

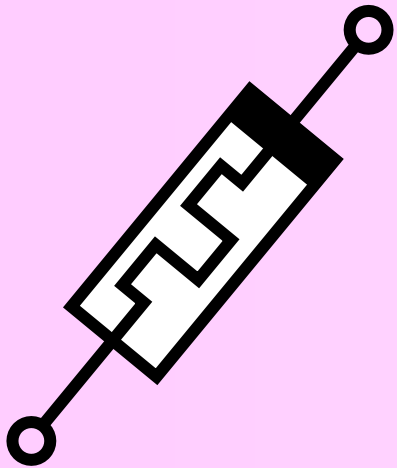


OSAKA UNIVERSITY



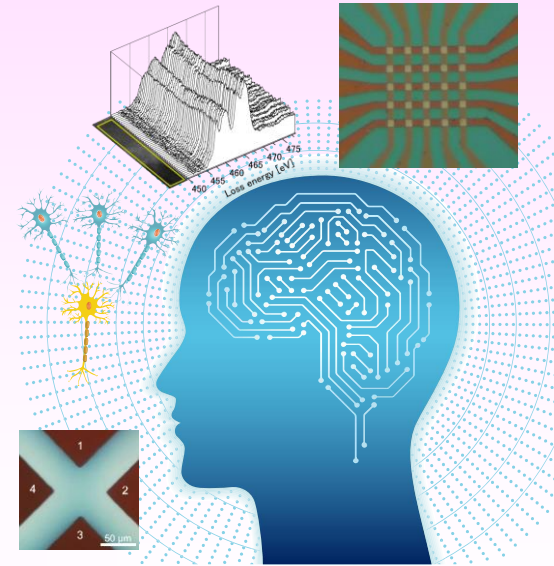
ENGINEERING SCIENCE

メモリスタ：半導体メモリデバイスの新展開



酒井 朗

大阪大学 大学院基礎工学研究科



大阪大学エマージングサイエンスデザインR³センター
ナノ社会人教育プログラム 特別講義
@大阪大学中之島センター10階 ホール3・4

令和6年10月4日（金）

サイバー空間と物理空間を統合するシステムによる人間社会

Society 5.0

Internet of Things (IoT)

connecting all people and things
sharing various knowledge and information
creating new values never seen before

Artificial Intelligence (AI)

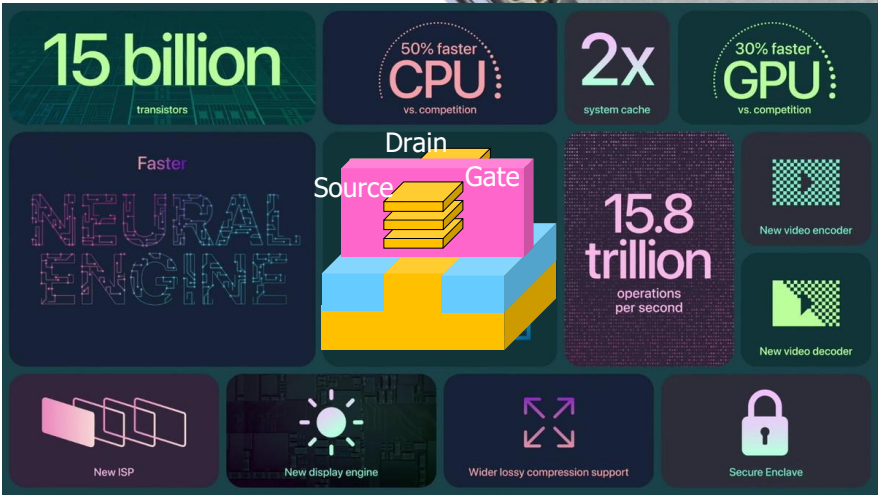
analyzing huge amount of big data
providing necessary information when needed

Semiconductor devices are at the core of these systems

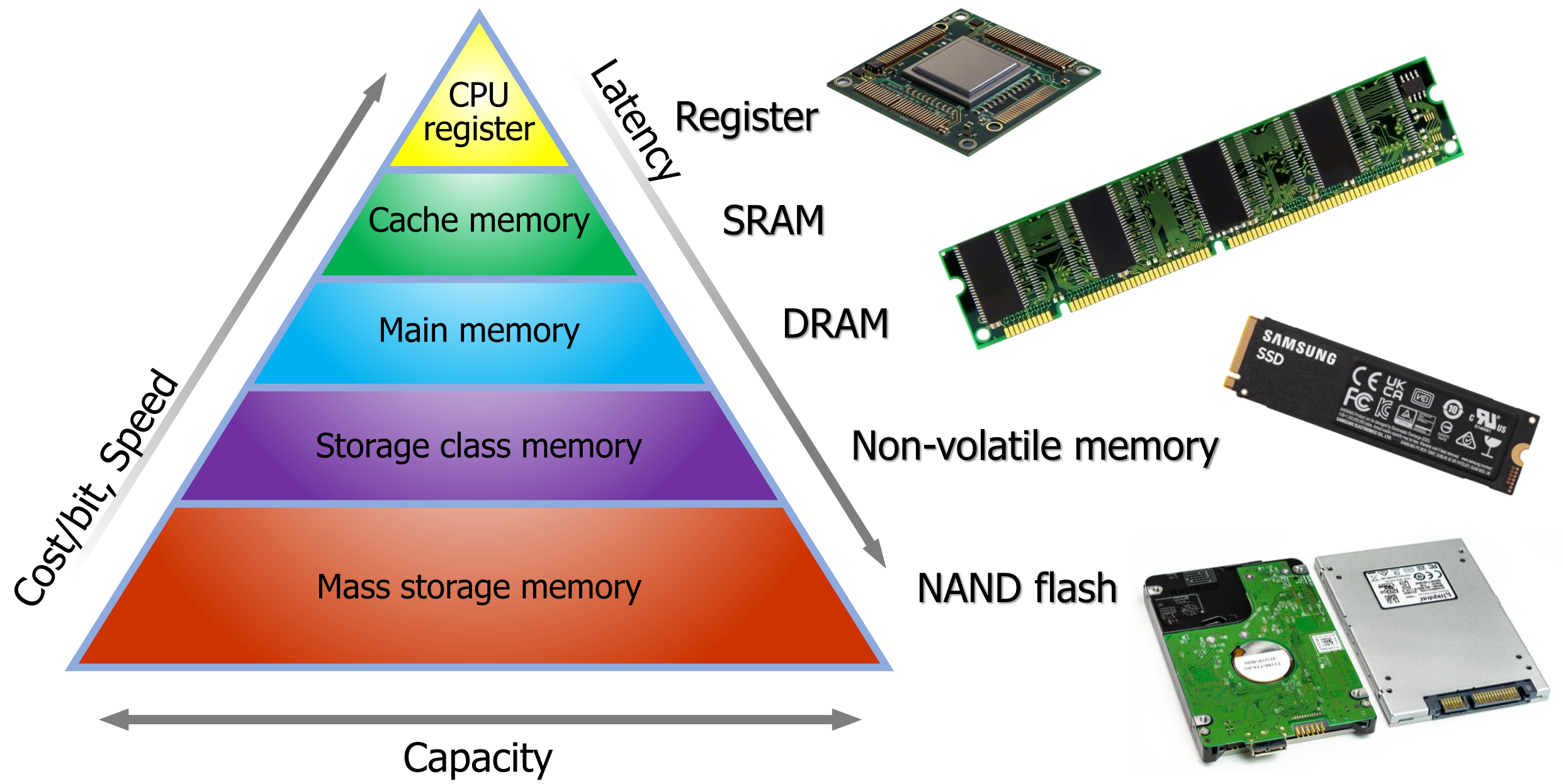
Robots and autonomous vehicles, etc.
overcoming social issues



