An optimization approach to understand biological search

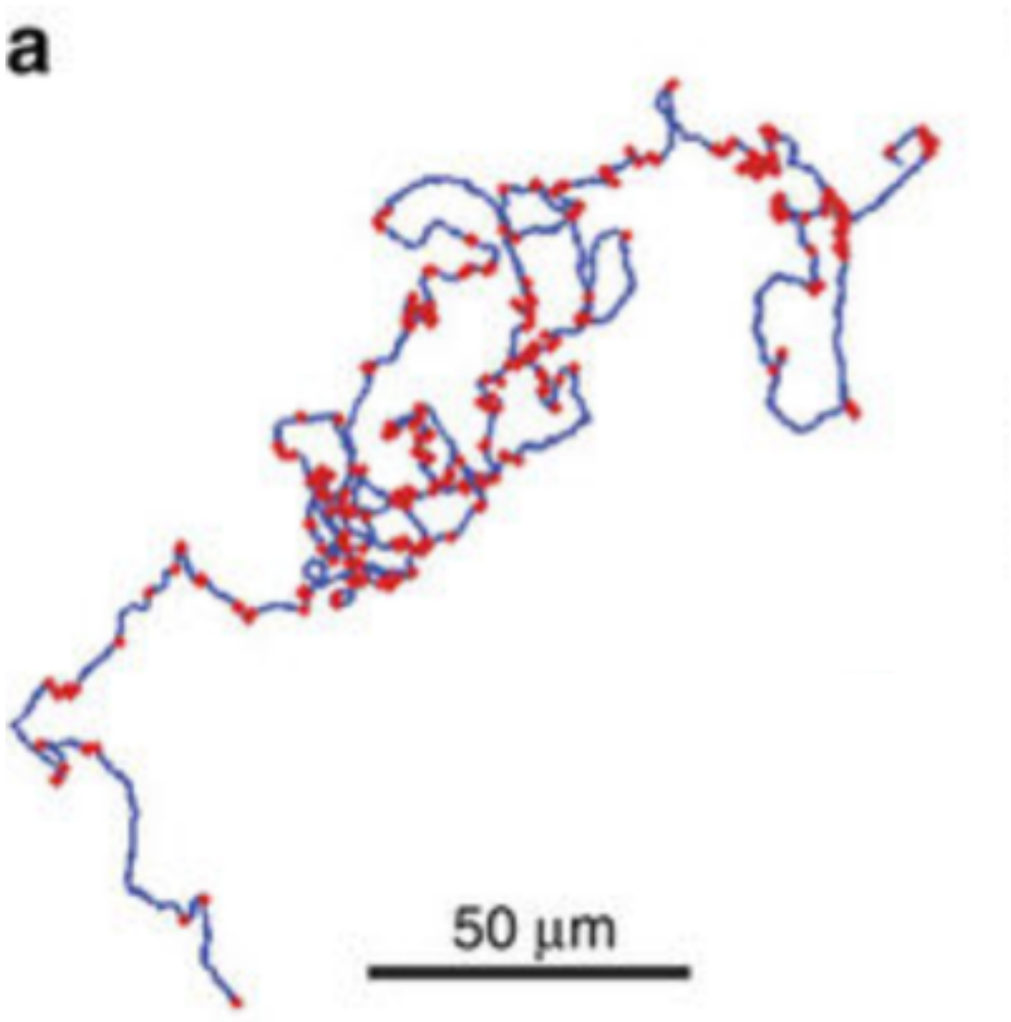
Taro Toyoizumi
RIKEN Center for Brain Science

Darwinian view of life

http://deskarati.com/

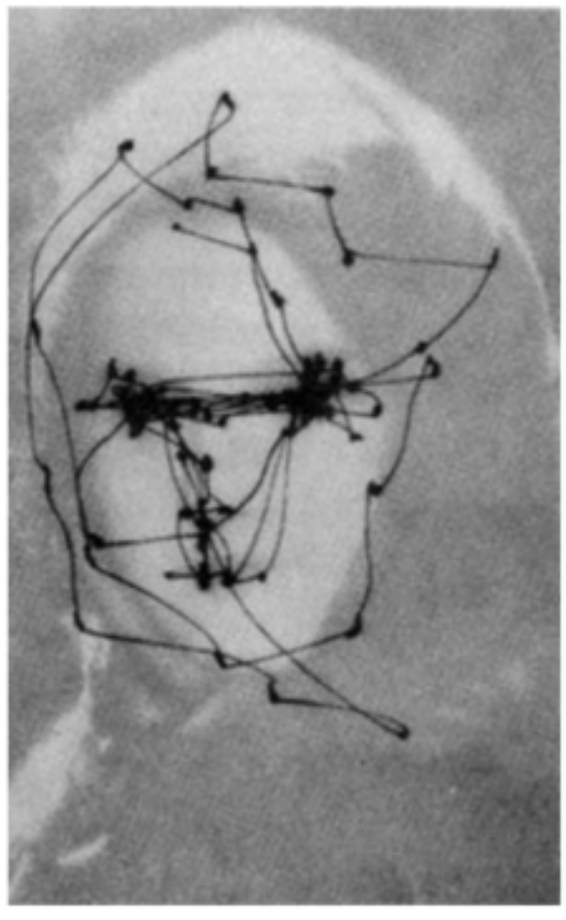
An evolutionally conserved search strategy: Levy flight

Bacteria



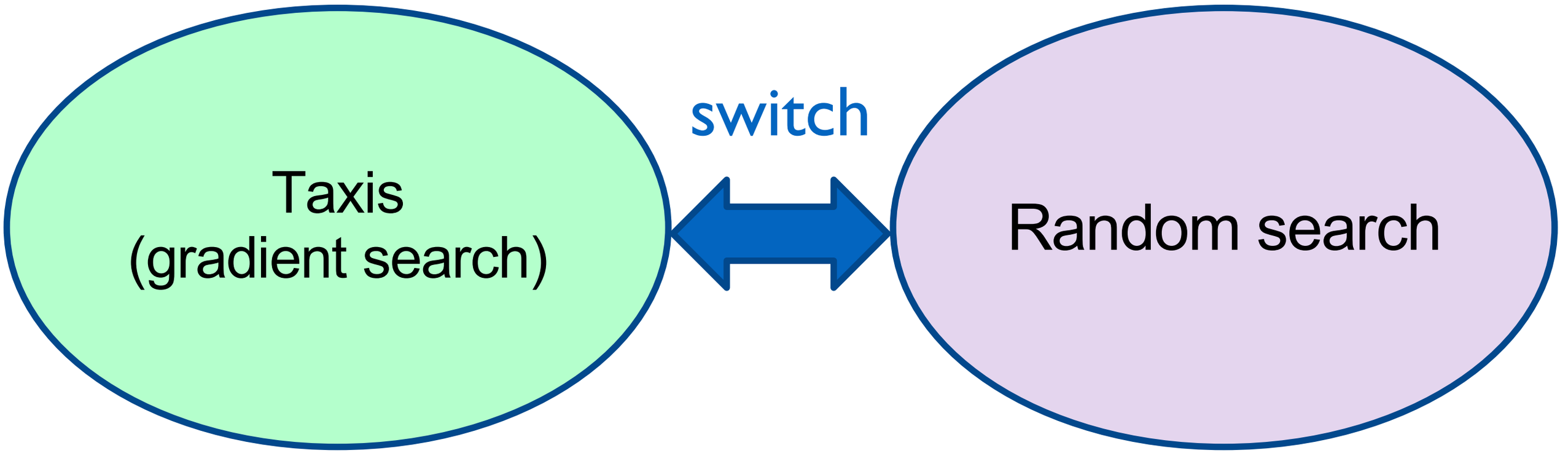
Ariel et al. 2015

Eye position



http://www.androidblues.com/visualperception.html

Conventional view



Levy flight emerges from the 2nd-order stochastic optimization

smoothness gradient

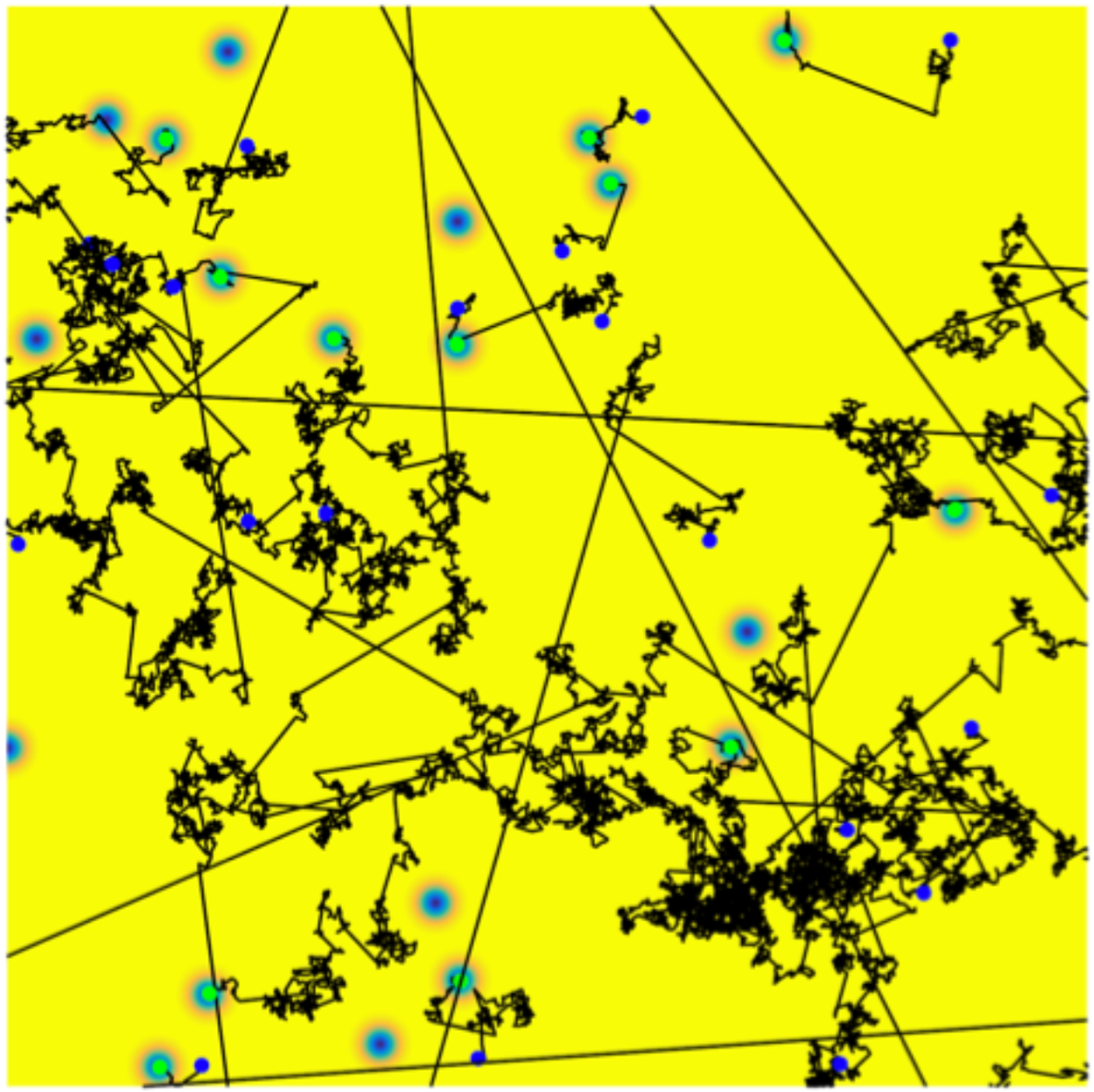
$$\Delta \propto (H + \beta I + \xi_H)^{-1} (\vec{g} + \vec{\xi}_G)$$

