脳機能分化・機能分割を実現する情報論的機構仮説
A hypothesis on the information-theoretic mechanism realizing the functional differentiation and parcellation

津田一郎
Ichiro Tsuda
札幌市立大学AITセンター
AIT center, Sapporo City University

CHAIN セミナー@北大学術交流会館第３会議室
２０２４年6月２８日

CREST Symbiotic Interaction

An exploration of the principle of emerging interactions in spatiotemporal diversity (2017－2022+2022－2023)

PI: Ichiro Tsuda

Co-PI: Yuji Kawai, Takashi Ikeda, Ikki Matsuda, and Tatsuya Kameda

Principle of Emerging Interactions

JST CREST
An exploration of the principle of emerging interactions in spatiotemporal diversity

Summary of Achievements via Collaborative Works

Application of reBASICS to rapid adaptation of robot arm

Application of ERC for a genesis of functional differentiation

Tsuda G・Kawai G

Evolutionary Reservoir Computers （ERC)
reBASICS

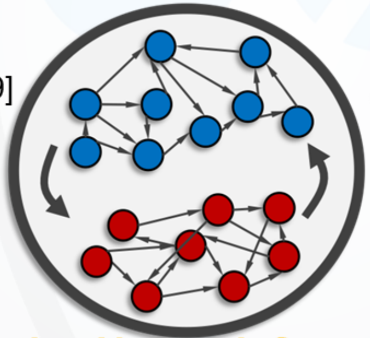
[Kawai et al 2022; Yamaguti et al 2019; Kawai et al 2019]

Small-World Networks

[Kawai et al., 2019]

Transfer Entropy Maximum

[Yamaguti & Tsuda, 2019]

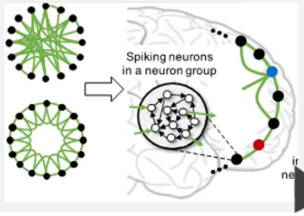


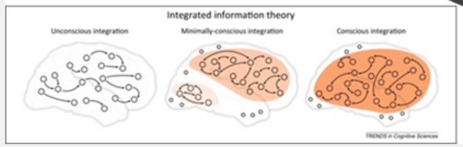
Effective Network Structure

Ikeda G

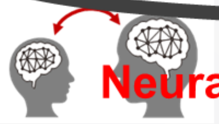
Autistic Brains
Networking in the Brain

[Park et al., 2019]





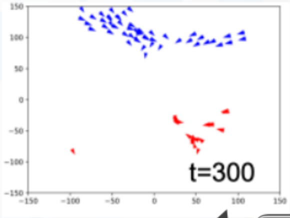
(with Tsuda G)

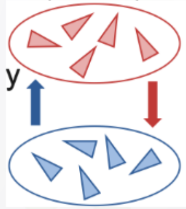


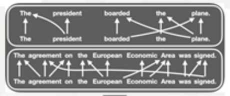
Matsuda G
Kameda G

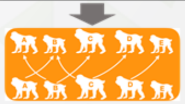
Emergence of Collective Intelligence
Evolutionary and Social conditions

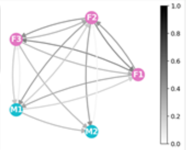
[Suganuma et al., 2019;Morita et al, 2020]







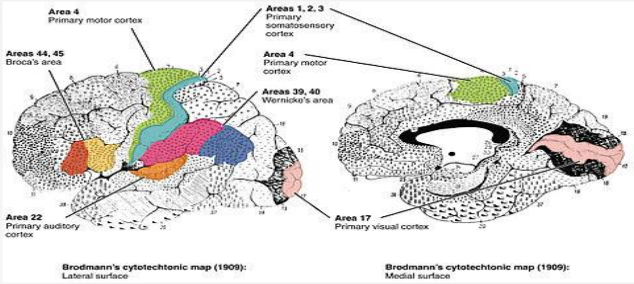




Neural Mechanism of Functional Differentiation based on Constrained Self-Organization

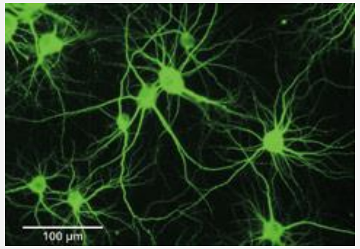
Research questions on functional differentiation

a. How was ***functional differentiation*** organized in the brain?
What is an organization mechanism?
Can it be described by conventional ***self-organization*** theories?
⇒ To answer these questions, ***constraint*** is a key concept.

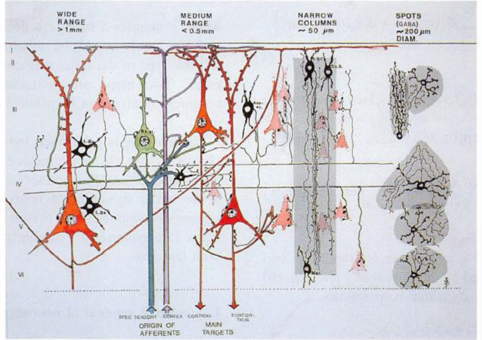


Brodmann areas (functional map)

b. How were ***neurons*** and ***neural modules*** generated?

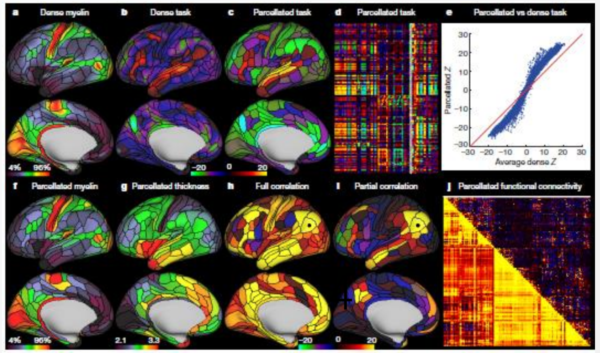


Gohara 2010



Szentagothai 1992

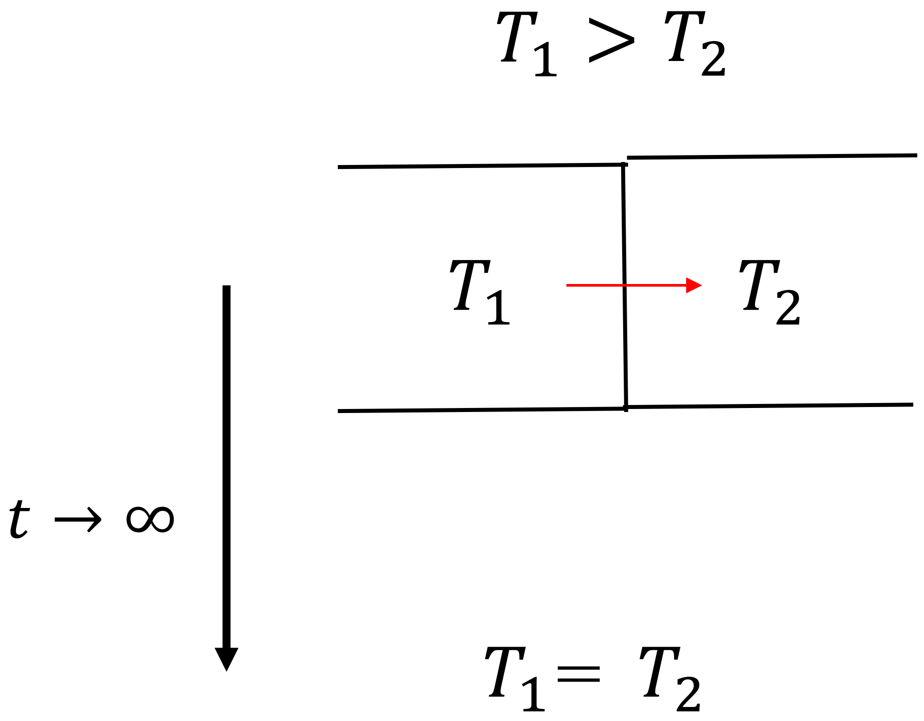
c. How is dynamic organization of function temporarily generated? ***functional parcellation***



Glasser et al., 2016

Thermal contact of two materials with different temperatures

This is the case of **no constraint**: entropy increases in time, finally an isothermal state occurs



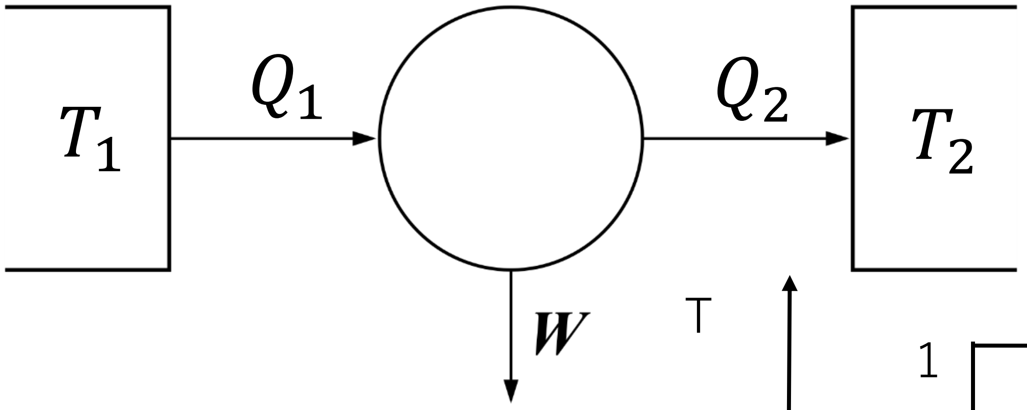
In a closed system, entropy increases
according to the second law of thermodynamics.

Then, **only prediction** is possible **for macroscopic states**.

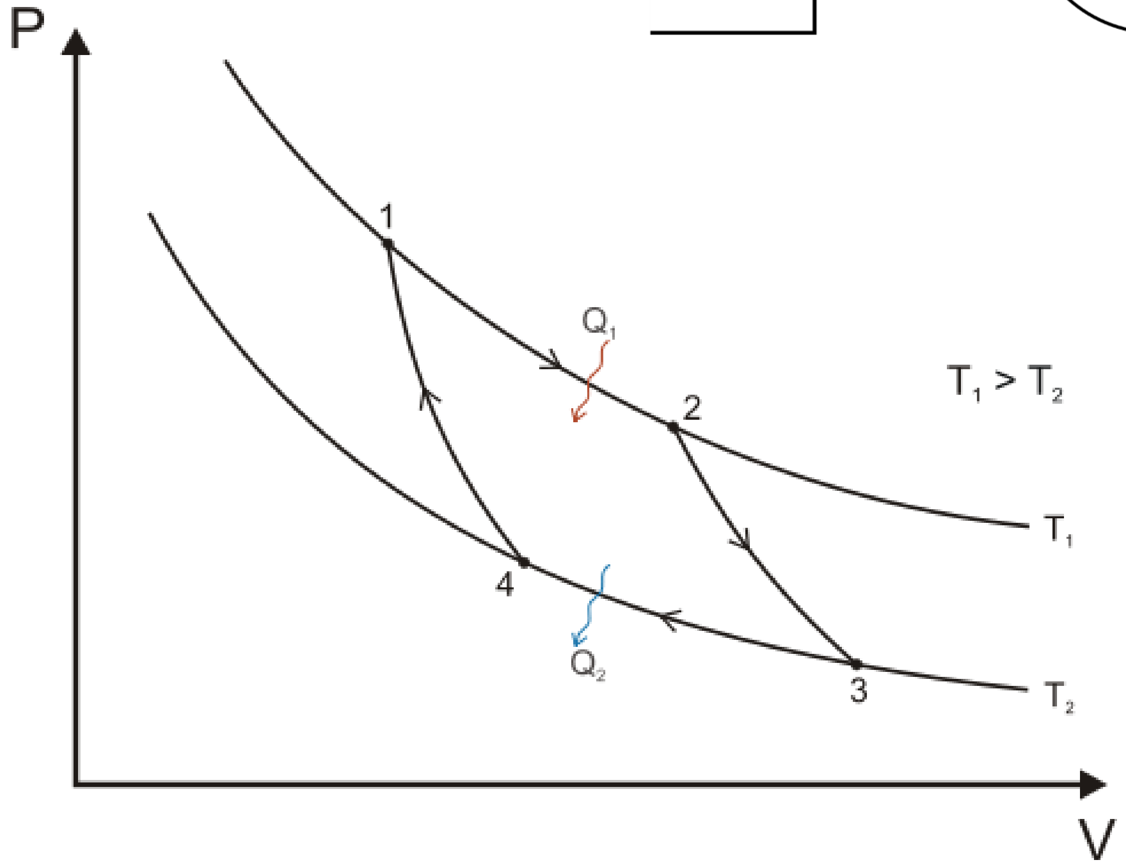
Carnot Cycle

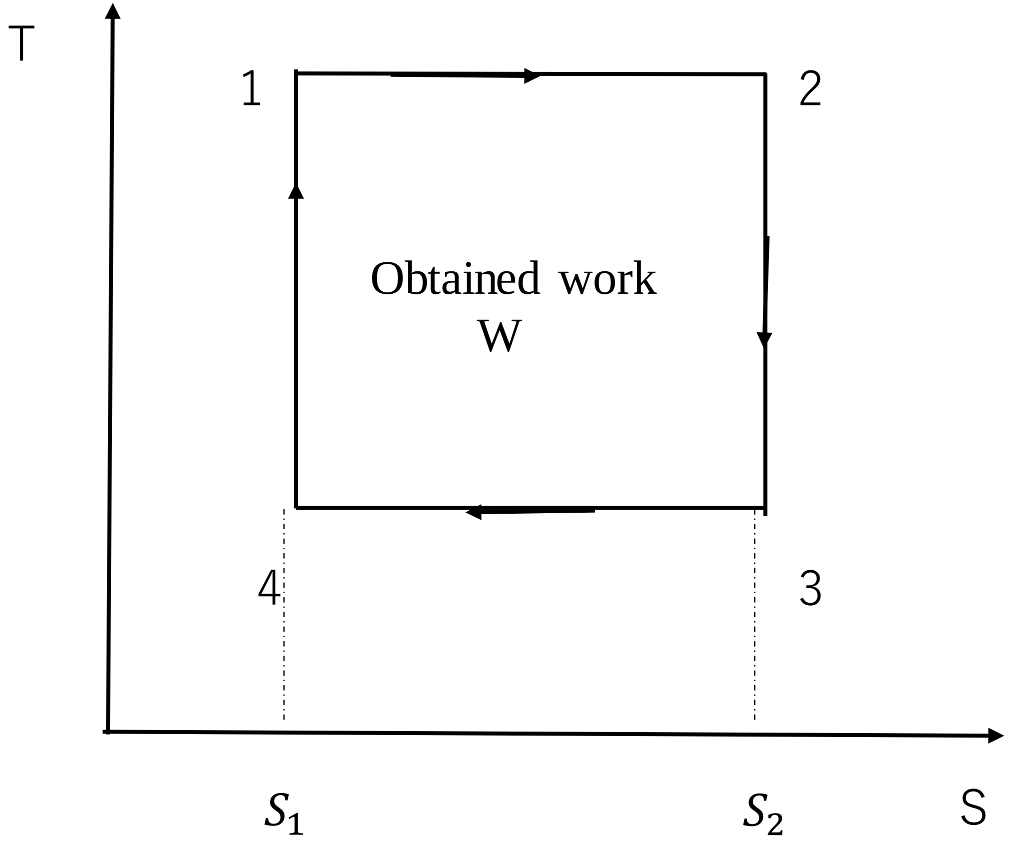
Adiabatic contraction →
Isothermal expansion →
Adiabatic expansion →
Isothermal contraction

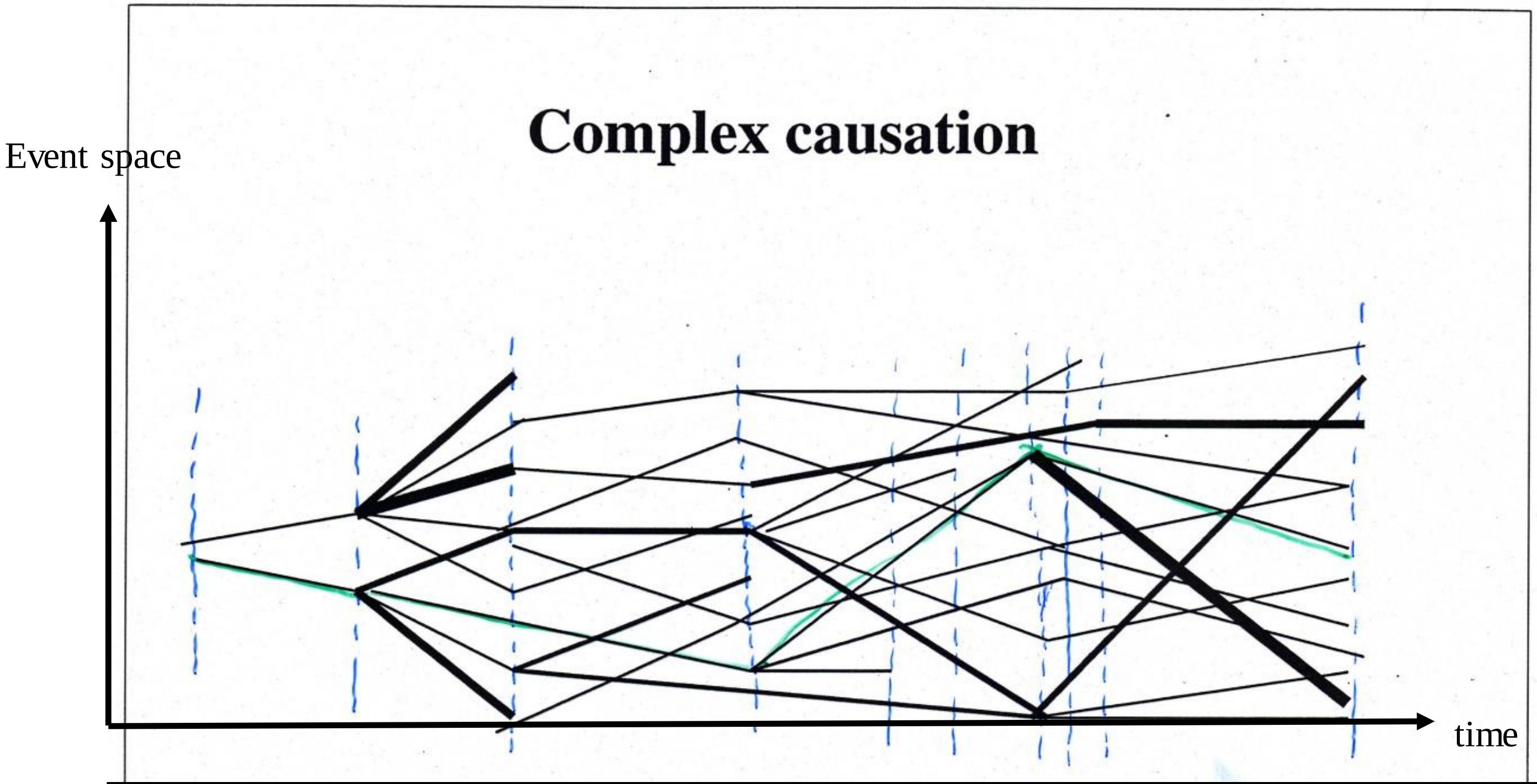
: **Constraints**



In this cycle, **entropy change** should be **zero**.