半導体デバイスはIoT, AI, Roboticsの要である



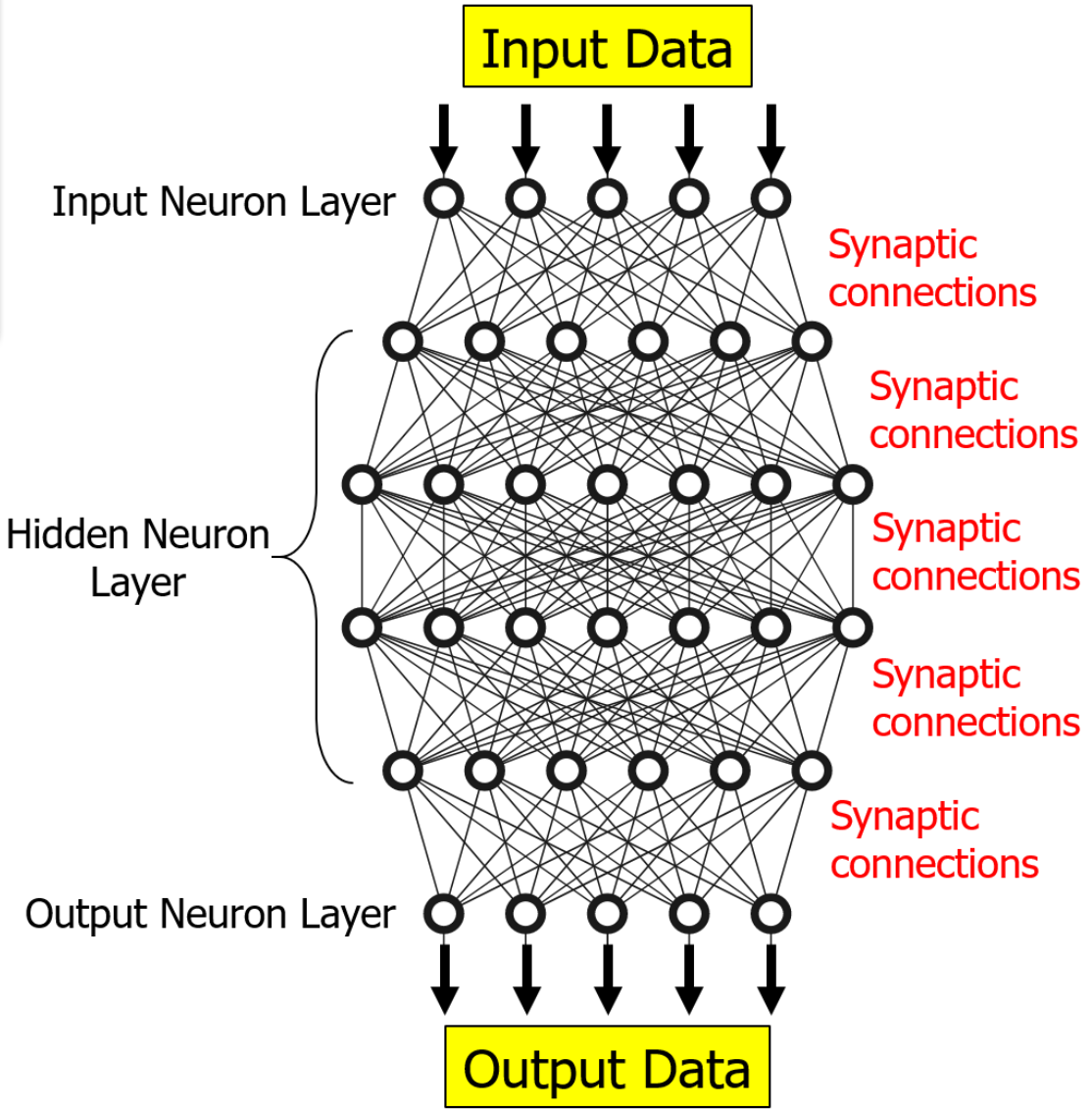
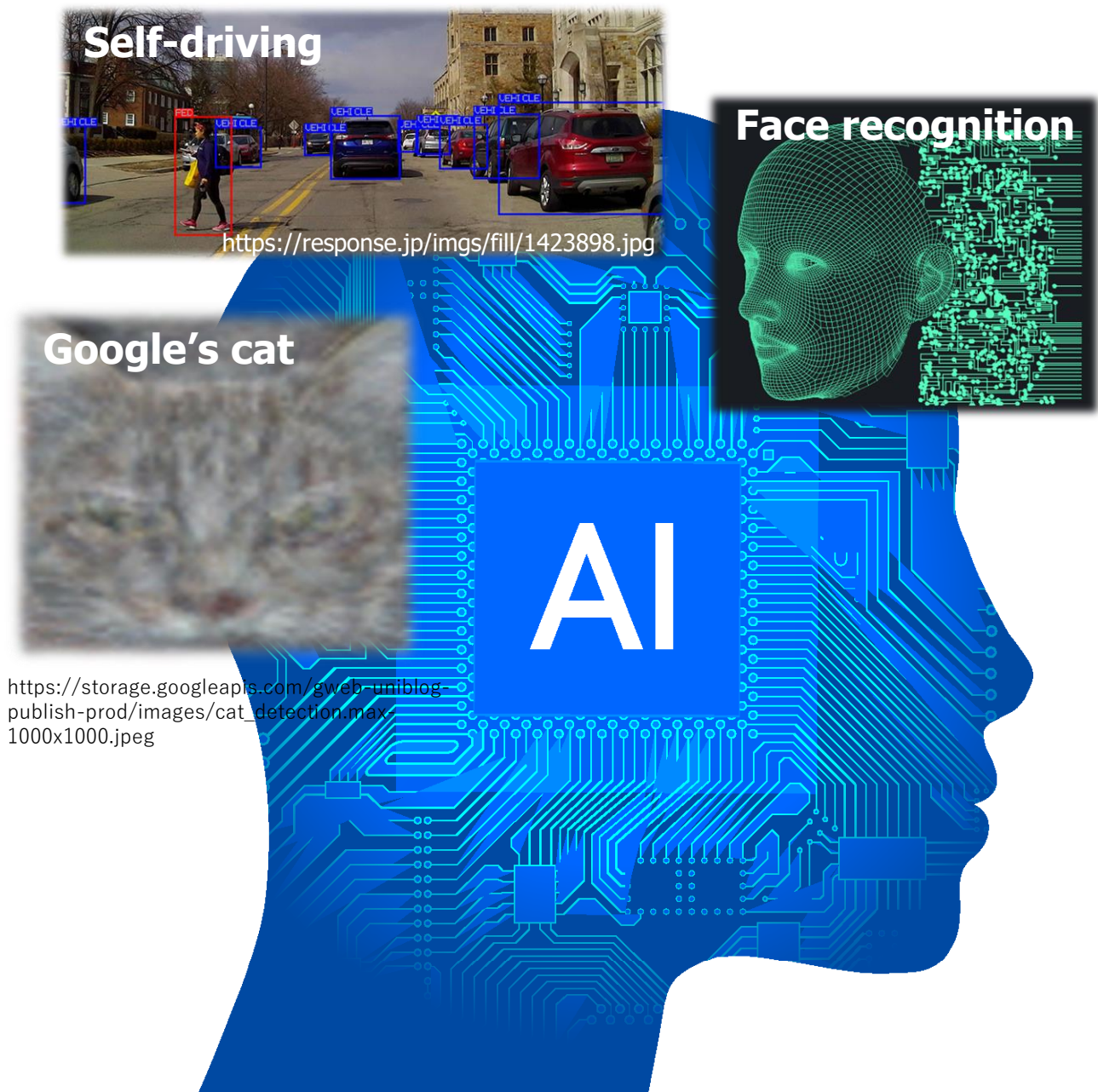
半導体メモリ階層構造



AIとしての人工ニューラルネットワーク (ANN)