Step-size

$$\Delta = \frac{\xi_G}{\xi_H}$$

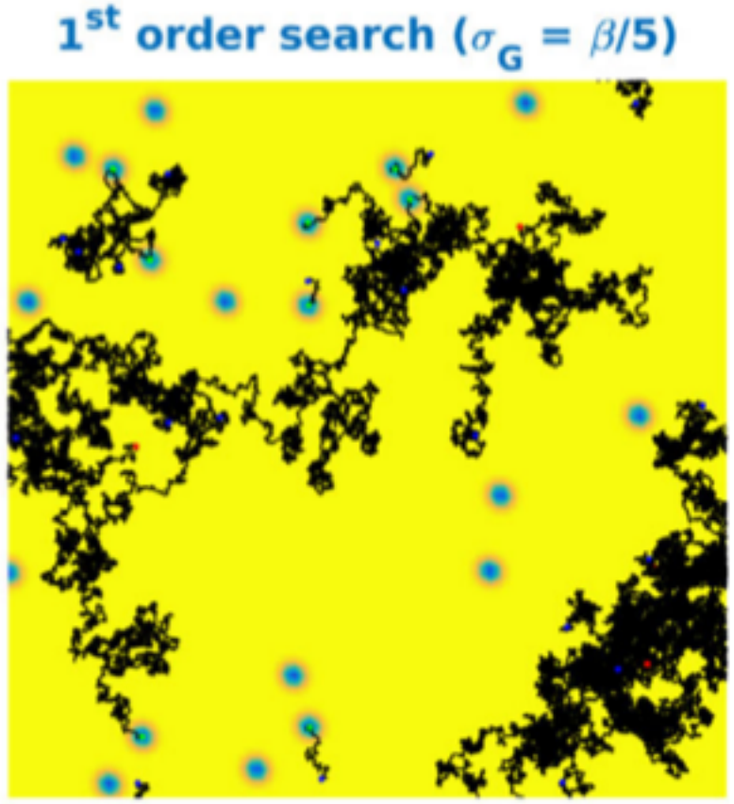
Step-size distribution

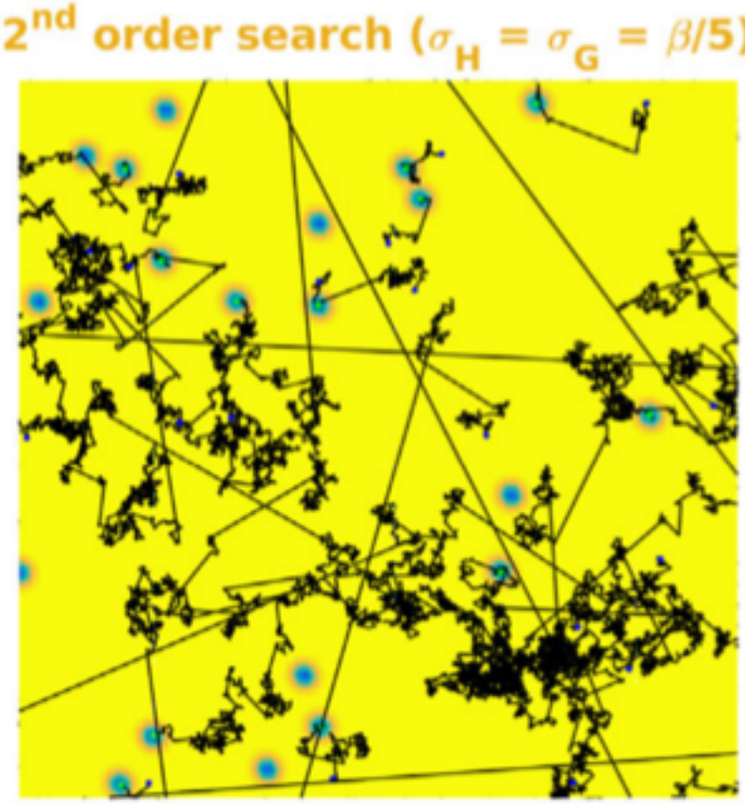
$$p(\Delta) = \frac{1}{\pi} \frac{1}{1 + \Delta^2}$$

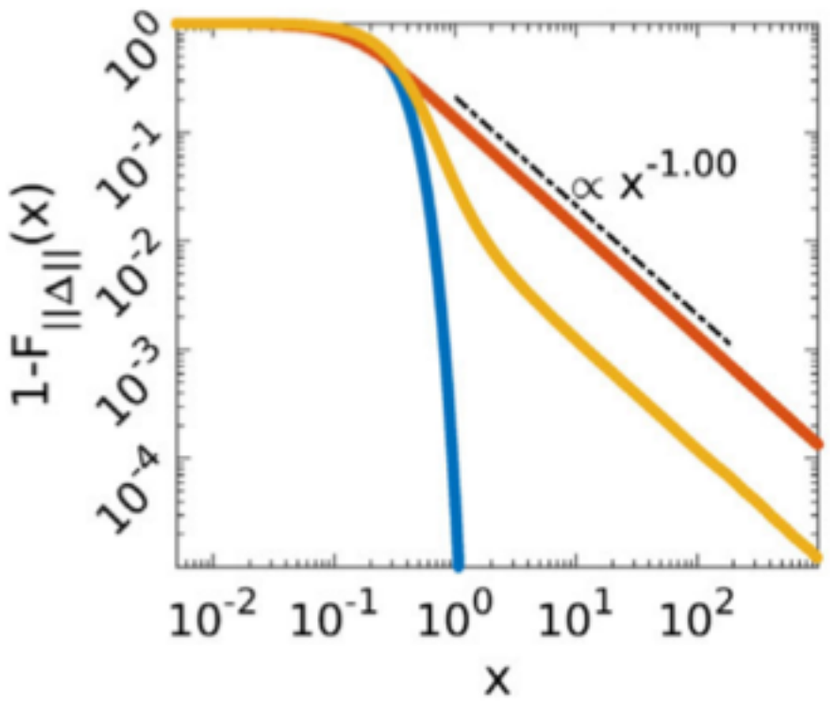


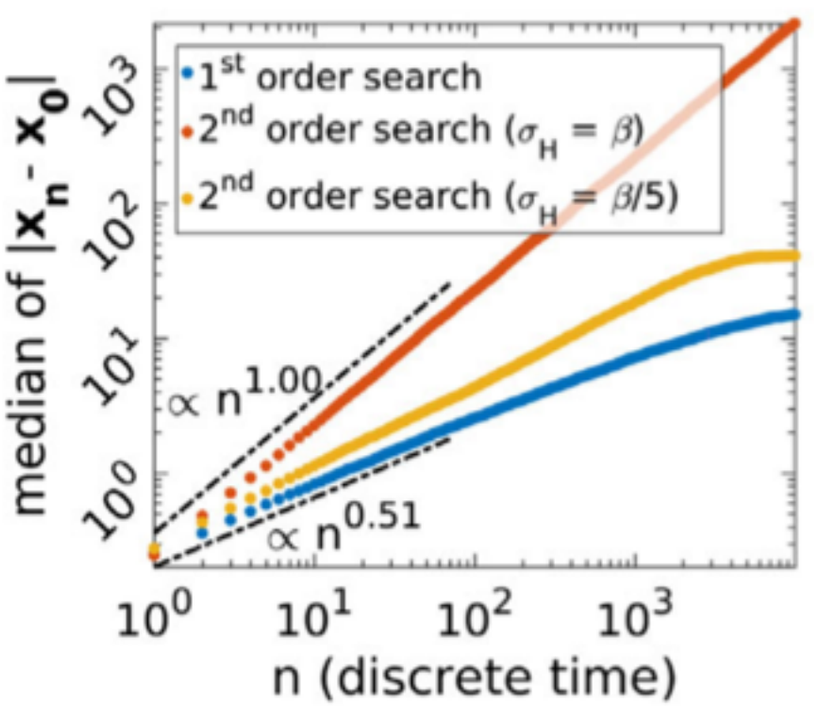
Kusmierz and Toyoizumi 2017

The 2nd-order stochastic optimization: The middle between Brownian and Levy searches



