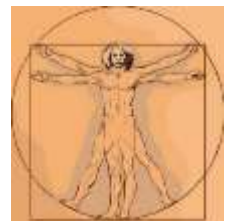


Information theoretical approaches to model synaptic plasticity

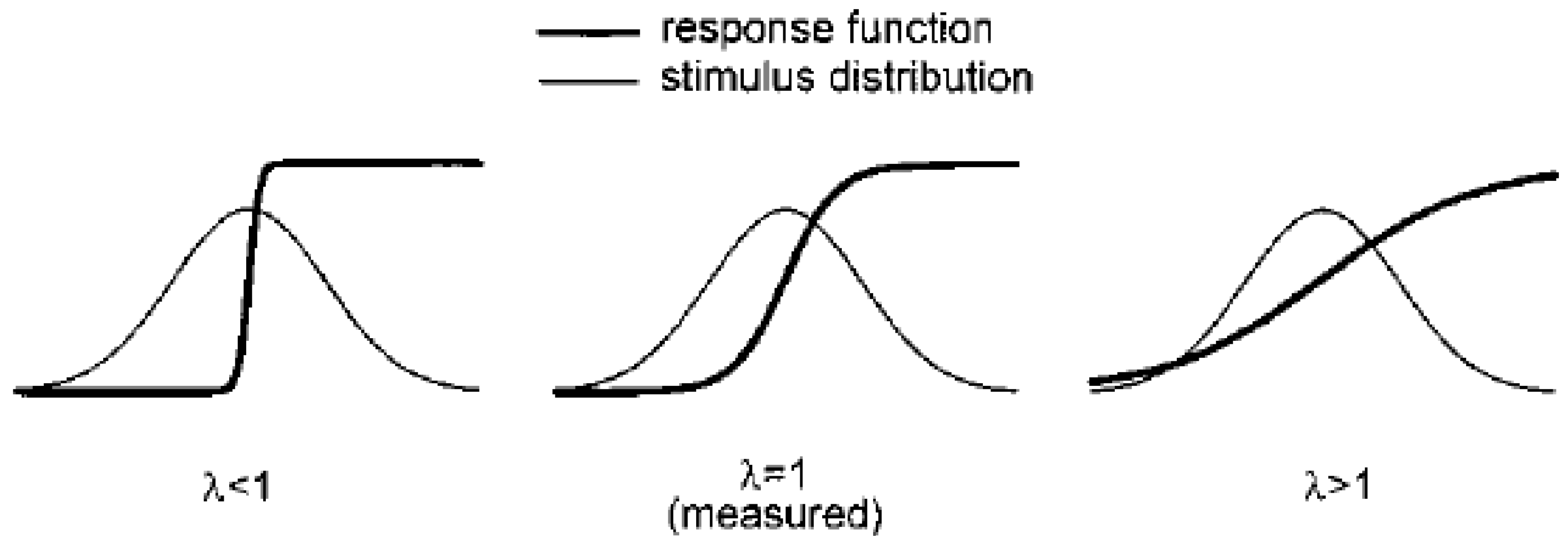
Taro Toyoizumi
RIKEN Center for Brain Science

Adaptation



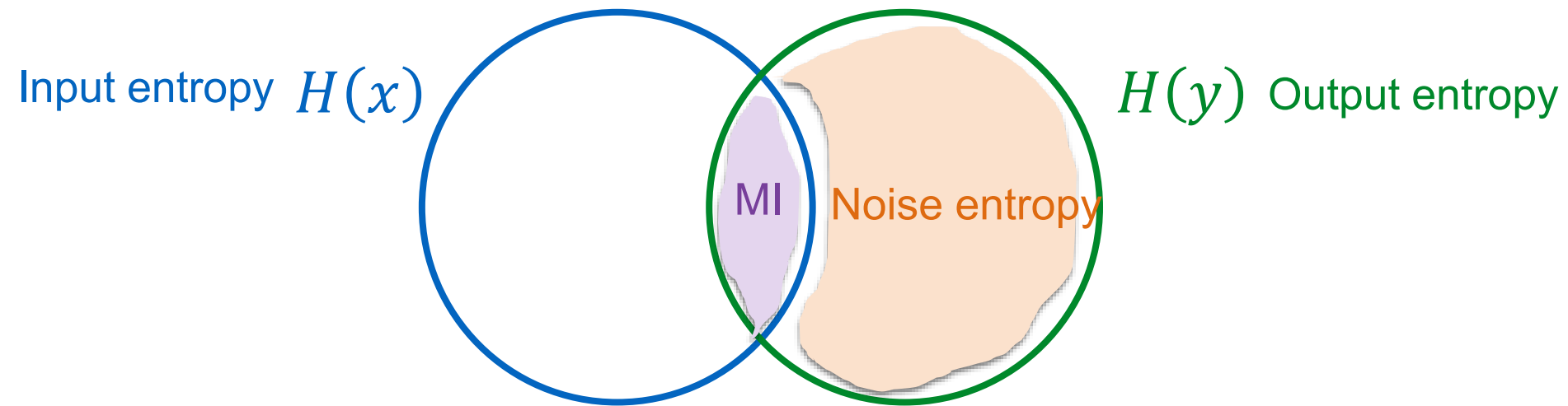
Optimal information representation

Which of the following setups is suitable for information representation?



Mutual information maximization

Visualizing the variability of random variables

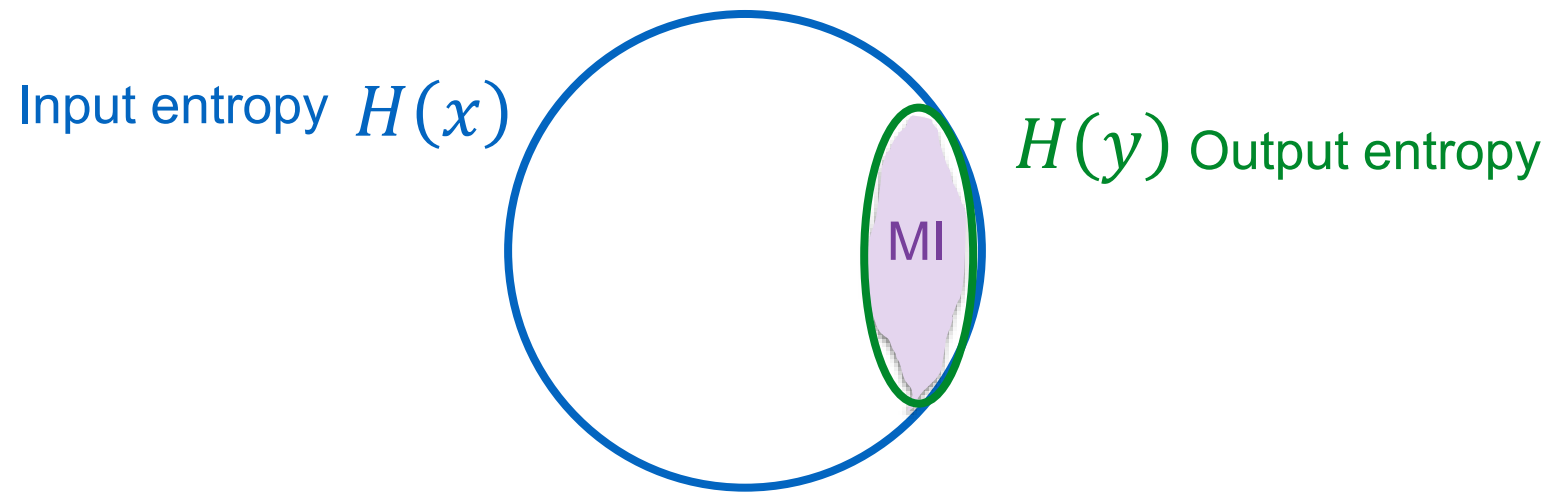


$$I(x; y) = H(y) - H(y|x)$$

The mutual information I quantifies the amount of information about y gained by observing x .

Mutual information maximization

Visualizing the variability of random variables



$$I(x; y) \approx H(y)$$

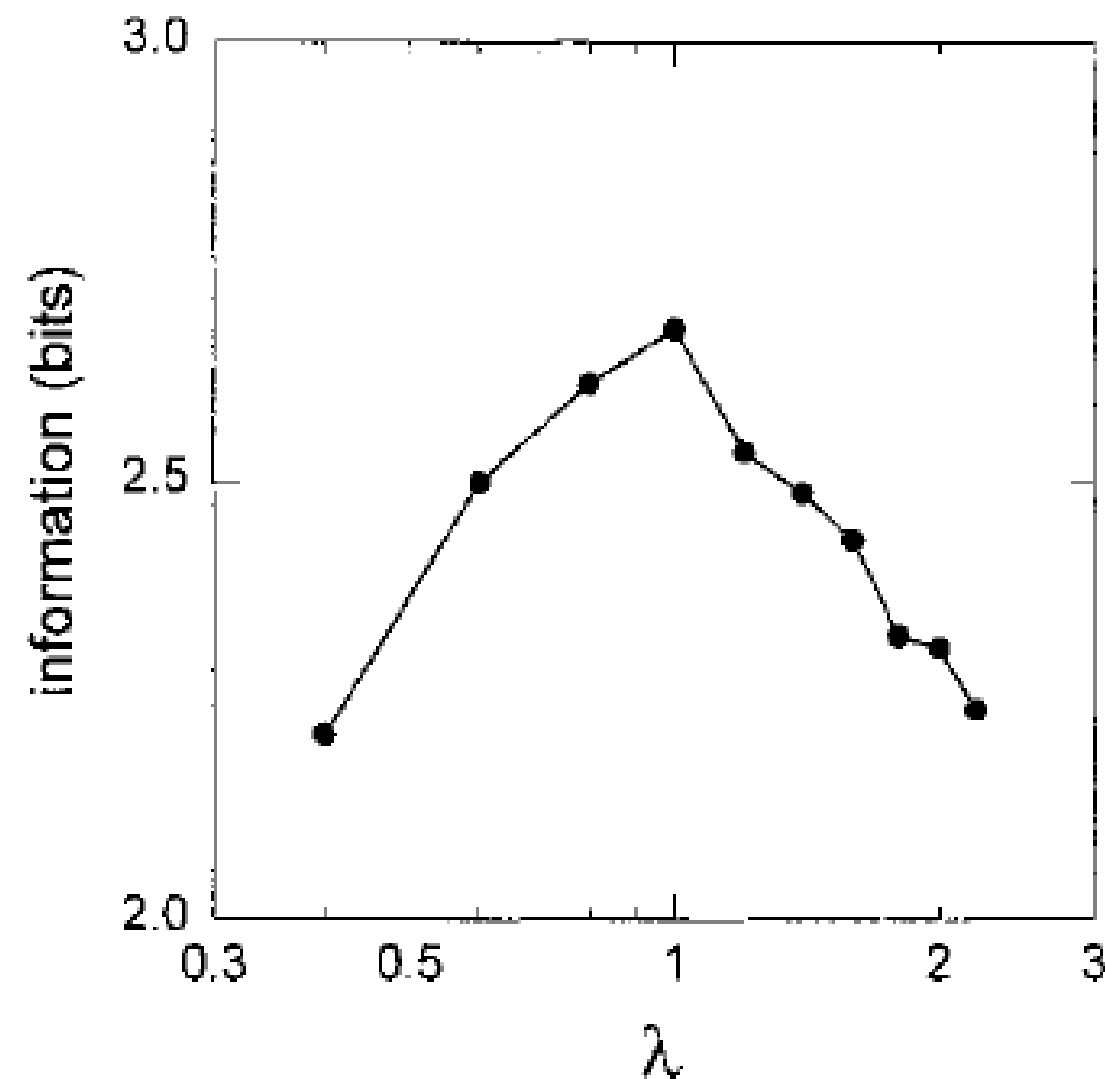
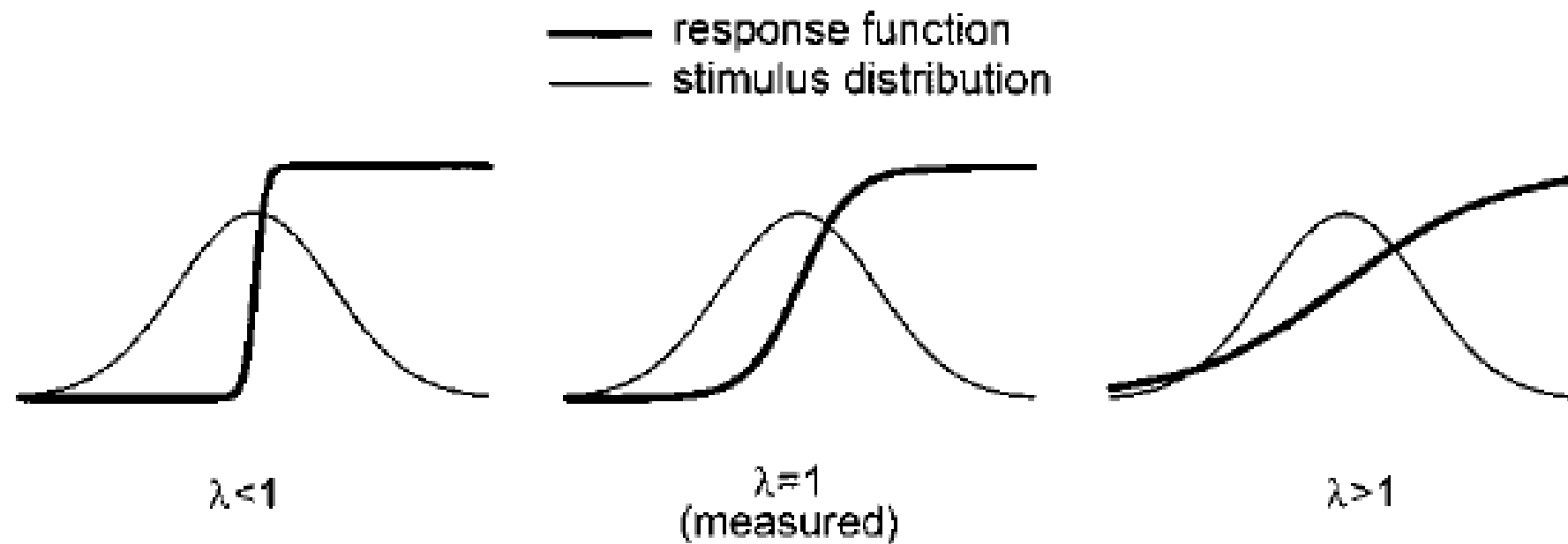
The mutual information I quantifies the amount of information about y gained by observing x .