ANN: Artificial Neural Network

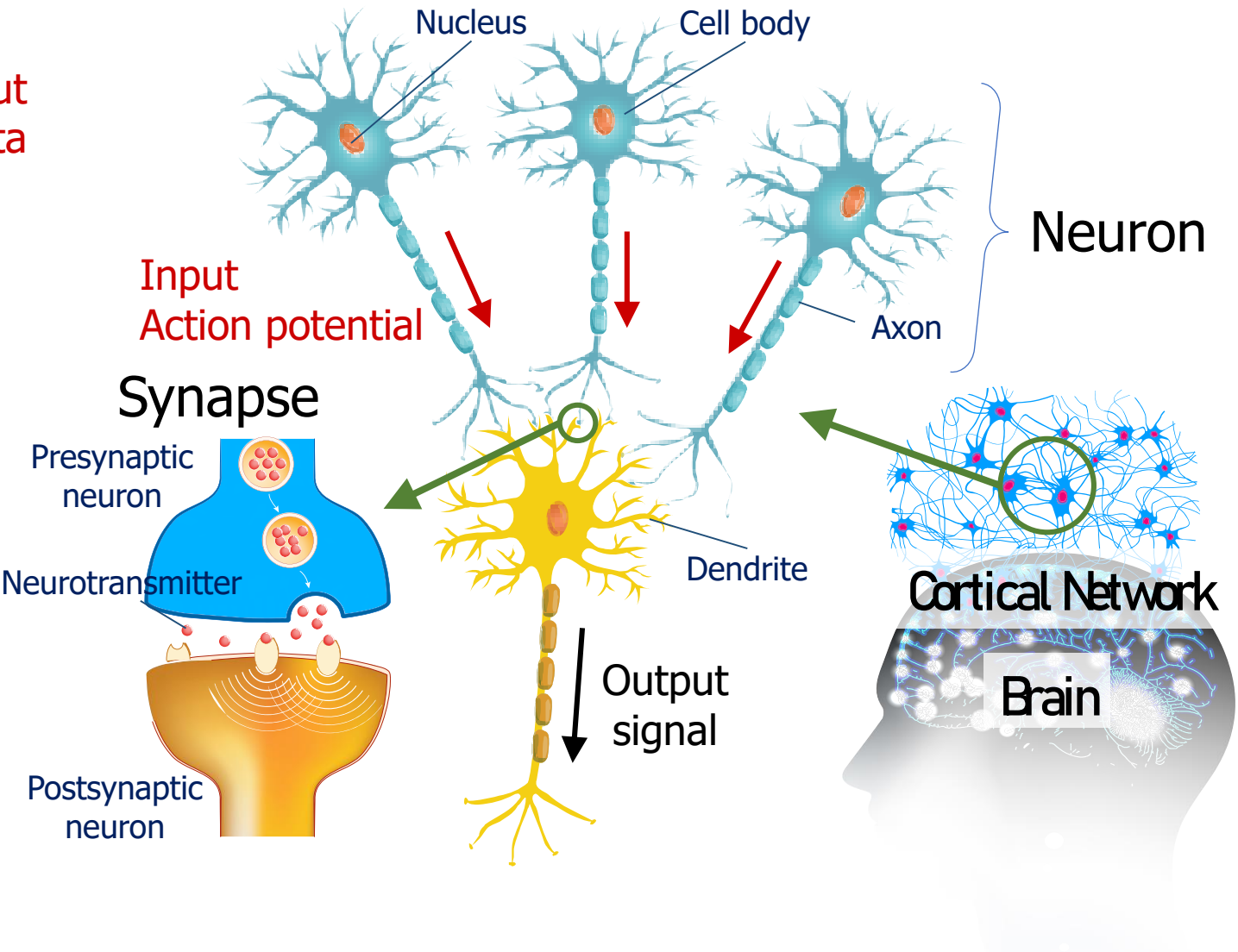
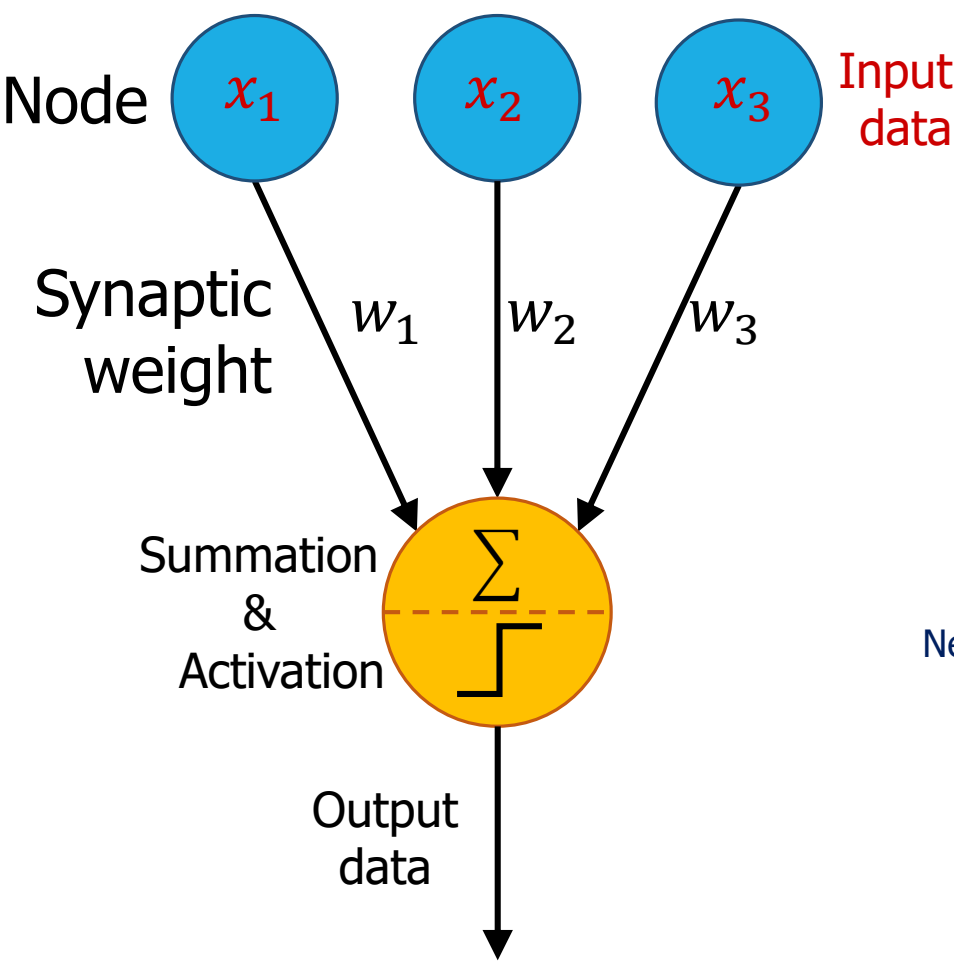
Deep Neural Network