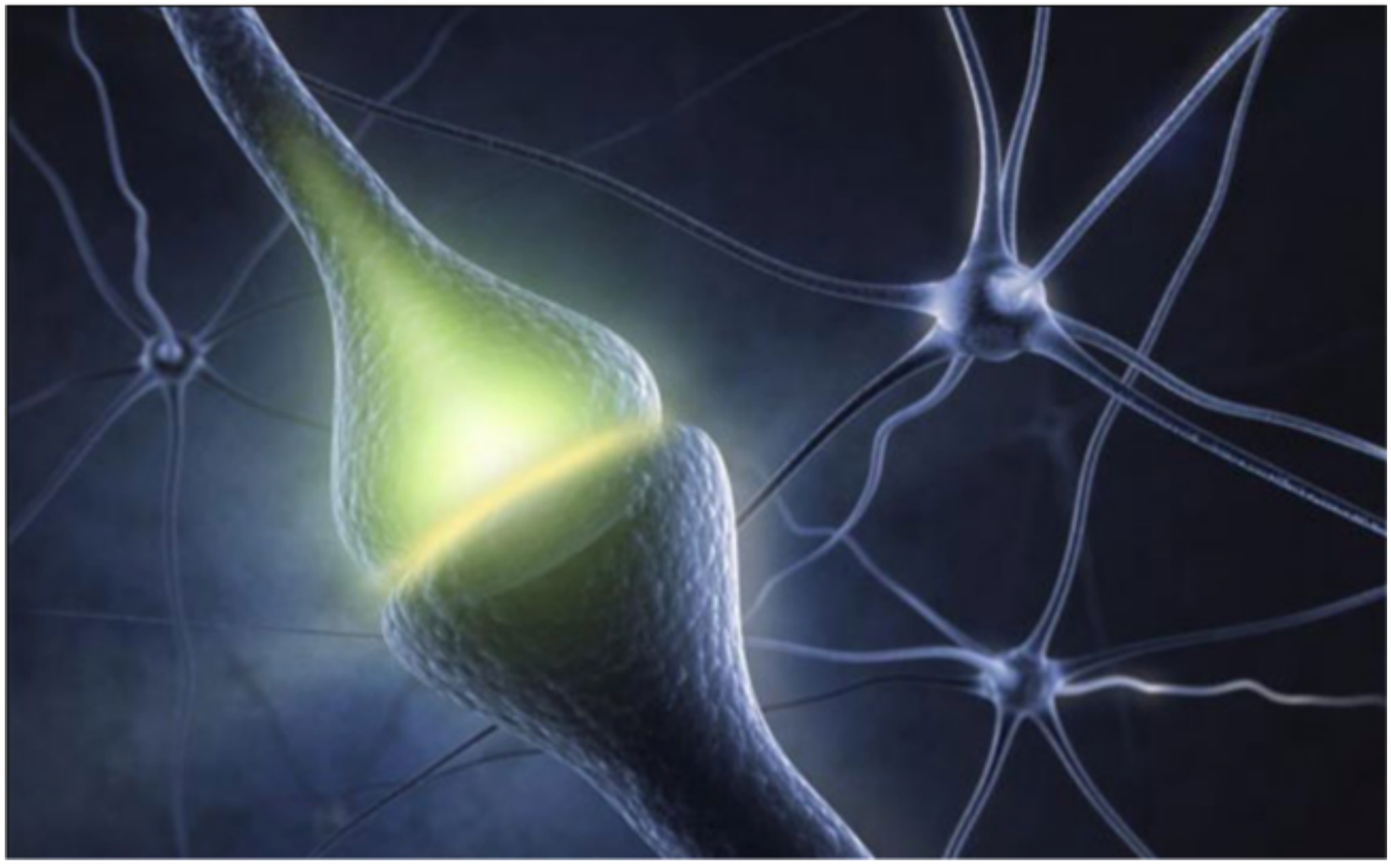






Kusmierz and Toyoizumi 2017

Synaptic plasticity

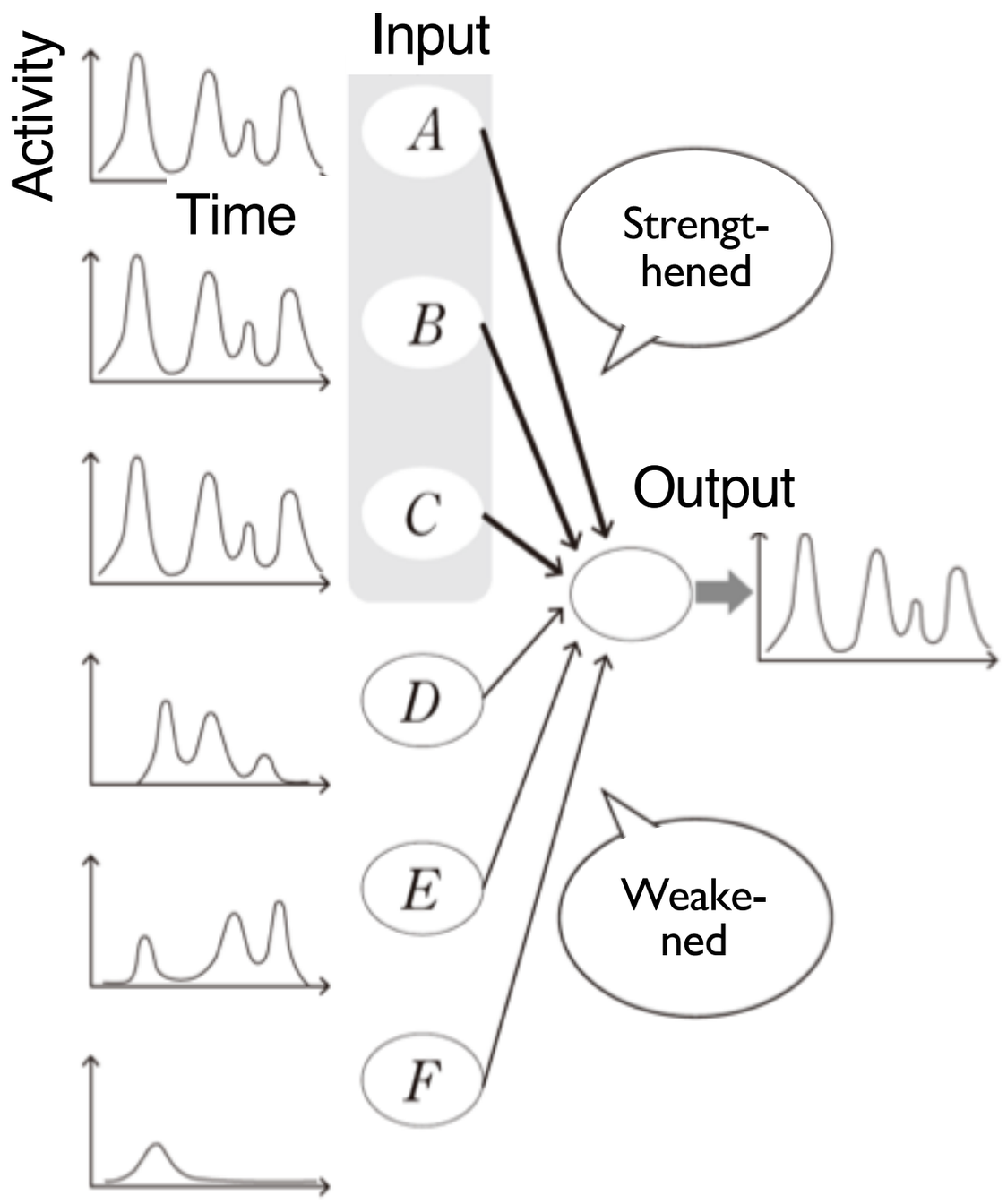


scitechdaily.com

Optimal synaptic plasticity:

A synaptic plasticity rule for maximizing information transmission

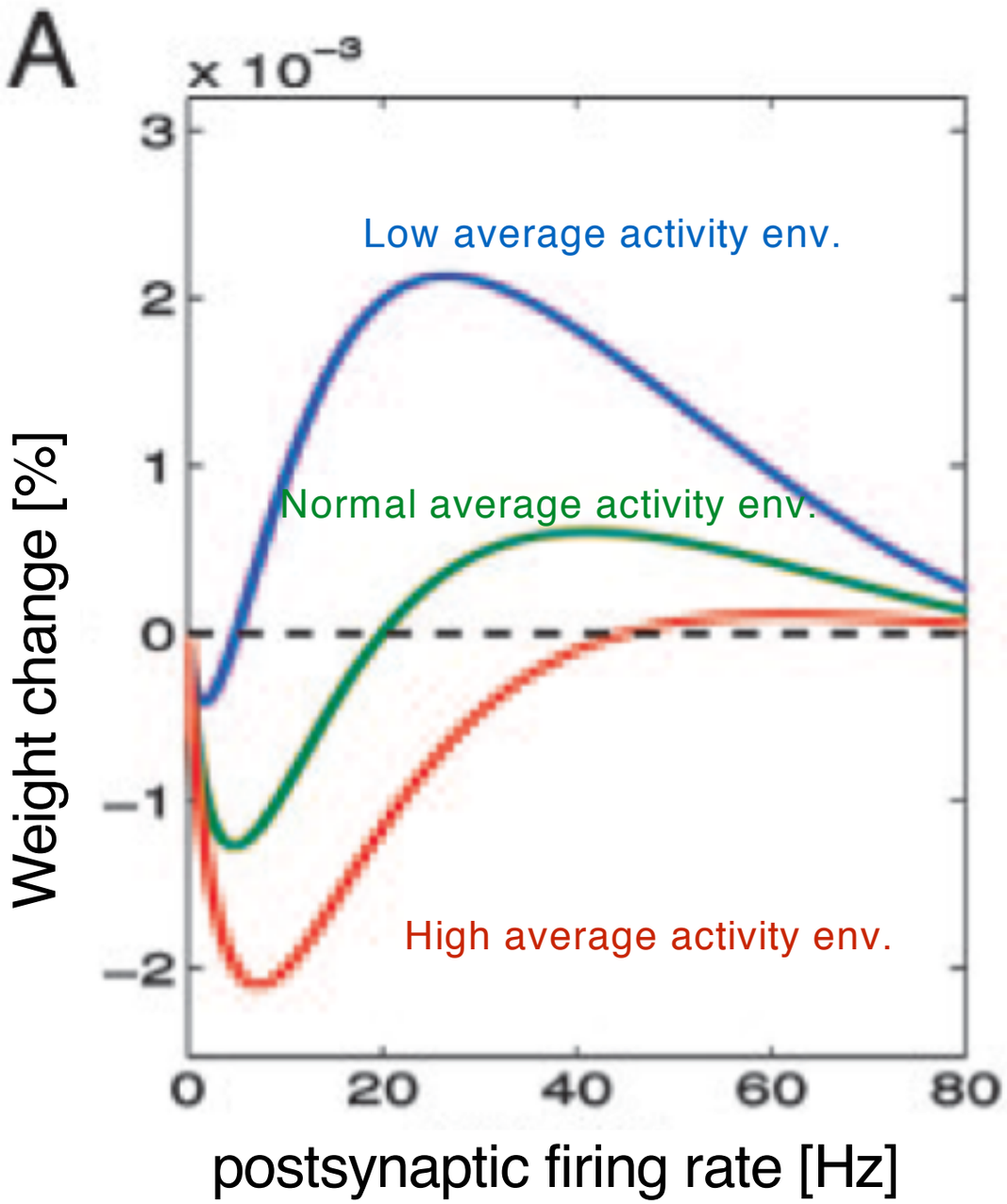
(Linsker 1988)