If y is deterministically set by x (i.e., $y = f(x)$), the noise entropy $H(y/x)$ is negligible.

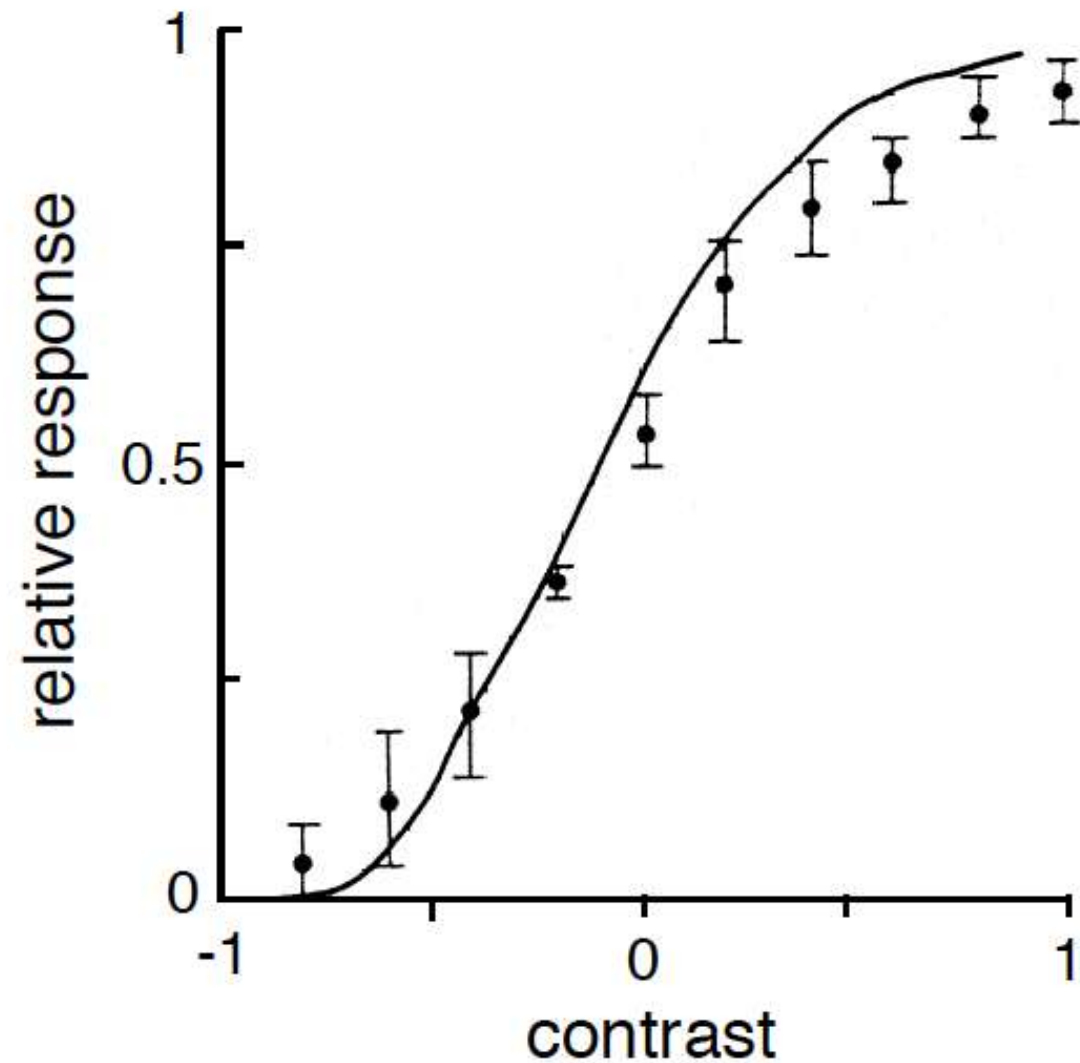
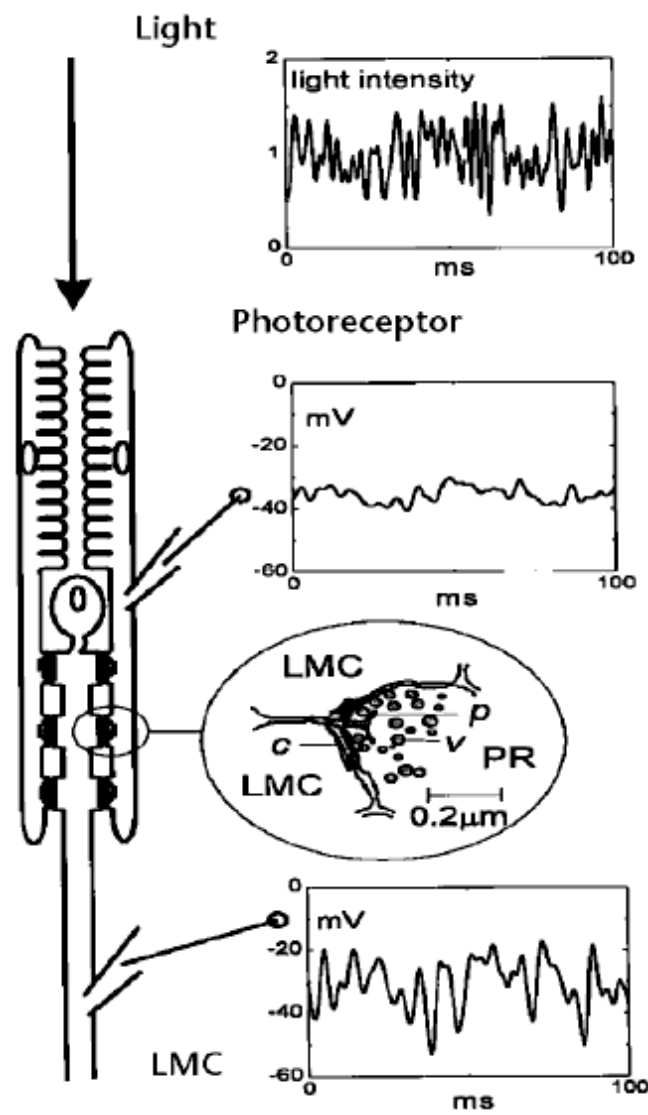
$$\begin{aligned} I(x; y) &\approx H(y) \\ &= -\int dy \, p(y) \log p(y) \\ &= -\int dx \, p(x) \log \left[\left| \frac{dx}{dy} \right| p(x) \right] \\ &= -\text{KL}[p(x) || f'(x)] \end{aligned}$$

Histogram equalization: The optimal nonlinearity is the cumulative distribution of x so that $p(y)=\text{const.}$

The histogram equalization

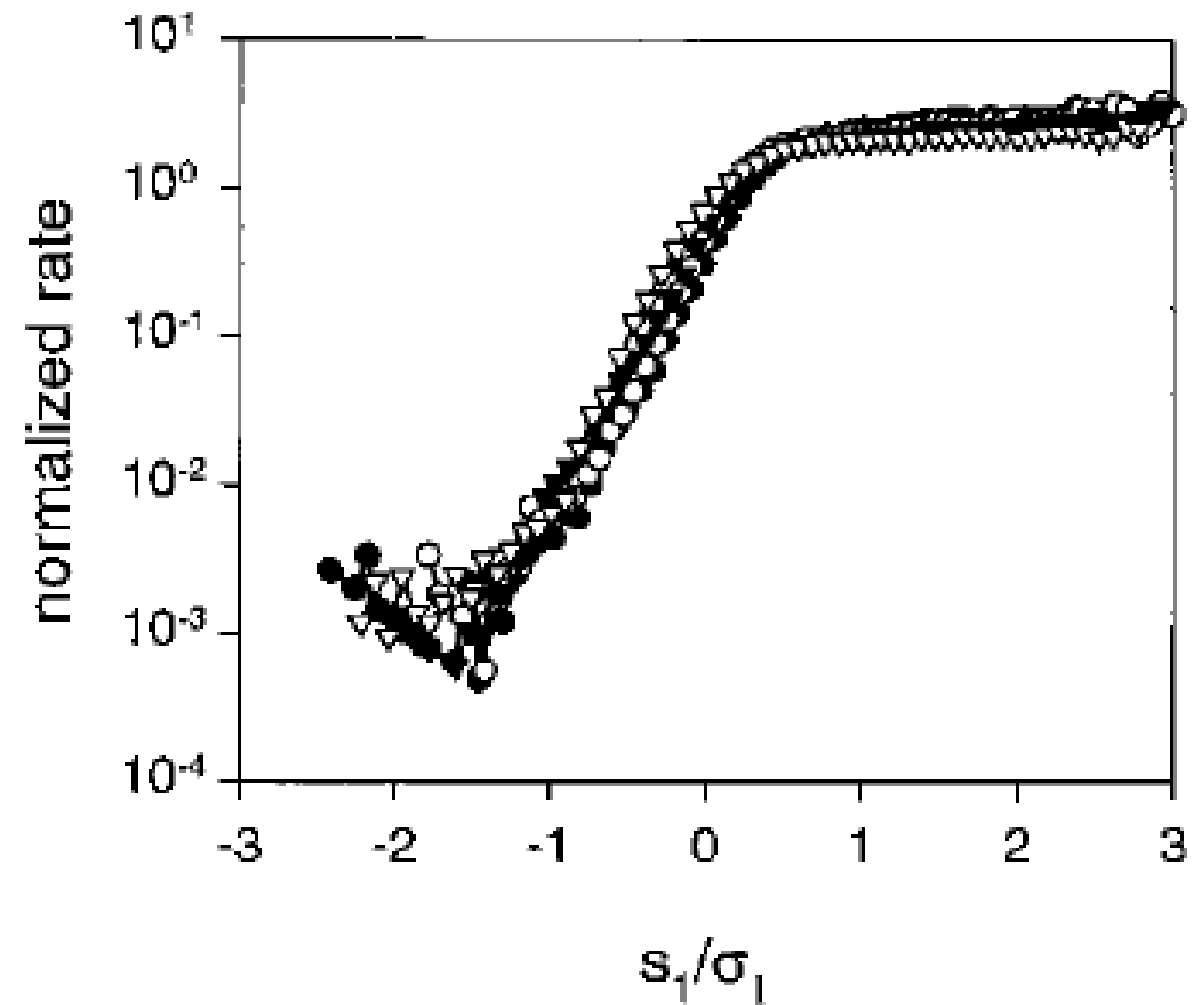
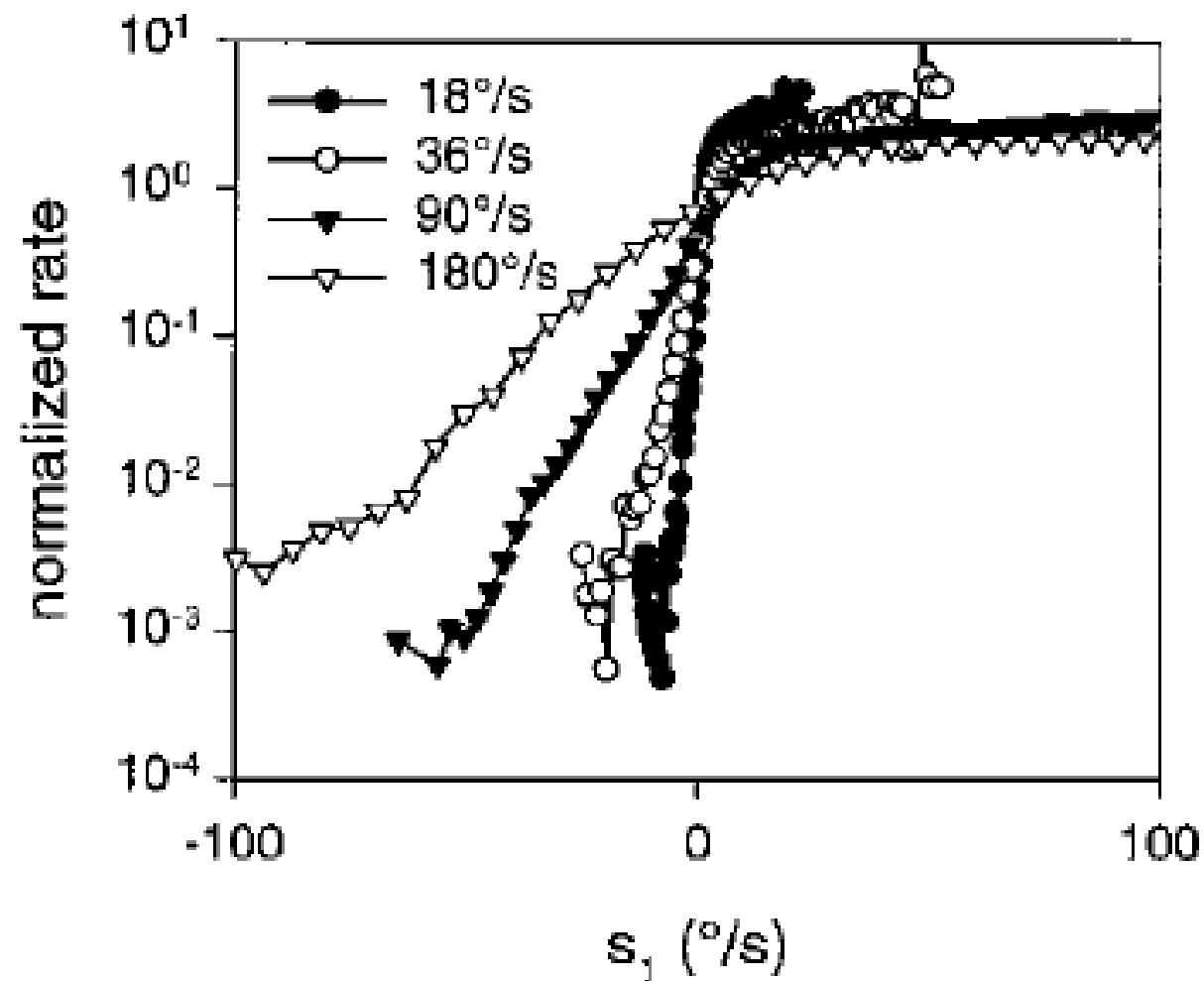