ANNは脳内皮質ネットワークを模倣している

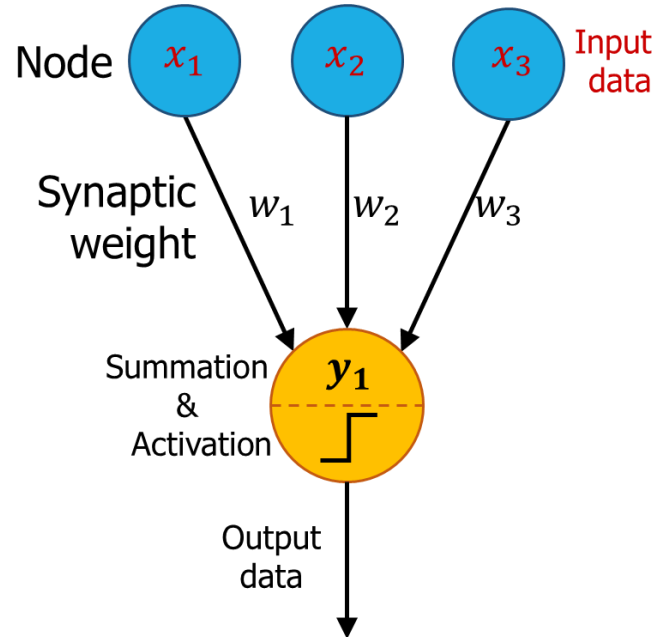
Single layer perceptron

Biological neural network



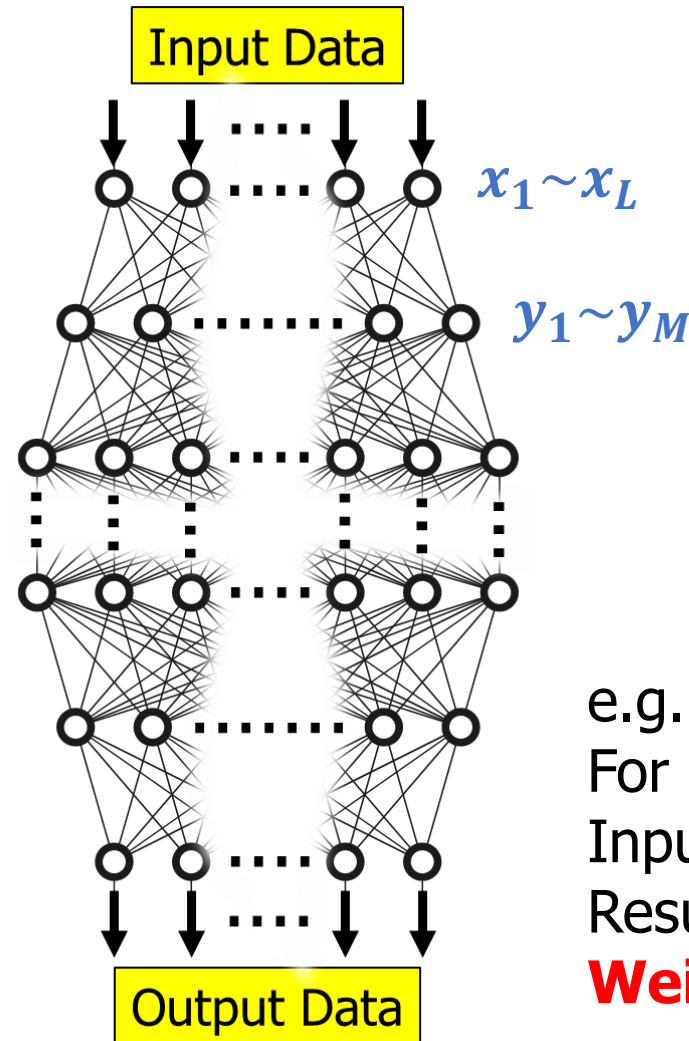
積和演算 (MAC : Multiply–Accumulate operation)

Single layer perceptron



$$y_1 = w_1x_1 + w_2x_2 + w_3x_3$$

Deep neural network



$$[x_1 \ x_2 \ \cdots \ x_L] * \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^M \\ w_2^1 & w_2^2 & \cdots & w_2^M \\ \vdots & \vdots & \ddots & \vdots \\ w_L^1 & w_L^2 & \cdots & w_L^M \end{bmatrix}$$

$$= \begin{bmatrix} y_1 = w_1^1x_1 + w_2^1x_2 \cdots + w_L^1x_L \\ y_2 = w_1^2x_1 + w_2^2x_2 \cdots + w_L^2x_L \\ \vdots \\ y_M = w_1^Mx_1 + w_2^Mx_2 \cdots + w_L^Mx_L \end{bmatrix}$$

e.g.

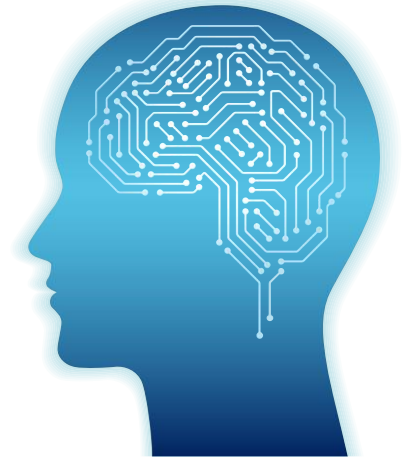
For 1,000 inputs and 1,000 nodes

Input data : a vector of 1000 elements

Results : a vector of 1000 elements

Weight : a matrix of 1,000,000 elements!

フォン・ノイマンアーキテクチャの深層ニューラルネットワーク

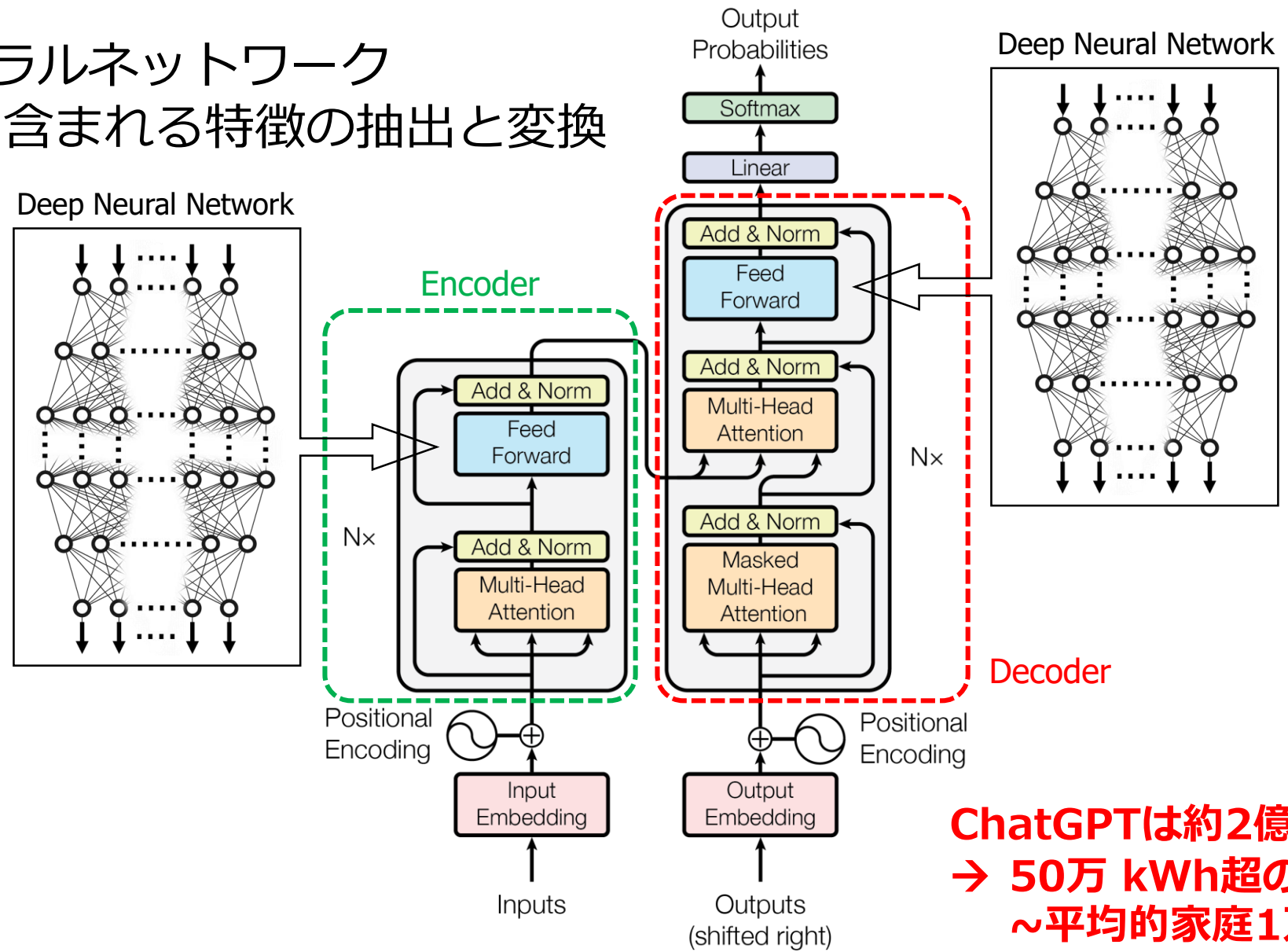
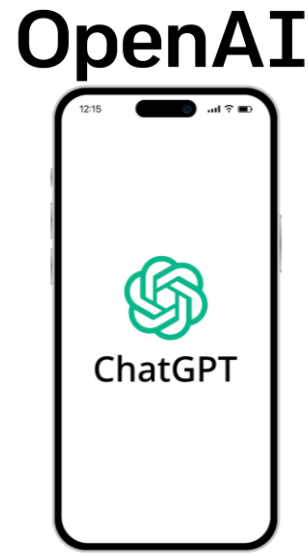


250,000 W (1,202 CPU + 176 GPU) ← Power consumption → 20 W

Multiply–Accumulate (MAC) operation is mostly power and time consuming !