Firing rate dependency and meta-plasticity

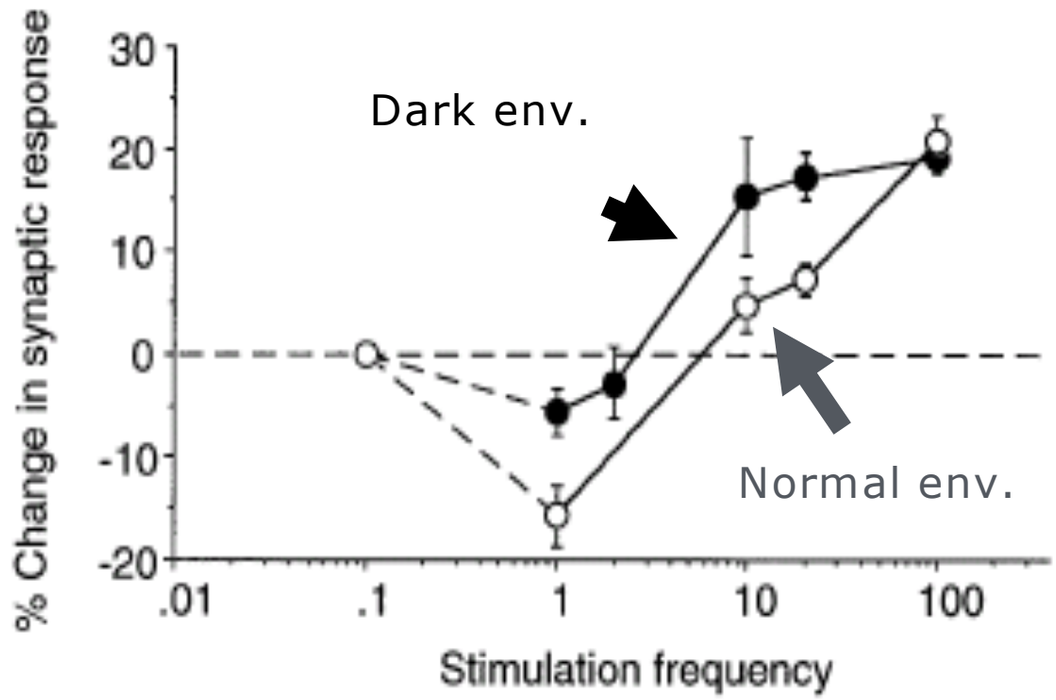
With an energy constraint….

Theory



Toyoizumi et al. 2005

Experiment

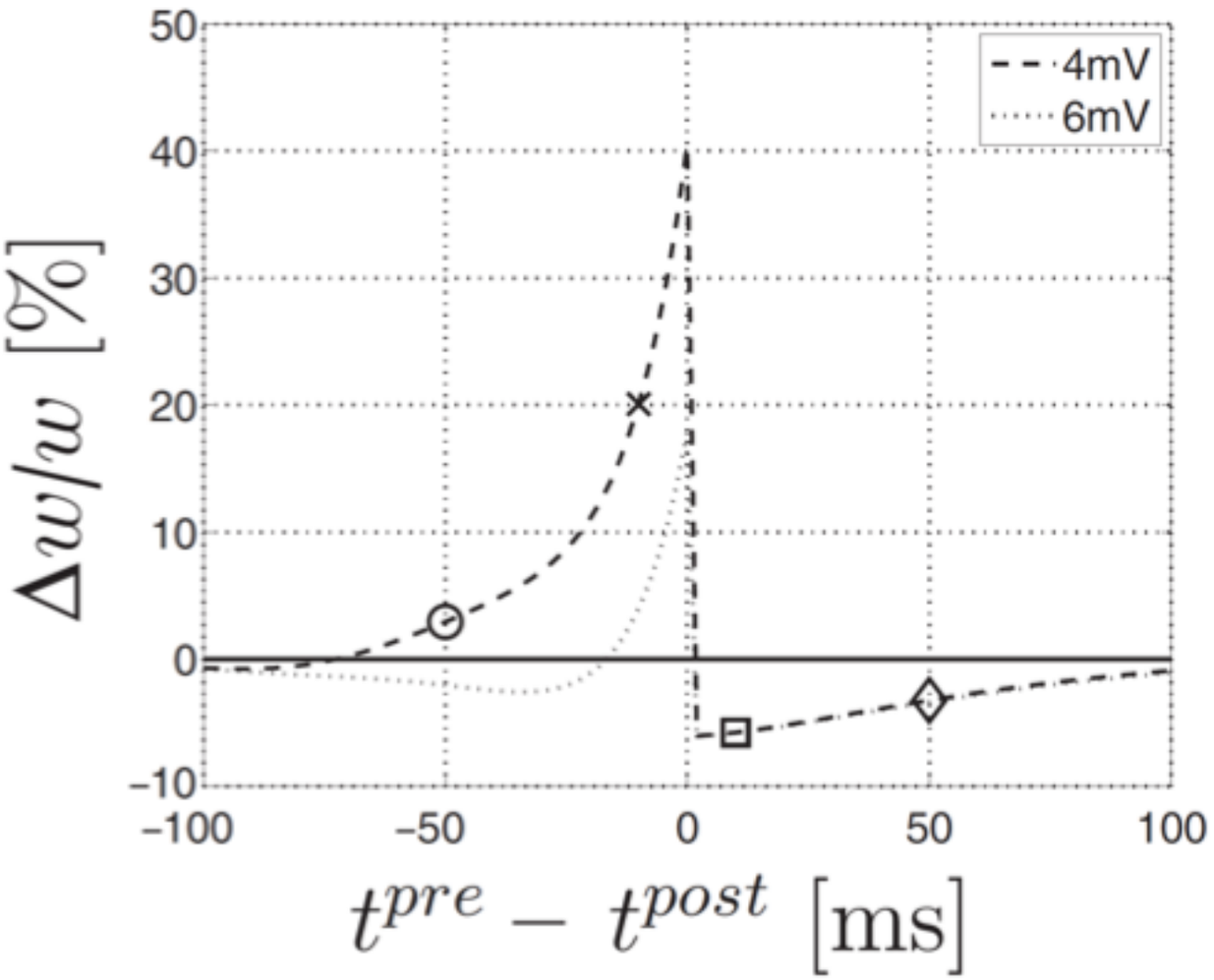


Kirkwood et al. 1996

STDP and weight-dependency

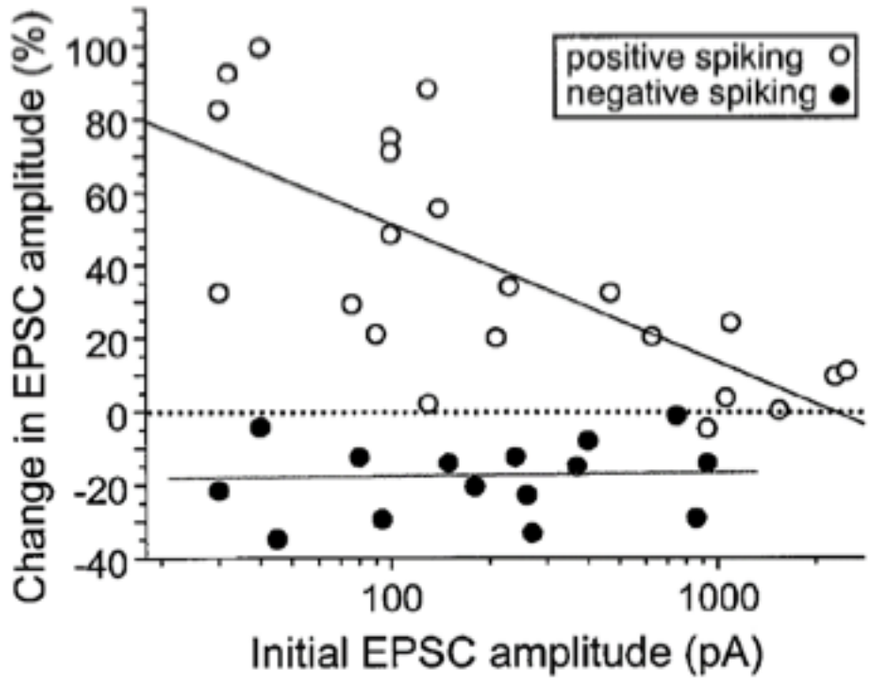
With a cost of maintaining synapses….