Flies efficiently use neuronal dynamic range



Adaptive rescaling maximizes information transmission

- Recordings from fly H1 neuron

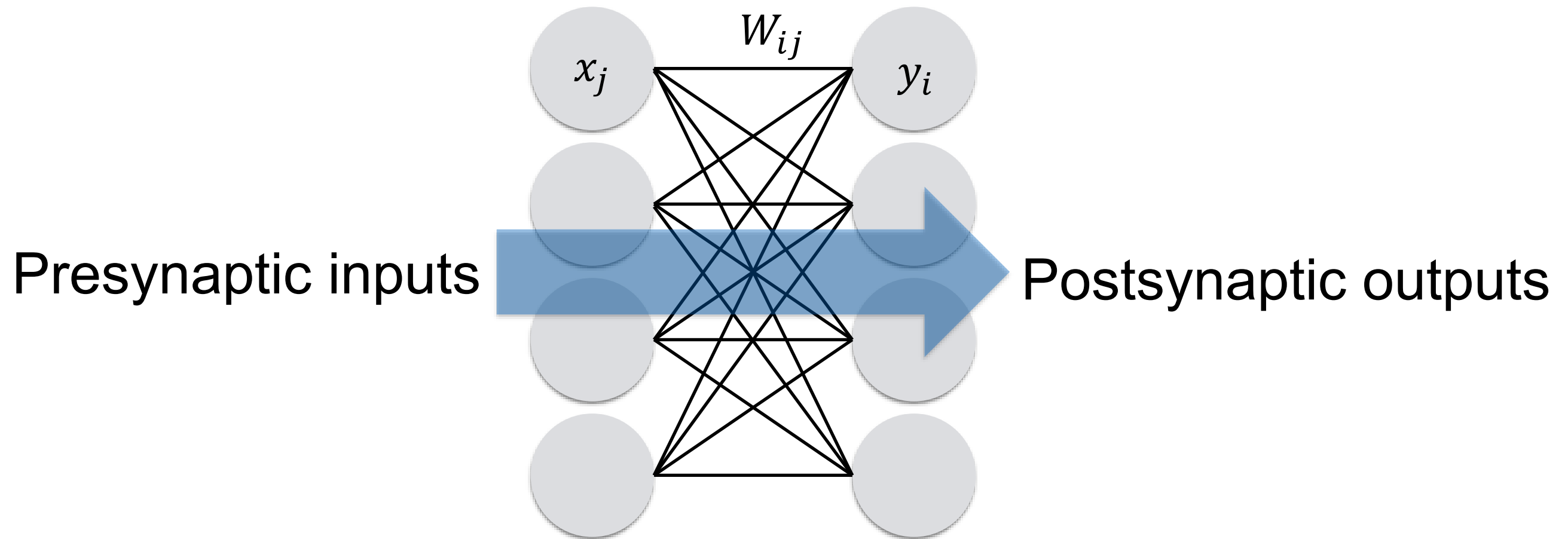


Brenner, Bialek, and de Ruyter van Steveninck, 2000

Information maximization (Infomax):

A synaptic plasticity rule for maximizing information transmission

(Linsker 1988; Bell and Sejnowski 1995)

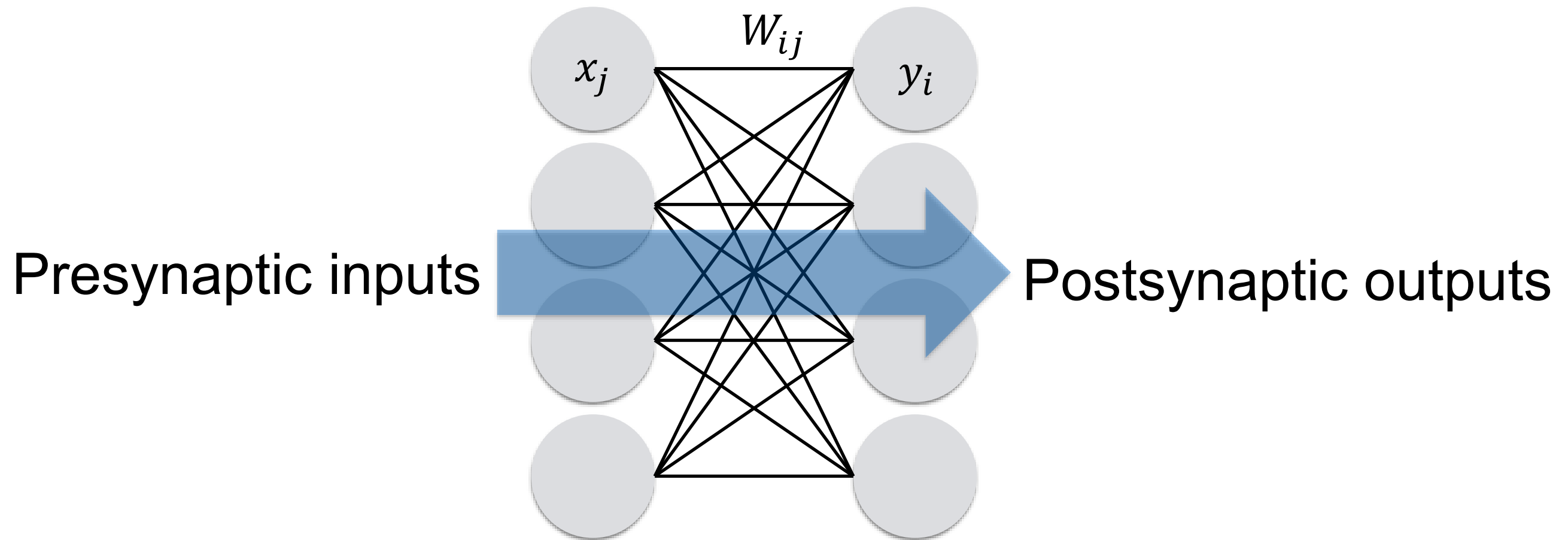


The Information transfer is maximized in the presence of small noise when individual neurons encode independent features (Bell and Sejnowski 1995).

Information maximization (Infomax):

A synaptic plasticity rule for maximizing information transmission

(Linsker 1988; Bell and Sejnowski 1995)



The Information transfer is maximized in the presence of small noise when individual neurons encode independent features (Bell and Sejnowski 1995).

But the learning rule is nonlocal, and its biological implementation is unknown.

Error-gated Hebbian rule (EGHR)

$$u = Wx$$

Bell and Sejnowski 1995

$$\dot{W}_{ij} \propto (W^{-\top})_{ij} - g(u_i)x_j$$



Approximation near the ICA solution
(quadratic approximation)

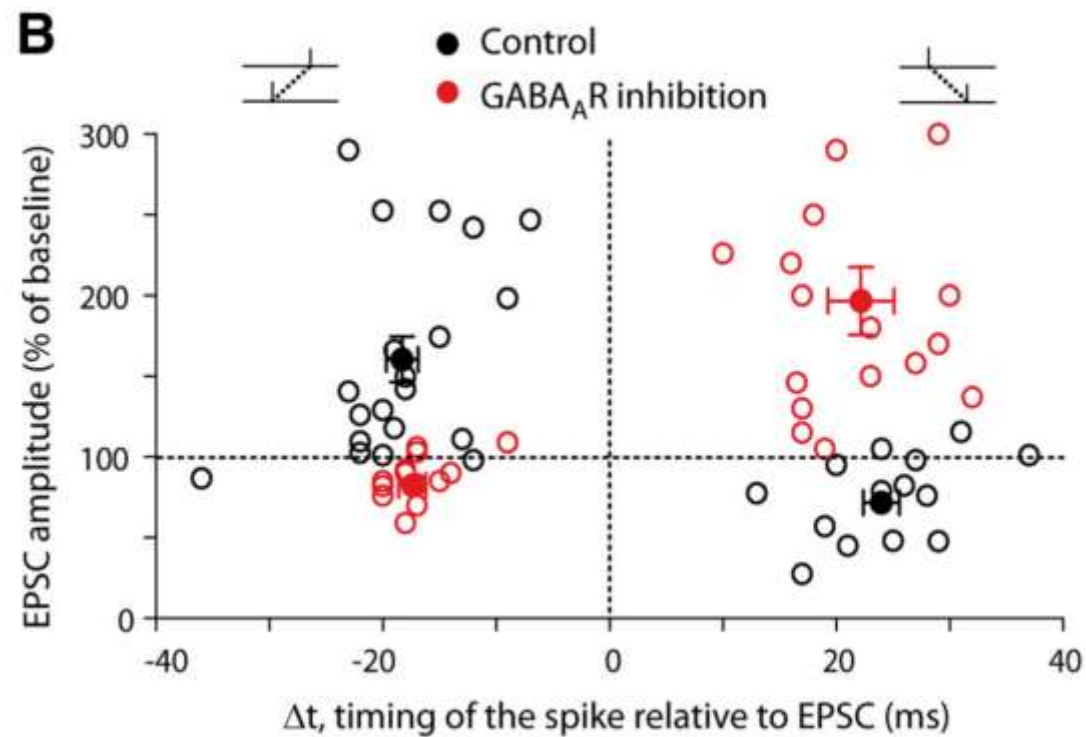
Isomura and Toyozumi 2016

$$\dot{W}_{ij} \propto E(u)g(u_i)x_j$$

$$E(u) \equiv E_0 - \sum_i G(u_i)$$

$$G'(u) \equiv g(u)$$

A local independent component analysis (ICA) rule: Modulation of Hebb's rule by a global third factor



Paille et al. 2013

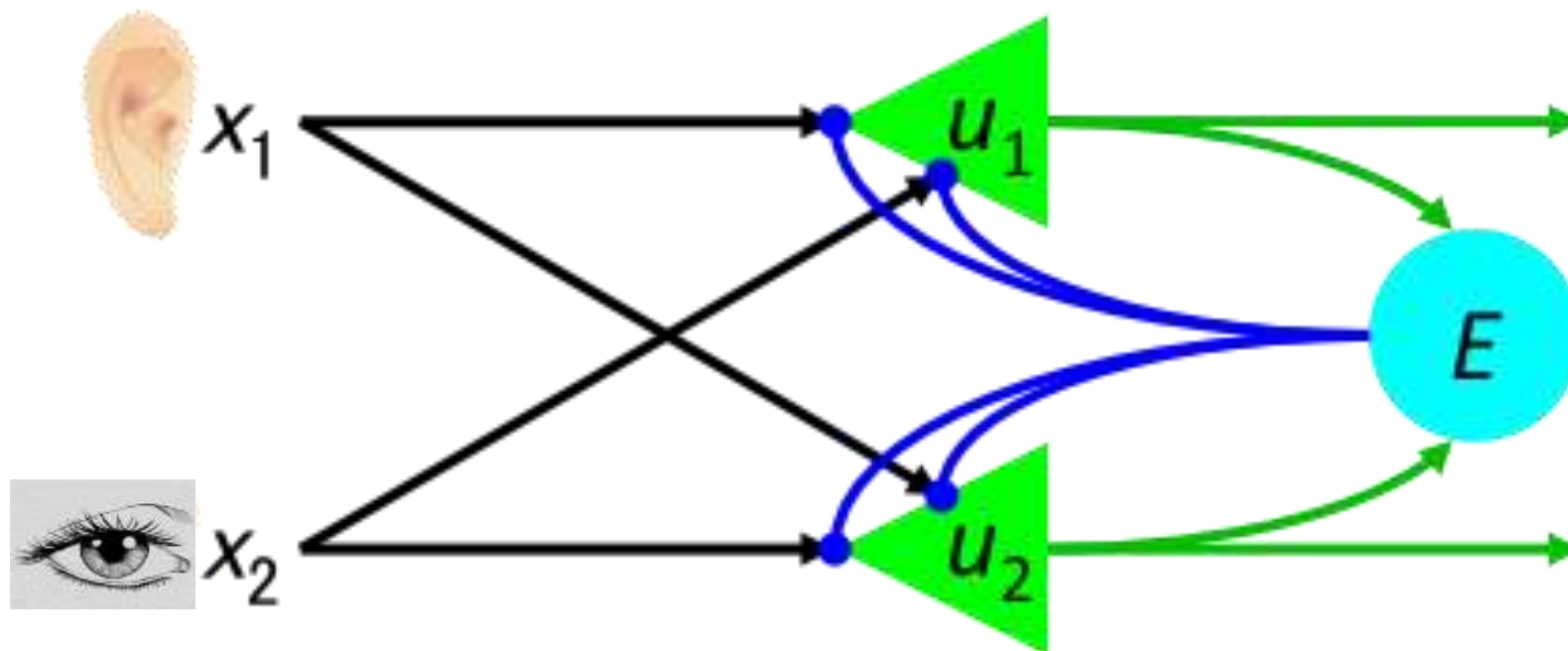
$$\dot{W}_{ij} \propto E \cdot y_i \cdot x_j$$

global scalar-factor
(e.g. GABA, neuromodulator, glial factor, etc.)