生成AI（Transformer）のモデルアーキテクチャ

深層ニューラルネットワーク
→ データに含まれる特徴の抽出と変換



**ChatGPTは約2億プロンプト/日
→ 50万 kWh超の電力量
~平均的家庭1万7000世帯以上**

フォン・ノイマン型コンピュータ

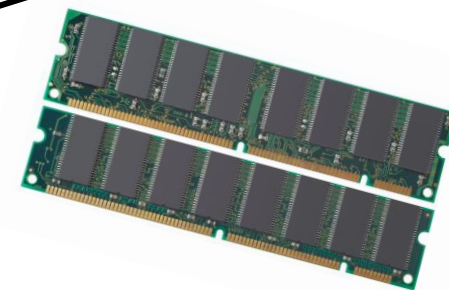


von Neumann

<http://www.lanl.gov/history/atomicbomb/images/NeumannL.GIF>

CPU
(中央演算素子)

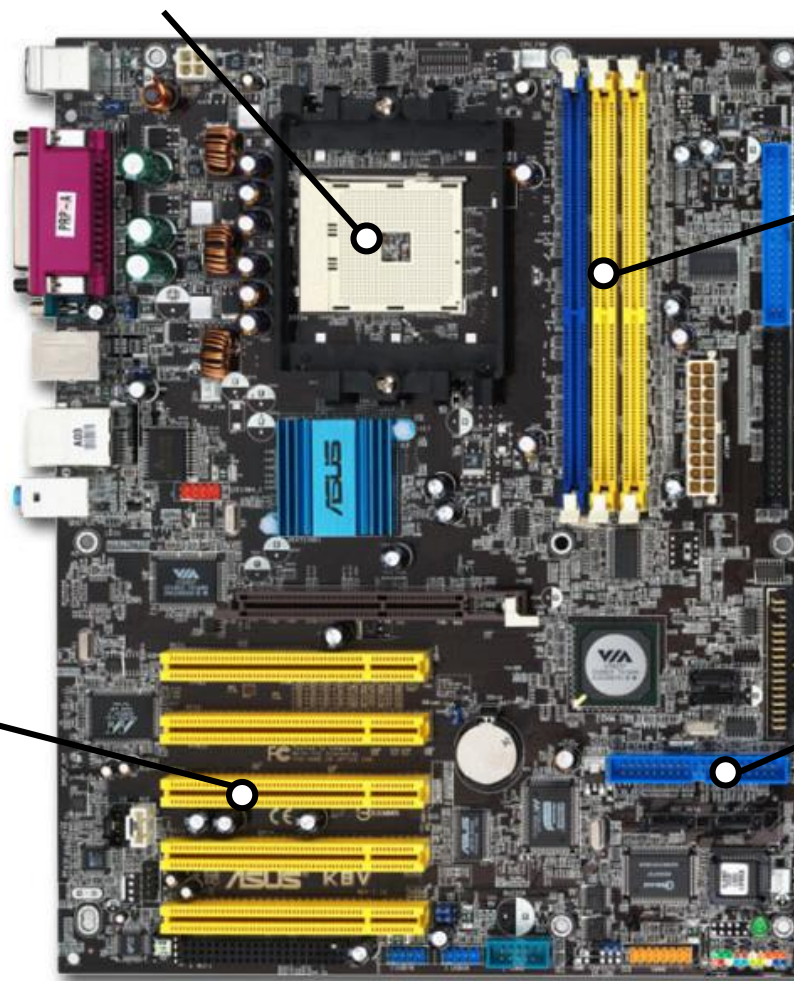
メモリスロット
(メモリデータ入出力)



SATAスロット
(ストレージ入出力)



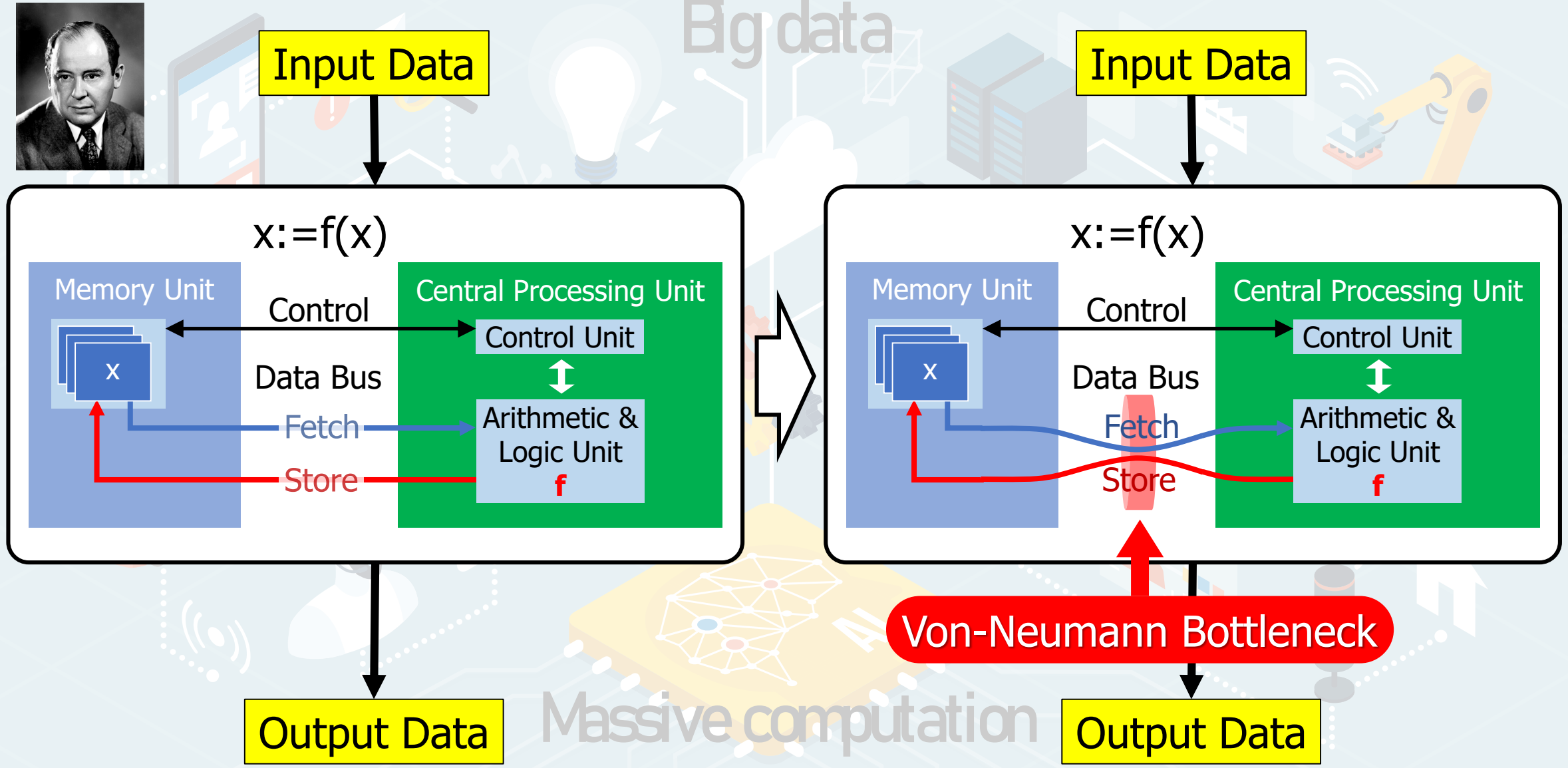
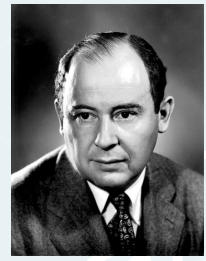
PCIバススロット
(ビデオカード入出力)