Theory



Toyoizumi et al. 2007

Experiment



Bi and Poo 1998

A Local Learning Rule for ICA and PCA

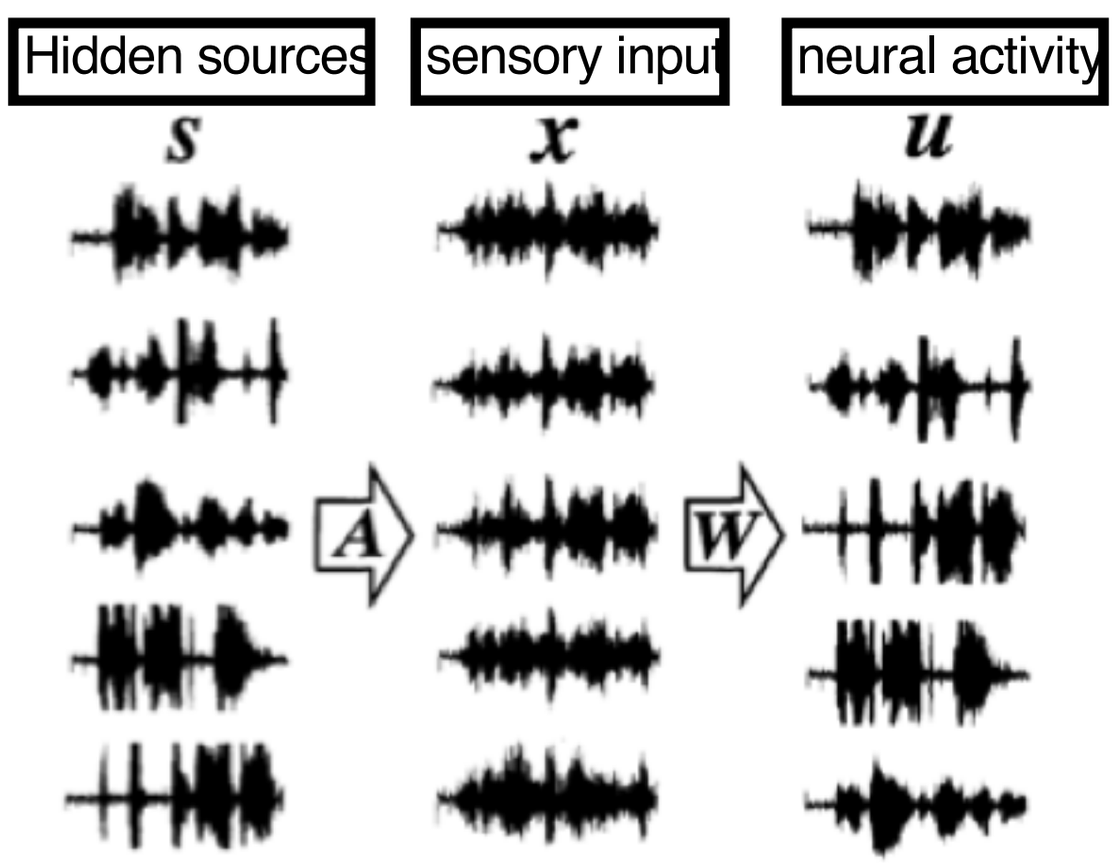
Information maximization using multiple output neurons

Bell and Sejnowski 1995

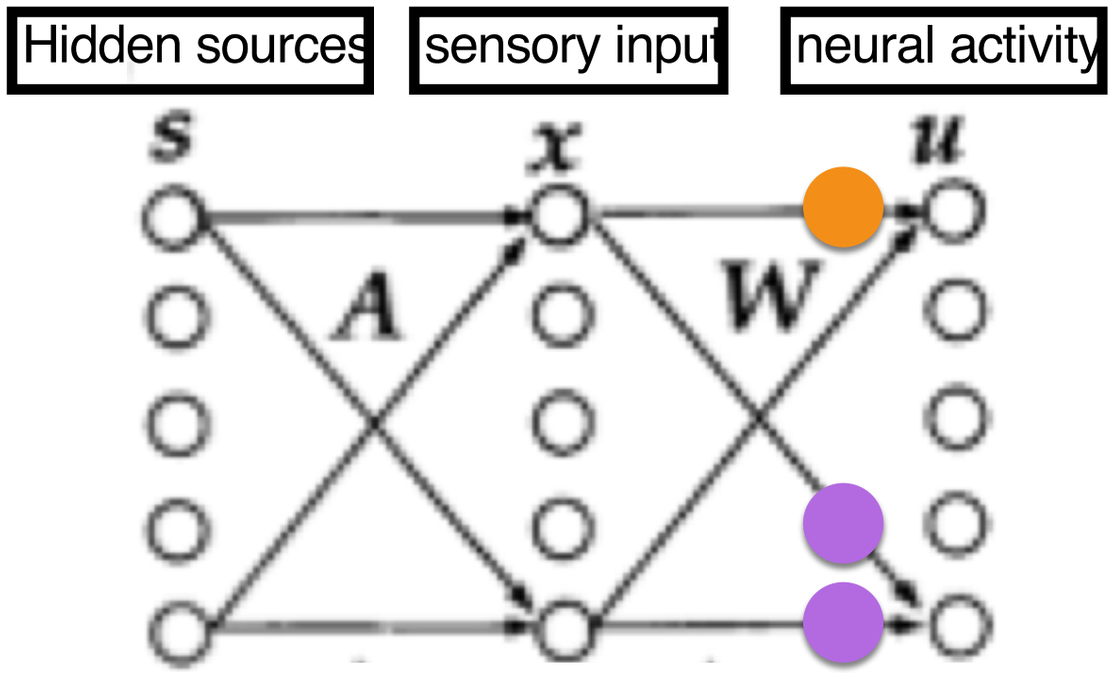
Independent component analysis (ICA)

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

Non-local engineering rule



A typical setup for ICA



$$\boldsymbol{x} = A\boldsymbol{s}$$

$$\boldsymbol{u} = W\boldsymbol{x}$$

$$g(u_i) = -(\log p_s(u_i))'$$

Bell and Sejnowski 1995

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

Non-local learning rule

Error-gated Hebbian rule (EGHR)

Bell and Sejnowski 1995 $\dot{W}_{ij} \propto (W^{-T})_{ij} - g(u_i)x_j$

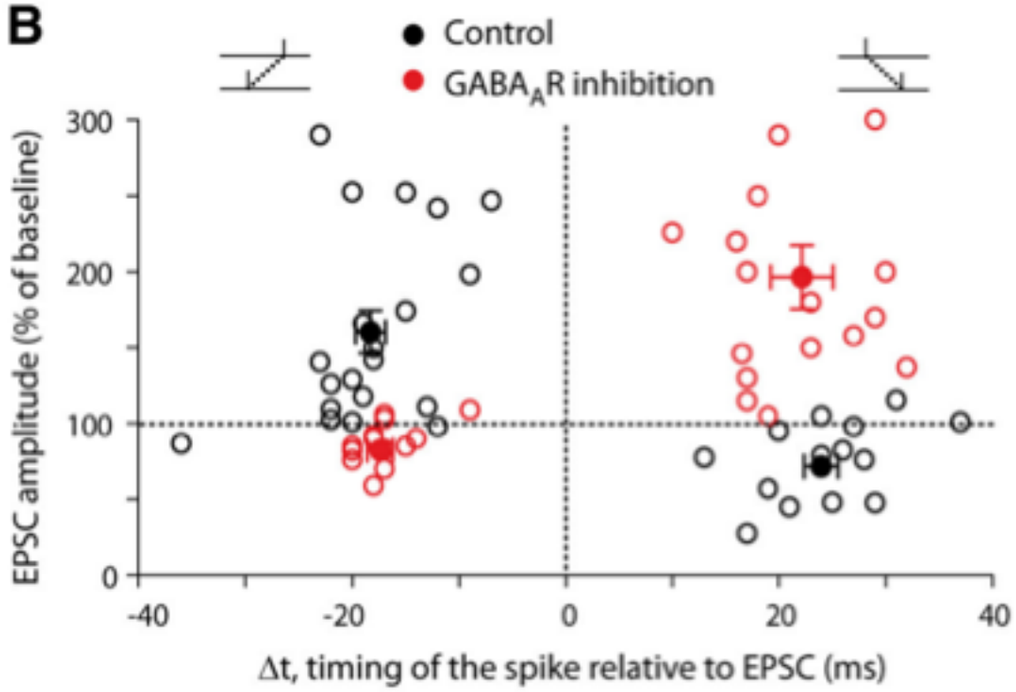
$\langle F(u)g(u)\rangle \approx \langle F'(u)\rangle$

holds around the ICA solution.