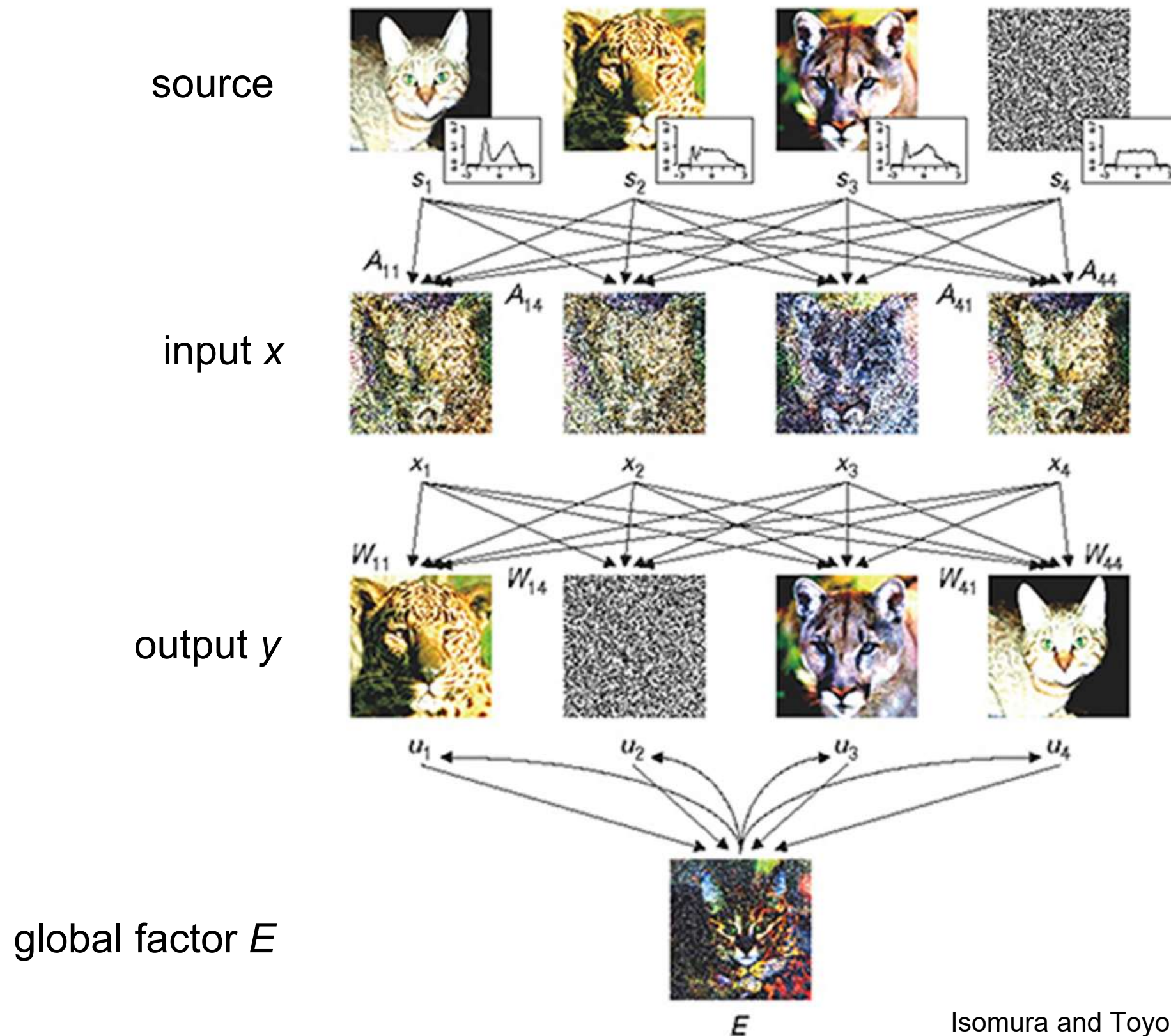
presynaptic
postsynaptic

$$E \equiv E_0 - \sum_i f(y_i)$$

Isomura and Toyoizumi 2016



Separating mixtures of natural images



Our learning rule does not require W to be a square matrix.
(The number of output neurons can be different from the input dimensions.)

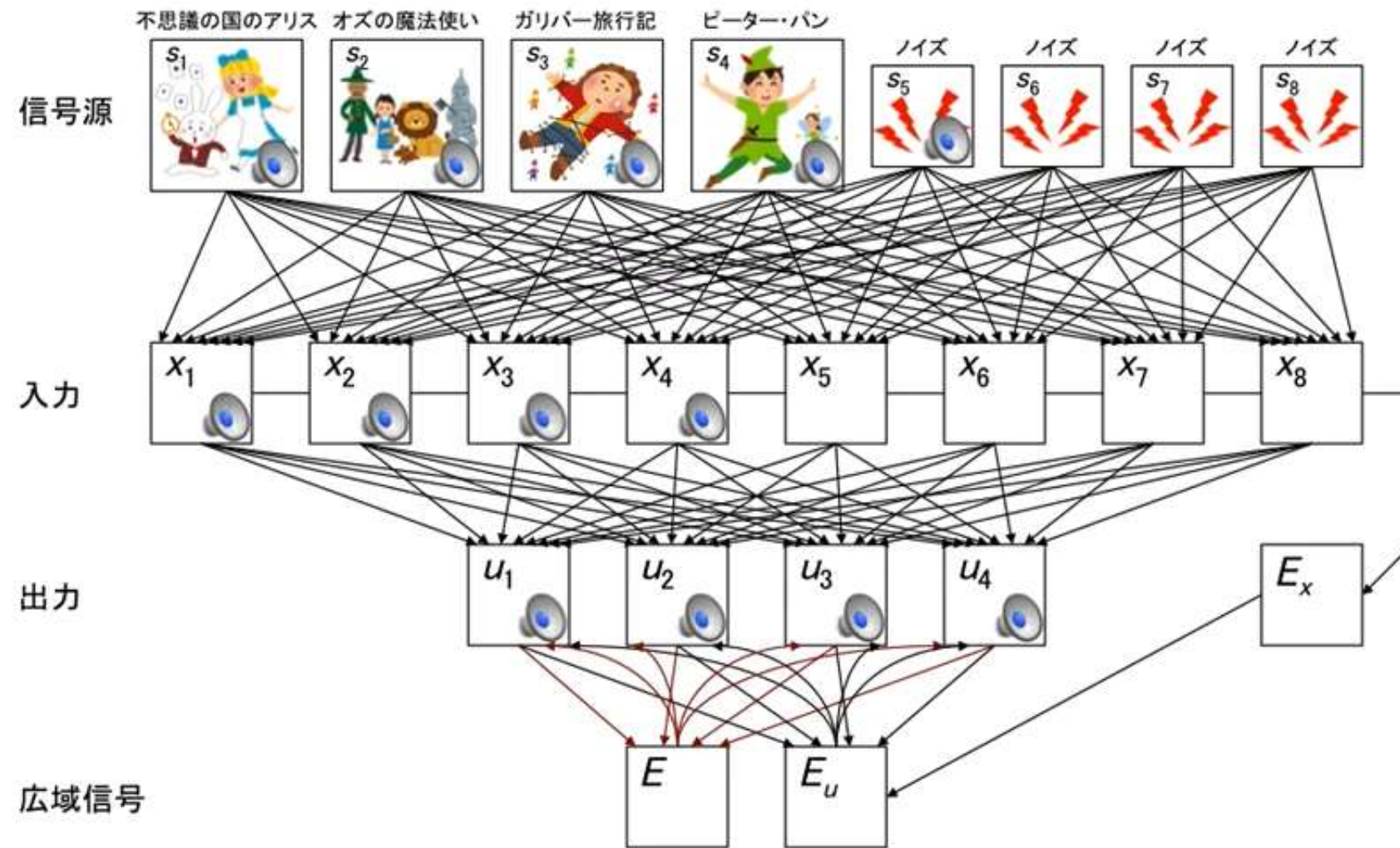
Three-factor learning not only separates sources but can also perform principal component analysis (PCA) with

$$E_{PCA}(x, u) \equiv \sum_j (x_j^2 - \langle x_j^2 \rangle) - \sum_i (u_i^2 - \langle u_i^2 \rangle)$$

without needing activity backpropagation like Oja 1989.

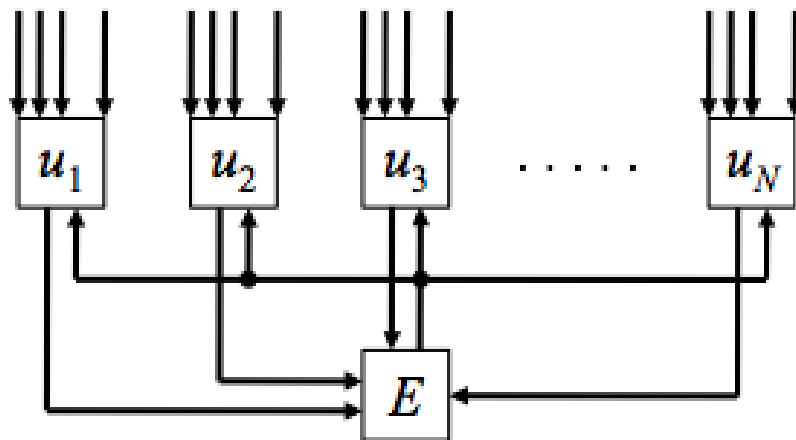
ICA + PCA

$$\dot{W}_{ij} \propto (1 - \beta)E(u)g(u_i)x_j + \beta E_{PCA}(x, u)u_ix_j$$

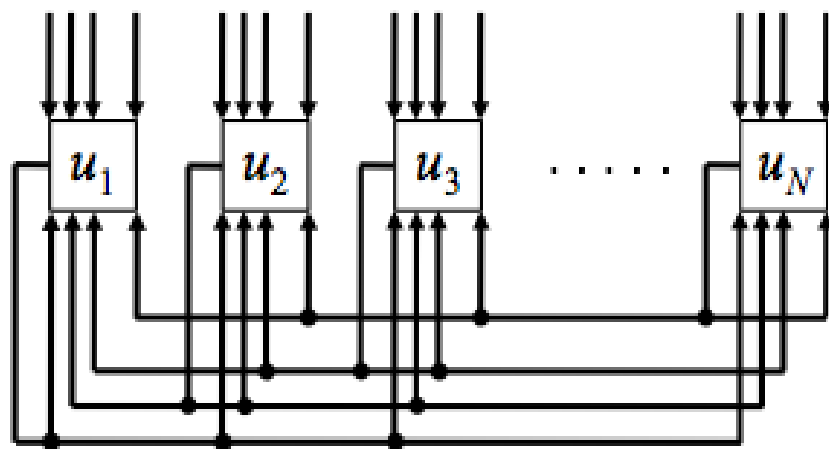


Parallel and energy-efficient neuromorphic computation

Proposed rule (parallel)



Conventional rules (serial)