フォン・ノイマンアーキテクチャは重大な問題を抱えている

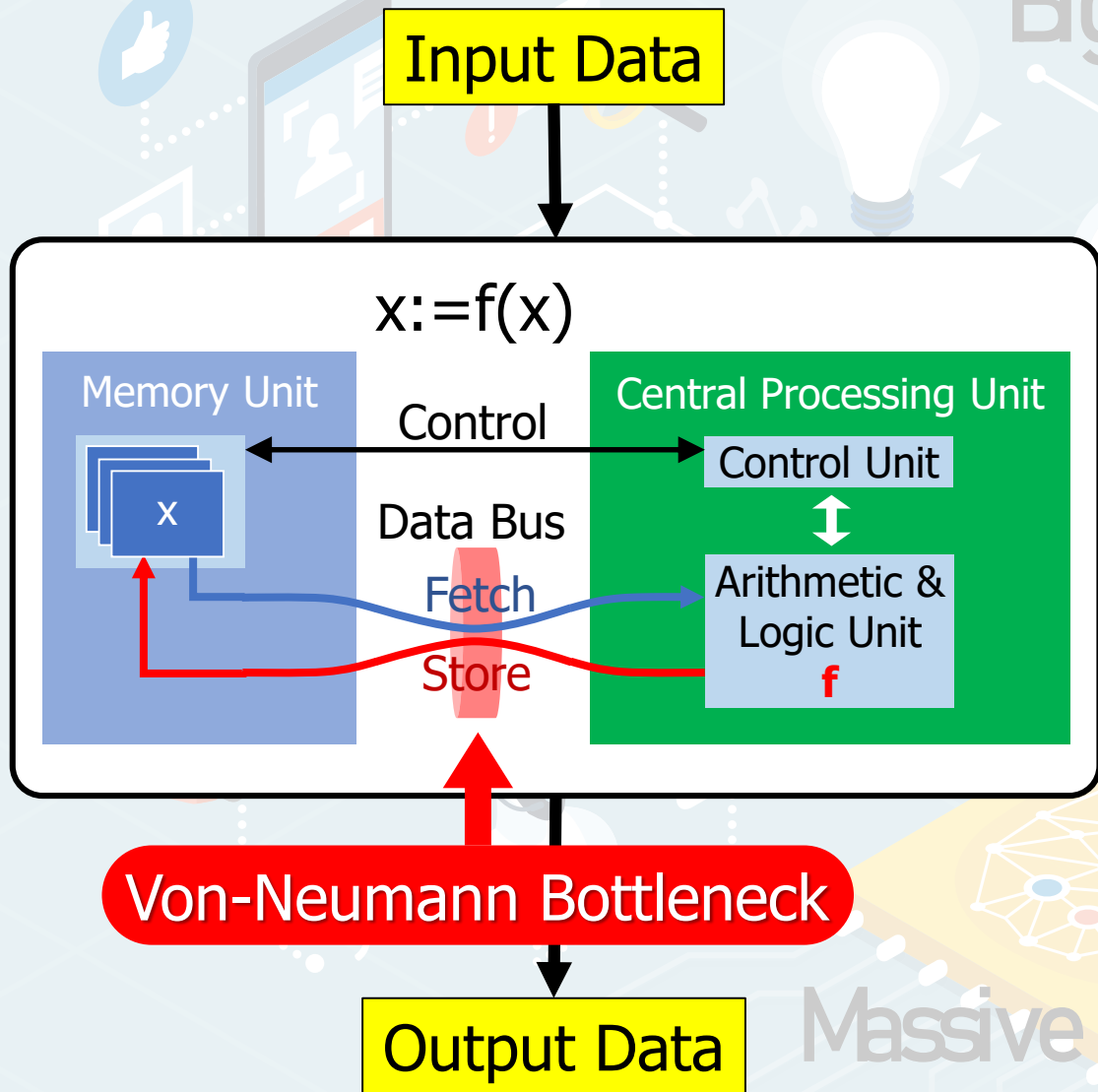
<http://www.lanl.gov/history/atomicbomb/images/NeumannL.GIF>