Isomura and Toyoizumi 2016 $E(u) \equiv E_0 - \sum_i \int^{u_i} g(z)dz$

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

A local ICA rule:
Modulation of Hebb's rule by a third factor



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

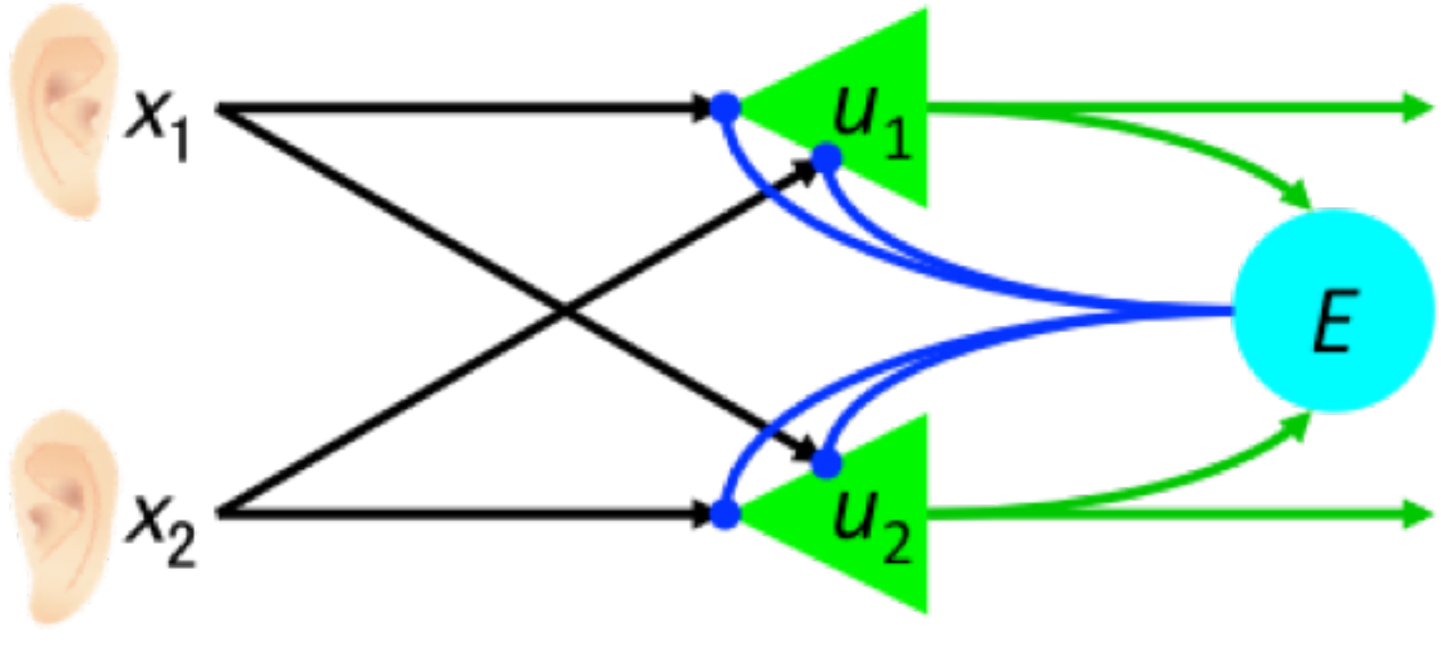
presynaptic

postsynaptic

global third-factor

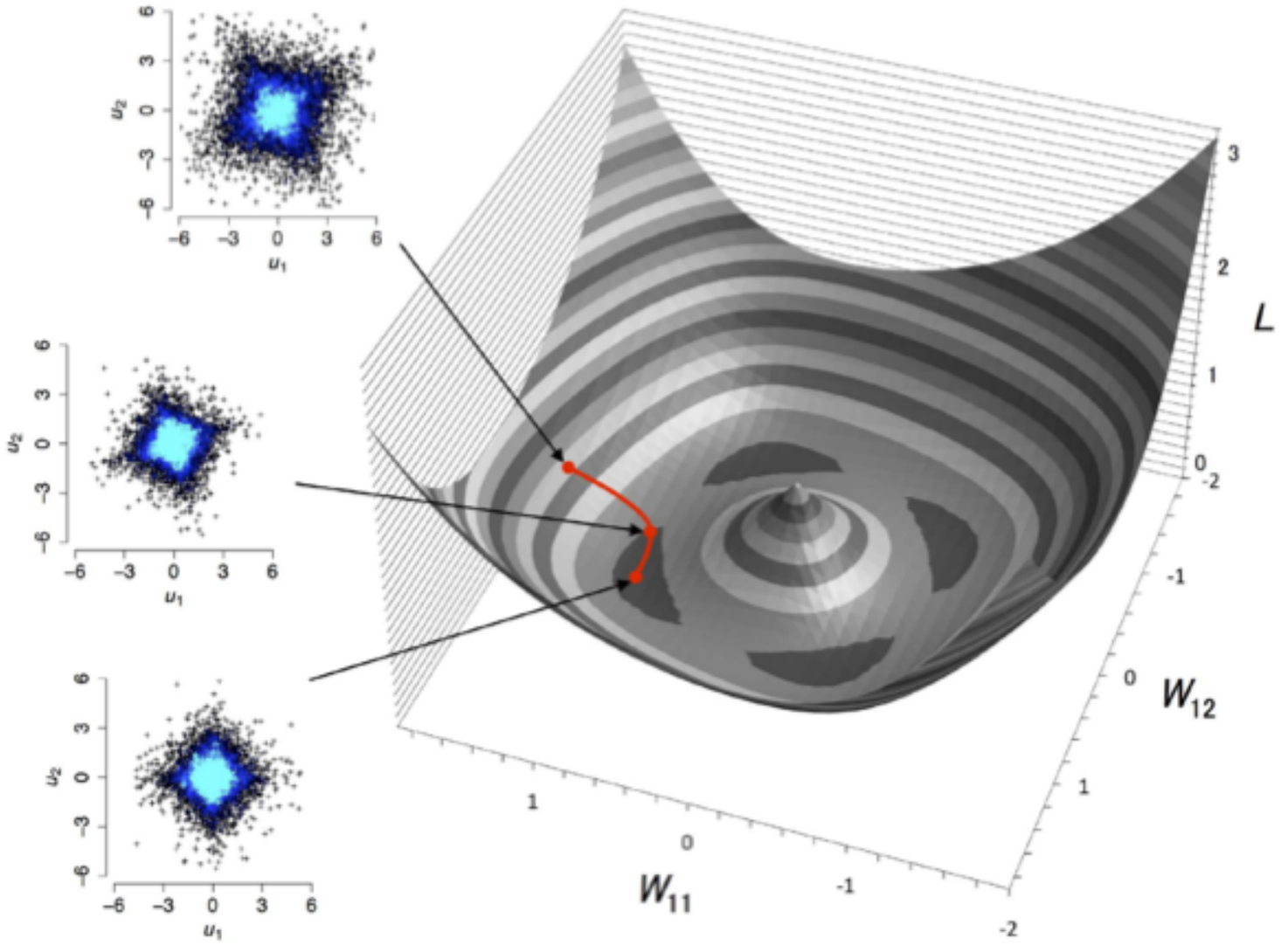
(e.g. GABA, neuromodulator, glial factor, etc.)

Paille et al. 2013



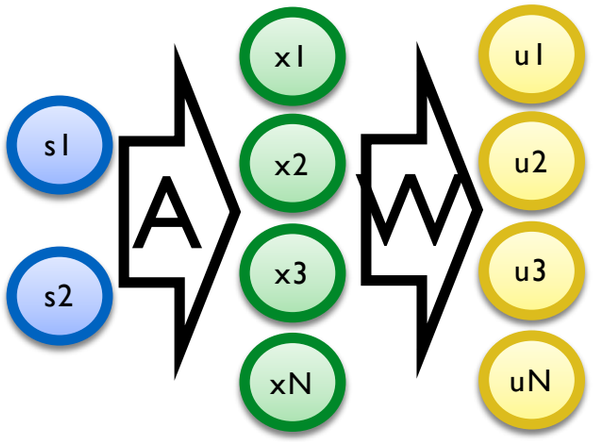
Dynamics of EGHR

$$\dot{W}_{ij} \propto \frac{\partial \langle E(u)^2 \rangle}{\partial W_{ij}}$$

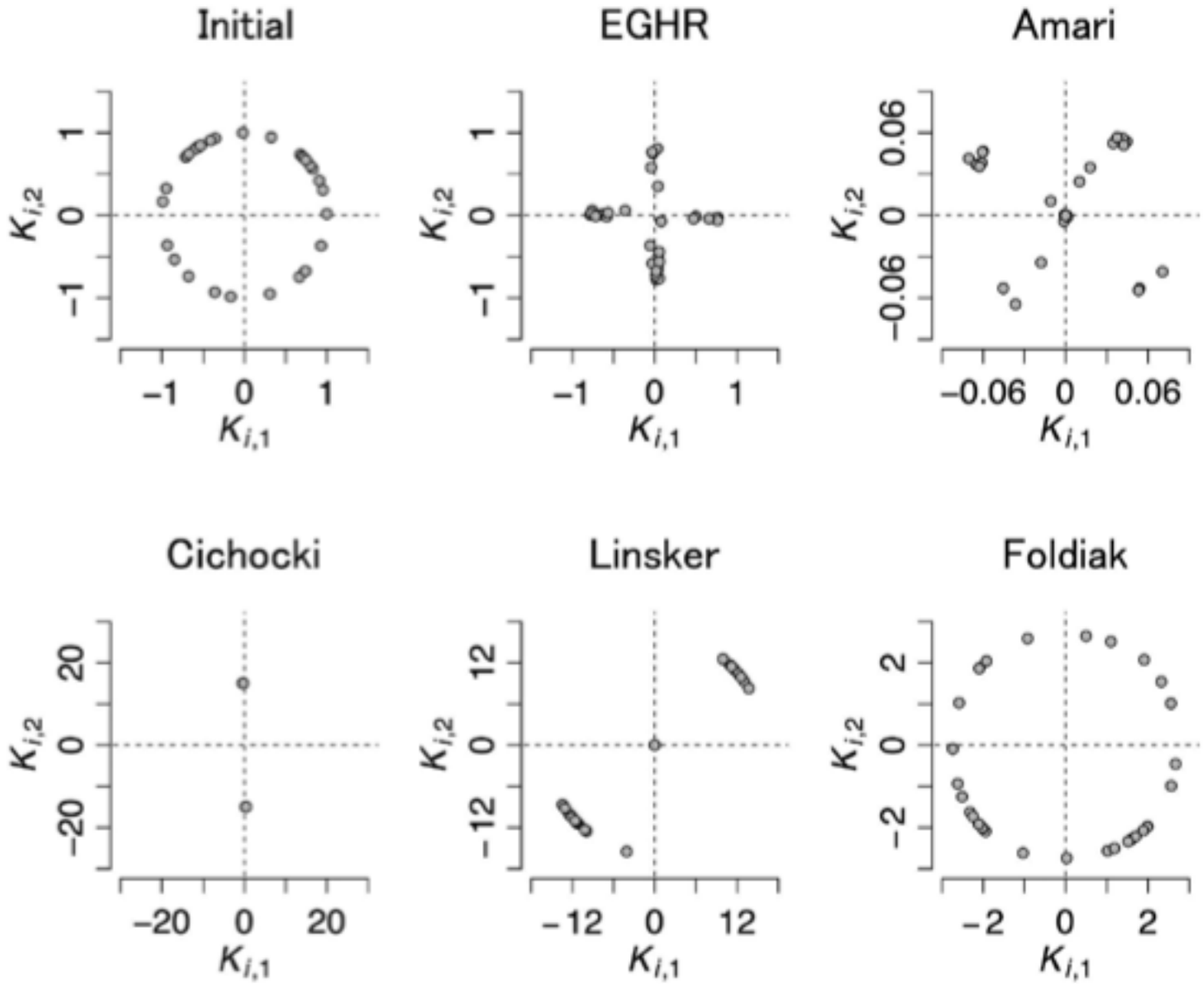


- EGHR is an approximation of the Infomax rule around the solution. Stability of the solution is roughly the same.
- No spurious solutions if the source distribution is near Gaussian.
- Unlike some heuristic rules, no preprocessing (e.g. whitening) is necessary.