Isomura and Toyozumi 2016

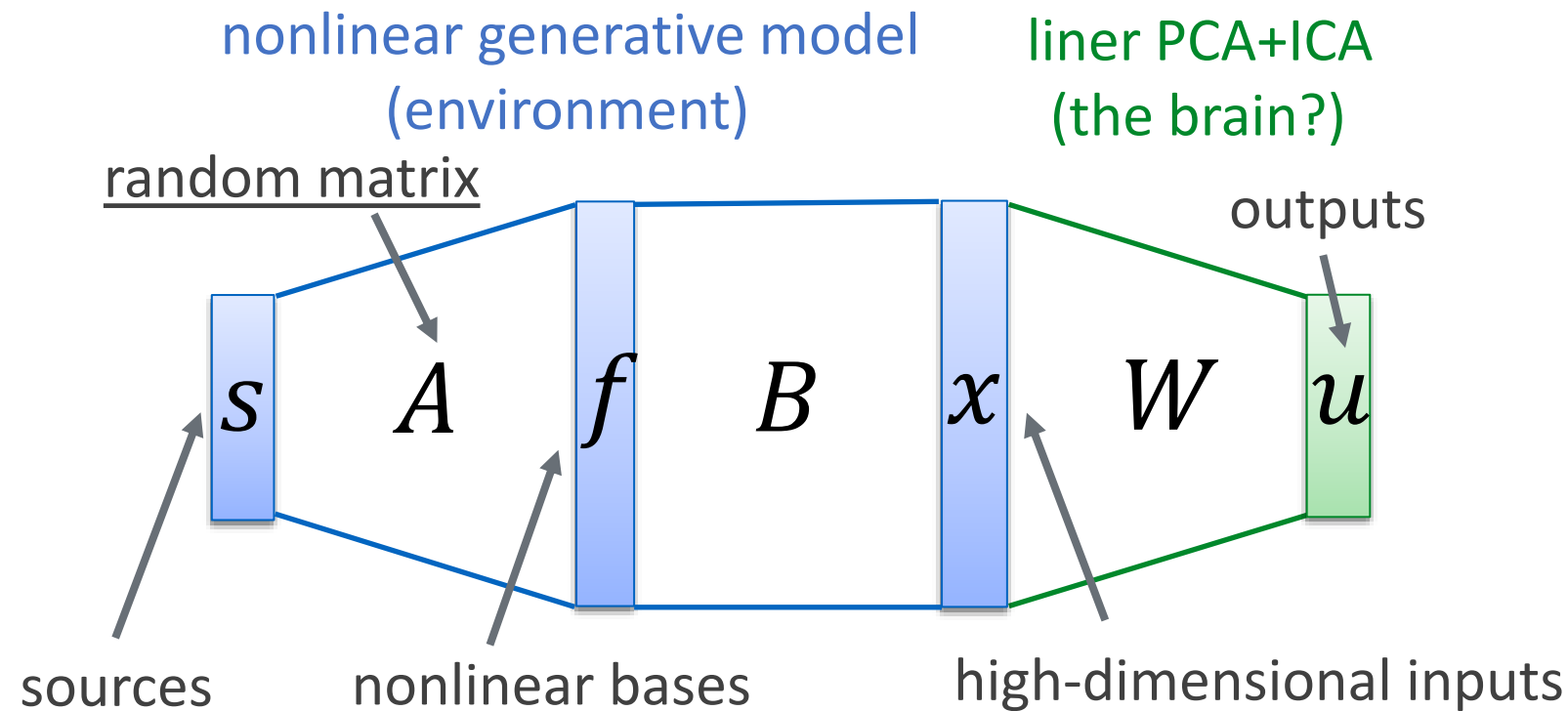
Recently, Isomura and Toyozumi have proposed an interesting biological plausible learning rule called Error-Gated Hebbian Rule (EGHR) [10] that enables local and efficient learning to find the independent components. But, the expensive part in this algorithm is the matrix-vector multiplication which can be calculated using RRAMs-based crossbar array offering a low power and efficient solution. In this letter, a RRAM-based hardware realization for ICA is investigated using EGHR for online evaluation of the weights. We demonstrate that this learning rule is capable of local and efficient learning, even when taking into consideration the RRAM non-idealities, the asymmetric nonlinear conductance, and device variability.

Fouda et al. 2018

But, in the real-world, signals nonlinearly mix.

Nonlinear ICA can be achieved by linear PCA+ICA

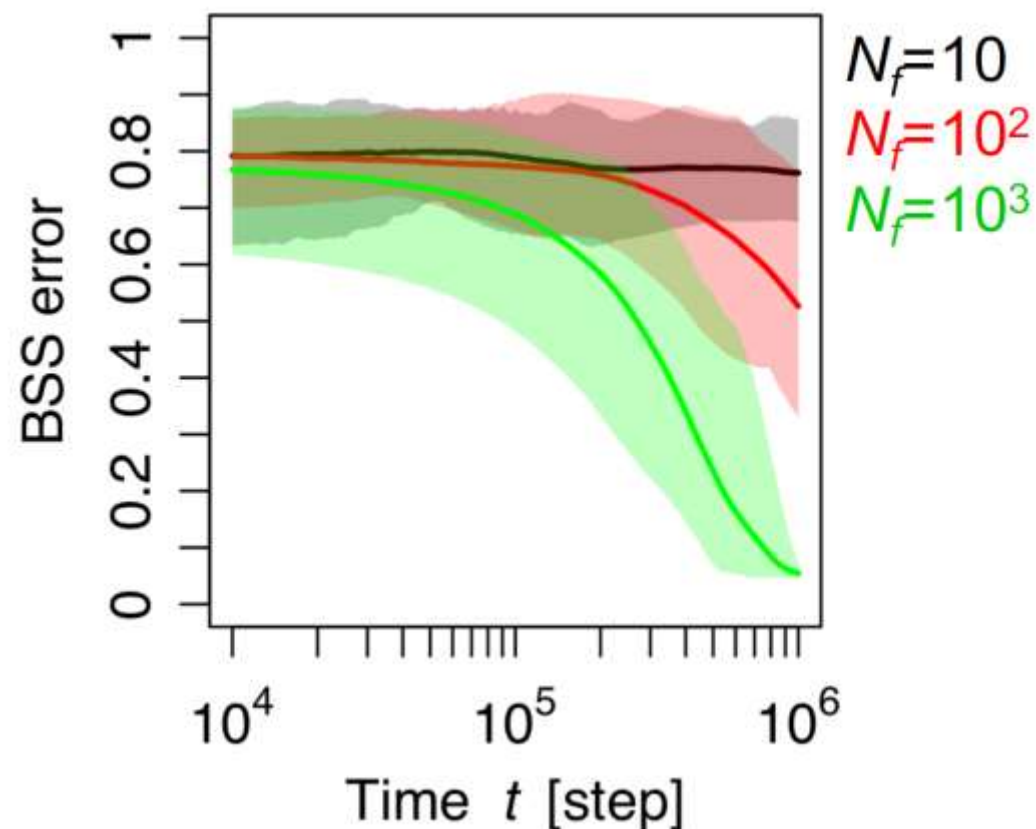
Isomura and Toyozumi, 2021a



$$x = Bf(As + a)$$

$$u = Wx$$

where $a \sim N(0, 1/N_s)$



The linear component dominates if there are many independent sources and bases per sources involved.

$$x \approx B\bar{f}'As$$

In this case, linear PCA can eliminate nonlinear components. Then, linear ICA can separate sources.

Asymptotically, zero source-separation error is achieved in the limit of large numbers of sources and bases per source.

In practice, source identification benefits from focusing on a relevant feature of inputs.

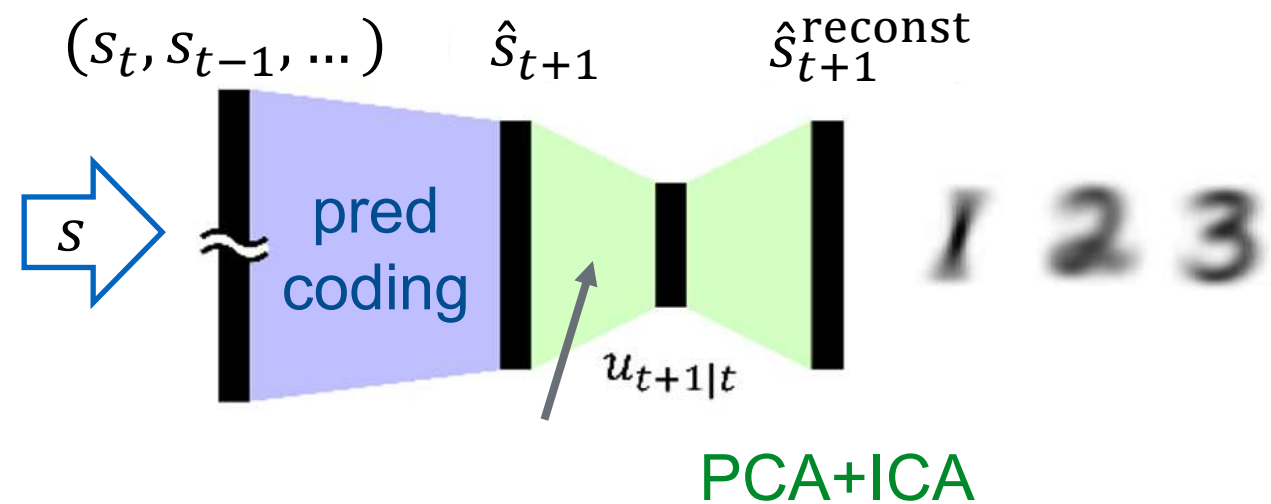


How to extract a relevant feature?

Nonlinear blind-source separation with high-dimensional sensors

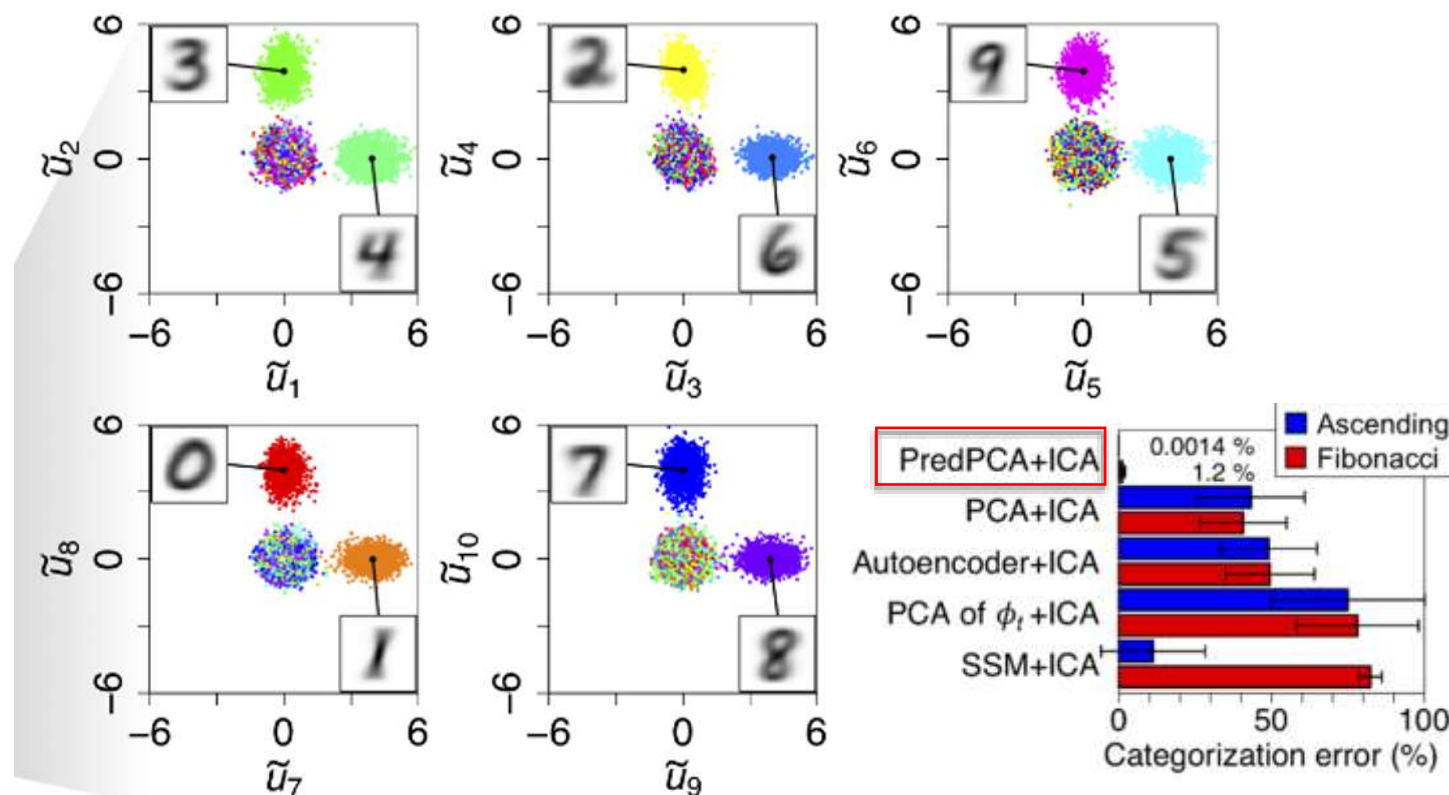
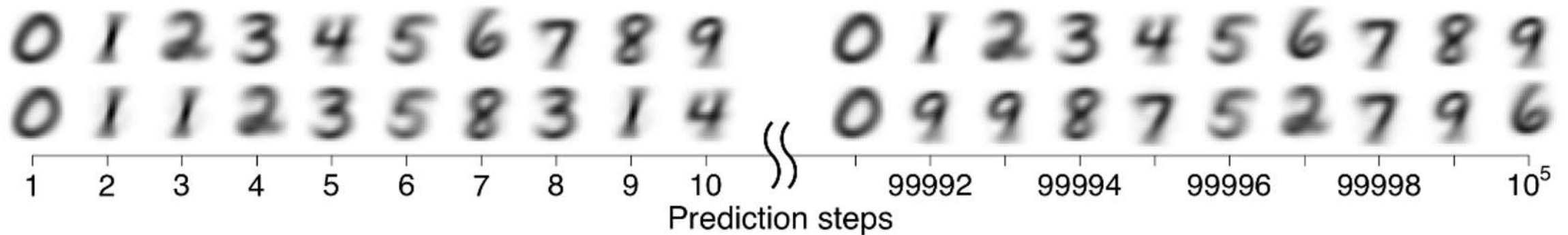
Dynamic environment

1 2 3 ...



Ascending

Fibonacci



Summary

- We have developed a local PCA&ICA algorithm, Error-gated Hebbian rule (EGHR), that requires only global information sharing by a third plasticity factor.
- EGHR is suitable for parallel blind-source separation and feature extraction by neuromorphic hardware.
- Some of the non-linear source separation problems can be solved by linear PCA&ICA.
- By focusing on predictable features, independent sources that underlie the dynamics can be efficiently separated.