At each bifurcation (multi-furcation) point, events with high probabilities do not always happen, so that *rare events* as a whole series of events with low probabilities can happen with probability 1, which is proved by **principle of large deviation**: This principle assures the appearance of low entropy events.

If there is **no constraint**, What happens?

Entropy increases in time.

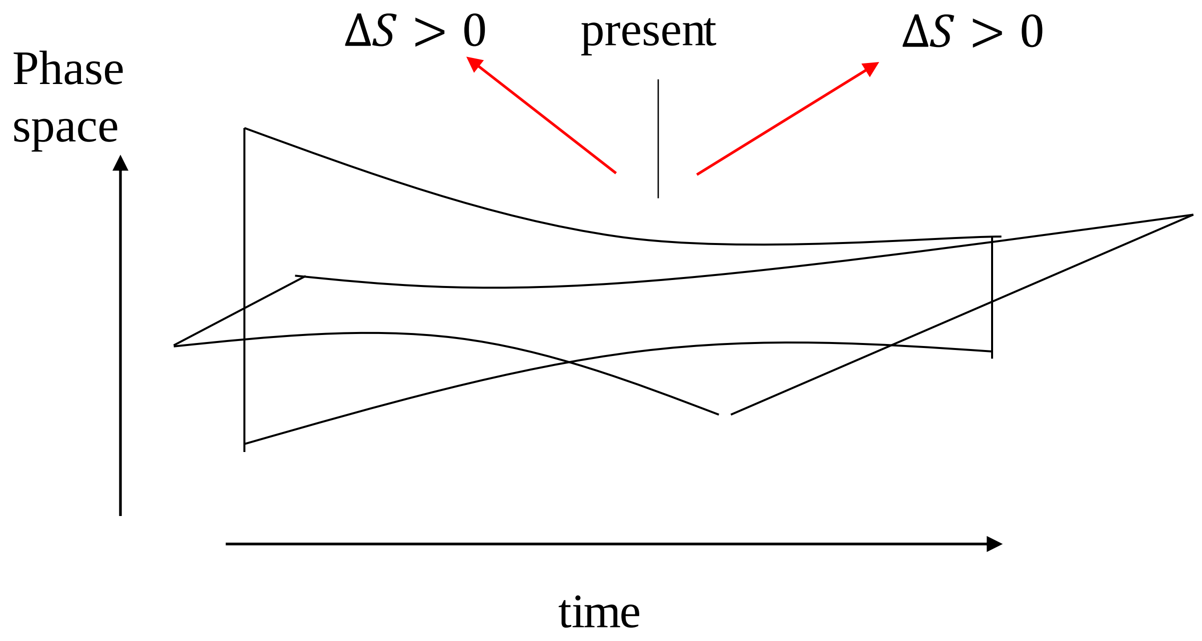
We can arbitrarily select initial conditions at any time because of the increase of entropy.

If a certain **constraint** is provided at a certain future time, What happens?

Entropy decreases in time.

We cannot arbitrarily select initial conditions; we have to find good i.c. to satisfy the constraint.

Life is more than a **Carnot cycle engine**, which obtains work by realizing zero entropy change, namely a reversible process, in the environment where entropy increases, whereas living organisms make far-from-equilibrium conditions and produce **order out of chaos**, which is an irreversible process.



In **chaotic dynamical systems**, "entropy" increases in both future and past direction of time.

$\longrightarrow$

Both **prediction** and **past estimate** are possible: $\boldsymbol{Bayesian\ inference}$ holds.

$\Rightarrow$ *Implying that life can have an ability of prediction and past estimate only in chaotic environments*

For cause A and effect X, $A \rightarrow X$

$$p(A|X) = \frac{p(X|A)p(A)}{\sum_A p(X|A)}$$

(a) What is the difference of *self-organization with constraints*(SOC) from conventional self-organization (SO)?

SO:

Macroscopic orders are generated via interactions of atoms and molecules at microscopic levels

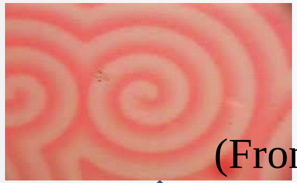
Nicolis, G., Prigogine, I. (1977); Haken, H.(1977)

Fixed boundary conditions!

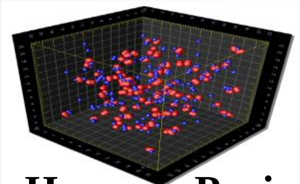
Macroscopic order

↑

Microscopic interactions



(From Kitahata http://www.chem.scphys.kyoto-u.ac.jp/nonnonWWW/kitahata/bz_1.html)



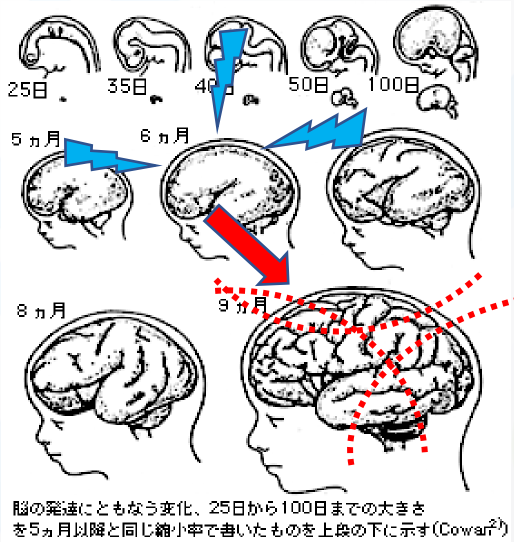
(**From Honours Project: MolDyn - Molecular Dynamics**
Veselin Dikov, Niko Manopulo, Darya Popiv, Ilya Saverchenko, Levi Valgaerts)

SOC:

Systems elements and subsystems are generated via constraints acting on the system

Open boundary conditions!

(ex)



(from http://www.yuchan.net/yuchan/dictionary/new_ikuji/brain004.html)

Constraints acting on a whole system

↓

Functional differentiation: **self-organization of functional elements**

Our view: **functional differentiation** in the brain should be formulated within the framework of **self-organization with constraints**, where the functional elements (or components, or subsystems) are produced by constraints that act on a whole system.

Tsuda, I. *Prog. Theor. Phys*. 1984;
Rosen, R. *Life Itself* , Columbia Univ. Press, 1991;
Freeman, W. J., *How Brains Make Up Their Minds*, A Phoenix Paperback, 1999;
Tsuda, I. *Behav. Brain Sci*. 2001;
Freeman, W. J., *Biol. Cybern*. **92**, 350−359 (2005).
Tsuda, I. et al. *Entropy* 2015;
Shimizu, H.: http://www.banokenkyujo.org/?page_id=48

Self-organization theory with constraints in *neural networks*

Ex)

・C. Von der Malsburg, *Kybernetik* 1973

A model for the primary visual cortex: a construction of orientation sensitive cells

Constraint : a sum of the connected weights to each neuron should be constant.

・S. Amari, *Bull. Math. Biol.* 1980

Topographic mapping via competitive neural networks by mutual inhibitions

Constraint : winner-take-all

・T. Kohonen, *Biol. Cybern.* 1982

SOM can be viewed as Expectation Maximization algorithm (EM)

Constraint : maximization of expectation

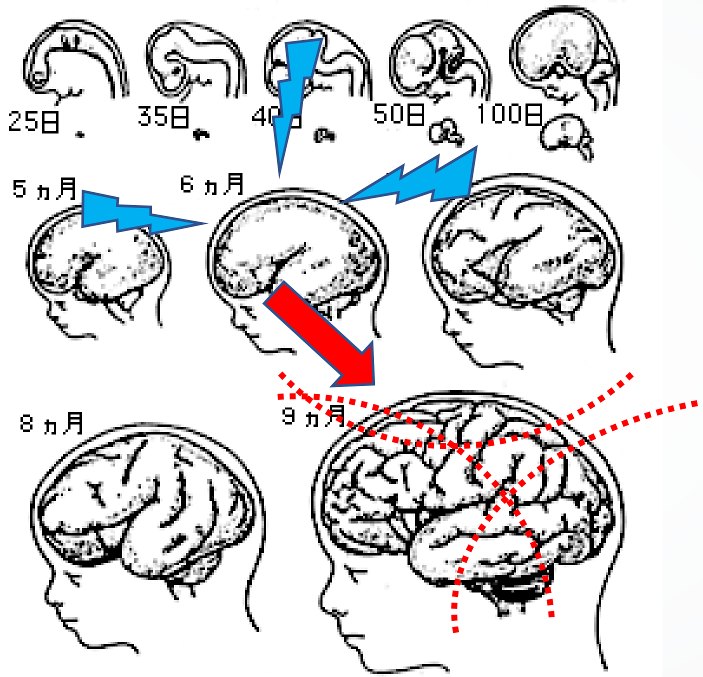
・H. Haken and J. Portugali, *Information Adaptation* (Springer, 2015)

Pattern recognition based on pattern formation

Constraint : attention parameters

Assertion(Emerging Science Principle)（創発原理）：

In living systems, self-organization with constraints is a blueprint, thereby ”functional differentiation=a genesis of functional elements”

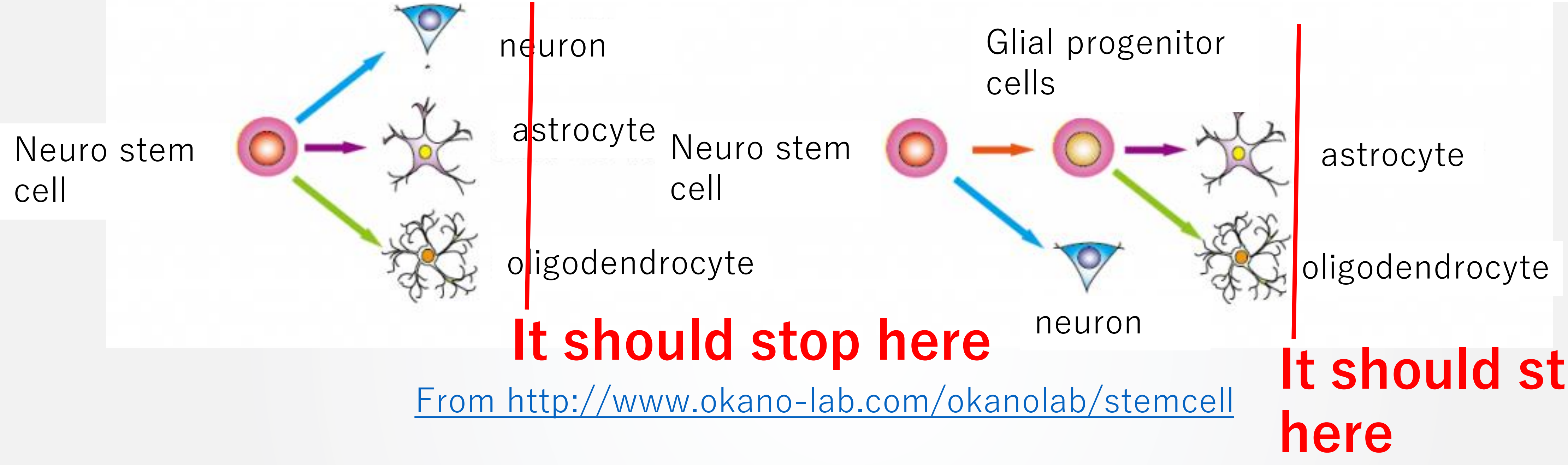


脳の発達にともなう変化、25日から100日までの大きさを5ヵ月以降と同じ縮小率で書いたものを上段の下に示す(Cowan[2])

Functional differentiation/percellation and cell differentiation follows developmental dynamical systems

Direct differentiation model

Intervening progenitor cells model



From http://www.okano-lab.com/okanolab/stemcell

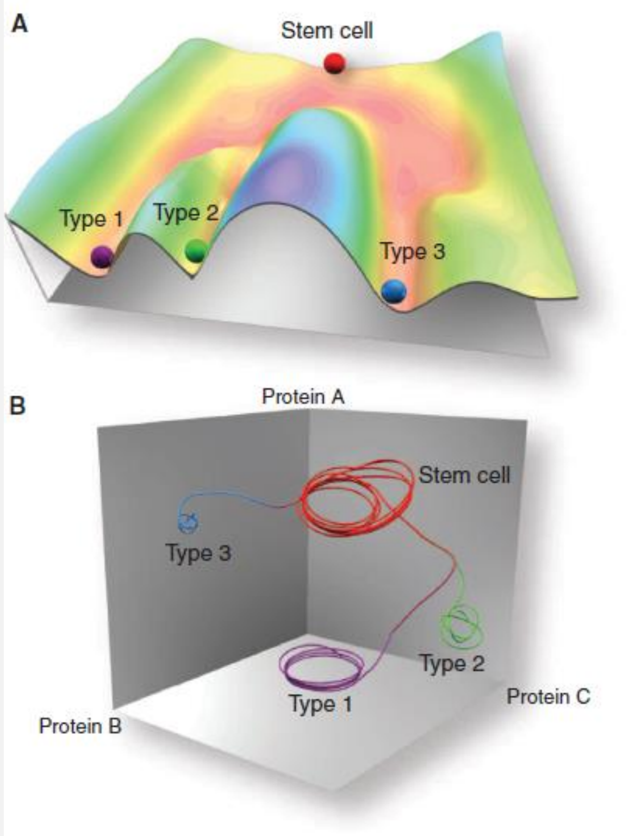
The existence of **neuro stem cells** in the third ventricle⇒**Neurogenesis** in dentate gyrus, side subventricular zone, etc. ⇒

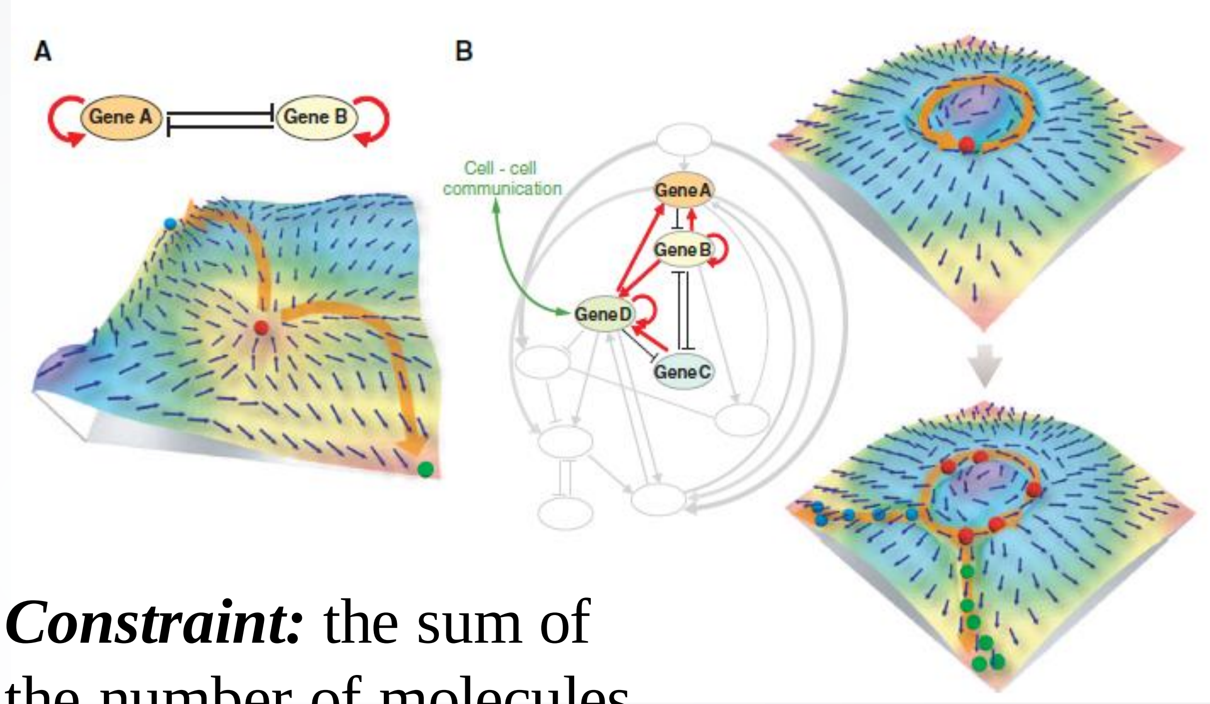
・Treatment model of brain injury via **acceleration of differentiation** by stimulation in the growth factor of progenitor cells

・Treatment model of brain injury via nerve graft with neuro stem cells stemming from fetus brains, stem cells, and iPS cells

⇒Finding the mathematical principle of neuronal differentiation is important.