EGHR is robust to the number of sources



Each output neuron describes a source if a marker is located on one of the axes.



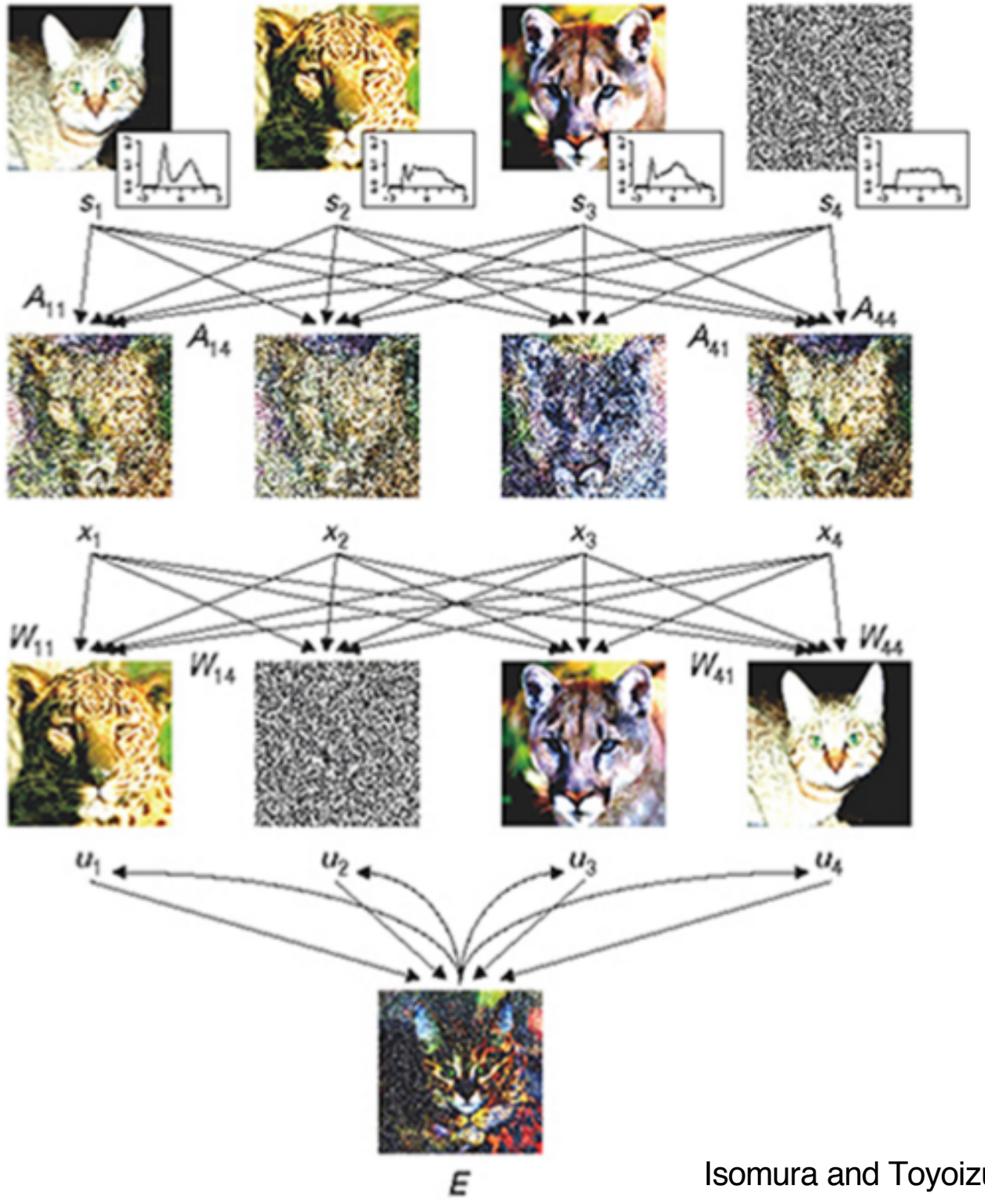
Separating mixtures of natural images

source

input

output

error



Isomura and Toyoizumi 2016

Separating movie clips

Source

Mixed input

Output

Isomura and Toyoizumi 2016

Not only separating sources but also extracting major features by PCA

A local learning rule for PCA

Oja 1989

$$\dot{W}_{ij} \propto u_i(x_j - \sum_k u_k W_{kj})$$

$$\langle F(u)g(u)\rangle \approx \langle F'(u)\rangle$$

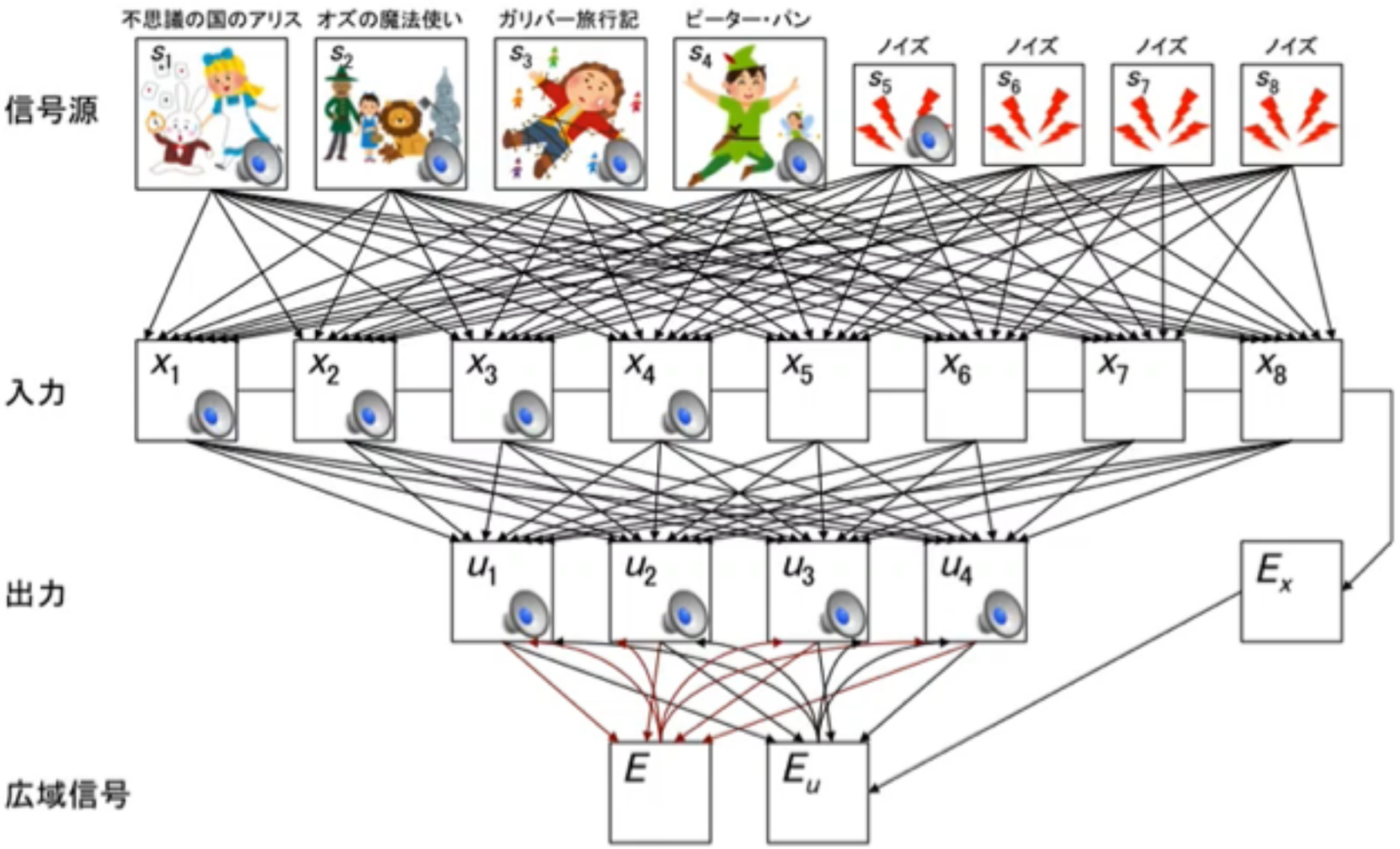
Isomura and Toyoizumi, *submitted*

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

$$\dot{W}_{ij} \propto E_{PCA}(x,u)u_i x_j$$

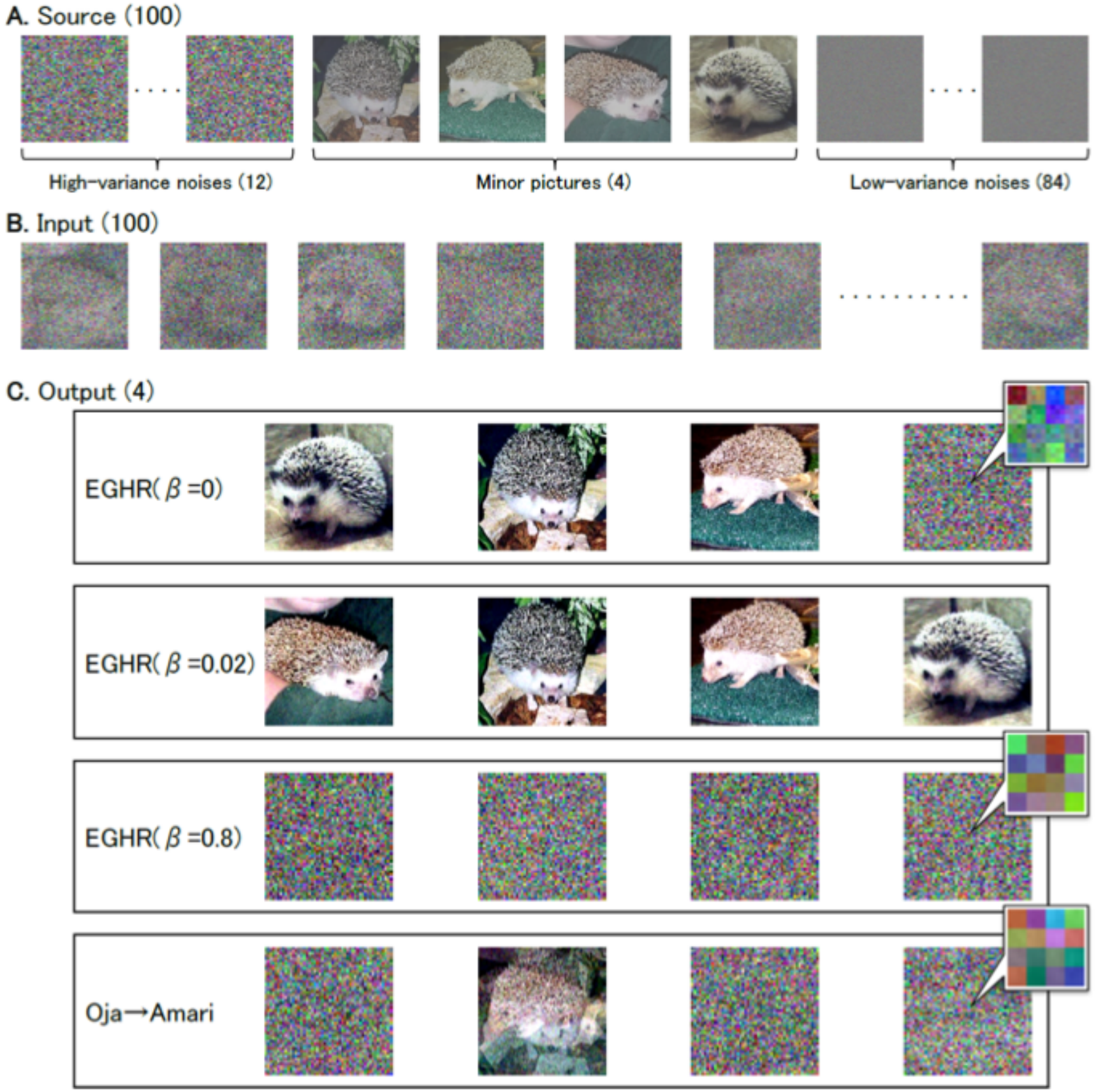
PCA + ICA

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



https://www.youtube.com/watch?v=tyRgk4AerTw&feature=youtu.be