Von-Neumann computing

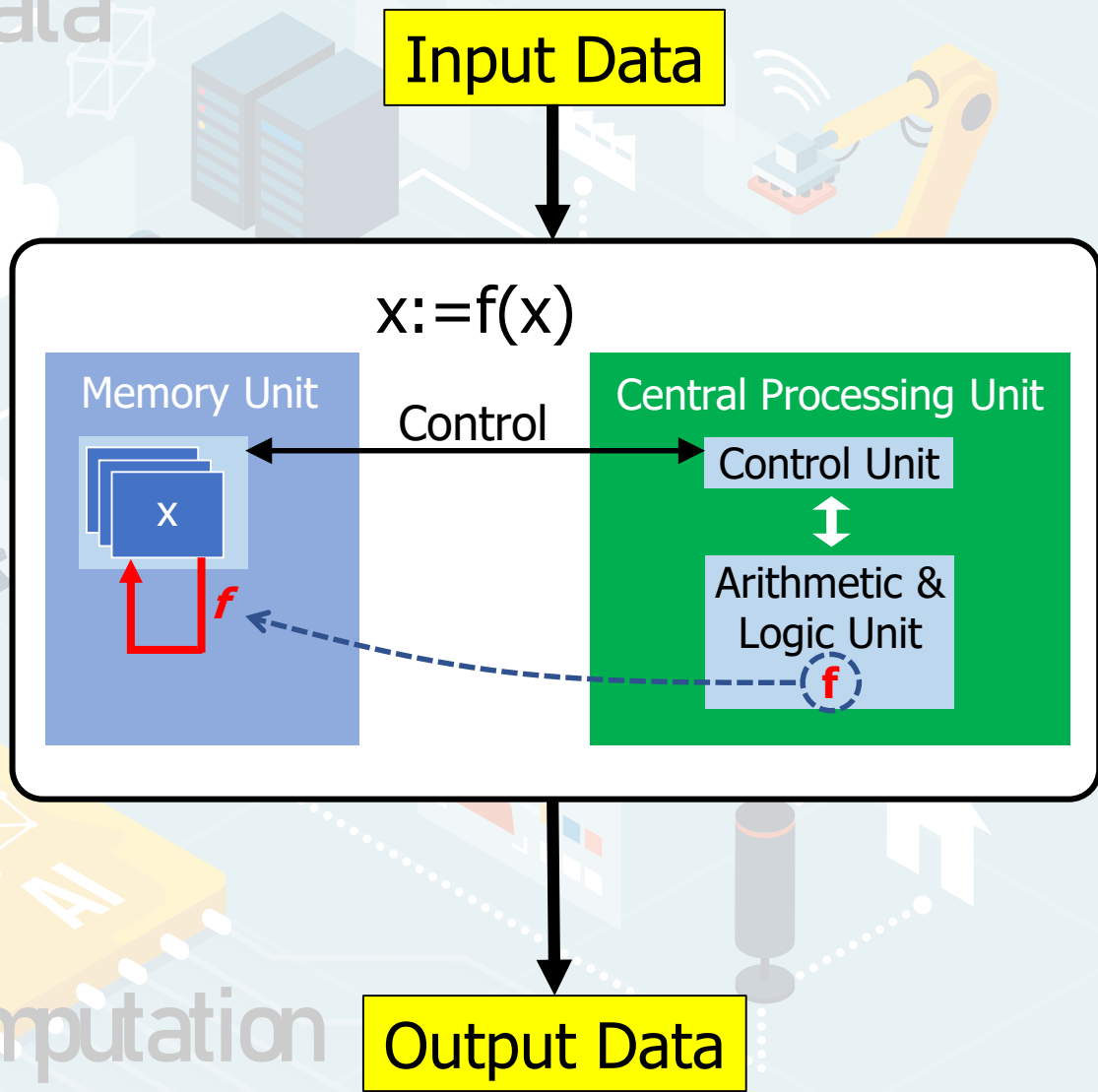


フォン・ノイマン型からイン・メモリコンピューティングへ

Von-Neumann computing



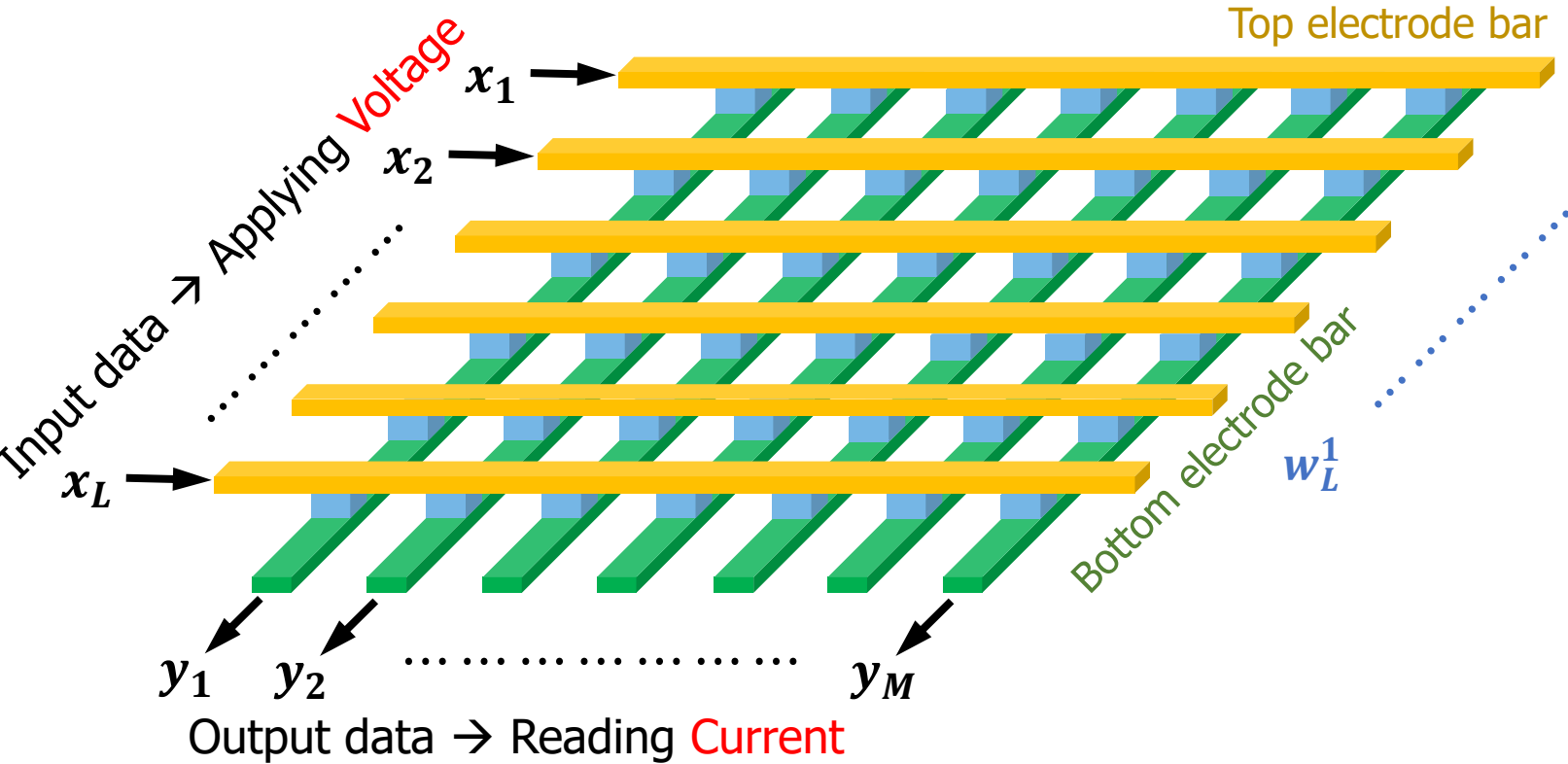
In-Memory computing



Big data

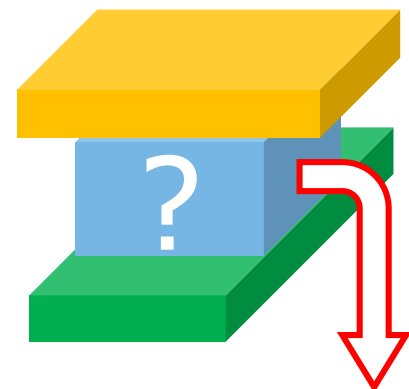
Massive computation

クロスバーアーキテクチャは積和演算問題を解決する



$w_1^1 \quad w_2^1 \quad \dots \dots \dots w_1^M$
 w_2^1
 w_L^1

Synaptic weight $L \times M$ matrix
→ Conductance of material



The material that can vary the conductance (resistance) as an artificial synapse is needed !

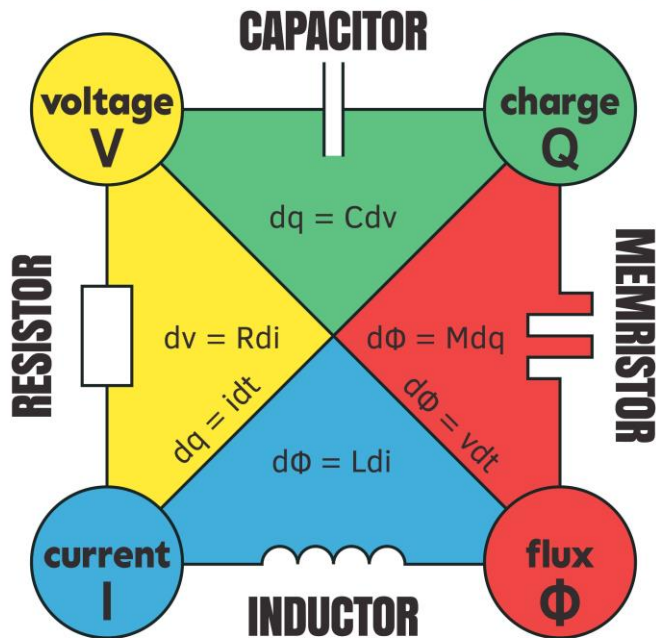
$$[x_1 \ x_2 \ \dots \ x_L] * \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_1^M \\ w_2^1 & w_2^2 & \dots & w_1^M \\ \vdots & \vdots & \ddots & \vdots \\ w_L^1 & w_L^2 & \dots & w_L^M \end{bmatrix} = \begin{bmatrix} y_1 = w_1^1 x_1 + w_2^1 x_2 \dots + w_L^1 x_L \\ y_2 = w_2^1 x_1 + w_2^2 x_2 \dots + w_L^2 x_L \\ \vdots \\ y_M = w_1^M x_1 + w_2^M x_2 \dots + w_L^M x_L \end{bmatrix}$$

Voltage × Conductance = Current

MAC can be perform in a hardware with cross-bars and cross-points !