Dynamical systems approach is useful for cell differentiation





Constraint: the sum of the number of molecules in each cell is fixed

C. Furusawa and K. Kaneko, *Science*, 2012

cf) Chambers et al., *Nature* 2007; Hayashi et al, 2008

Constrained mechanics

1. Holonomic：

・The degree of freedom of infinitesimal change equals the one of global change: Integrable

2. Nonholonomic：

・The former degree of freedom does not equal the latter: nonintegrable

・d'Alembert's principle(principle of virtual work)

・deleting Lagrange multipliers⇒feedback control (selecting a solution following constraints among extrema of functional)

・causal

3. Vakonomic mechanics (by V.V.Kozlov)：

・Lagrange multipliers are independent variables, depending on both initial and final states

⇒ Final state sensitivity

・Optimal control theory (Andronov-Pontryagin)
（finding an extremum, giving constraints on variational manifold)

・noncausal

CF) The case of constrained Hamiltonian systems ⇒for example, Dirac method

I. Tsuda, Y. Yamaguti, H.Watanabe
Entropy 2016, 18, 74

Step1. on $\Omega(\mathrm{n}) \times R^l$

$$\frac{dx}{dt} = f(x;\lambda)$$

The description of neural dynamics

Step2. on $\Omega(\mathrm{n}) \times R^l \times \mathrm{W}$

$$\frac{dx}{dt} = f(x,\lambda) + G(x,t)$$

The description of the interactions between neural systems and environments

Step 3.

The description of constrained dynamics

$$\delta L = \delta \int_0^T \{C + \mu(\frac{dx}{dt} - f(x,\lambda) - G(x,t))\}dt = 0,$$

C: ***intentional*** constraints

↓

3.1 In the case that C is quantified, on $\Omega(\mathrm{n}) \times R^l \times \mathrm{W} \times R^n$

$$\begin{cases} \frac{dx}{dt} = f(x,\lambda) + G(x,t) \\ \frac{d\mu}{dt} = \mathrm{h}(\mu, x, \dot{x}) \end{cases}$$

μ: Lagrange multiplier

The framework of self-organization with constraints for yielding functional elements

$$\delta \int_0^T \{C(y(x),t) + \mu(x,\dot{x},t)\left(\frac{dx}{dt} - f(x,\lambda) - G(y(x),t)\right)\}dt = 0$$

$C(y(x),t)$: ***Intentional and Informational Constraints***

cf) Pattee, *J. Social Biol. Struct.* 1978;
Tschacher and Haken, *New Ideas in Psychol.* 2007

$\frac{dx}{dt} - f(x,\lambda) - G(y(x),t) = 0$: ***Dynamical Constraints***

Note: Lagrange multipliers are a function not only of a state variable x, but also of its time-derivatives $\dot{x}$ and time t, and their equations of motion are derived

$\Rightarrow$ *Vakonomic dynamics* (Kozlov, V.V. 1983)

$\Rightarrow$

Theorem 1
The vakonomic dynamical systems, derived from any differentiable dynamical systems are *linearly unstable* on subspace of Lagrange multiplies.

Optimization with Exponential Discount

$$\begin{cases} \text{Minimize} & \int_0^\infty e^{-\tilde{\rho}t}|p(t)|^2/2dt \\ \text{subject to} & \dot{z}(t) = F(z(t)) + p(t), z(0) = z_0, \lim_{t\to\infty} z(t) = z_\infty. \end{cases} \quad (10)$$

Give the Lagrangian with a discount term $e^{-\tilde{\rho}t}$ by

$$\mathcal{L}(z, p, \mu) = e^{-\tilde{\rho}t}|p|^2/2 + \mu^T(\dot{z} - f(z) - p),$$

and by the variational principle, we have

$$\begin{cases} \dot{z} = F(z) + pe^{\tilde{\rho}t} \\ \dot{p} = -D_F(z)^T p. \end{cases}$$

Changing the variables $q(t) = p(t)e^{\tilde{\rho}t}$ leads the autonomous system on $\mathbb{R}^{2n}$,

$$\begin{cases} \dot{z} = F(z) + q \\ \dot{q} = \tilde{\rho}q - D_f(z)^T q. \end{cases} \quad (11)$$

The eigenvalues of the Jacobian matrix for the fixed point $(z_*, 0)$ for the system (11) become λ_i and $-\lambda_i + \tilde{\rho}$. Thus, when $\lambda_i < 0$ for all i, the point $(x_*, 0)$ can become stable fixed point by taking $\tilde{\rho} < 0$ satisfying $+\lambda_i - \tilde{\rho} < 0$ for all i. By substituting $\rho = -\tilde{\rho}$, we obtain the system $(*)$. Therefore, we can expect that it is possible to realize the stable control by considering infinite horizon optimal control problem with "negative" discount.

Note that, by Legendre transformation, the system (11) can be rewritten as

$$\begin{cases} \dot{z} = \frac{\partial H}{\partial q} \\ \dot{q} = -\frac{\partial H}{\partial z} + \tilde{\rho}q \end{cases} \quad (12)$$

where the Hamiltonian H is given by

$$H(z, q, \mu) := e^{-\tilde{\rho}t}|q|^2/2 + \mu^T(F(z) + q)$$

Particularly, applicable to the problems of convex constraints such as least energy consumption

Theorem 2

There exist positive-measure initial conditions to reach a given final state.

Example: Finding optimal perturbations of ***Bonhoeffer-van der Pol equations***

$$\dot{x} = c\left(y + x - \frac{x^3}{3} - r\right) + I$$

$$\dot{y} = -(x - a + by)/c$$

Forger, D.B. et al, JTB 230(2004)521-32

$a = 0.7, b = 0.8, c = 3.0, r = 0.342, I = 0$: *subcritical Hopf bifurcation*

- Changing the state from a limit cycle to a stable fixed point by adding an external signal
- Choosing an optimal external signal (variational principle)

External signal: $z(t)$

boundary conditions : $\left(x(t_i), y(t_i)\right)$ *on limit cycle*; $\left(x(t_f), y(t_f)\right)$ *on fixed point*

$$L = z(t)^2 + \mu_x(t)\left\{\dot{x} - c\left(y + x - \frac{x^3}{3} - r\right) - I - z(t)\right\} + \mu_y(t)\{\dot{y} + (x - a + by)/c\},$$

where $\mathrm{C} = z(t)^2$

$$\delta \int_{t_i}^{t_f} dt\, L\left(x, y, z, \mu_x, \mu_y\right) = 0$$

Euler-Lagrange equations: $\frac{dL}{dq_i} - \frac{d}{dt}\frac{dL}{d\dot{q}_i} = 0 \left(q_i = x, y, z, \mu_x, \mu_y\right)$

- $q_i = \mu_x, \mu_y \rightarrow BvP\ eq$
- q_i=z → $z = \mu_x/2$
- q_i=x → $\frac{d\mu_x}{dt} = -c(1 - x^2)\ \mu_x + \mu_y/\mathrm{c}$
- q_i= y → $\frac{d\mu_y}{dt} = -c\ \mu_x + b\mu_y/\mathrm{c}$

As $\left(\mu_x(0), \mu_y(0)\right)$ *are not given, start from appropriate values*

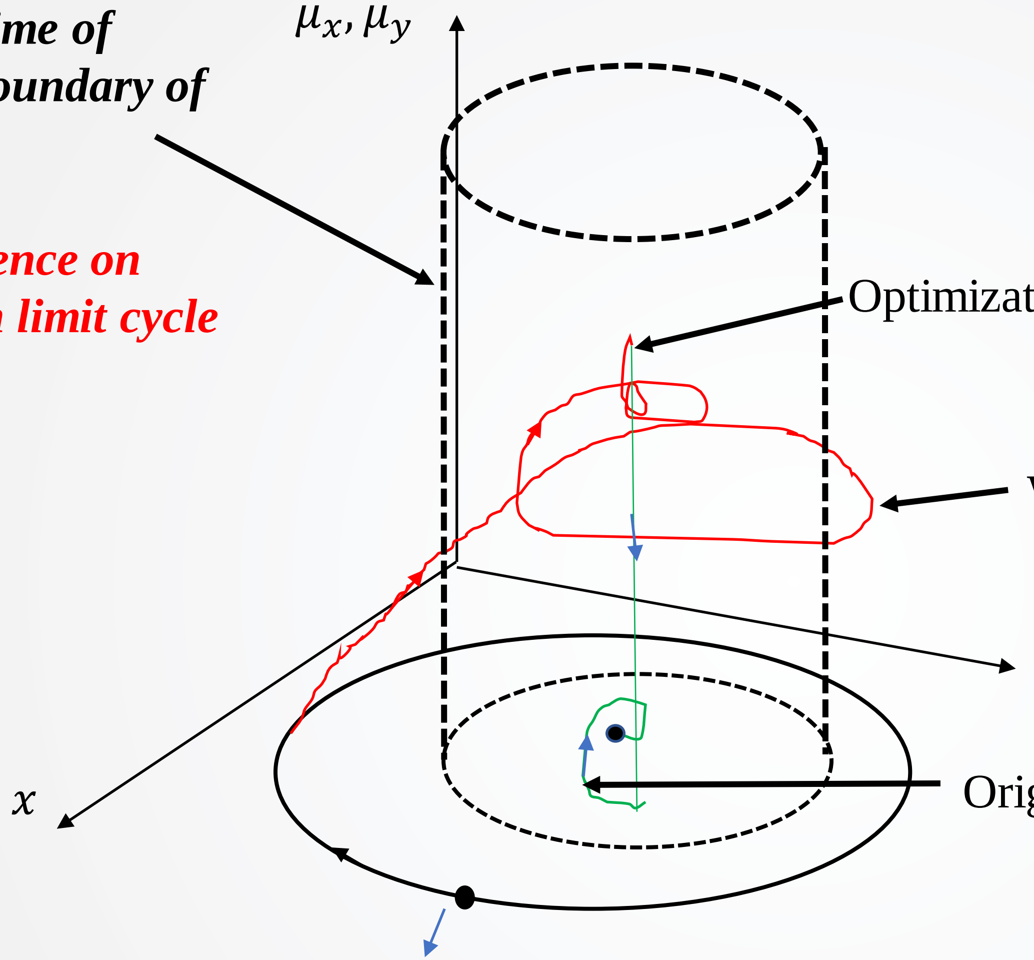
As a conservation quantity, Hamiltonian can be defined.

$Eigenvalues\ of\ Jacobi\ matrix\ at\ (x_{fp}, y_{fp}, 0, 0)$:

$$\lambda_{1,2} \sim -0.01103 \pm 0.96677i$$
$$\lambda_{3,4} \sim 0.01103 \pm 0.96677i$$

Obtain the arrival time of trajectories at the boundary of cylinder

$\Rightarrow$ ***sensitive dependence on initial conditions on limit cycle***



$\Rightarrow$ Another solvable method: exponential discount: $\delta \int_{t_i}^{t_f} dt\, e^{\rho t} L(x, y, z, \mu_x, \mu_y) = 0$

In order to obtain global stable states, we adopted a genetic algorithm.

Evolutionary and developing dynamics via genetic algorithm

Dynamical and other constraints

$$\begin{cases} \dfrac{dx}{dt} = f(x;b) \\ \text{with } C \end{cases}$$

Dynamical Systems：{state space、dyn. rule（transition rule of states}
Family of D.S.：{D.S., bifurcation parameters}

- **Genes** **defined by** **bifurcation parameters** b
- **Cells' states** **（via activated proteins and/or electric activity）** **defined by** **dynamical variables** x
- **A principle of evolution-development yielding functional differentiation** **defined by** **self-organization with constraints** C

Changing dynamical systems $f(x;b)$ by changing parameters b to satisfy constraints C.
⇒Finding a set of initial conditions satisfying constraints on (subspaces) of function space.

(b), (c) Applications of the variational principle

1. Mathematical modeling of ***functional modules***

1.1 Developed a mathematical model in terms of coupled neural oscillators

- The dynamics of each oscillator is defined by

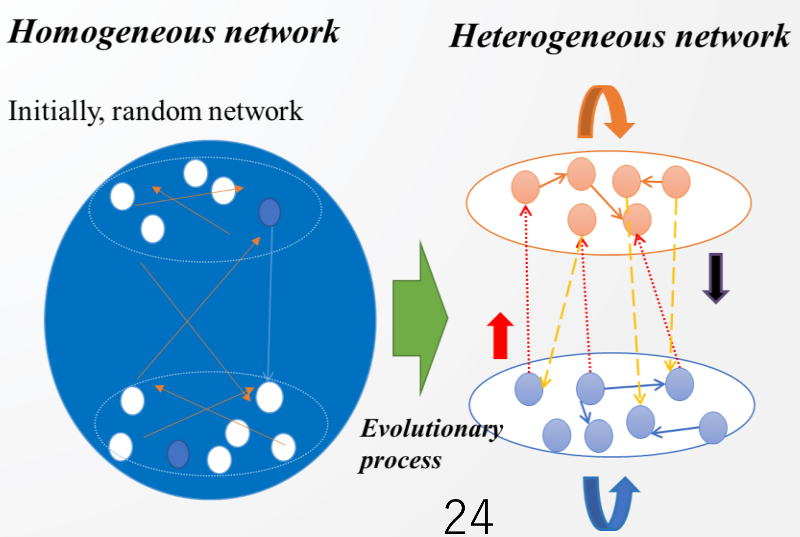
$$\theta_{t+1}^{(i,k)} = \omega^{(i,k)} + \theta_t^{(i,k)} + \frac{\alpha}{Np^c} \sum_{(j,l)\in G^{(i,k)}} \sin\left(\theta_t^{(j,l)} - \theta_t^{(i,k)} - \psi_{kl}^{ij}\right) + \sigma_\beta \beta_t^{(i,k)}$$

for k th oscillator in i th module

Constraint: **maximum transfer entropy**

Evolutionary dynamics yields functional modules

Y. Yamaguti and I. Tsuda, *Neural Networks*, 2015; in preparation 2022



I.Tsuda, Y.Yamaguti, H. Watanabe, *Proc. of ICCN*;
Y. Yamaguti, I Tsuda, *Neural Networks*;
Y. Yamaguti, I. Tsuda, Y. Takahashi, *Cogn. Neurodyn.*

1.2 Functional Differentiation by Minimization of Mutual Information between Neural Groups

(Y. Yamaguti & I. Tsuda, in preparation 2024)

- Developing functionally differentiated structures by minimizing the amount of mutual information between the states of neural groups in RNN by gradient learning.

Theorem and definition used in the model

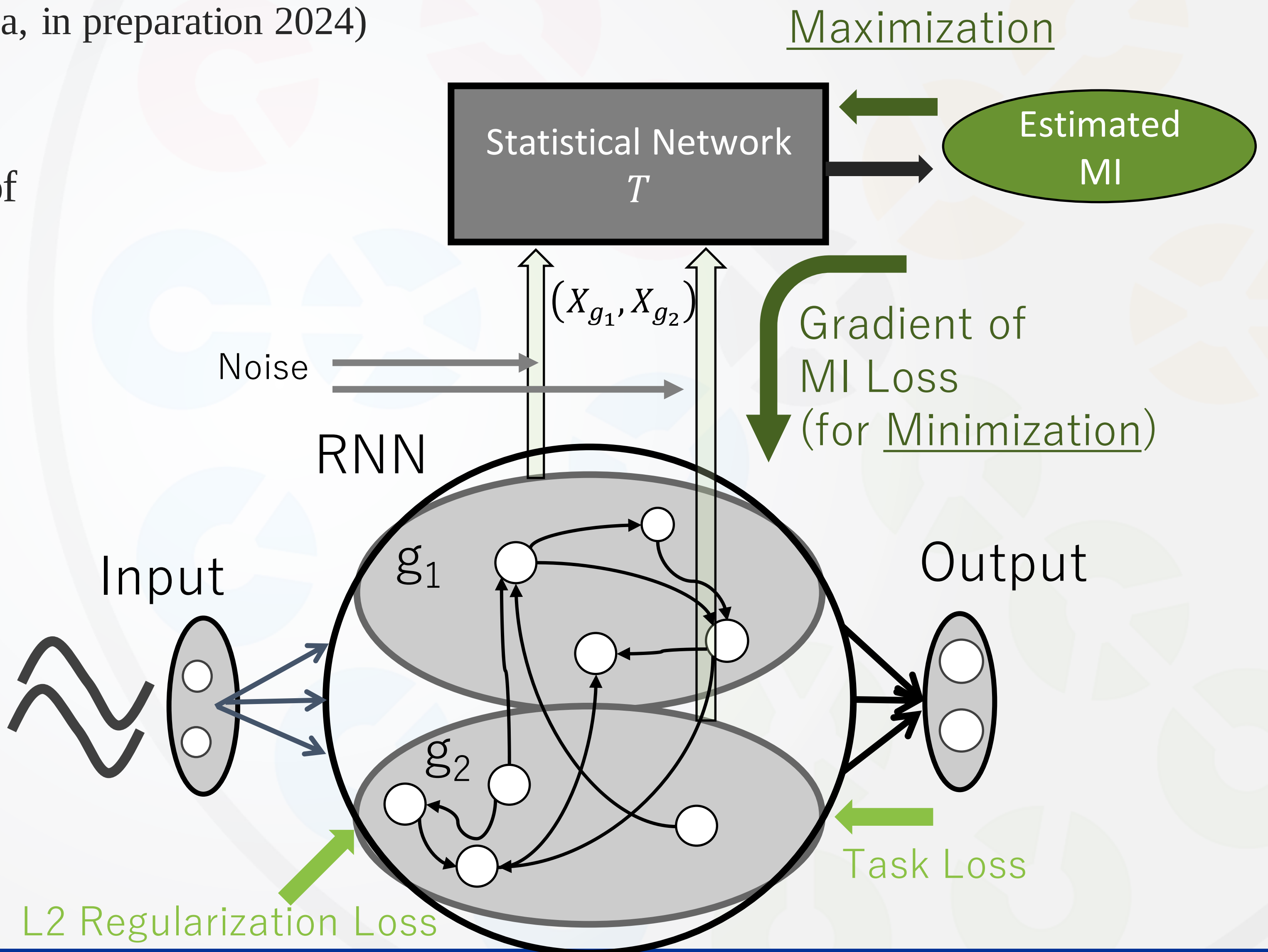
Theorem (Donsker-Varadhan representation). *The KL divergence admits the following dual representation:*

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T]),$$

where the supremum is taken over all functions T such that the two expectations are finite.