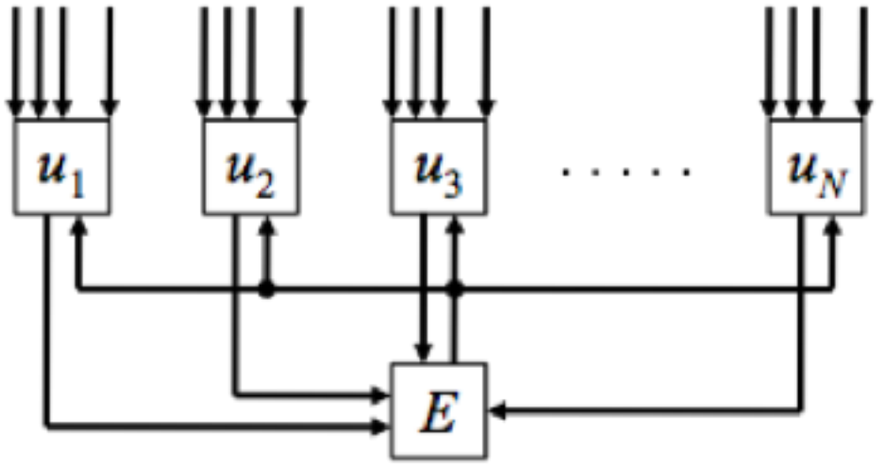
EGHR vs a cascade of PCA & ICA



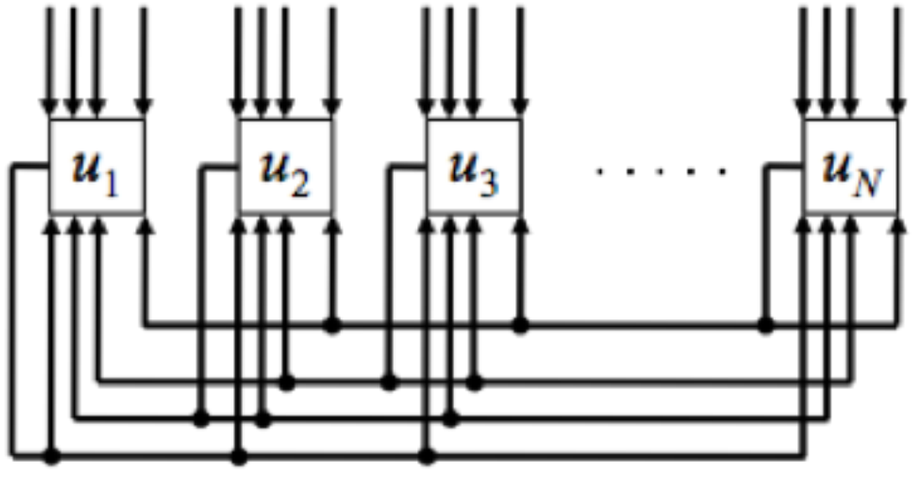
Isomura and Toyoizumi 2018

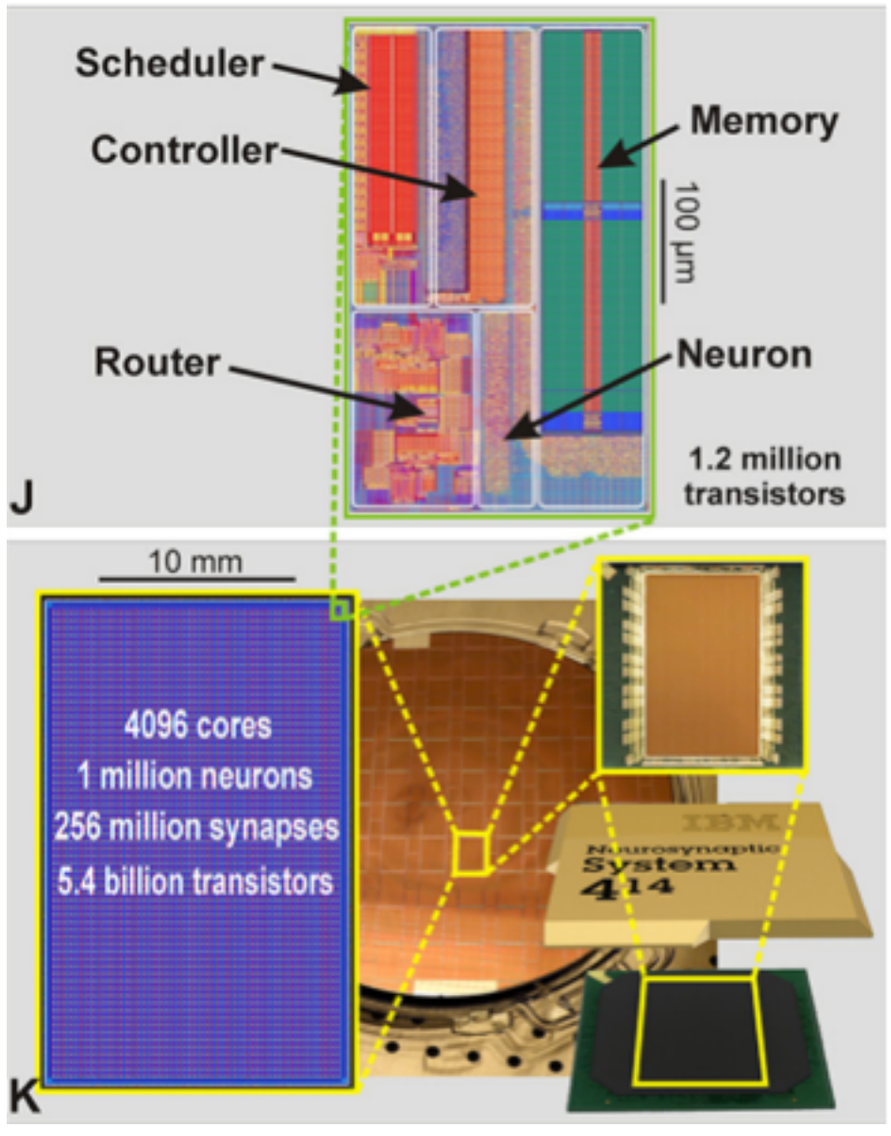
Large-scale and parallel neuromorphic computation

Proposed rule (parallel)



Conventional rules (serial)





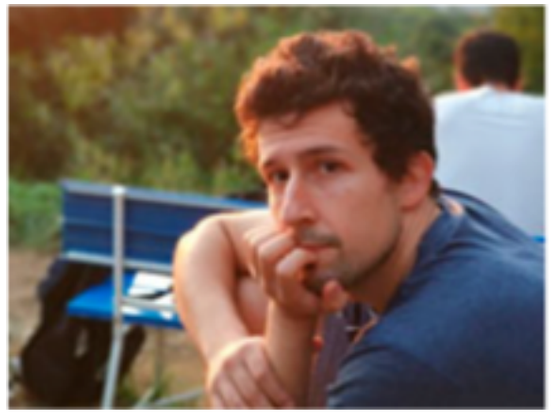
Merolla et al. 2014

Summary

- We have developed a local ICA&PCA algorithm, EGHR, that requires only global information sharing by a third plasticity factor.
- EGHR is more robust than conventional local ICA rules.
- EGHR is suitable for parallel blind-source separation and feature extraction by neuromorphic hardware.
- Generally, information decoding is hard if neural activities are correlated. The independent coding by EGHR can work with a simple decoding scheme.

Collaborator

- Lukasz Kusmierz (RIKEN)



- Takuya Isomura (RIKEN)

Funding

Brain/MINDS

Brain Science Institute RIKEN

Thank you