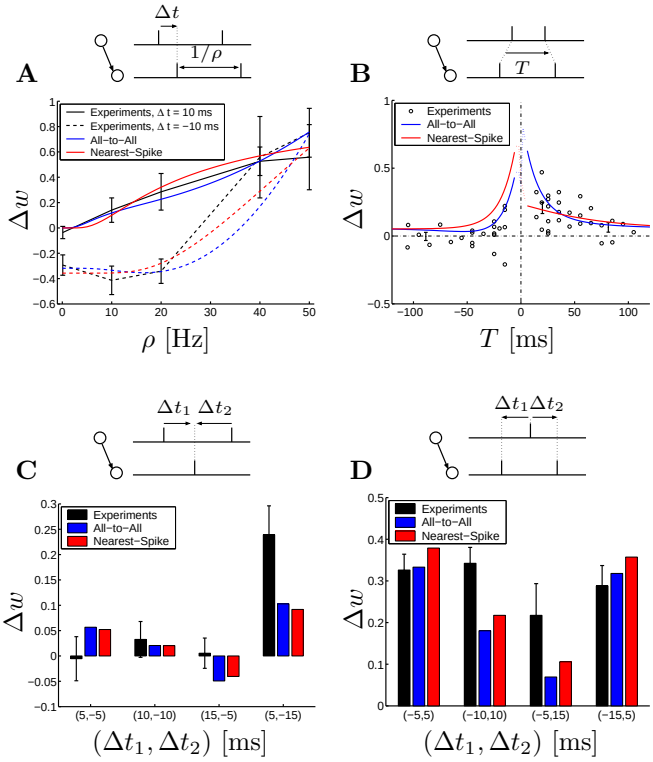

Figure 5: Minimal triplet learning rules are almost as good as full triplet learning rules. In all four subgraphs, black line or circle denote experimental data, blue lines correspond to the All-to-All pair model and the red lines to the Nearest-Spike pair model. **A**. Weight change in a pairing protocol as a function of the frequency ρ (*solid lines*: $\Delta t = 10$ ms, *dashed lines*: $\Delta t = -10$ ms) . Black lines and data points (with errors) redrawn from Sjöström (2001). **B**. Quadruplet protocol. Black circles are redrawn from Wang et al. (2005). **C**. Triplet protocol for the *pre-post-pre* case and **D** for the *post-pre-post* case. Black dots in **B** and black bars (and standard errors) in **C** and **D** from Wang et al. (2005).

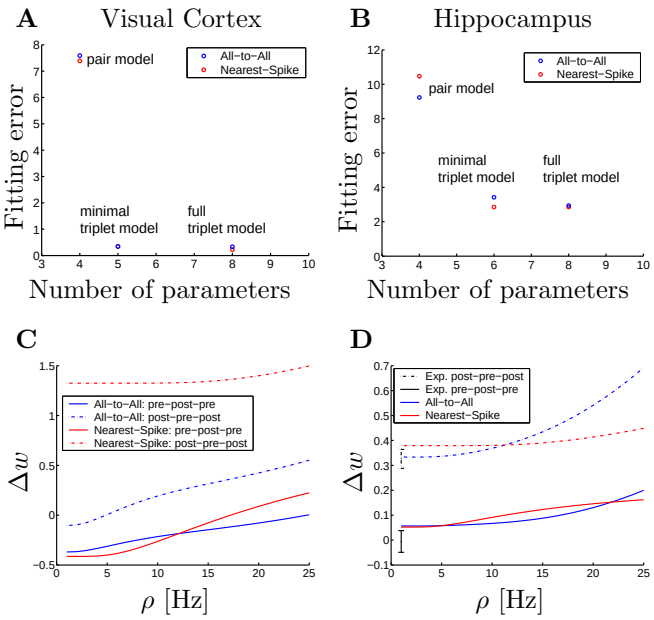

Figure 6: **A** and **B**: Comparison between the pair and triplet models. **C** and **D** Predictions of the triplet models. **A**. Fitting error (c.f. Eq. 5) for the visual cortex data set of Sjöström et al. (2001) as a function of the number of parameters in the model. The minimal model has only one extra parameter compared to a pair-based model but performs more than 20 times better. **B**. Fitting error for the hippocampal data set of Wang et al. (2005). **C** Predicted weight change (*visual cortex*) of the triplet protocol (*solid lines*: pre-post-pre with $(\Delta t_1, \Delta t_2) = (+5, -5)$ ms, *dot-dashed lines*: post-pre-post with $(\Delta t_1, \Delta t_2) = (-5, +5)$ ms) with All-to-All interactions (*blue lines*) and with Nearest-Spike interactions (*red lines*). **D**. Same as in **C** but for the hippocampal culture data set. Black bars correspond to the experimental results also present in subplots C and D of Figs. 2, 4 and 5.

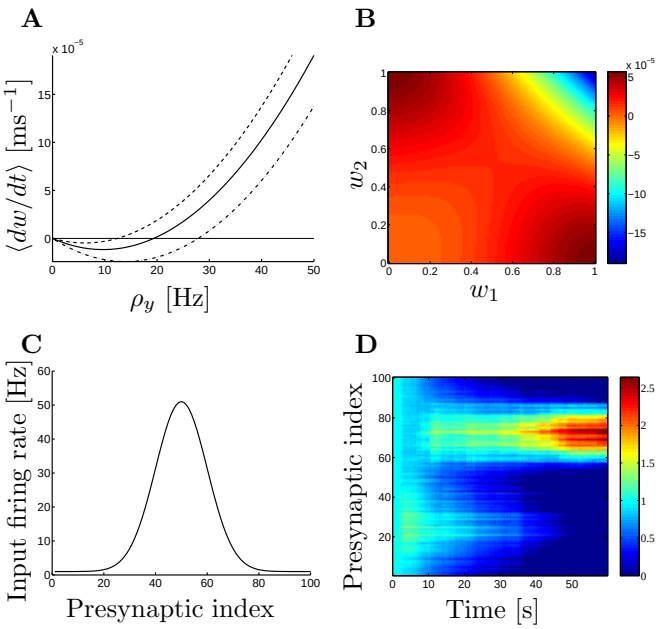

Figure 7: The triplet learning rule can be mapped to a BCM learning rule. **A**. Instantaneous weight change as a function of the postsynaptic frequency for a minimal triplet model. (compare Eq. 6 with $A_3^- = 0$). The pre- and postsynaptic spike trains are Poisson spike trains. The *dashed line* corresponds to $\lambda = \bar{\rho}_y^p/\rho_0^p = 0.64$, *solid line*: $\lambda = 1$ and *dot-dashed line*: $\lambda = 1.44$. **B**. Energy landscape produced by the minimal triplet learning rule (with $p = 2$ and $\rho_0 = 10$ Hz) in a two-input environment: $\rho_x^1 = (10\text{ Hz}, 0)^T$ and $\rho_x^2 = (0, 10\text{ Hz})^T$. The presence of two specialized (and stable) fixed point as well as two unspecialized (and unstable) fixed points is an essential feature of the BCM learning rule. **C**. Gaussian stimulation profile across 100 presynaptic neurons. The center of the Gaussian is shifted randomly every 200 ms to one of 10 random positions. Periodic boundary conditions are assumed. **D**. Evolution of the 100 weights as a function of time under the stimulation described in C. After one minute of stimulation, the postsynaptic neuron becomes sensitive to a stimulation centered around the 70^{th} presynaptic neuron. The parameters taken in the minimal model are those which correspond to the visual cortex (compare Tables 3 and 4).