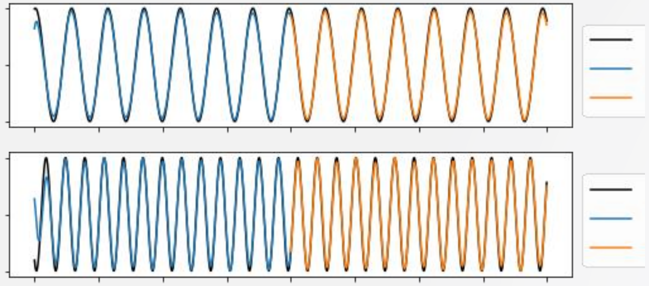
Definition (Mutual Information Neural Estimator (MINE)). Let $\mathcal{F} = \{T_\theta\}_{\theta\in\Theta}$ be the set of functions parametrized by a neural network. MINE is defined as,

$$\widehat{I(X;Z)}_n = \sup_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}^{(n)}_{XZ}}[T_\theta] - \log(\mathbb{E}_{\mathbb{P}^{(n)}_X \otimes \hat{\mathbb{P}}^{(n)}_Z}[e^{T_\theta}]).$$



Task : Multi-frequency sinusoidal signal prediction

Outputs of trained RNN



- The minimization of mutual information led to the formation of functionally differentiated modules.
- During the training, the development of "functional" (correlational) differentiation preceded the development of structural differentiation.

$Q = \sum_{i=1}^{N}(e_{ii} - a_i^2)$, where

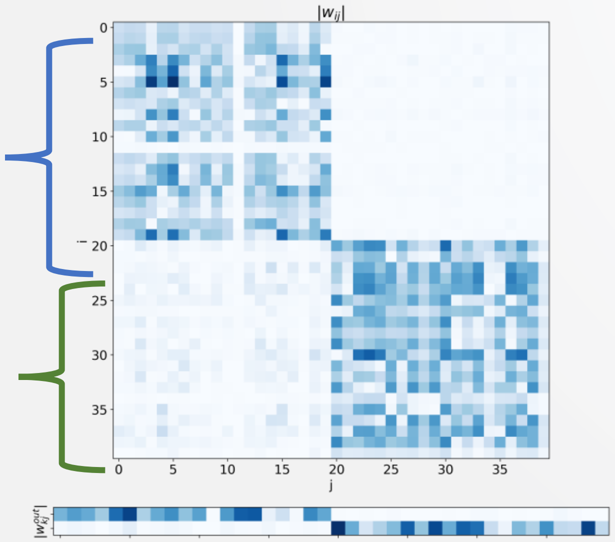
$\sum_j e_{ij} = a_i$ (for random connections)

$< e_{ii} > = a_i^2$

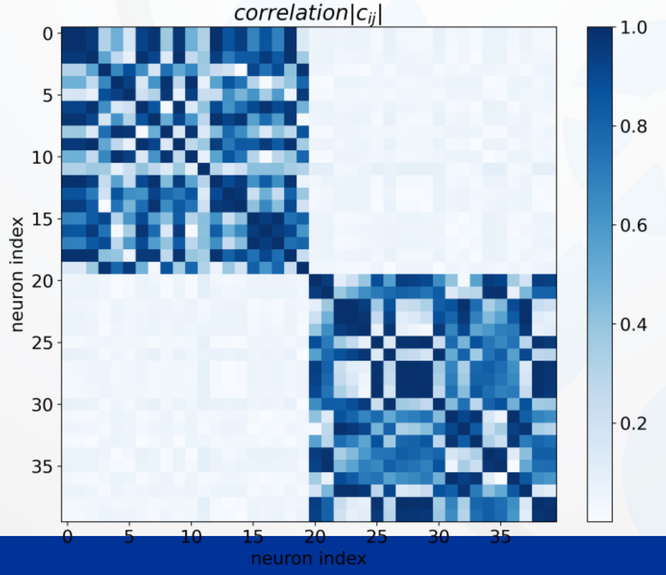
Weight matrix of RNN

Group 1

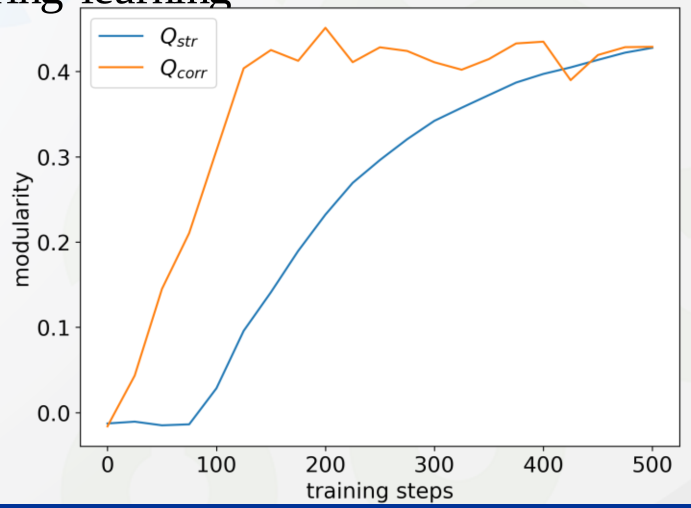
Group 2



Correlations
between Neural units



The development of modularity Q during learning



2. Mathematical modeling for the production of ***spiking neurons* and *glial cells***

・Watanabe, H., Ito, T., Tsuda, I., *Neuroscience Research* **156** (2020) 206-216.
・I. Tsuda, Y. Yanmaguti and H. Watanabe, *Entropy* 2016, **18**, 74:pp1-13.

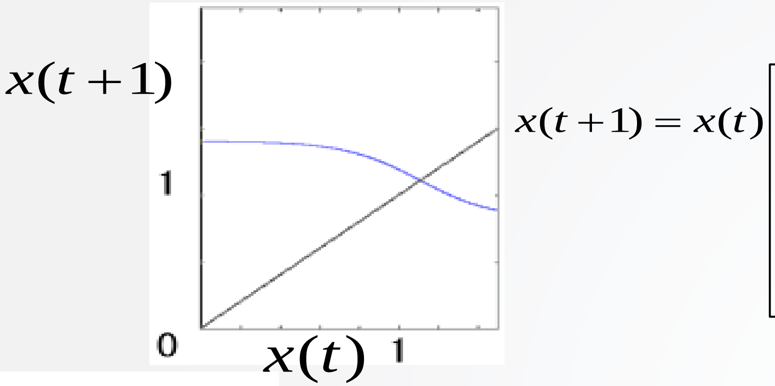
Coupled dynamical systems ⇒ parameters are changed to satisfy the constraint

$$x_0(t+1) = f_0\big(x_0(t)\big) + \sum_l w_{0l} x_l(t) + G(t) + \sigma$$

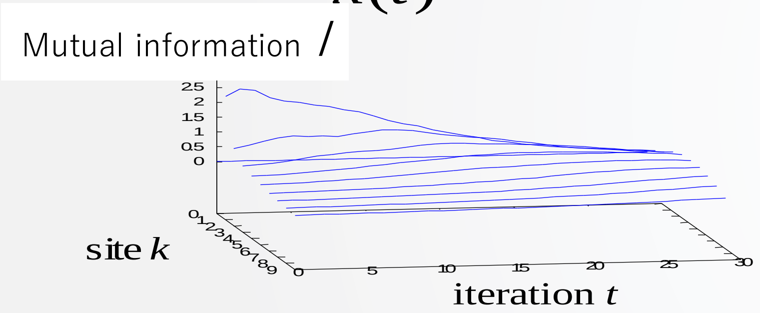
$$x_k(t+1) = f_k\big(x_k(t)\big) + \sum_l w_{kl} x_l(t) + \sigma$$

$$f_k(x) = a_1 \tanh\big(a_2(x - a_3)\big) - a_4 \tanh\big(a_5(x - a_6)\big) + b_k$$

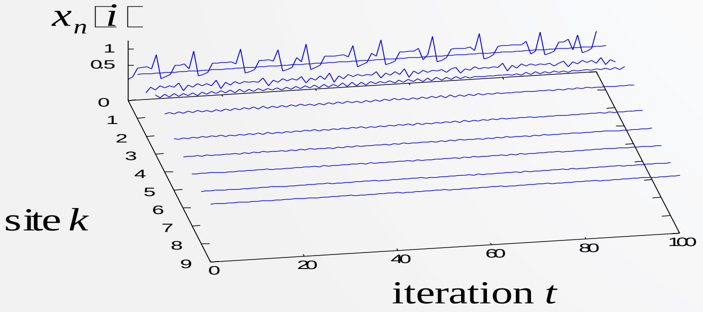
***Constraint*: maximum transmission of information of external signal**