Collaborator

Blind source separation

- Takuya Isomura (RIKEN)



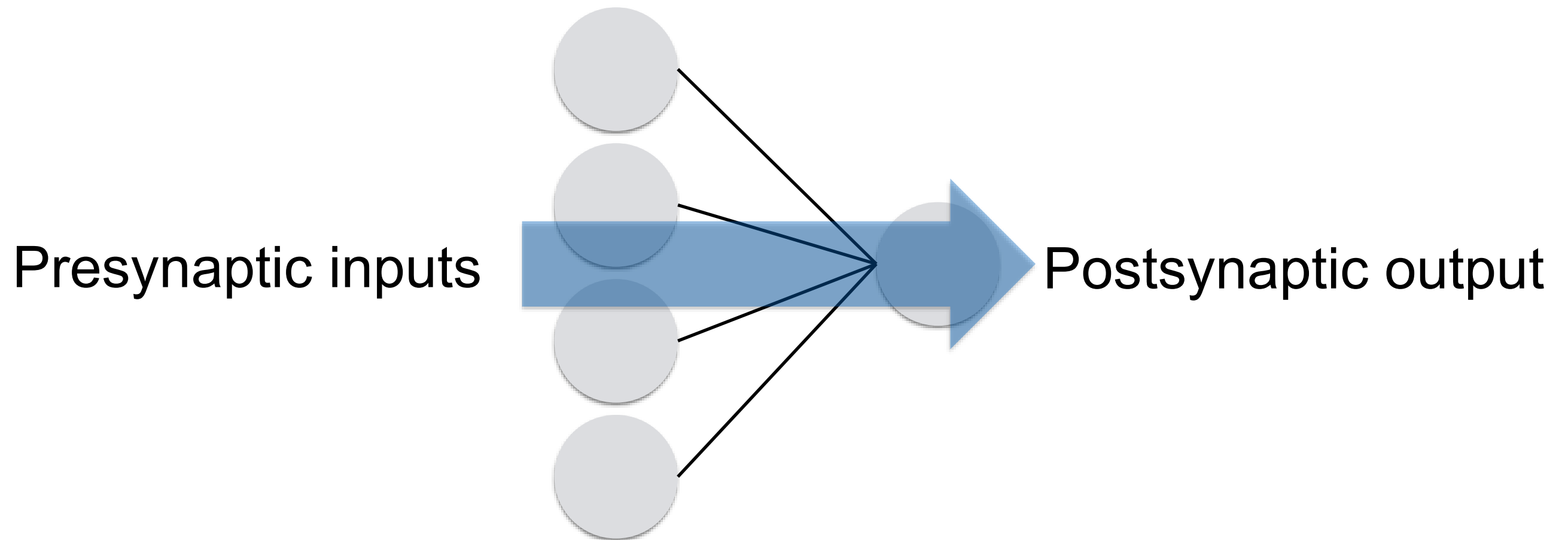
Funding



Brain/MINDS



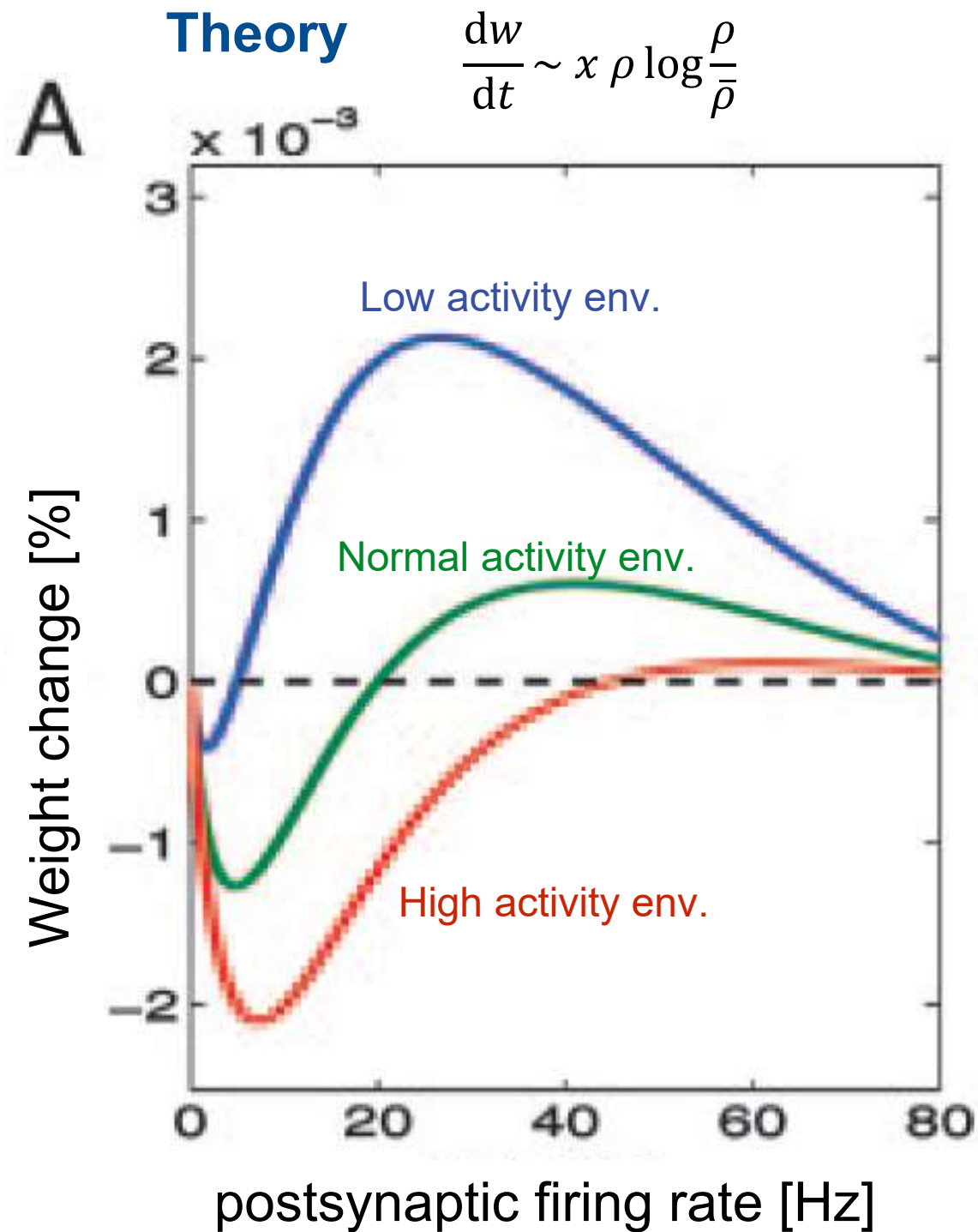
Infomax for a spiking output neuron: comparisons with experiments



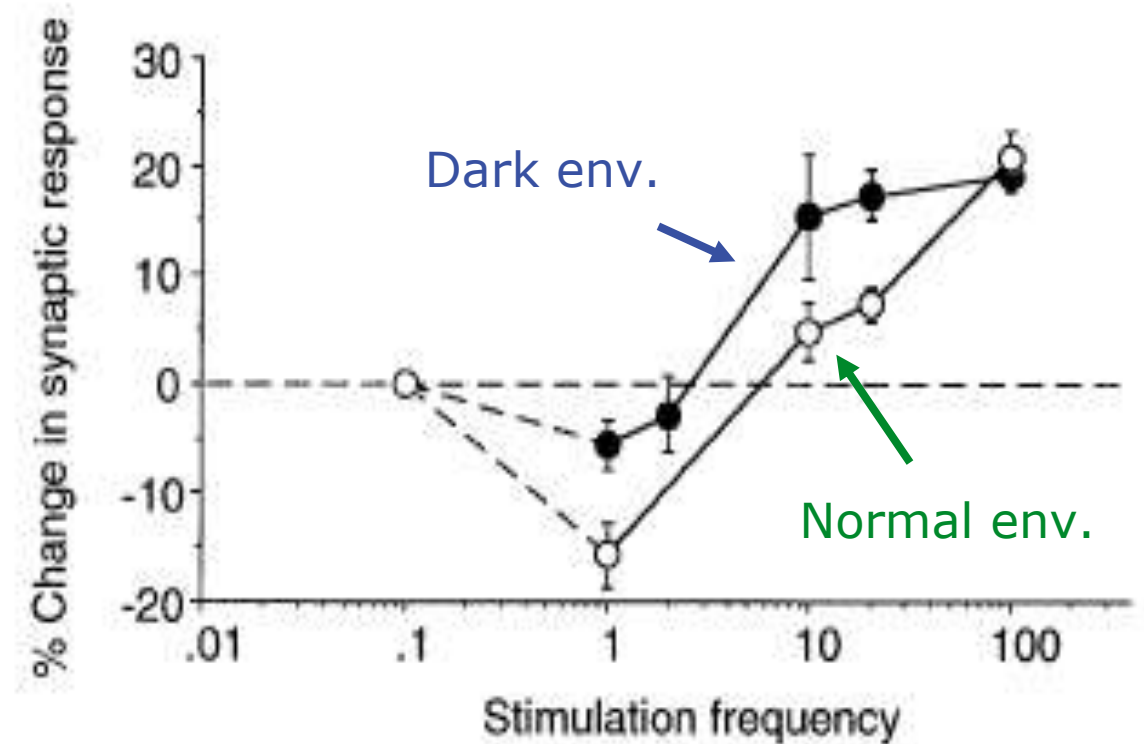
REINFORCE algorithm for a spiking neuron (x : input spike trains, y : output spike train, Δ : relative surprise)

$$\frac{dw}{dt} \propto \frac{d}{dw} \underbrace{\mathbb{E}_{x,y} [\Delta(y|x)]}_{\text{mutual information}} = \mathbb{E}_{x,y} \left[\frac{d \log p(y|x)}{dw} \Delta(y|x) \right]$$

Infomax under a firing rate constraint: the Hebbian property and meta-plasticity



Experiment



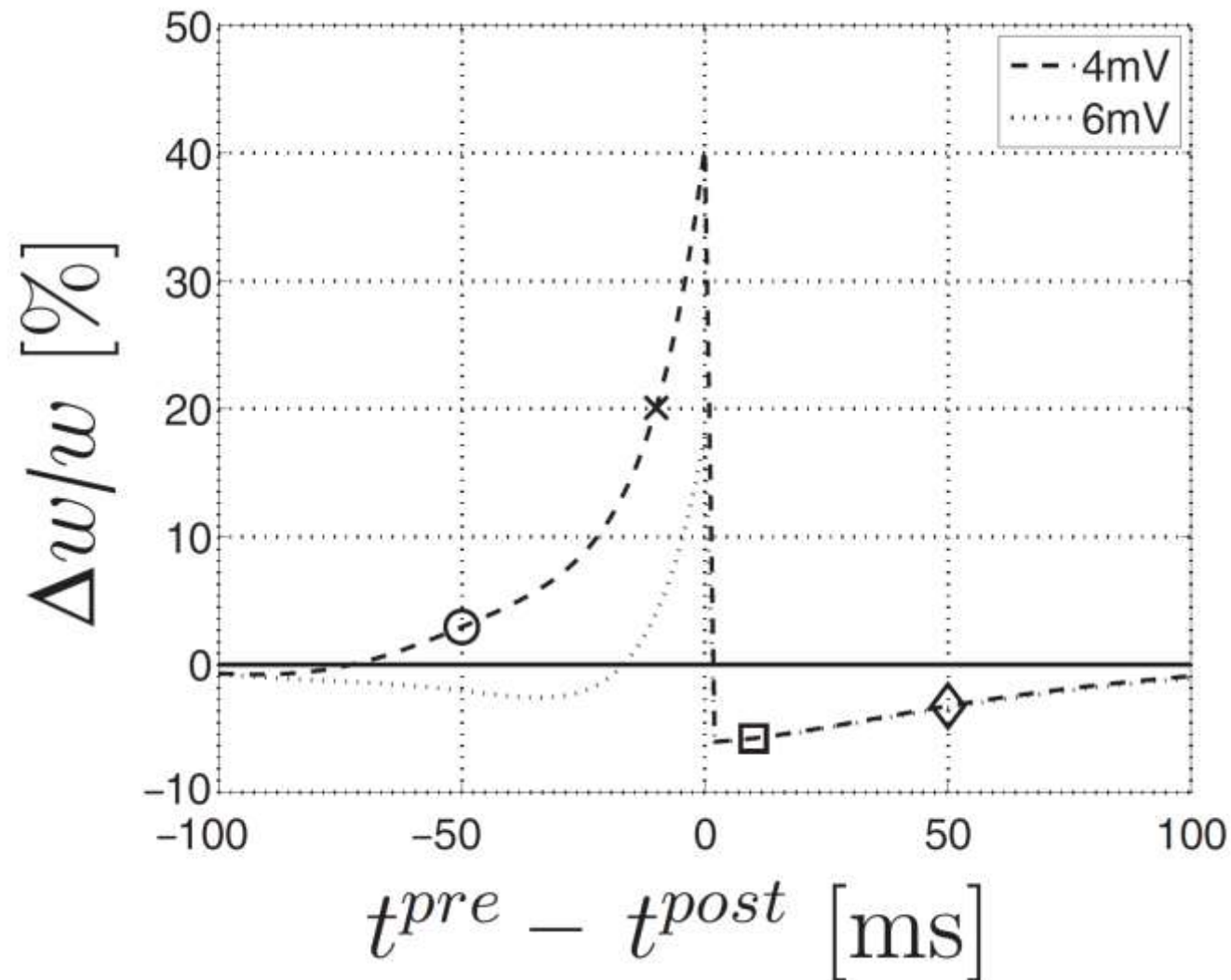
Kirkwood et al. 1996

Toyoizumi et al. 2005

Spike-timing-dependent plasticity (STDP) and its weight-dependency

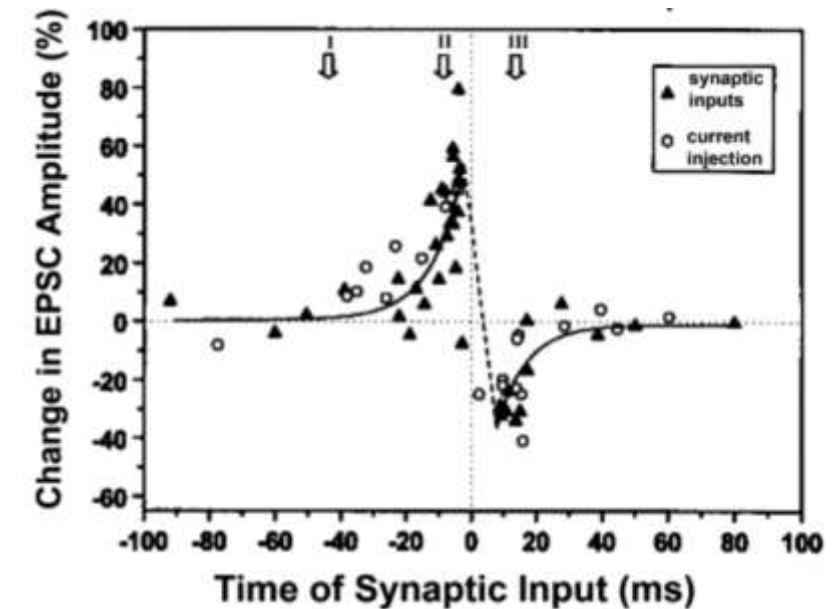
With a cost of maintaining synapses....

Theory

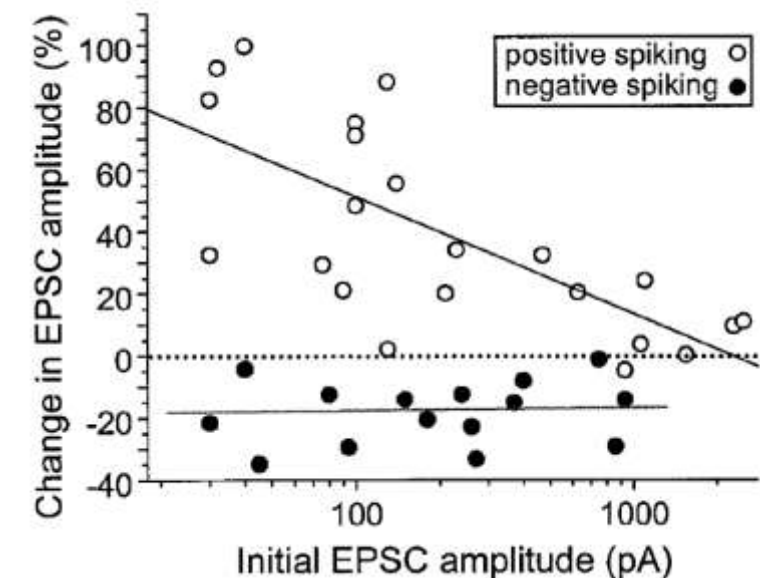


Toyoizumi et al. 2007

Experiment

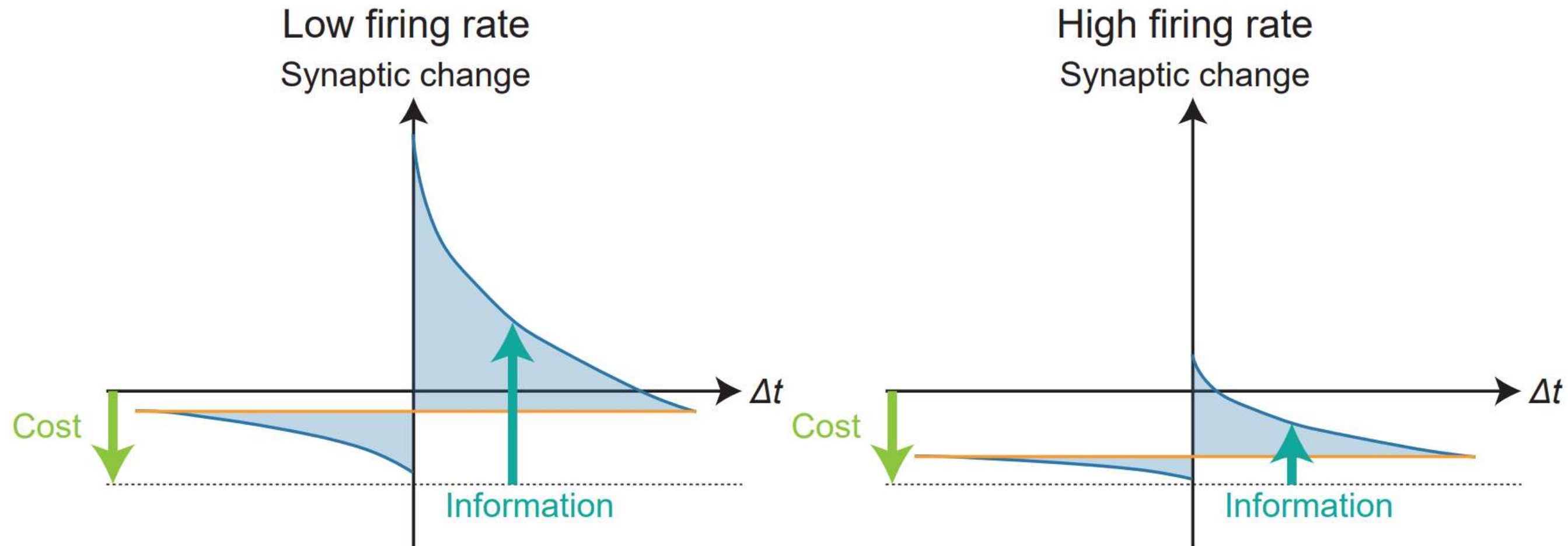


Zhang et al. 1998



Bi and Poo 1998

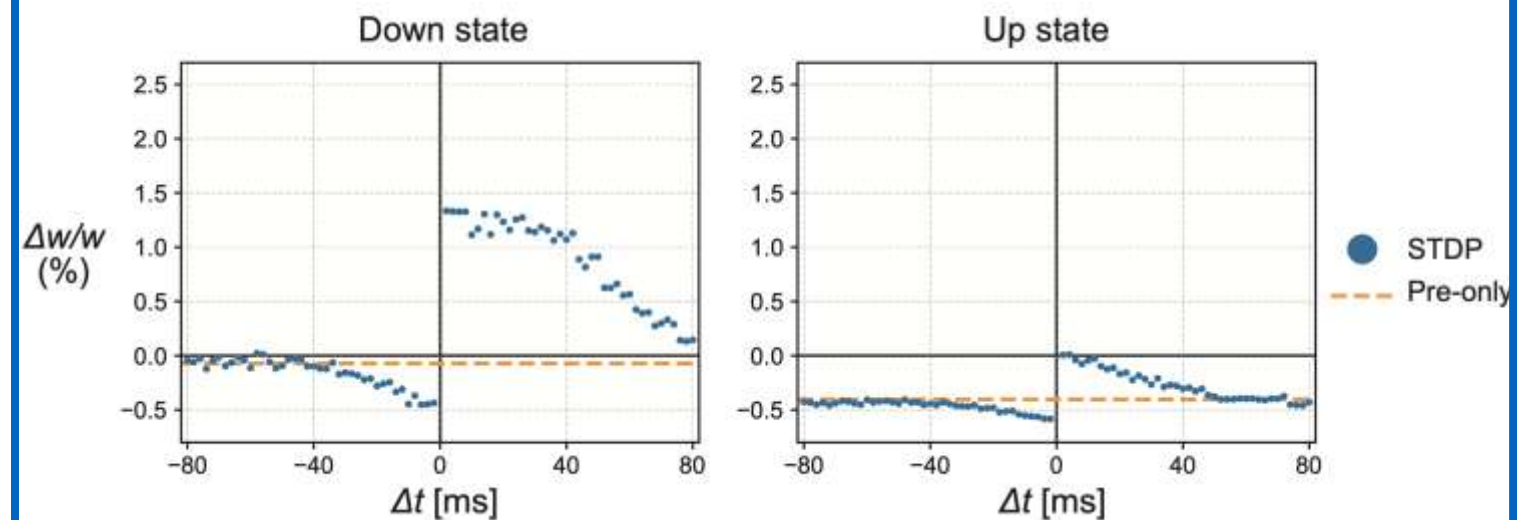
The Infomax STDP depends on the baseline firing rate of the postsynaptic neuron



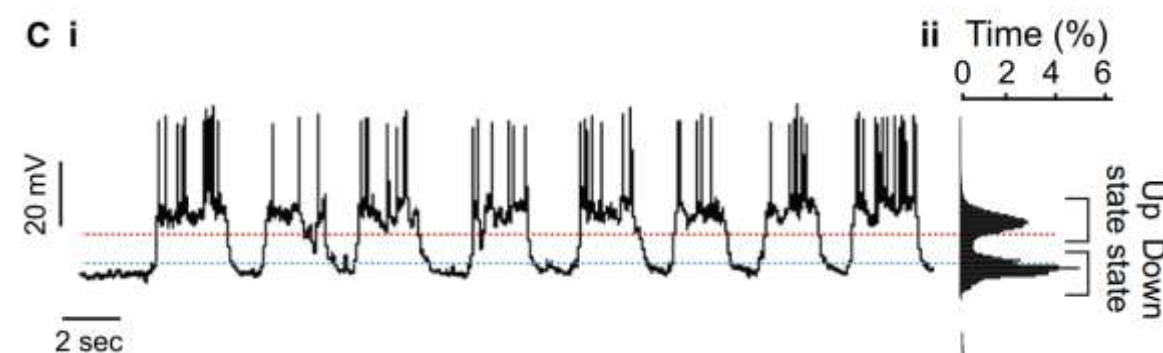
The information transfer is reduced if the baseline-firing rate is high because the postsynaptic neuron may emit spikes regardless of the transmission of a synapse.

The distinct baseline firing rates during up and down states of slow-wave sleep explain *in vivo* STDP results.

Theory

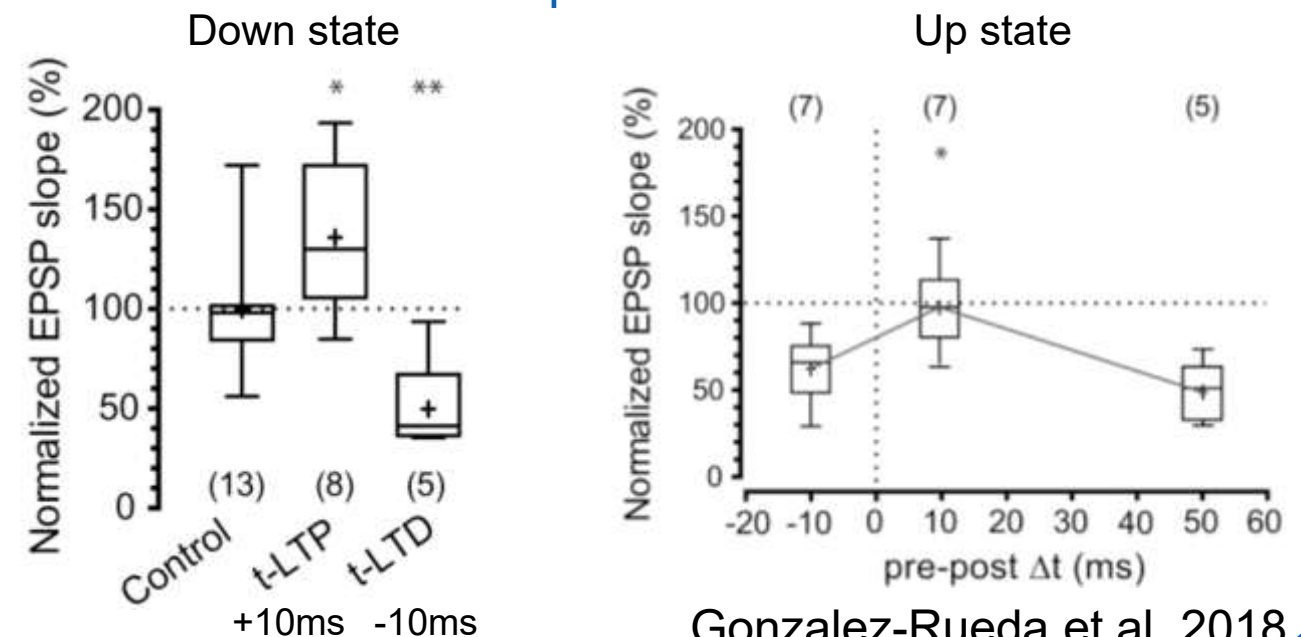


Yoshida and Toyozumi 2023



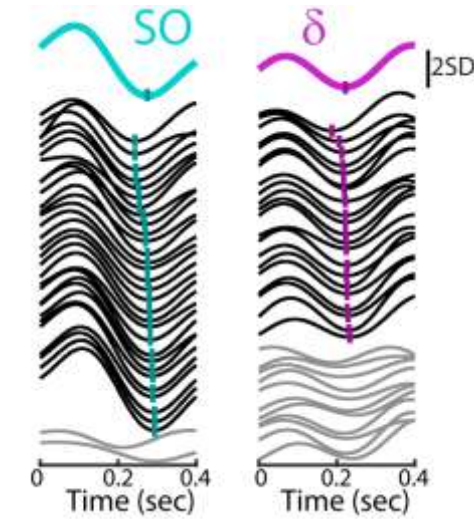
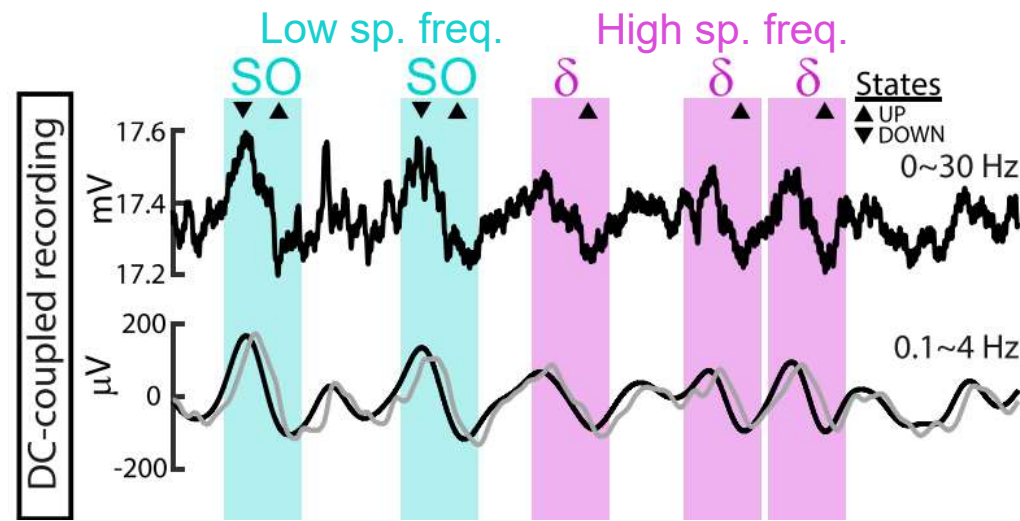
Gonzalez-Rueda et al. 2018