The rule of state change represented by map



Spatio-temporal change of information of external signal

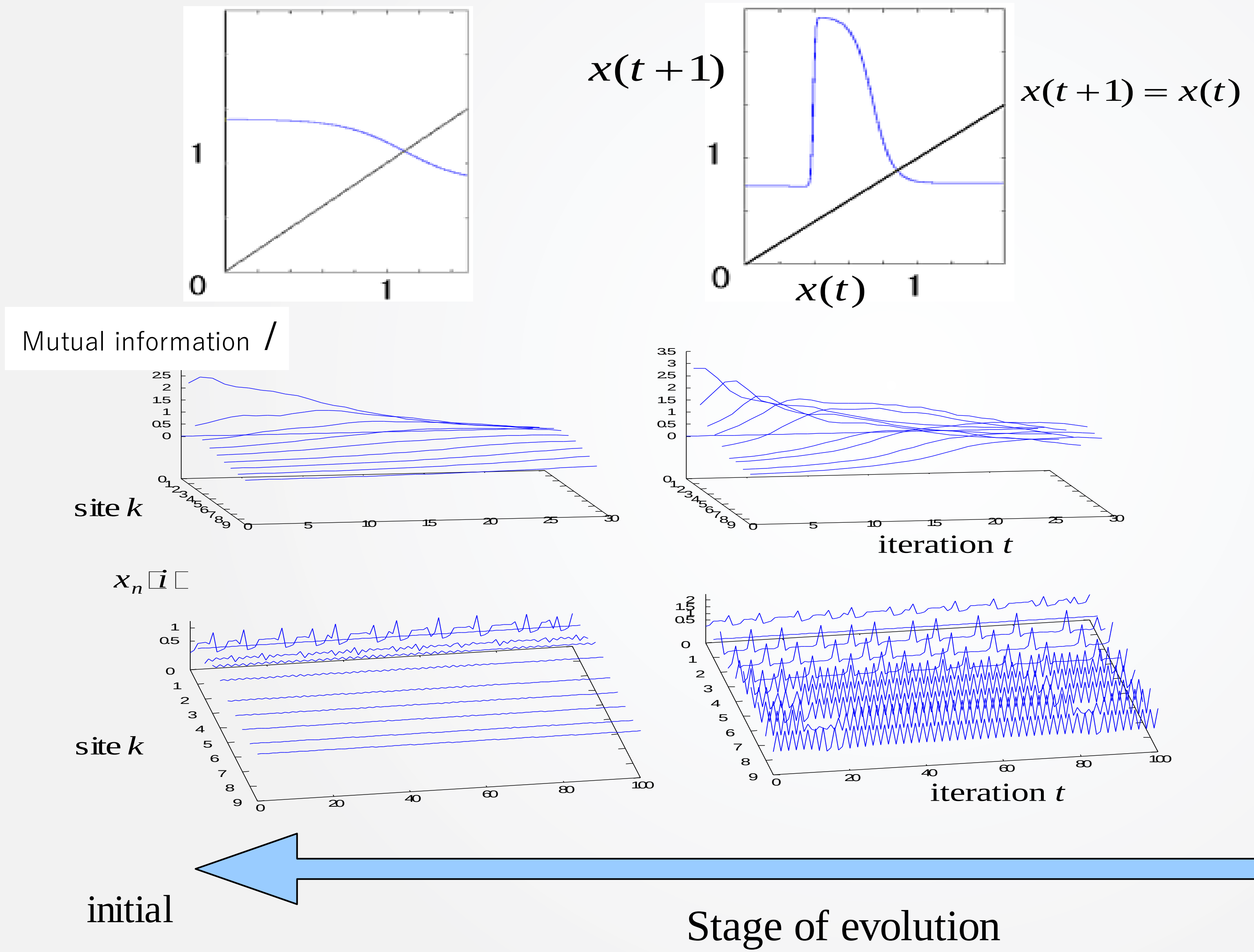


Time series of elementary dynamics

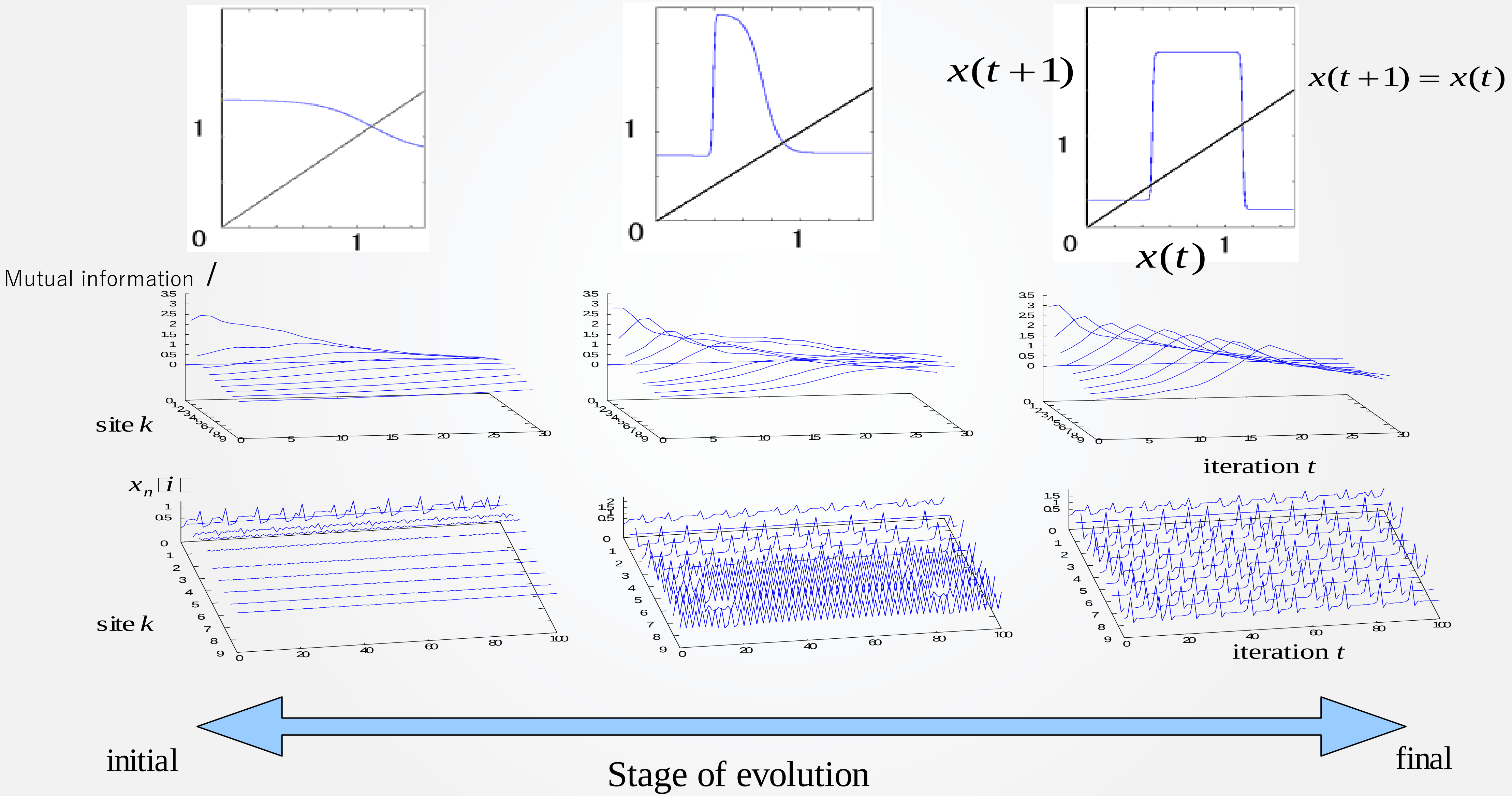
initial

Stage of evolution

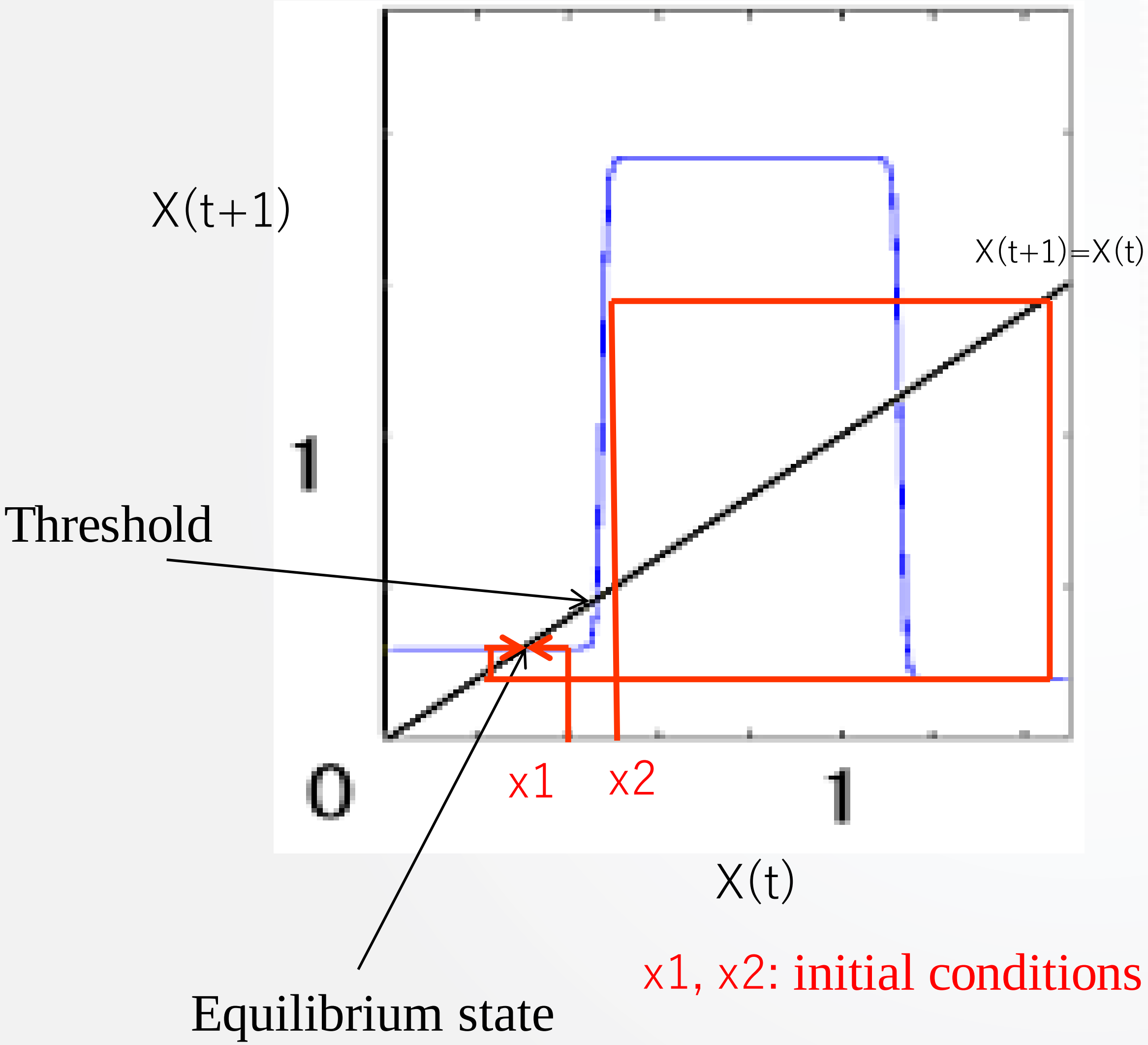
final



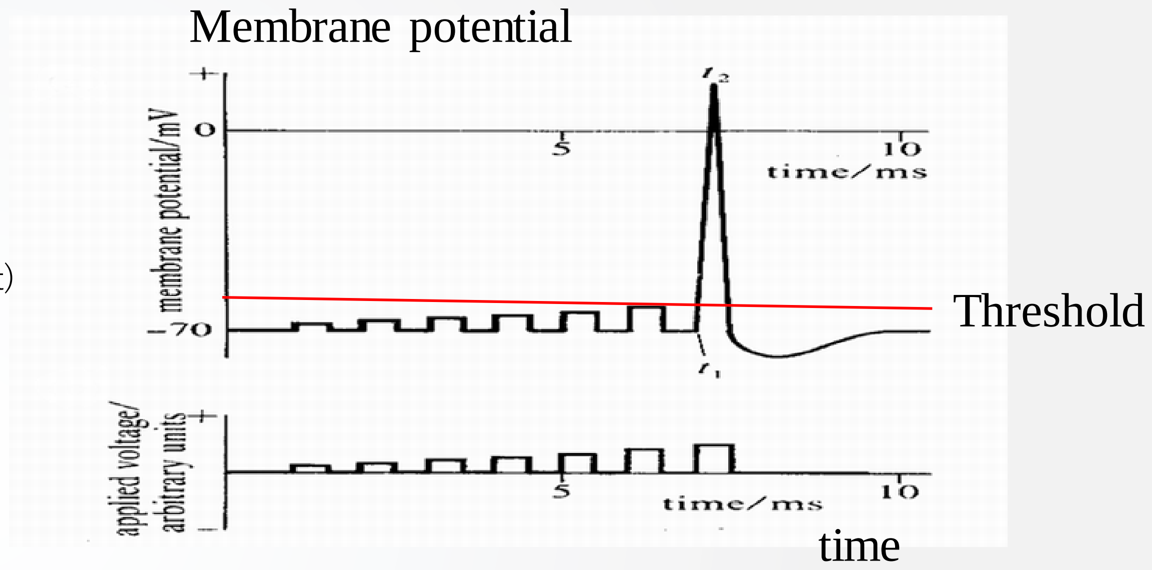
$x(t+1)$
$x(t+1)=x(t)$
$x(t)$
0
1
Mutual information I
site k
iteration t
$x_n(i)$
initial
final
Stage of evolution


x(t+1)
x(t+1) = x(t)
x(t)
Mutual information
site k
iteration t
site k
iteration t
initial
Stage of evolution
final

The excitable system emerges, which possesses characteristics of **neurons**



x1, x2: initial conditions



The resulting dynamics of the network is chaotic:
a neuron works and processes information in chaotic environment.

Coupled dynamical systems ⇒ parameters are changed to satisfy the constraint

$$x_0(t+1) = f_0(x_0(t)) + \sum_l w_{0l} x_l(t) + G(t) + \sigma$$

$$x_k(t+1) = f_k(x_k(t)) + \sum_l w_{kl} x_l(t) + \sigma$$

$$f_k(x) = a_1 \tanh(a_2(x - a_3)) - a_4 \tanh(a_5(x - a_6)) + b_k$$

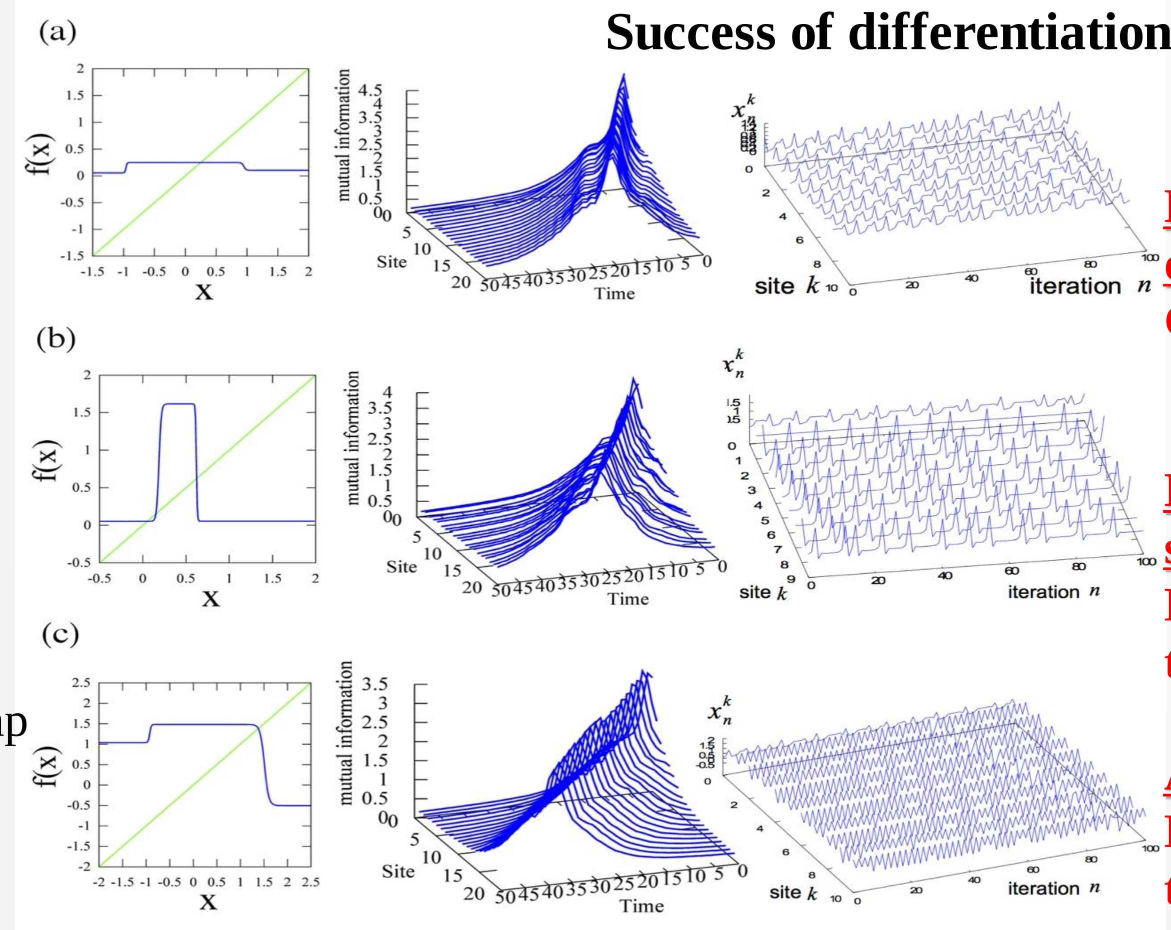
***Constraint*: maximum transmission of information of external signal**

Three types of dynamical systems were selected：
Success of differentiation of spiking neurons and glial cells

Identity map (via strong coupling)

Excitable map (via weak coupling)

Oscillatory map (whole ranges of coupling strength)



Passive depolarization: Glial cells-type

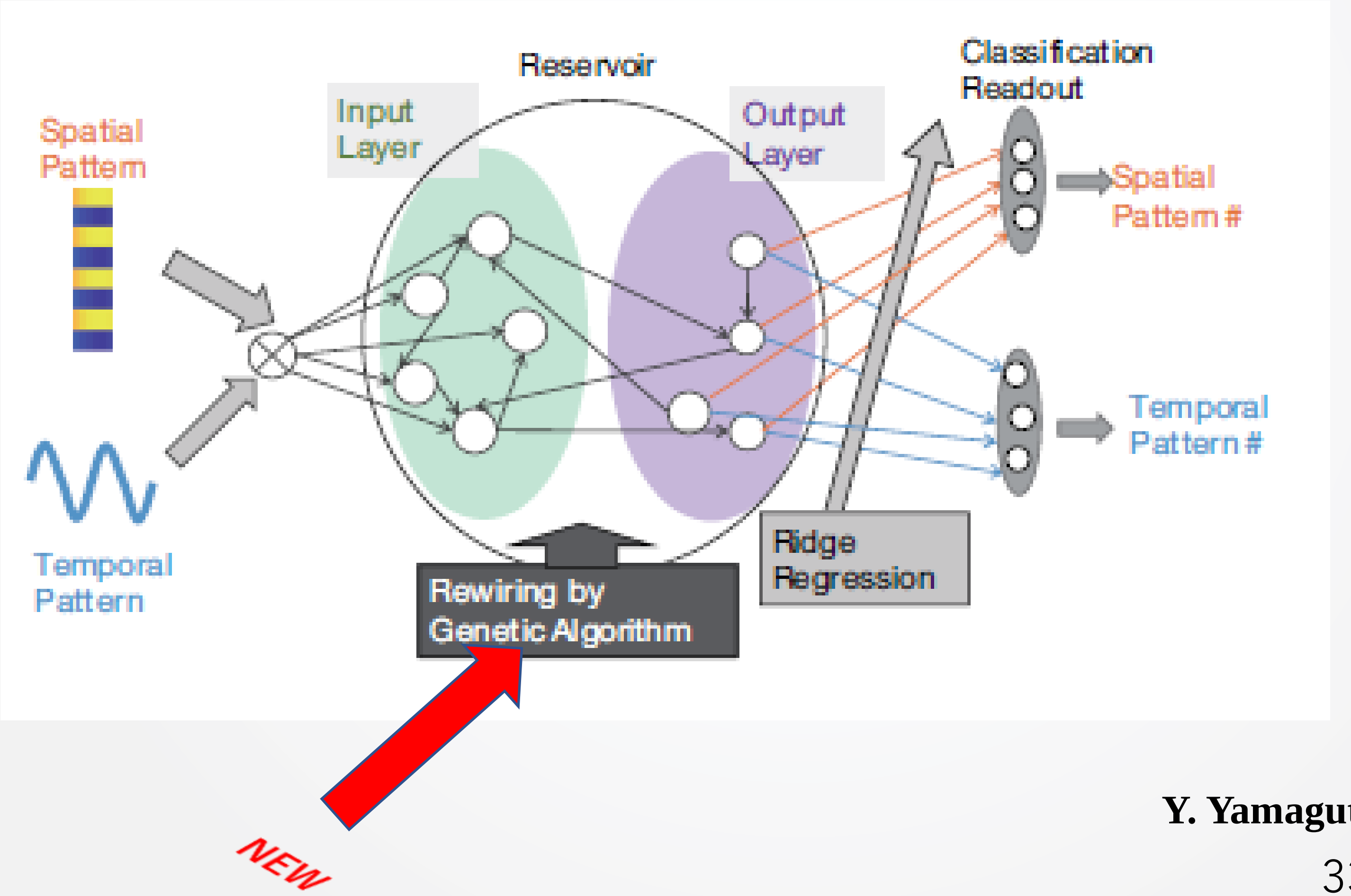
Neuronal spiking: Neuronal cells type

Active variations: Neuronal or Glial cell type

3. Mathematical modeling for

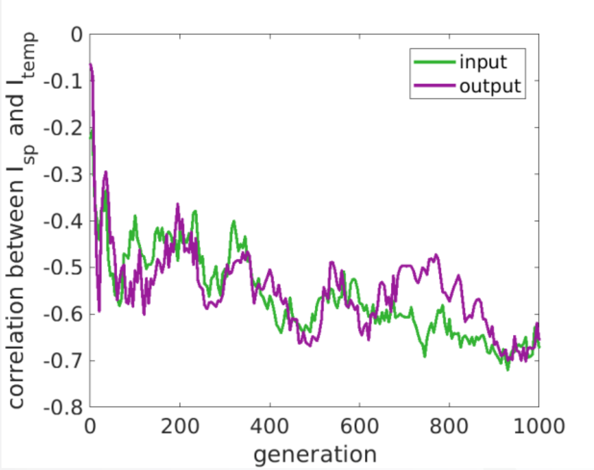
***Evolutionary Reservoir Computers (ERC)*:**

Success of Learning of Separation of Temporal and Spatial Patterns



Y. Yamaguti and I. Tsuda, Chaos 2021

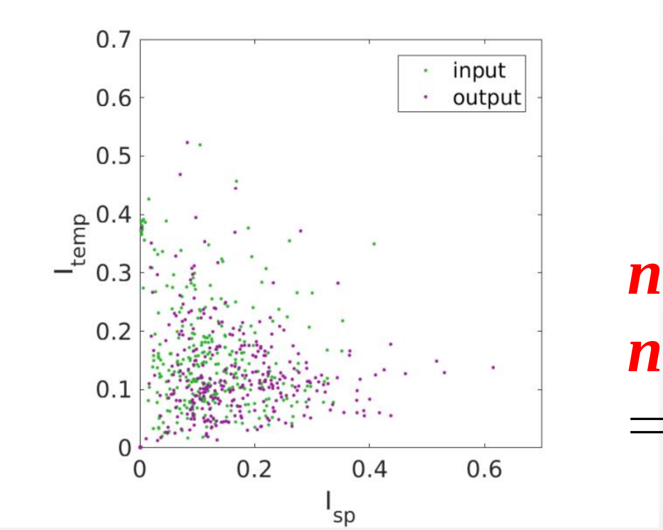
3.1 Investigating the Neural Specificity: Differentiation of Neurons



- Negative correlation appears between I_{sp} and I_{temp}.