Experiment



Gonzalez-Rueda et al. 2018

High and low spatial frequency slow waves

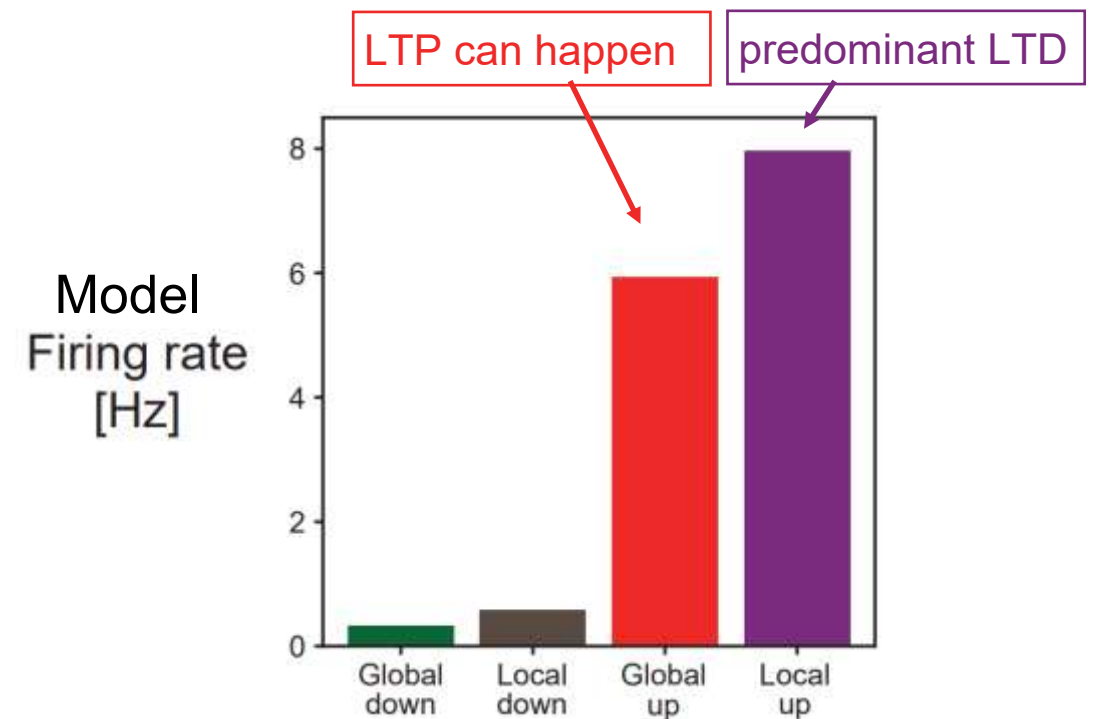


Kim et al. 2019

Surround suppression in the cortex:
(c.f. stabilized superlinear networks)

Low surround activity $\Rightarrow$ Central excitatory activity $\uparrow$
High surround activity $\Rightarrow$ Central excitatory activity $\downarrow$

Ozeki et al. 2009; Rubin et al. 2015; Ahmadian and Miller 2021

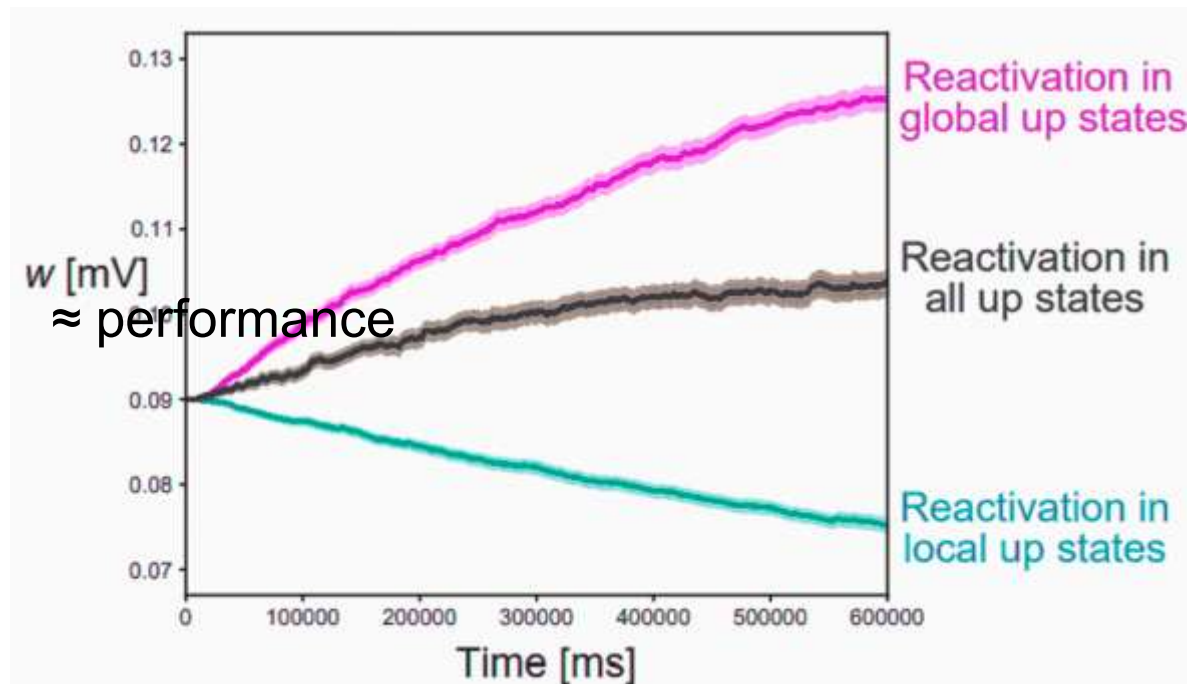
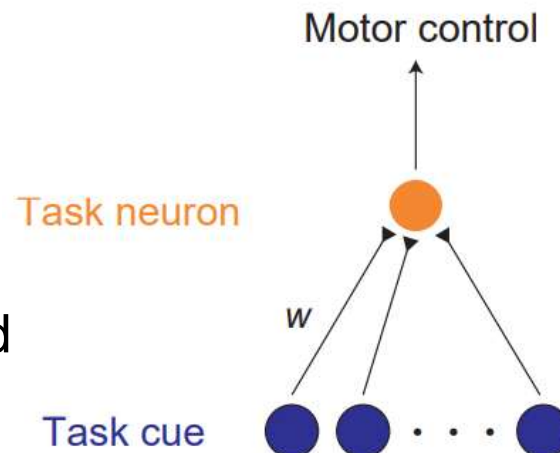


Yoshida and Toyozumi 2023

The model predicts memory consolidation during global up states and memory forgetting during local up states

Theory

- Task cue neurons are reactivated during the post-training non-REM sleep.
- Synapses change according to the infomax rule during global and local up states.

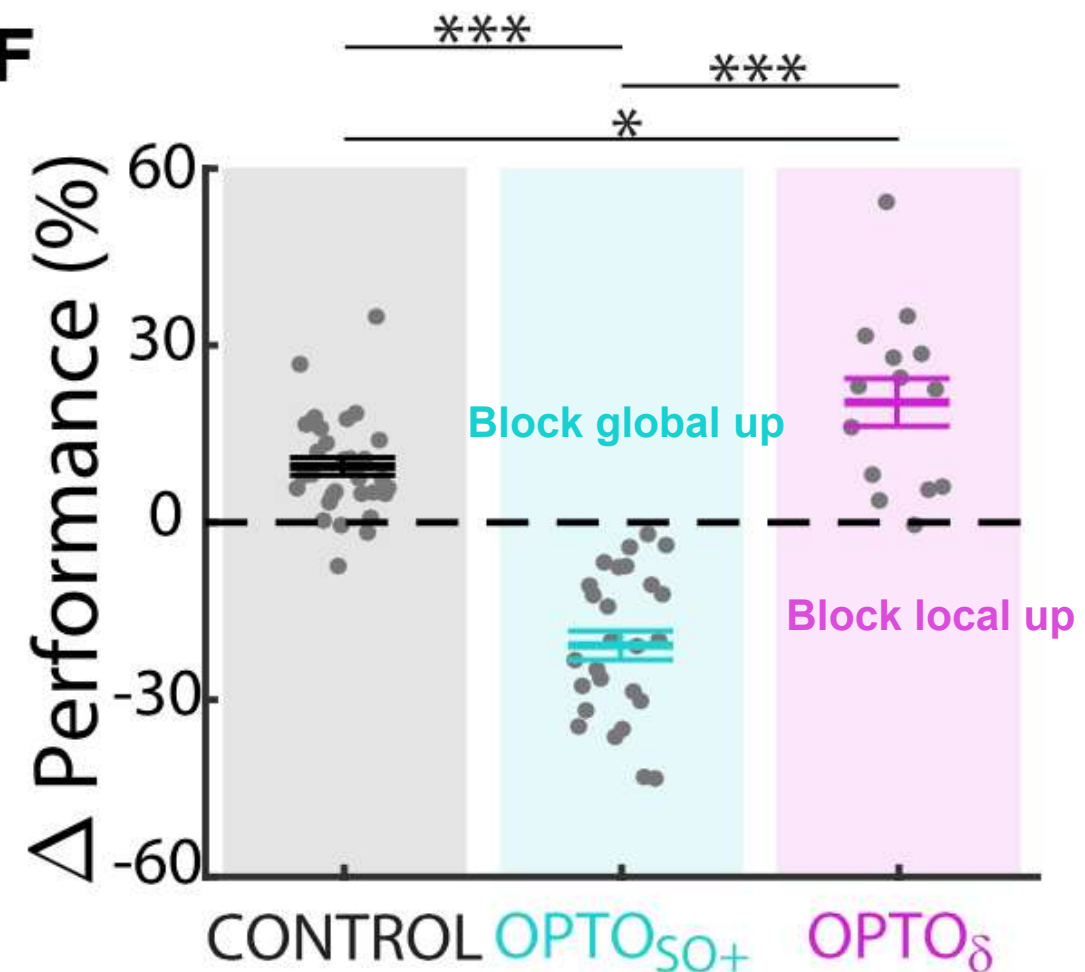


Yoshida and Toyoizumi 2023

Experiment



F



Kim et al. 2019

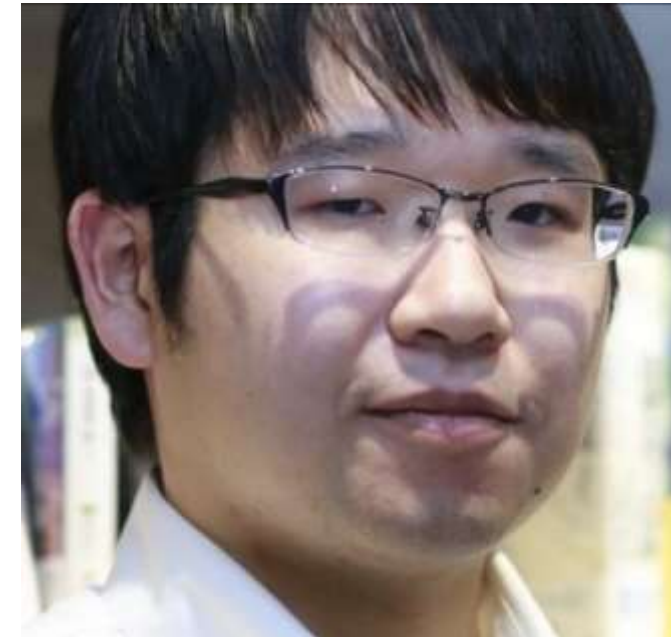
Summary

- Infomax STDP is biased toward LTD when the baseline firing rate is high.
- Based on the higher baseline firing rate in up states than down states during slow-wave sleep, the model reproduces the LTD-dominated STDP in up states.
- Based on the higher baseline firing rate in local up states than global up states, the model reproduces memory consolidation in global up states and memory forgetting in local up states.

Collaborator

Sleep infomax

- Kensuke Yoshida (RIKEN)



Funding



Brain/MINDS



Thank you!

toyoizumilab.brain.riken.jp