- Emergence of 'spatial neuron' and 'temporal neuron'

Initial network



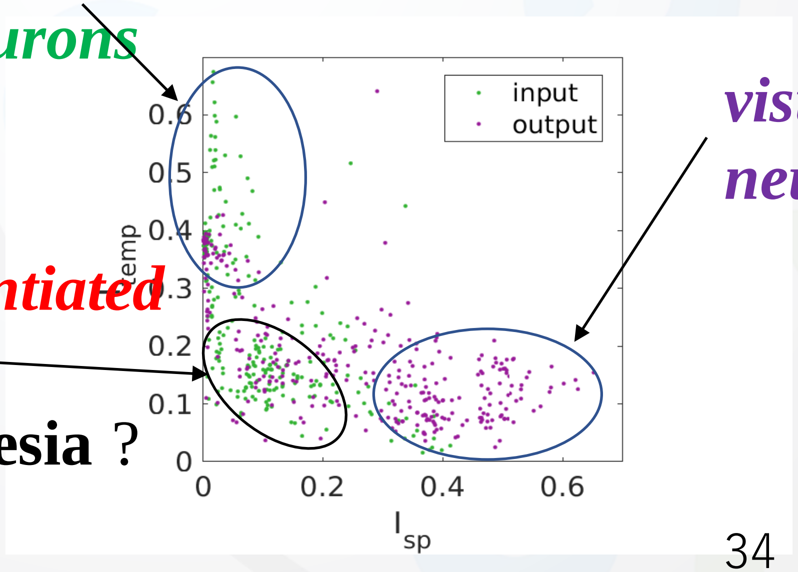
After evolution

auditory neurons

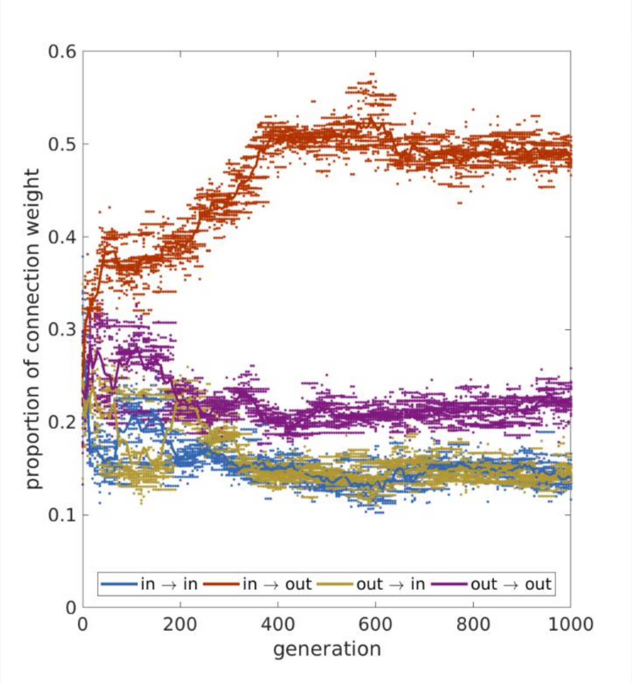
visual neurons

nondifferentiated neurons

⇒ synesthesia ?



3.2 Changes of the Network Architecture

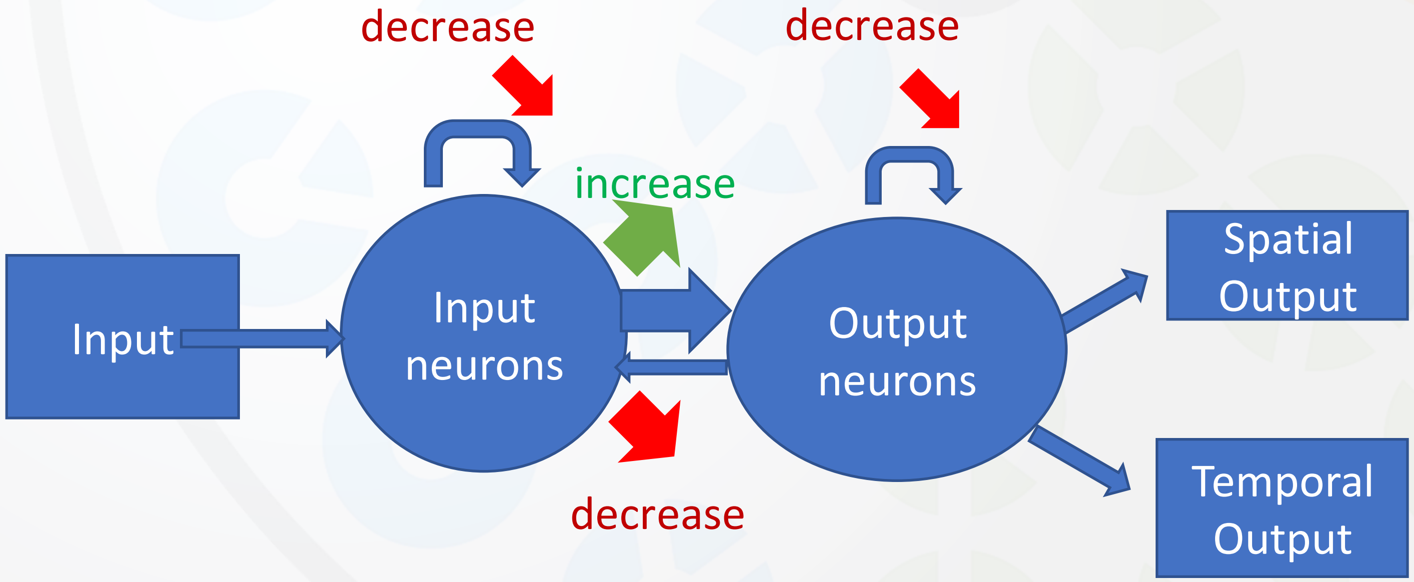


Random networks evolved to *feedforward networks with weaker recurrent networks*

⇒ Consistent with small-world network by Kawai et al

The feedforward connections are more strengthened than the feedback connections: about 5:1-3:2, depending on the evolution

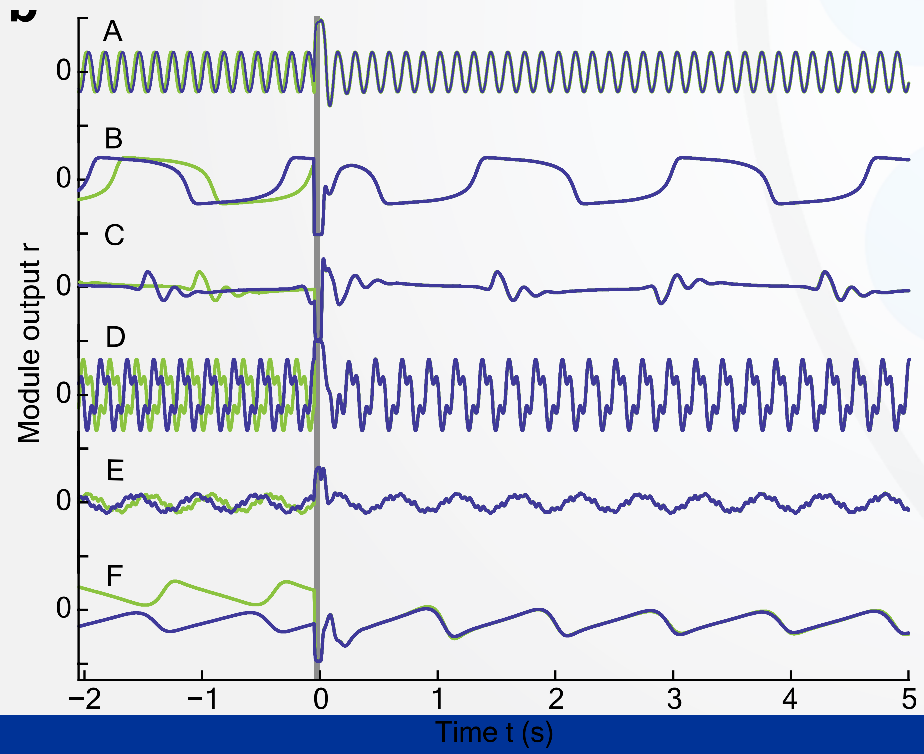
cf) Local Network in
Rats primary visual cortex ~10:1,
Human frontal and temporal cortices ~ 5:1-7:1
(Seeman et al 2018)

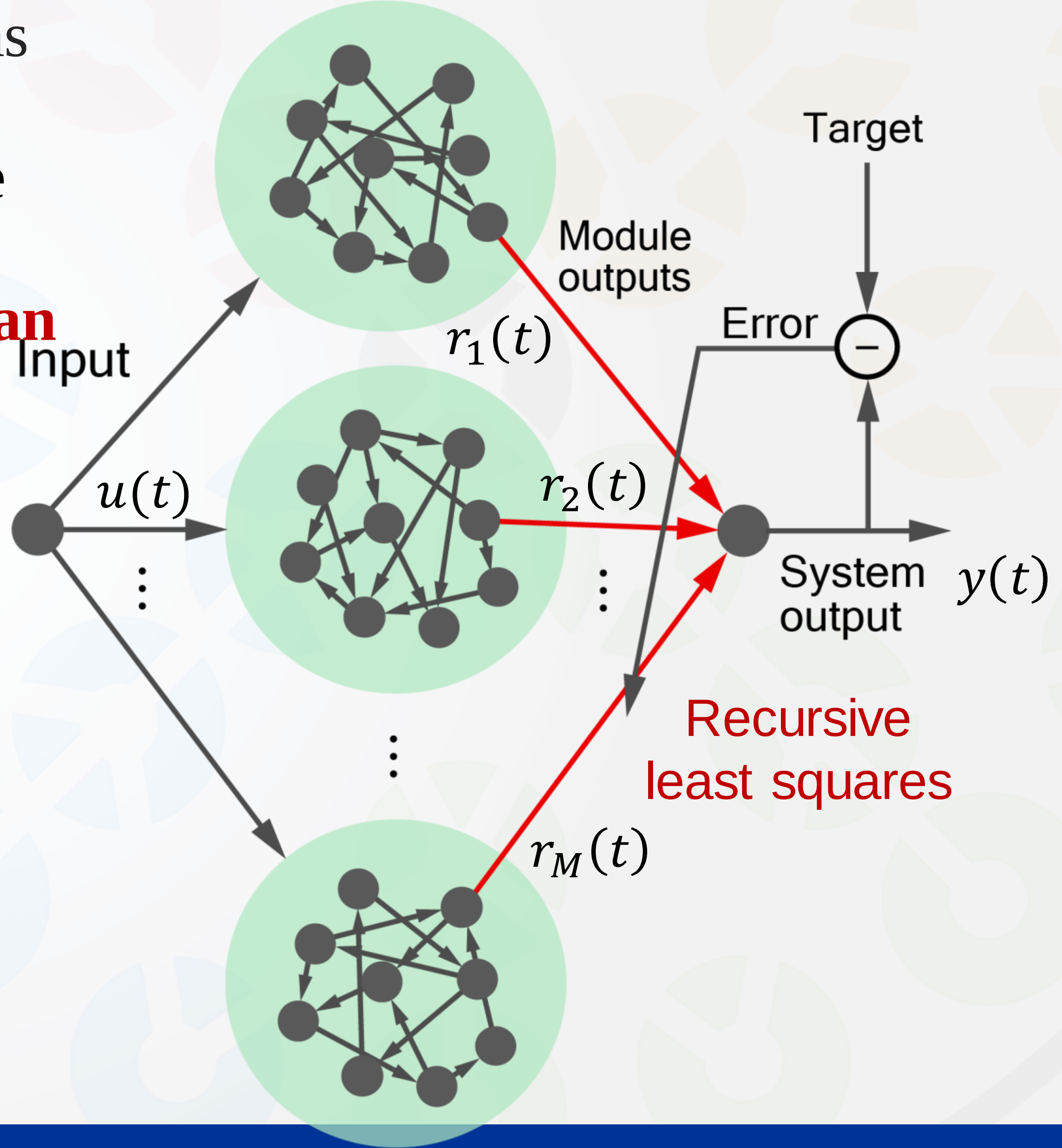


4. reBASICS: reservoir of basal dynamics

[Kawai, IT et al., *ICANN*, 2022; Kawai, IT et al. Neur. Netw. 2022]

- Several small random neural networks are used as modules and connected in parallel.
- Each module spontaneously produces stable time series with diverse phases and frequencies.
- **They are functionally differentiated to create an orthogonal basis.**

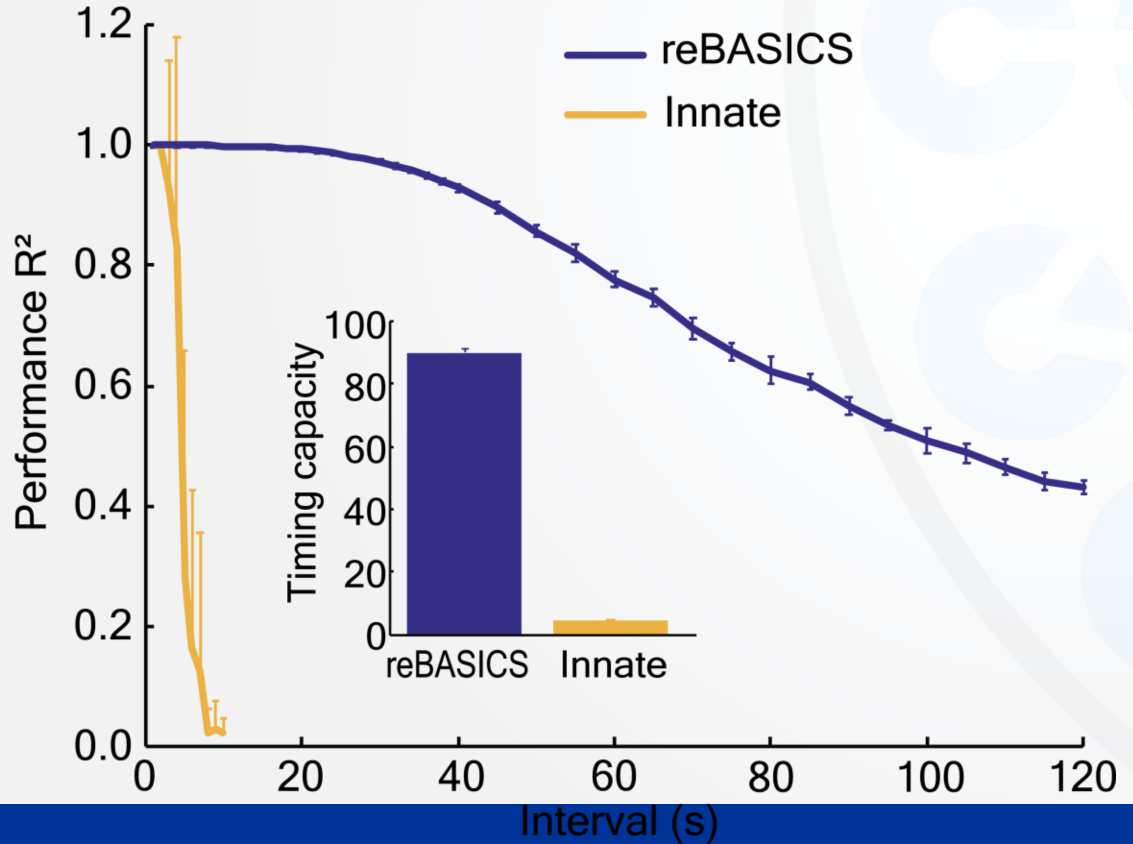


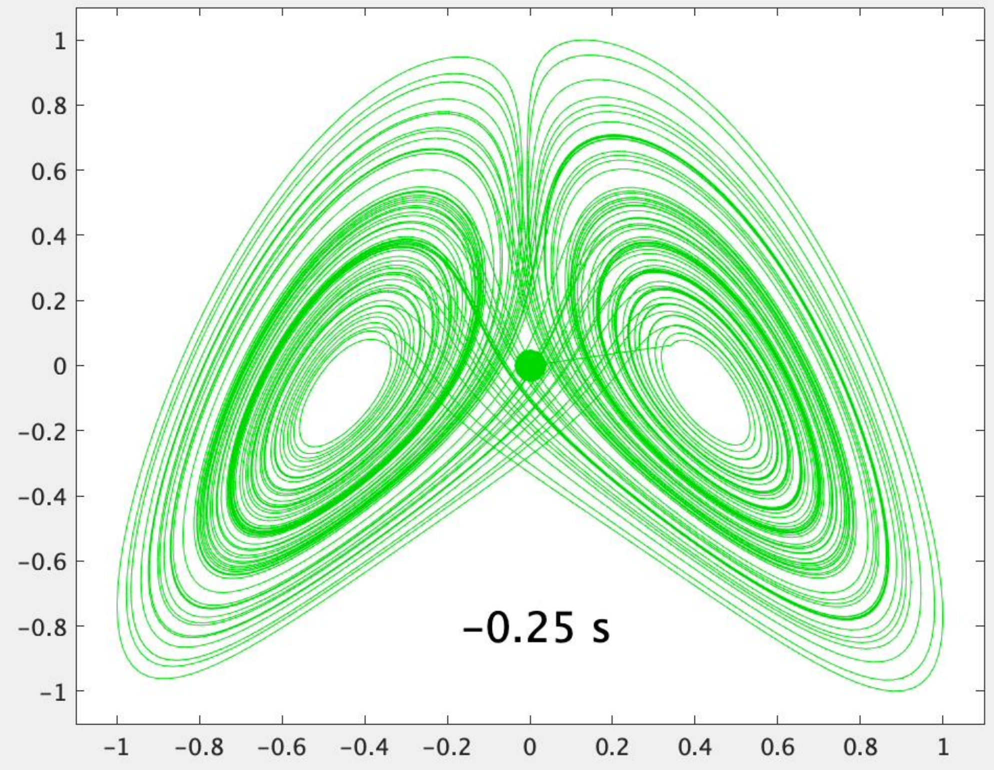


Very large timing capacity

[Kawai et al., *ICANN*, 2022; *Neural Networks* 2022]

- reBASICS could learn the timing even for very long intervals of one minute or more.
- The total performance (timing capacity) of reBASICS was more than **twenty times larger** than that of the innate training (existing approach).
- reBASICS could also learn the long-term Lorenz time series.

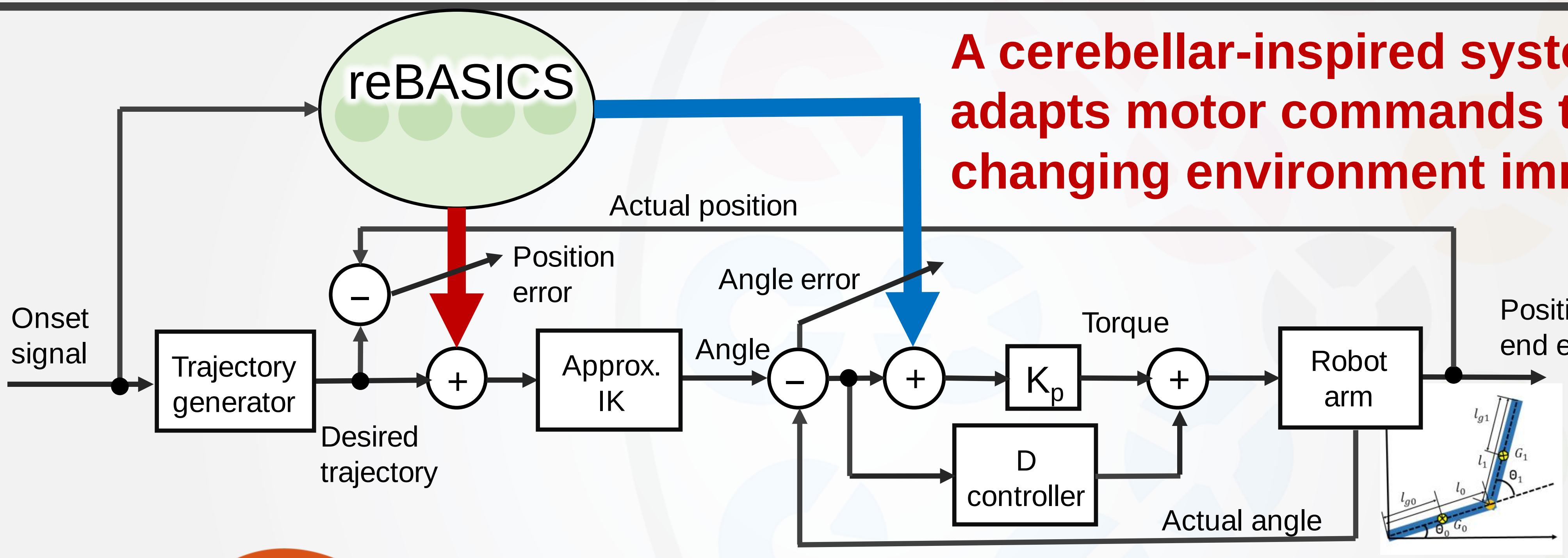


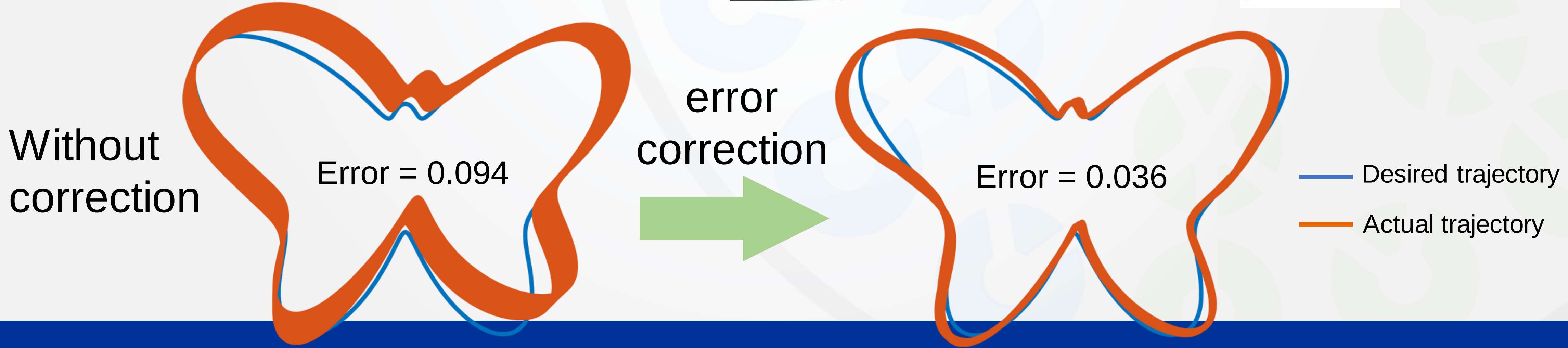


Correction of PD control of a robot using reBASICS

[Kawai & Asada, 2024]

A cerebellar-inspired system that adapts motor commands to a changing environment immediately

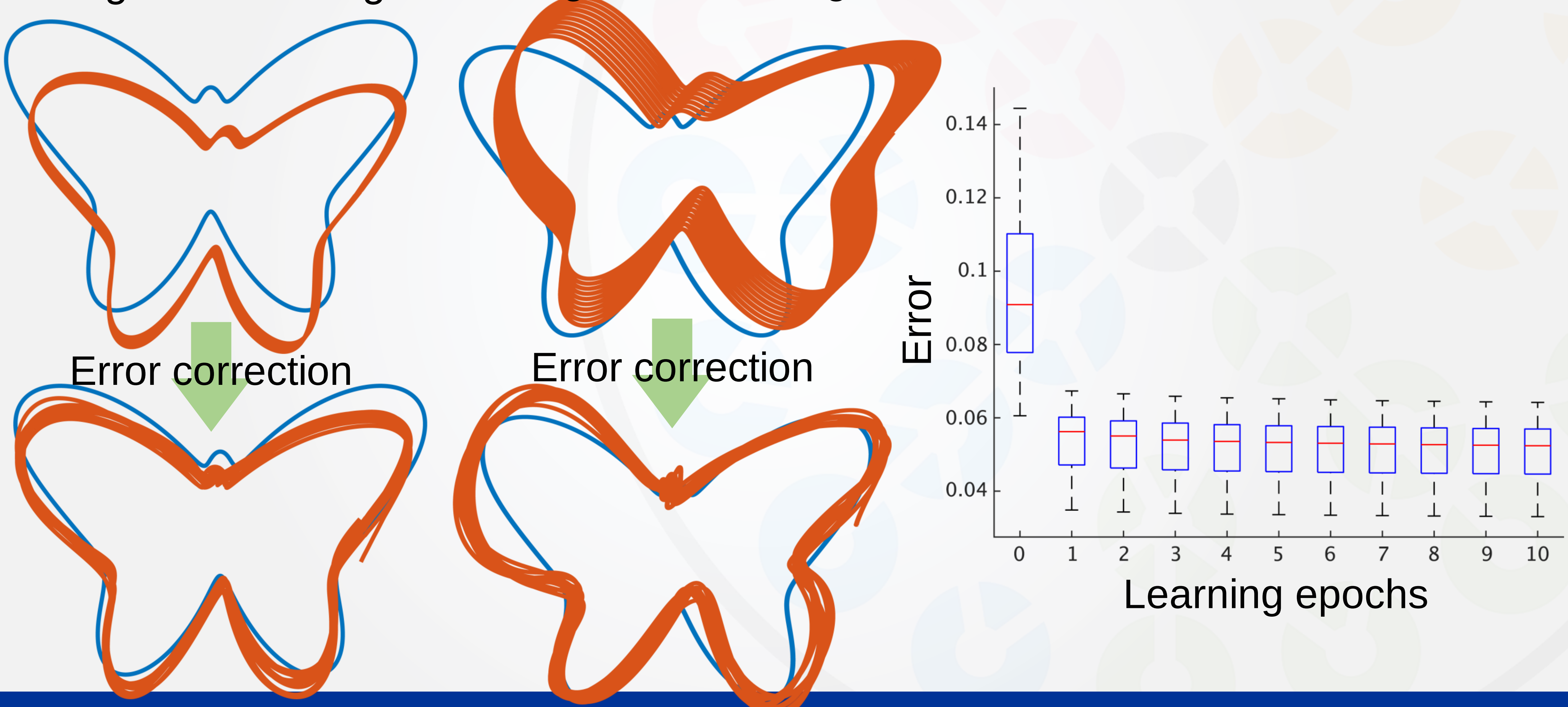




Rapid adaptation to the body's physical changes

[Kawai & Asada, 2024]

Change of links' length | Change of links' weight



What kind of chaos is adequate for information processing in the brain？

⇒Establishment of chaotic information processing

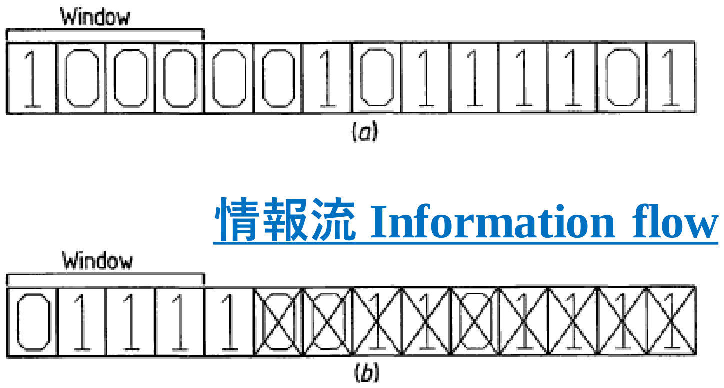
K. Matsumoto and I. Tsuda, 1985, 1987, 1988; R. Shaw 1982;
J. S. Nicolis 1982, 1991; J.S.Nicolis and I. Tsuda 1985, 1989

$$I(t) = I(0)e^{-\alpha t} \rightarrow \frac{dI}{dt} = -\alpha I \rightarrow$$

$$\frac{dI}{I} = -\alpha dt$$

同じ割合の情報損失
The same ratio of information loss

Kullback-Leibler divergence ⇒ Lyapunov exponent
⇒Mutual information⇒The definition of **time-dependent mutual information**



情報流 **Information flow**

Figure 1. Computer register: (a) indicates the noise-free case and (b) the noisy case. The window corresponds to the width of the observation.

BZ map

情報流の揺らぎ大
Large fluctuation of Information flow

Logistic map

情報流の揺らぎ小
Small fluctuation of Information flow

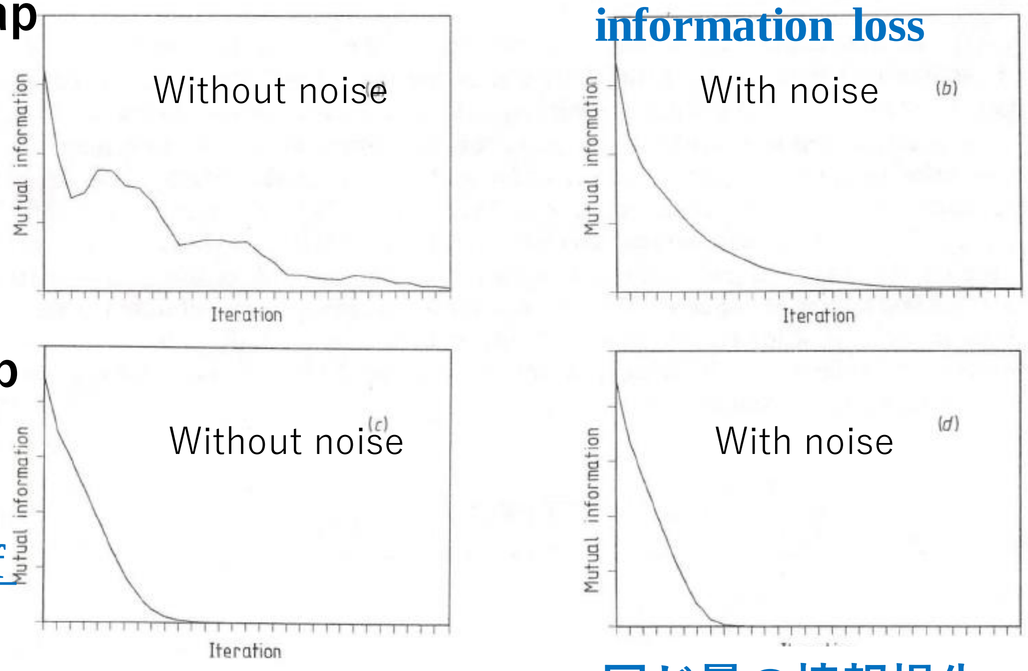


Figure 2. Mutual information: (a) and (... respectively in the BZ map. (c) and (d) sho... in the logistic map.

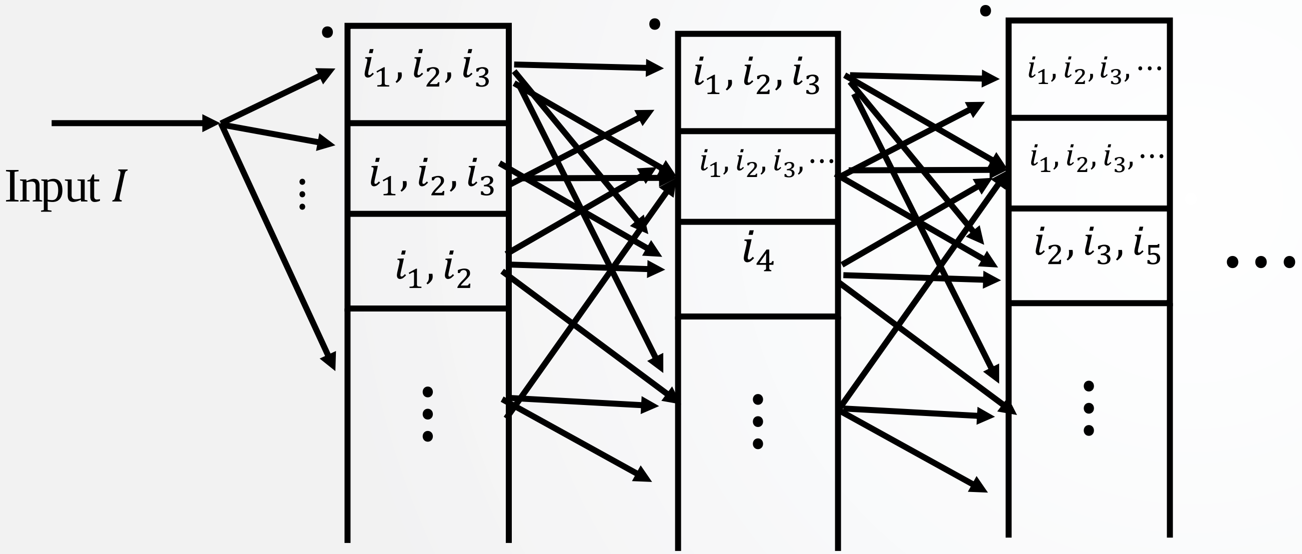
同じ量の情報損失
The same amount of information loss

$$I(t) = I(0) - \beta t \rightarrow \frac{dI}{dt} = -\beta \rightarrow dI = -\beta dt$$

The case of large fluctuation of information flow

Mutual information can be decomposed into bit-wise information.
K. Matsumoto and I. T, *J. Phys*. A 1988

Decomposition into bit-wise information
$I = (i_1, i_2, i_3, \cdots)$



情報の重ね合わせが可能
Superposition of information contents

入力情報はネットワークの中に保存される
Input information is dynamically stored in the network

生物学的カオスはほぼこの性質を満たす
Biological chaos satisfy this property

Shannon's channel シャノンの通信路

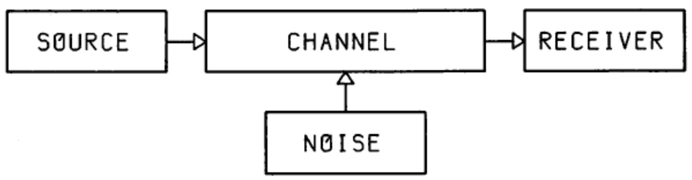


Fig. 1. A communication system considered in information theory.

相互情報量をビットごとの情報量に展開できる
Mutual information can be decomposed into bit-wise information quantities

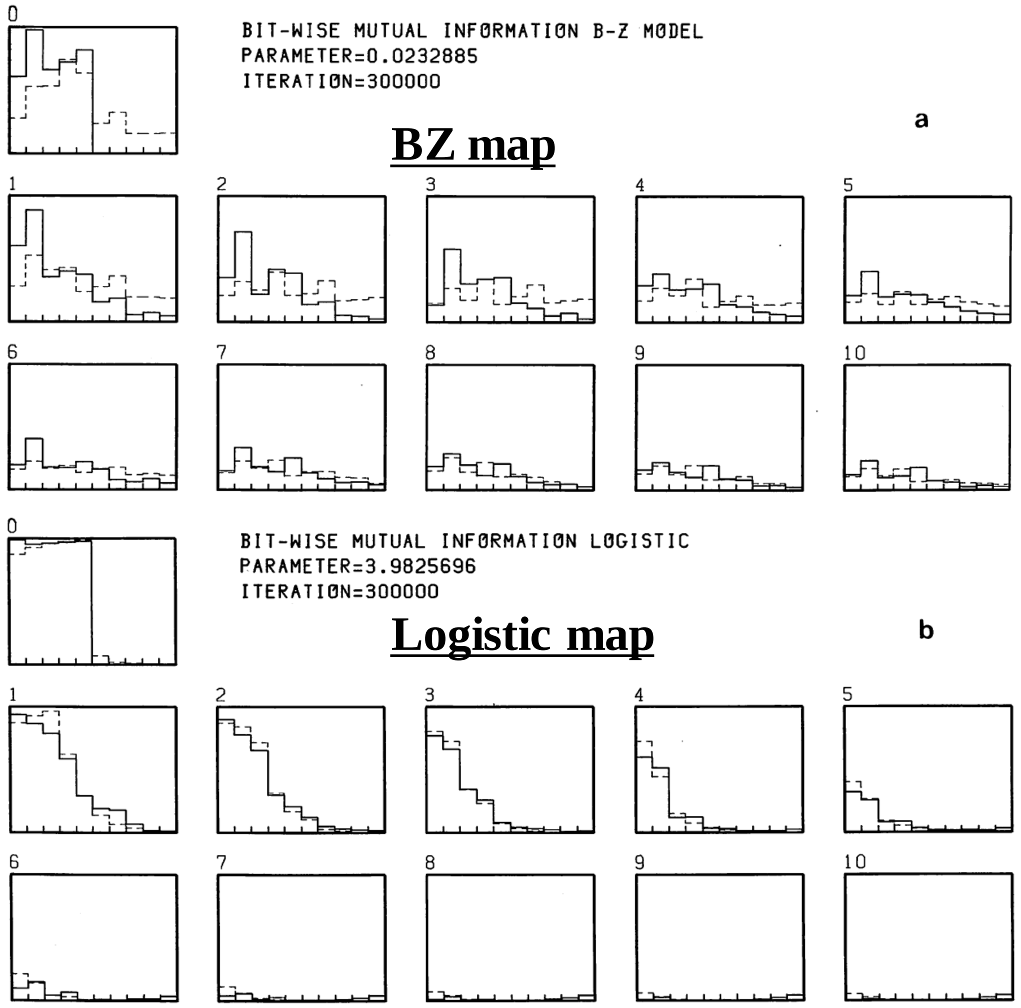


Fig. 5. Bit-wise mutual information for B–Z map (eq. 3.3) (a) and logistic map (b). The figures at the left shoulder of each box is the time interval between input and output. The abscissa of each box represents the output binary places the rightmost being the highest place. The real lines are calculated using eq. (2.5) and the dotted lines eq. (2.6).

結合カオス力学系（⇒カオスネットワーク）を情報の通信路とみなす（カオスは内的ノイズを有する）
Coupled chaotic dynamical systems (chaotic networks) are viewed as Information channel (chaos possesses internal noise)

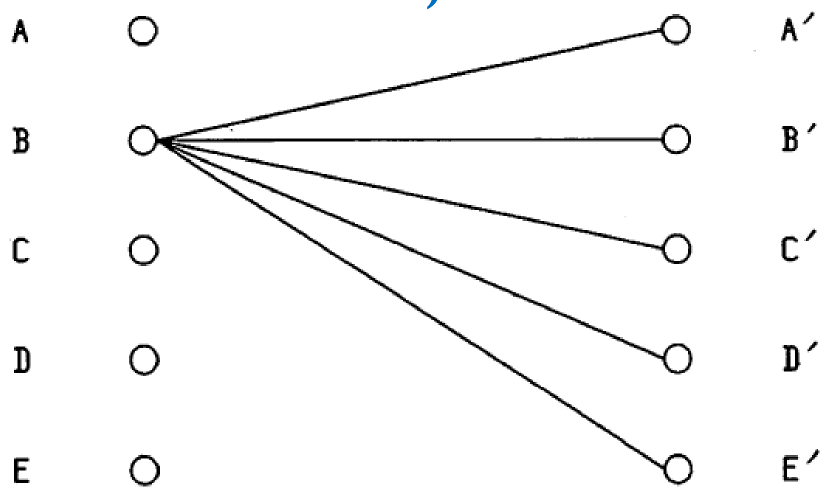


Fig. 2. Relation between input and output in a noisy channel. Input B does not necessarily leads to output B'.

Philosophy is necessary

$$\mathbb{Q} = \{ERC, reBASICS\} \underset{\looparrowright}{\overset{\hookleftarrow}{}} Env.$$

$\hookleftarrow$ Constraints

$\looparrowright$ Rapid adaptation after functional differentiation

We developed a new type of reservoir computers, which rapidly adapt to given environments and solve tasks, by realizing functional differentiation according to given constraints.

$$\mathcal{H}_1(\mathbb{Q}_1, \mathbb{Q}_2, \cdots, \mathbb{Q}_n) \underset{\looparrowright}{\overset{\hookleftarrow}{}} \mathcal{H}_2(\mathbb{Q}_1, \mathbb{Q}_2, \cdots, \mathbb{Q}_n)$$

What is emergent information in a society consisting of interactive agents $\mathbb{Q}$s ?

I personally hope, it should be *conscience*.

参考資料

大偏差原理(Large Deviation Principle: Donsker-Varadhan)

$\{X_n\}: i.i.d., E(X_1) = m, V(X_1) = \sigma^2 = v > 0$とする.
$S_n = \sum_{k=1}^{n} X_k$, $a > m$に対して$P(S_n > an)$を求めたい.
大数の法則（$\frac{S_n}{n} \to m, a.s.$）より、$P(S_n > an) \to 0\ (n \to \infty)$
中心極限定理より$P(S_n > a\sqrt{n} + mn) = P(\frac{S_n - mn}{\sqrt{n}} > a) \to \frac{1}{\sqrt{2\pi v}} \int_a^\infty e^{-(\frac{x^2}{2v})} dx$
したがって、$P(S_n > an) \to \frac{1}{\sqrt{2\pi v}} \int_{\sqrt{n}(a-m)}^\infty e^{-(\frac{x^2}{2v})} dx$
これが 0 に収束していくオーダーを与えるのが大偏差原理

大偏差原理（Cramér’s theorem）

$\{X_n\}: i.i.d., \forall t \in R, E[e^{t|X_1|}] < \infty$とする.
$\varphi(t) = E(e^{tX_1})$とおく．$\forall a > E(X_1) = m$に対して,
$\lim_{n\to\infty} \frac{1}{n} \log P\,(S_n > an) = -I(a)$,　すなわち$P(S_n > an) \sim e^{-nI(a)}$.
ここで$I(a) = \sup_t (at - \log \varphi(t))$, $\lim_{a\to\pm\infty} I(a) = \infty, I(a) \geq 0 = I(m)$であり、
$I(a)$は下半連続(lower semicontinuous:$\forall \varepsilon > 0, \exists \delta > 0, \forall y, |y - x| < \delta, I(y) > I(x) - \varepsilon$ （下から幅εで抑えられる）で、凸関数
つまり、$\boldsymbol{\log \varphi(t)}$の$\boldsymbol{t}$に関するルジャンドル変換である$\boldsymbol{I(a)}$が存在する.

Donsker-Varadhan表現

定理　For stochastic variables X, if stochastic distribution functions $p(x)$, and $q(x)$ are defined, the following equality holds.

$$D_{KL}(q||p) = \sup_{T:X\to R}(E_q[T(x)] - \log E_p\left[e^{T(x)}\right])$$

証明）$q(x)$の近似分布$h(x)$ を次のように考える（経験分布$h(x)$ を母集団分布$p(x)$で測る：
$h(x)$ は$p(x)$が属する空間と同じ空間からとるとする）。

$-T(x)$をエネルギー関数と考えて、ギブス分布（が存在すると仮定）
を近似分布とする。 $h(x) = \frac{e^{T(x)}p(x)}{\int p(x)e^{T(x)}dx} = \frac{e^{T(x)}}{E_p[e^{T(x)}]}p(x)$

$h(x) = \frac{e^{T(x)}}{E_p[e^{T(x)}]}p(x)$より、 $D_{KL}(q||p) = E_q\left[\log\frac{q(x)}{p(x)}\right], D_{KL}(q||h) = E_q\left[\log\frac{q(x)}{h(x)}\right]$

$$D_{KL}(q||p) - D_{KL}(q||h) = E_q[T(x)] - \log E_p\left[e^{T(x)}\right]$$

より

$$D_{KL}(q||h) = D_{KL}(q||p) - (E_q[T(x)] - \log E_p\left[e^{T(x)}\right]) \geq 0$$

ゆえに定理が従う。

定理（Ichiro Tsuda　2023） When q is defined properly for p, the equality $I(a) = D_{KL}(q||p)$ holds.

証明）

$$I(a) = \sup_t\big(at - \log\varphi(t)\big) \qquad \varphi(t) = E(e^{tX_1})$$

$I(a)$は$\varphi(t)$ のt に関するルジャンドル変換：$\big(t, \varphi(t)\big) \to \big(a, -I(a)\big) = \left(\frac{d(\log\varphi(t))}{dt}, -I(a)\right)$

X_1の分布を$p(X_1)$とする。$a = \frac{\varphi\prime(t)}{\varphi(t)} = \frac{E(X_1 e^{tX_1})}{E(e^{tX_1})}$ より、 a は分布$q(X_1) \equiv p(X_1)e^{tX_1}$ によるX_1の平均。

$$I(a) = \sup_t\big(at - \log\varphi(t)\big) = \sup_t\left(\frac{E\big(X_1 e^{tX_1}\big)}{E\big(e^{tX_1}\big)}t - \log E(e^{tX_1})\right)$$

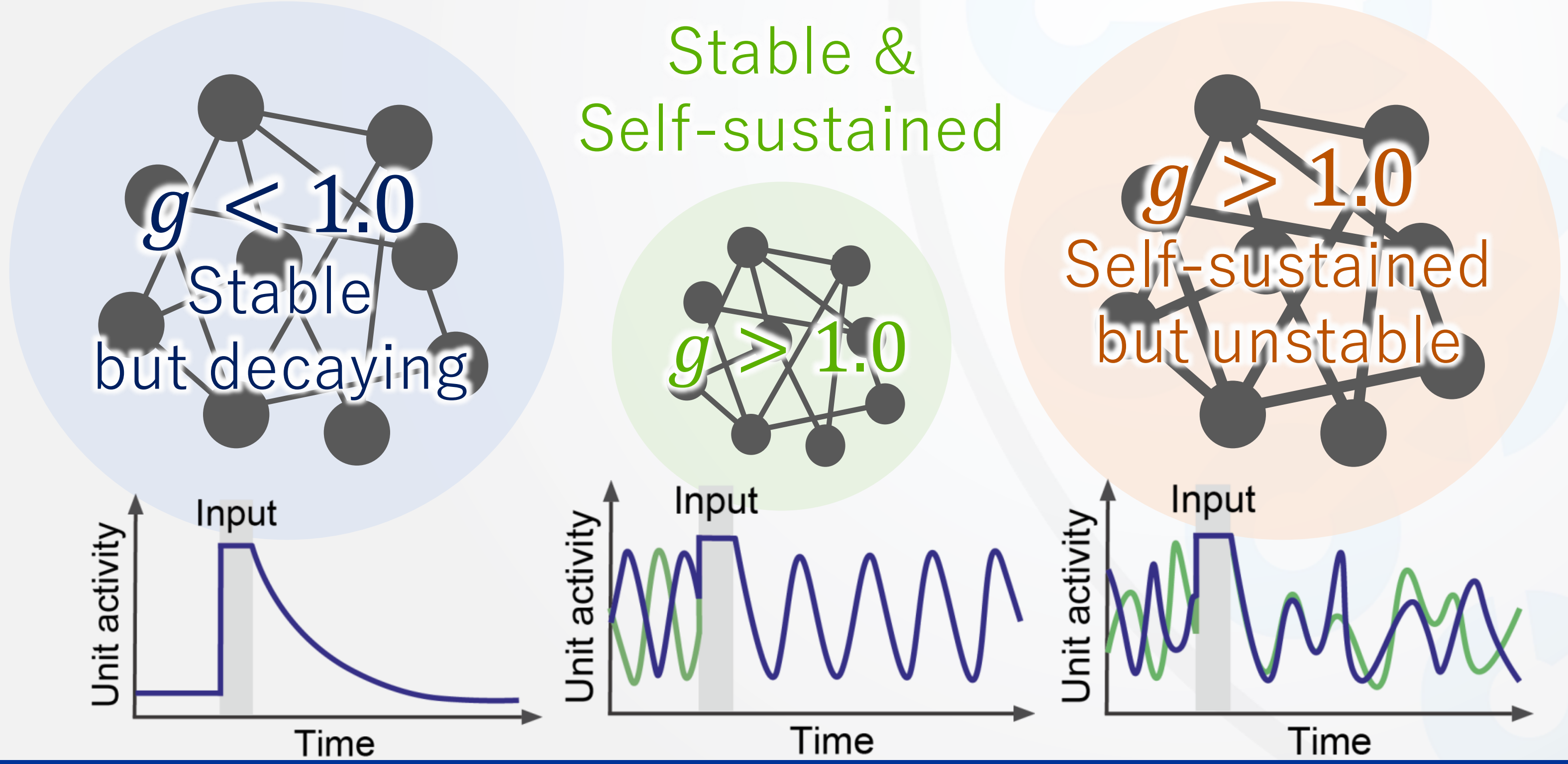
$$= \sup_t\Big(E_q(tX_1) - \log E_p(e^{tX_1})\Big)$$

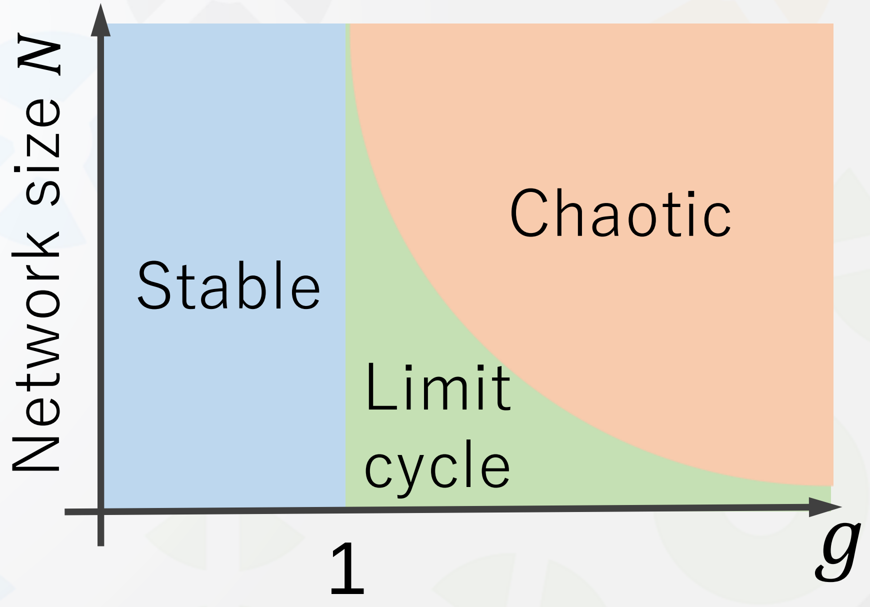
Donsker-Varadhan表現より、 $I(a) = D_{KL}(q||p)$

⇒　This leads to a reasonable interpretation that the estimator of MI with deep neural networks guarantees a sufficient sampling over even largely deviated from an average.

Network size and chaos

- In a network with a finite size N as g increases larger than 1, **asymptotically stable state → limit cycle state → chaotic state.**
  [Sompolinsky et al., *Physical Review Letters*, 1988; Doyon et al., *Acta Biotheoretica*, 1994]
- However, small N = low degrees of freedom & low orthogonality

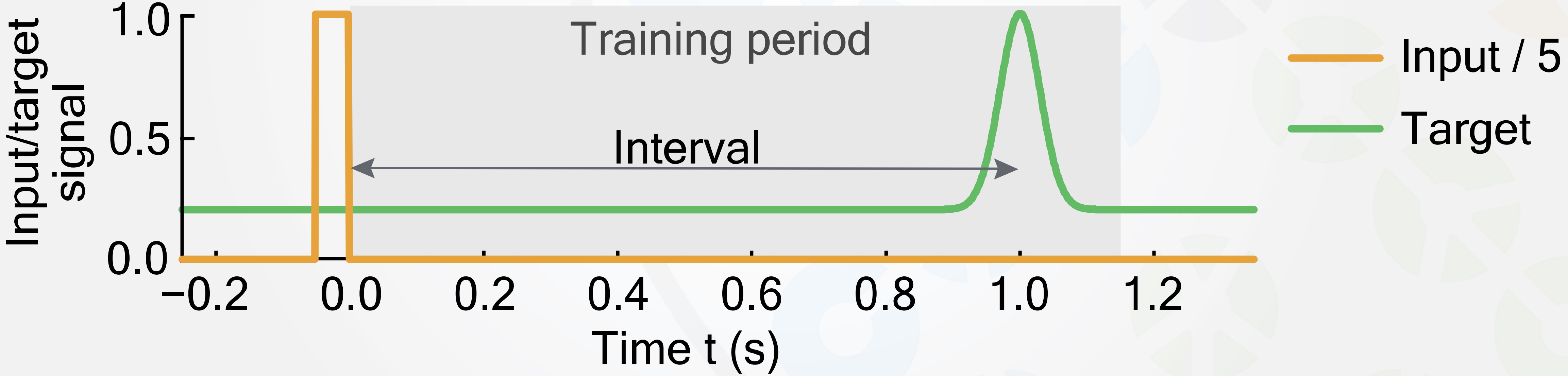




Motor timing learning

[Kawai et al., *ICANN*, 2022; *NN*, 2023]

- An example of the task with an interval of 1 s.
- Performance R^2: the square of the correlation coefficient between target and output



The number of modules M

[Kawai et al., submitted]