メモリスト：第4の受動素子

CONCEPTUAL SYMMETRY



Four fundamental circuit variables

i : electric current

v : voltage

q : charge

ϕ : magnetic flux

Fundamental passive circuit elements

R : resistance (dv/di) → **Resistor**

C : capacitance (dq/dv) → **Capacitor**

L : inductance ($d\phi/di$) → **Inductor**

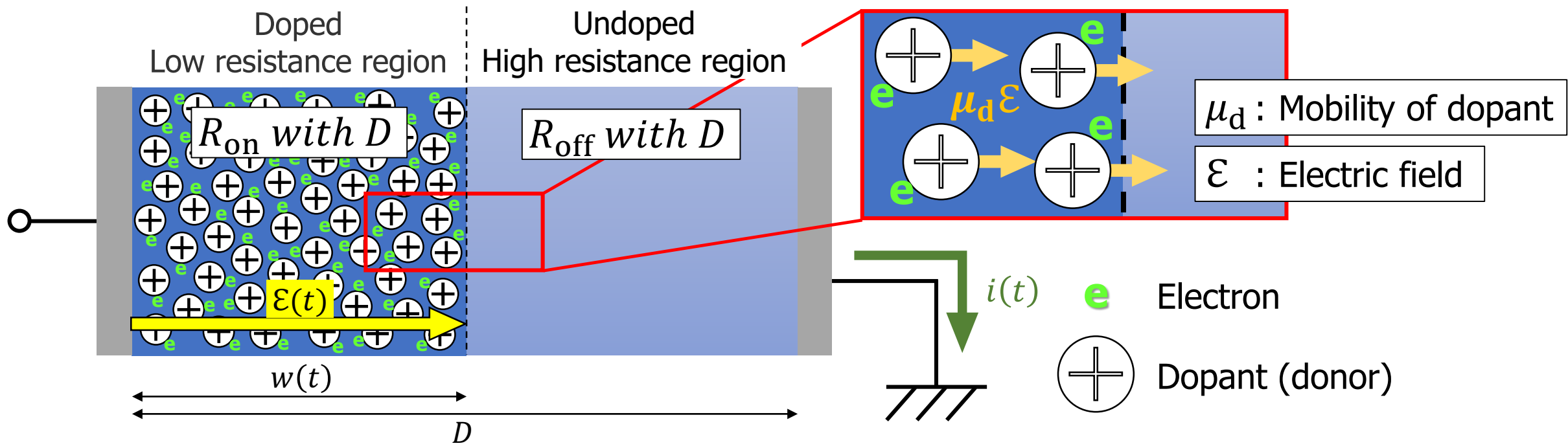
$$d\phi = M dq \Leftrightarrow \frac{d\phi}{dt} = M \frac{dq}{dt} \Leftrightarrow v = Mi$$

if $M = \text{const}$, $M = R$, being simple resistance

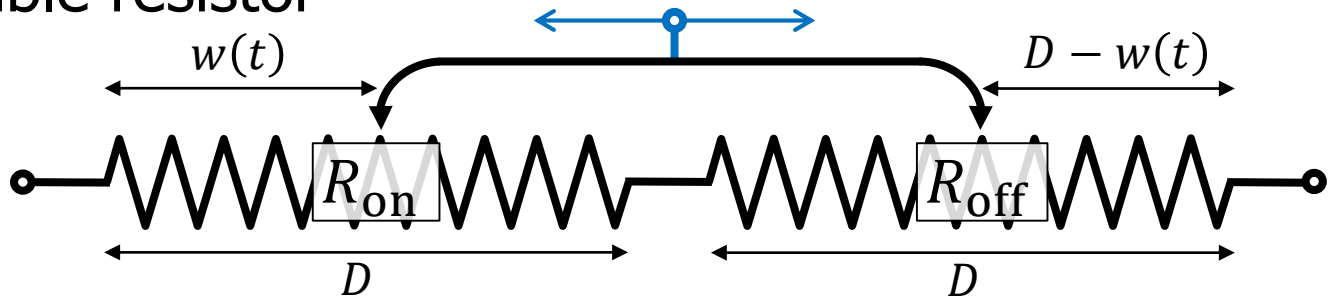
but if $M = M(q)$, a nonlinear element, no combination of R, C, L can duplicate M

メモリスタ：デバイス物理モデル

D. Strukov et al., Nature 453, 80 (2008).
"The missing memristor found"

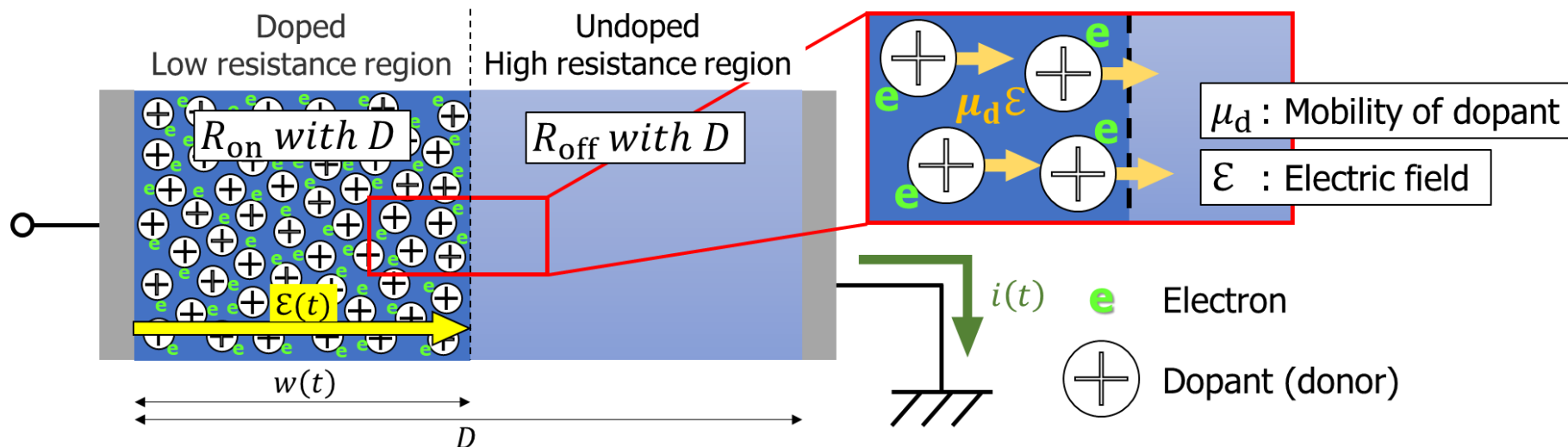


Similar to a variable resistor



Time-dependent total resistance :
$$R(t) = R_{on} \frac{w(t)}{D} + R_{off} \left\{ 1 - \frac{w(t)}{D} \right\}$$

メモリスタの『状態 (State) 』



$$R(t) = R_{\text{on}} \frac{w(t)}{D} + R_{\text{off}} \left\{ 1 - \frac{w(t)}{D} \right\}$$

When $x(t) \equiv \frac{w(t)}{D}$ (non-dimensional)
 $x(t)$ can represent a "state" of memristor at time t

Velocity of the boundary

$$\frac{dw(t)}{dt} = \mu_d E(t) = \mu_d \left\{ \frac{R_{\text{on}} \frac{w(t)}{D} \cdot i(t)}{w(t)} \right\} = \frac{\mu_d R_{\text{on}}}{D} \cdot i(t)$$

Voltage applied to the doped region

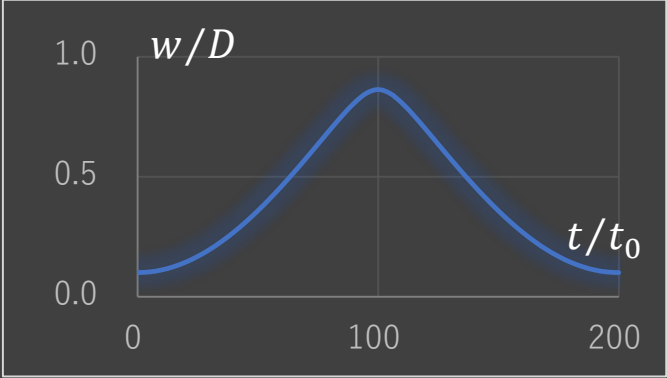
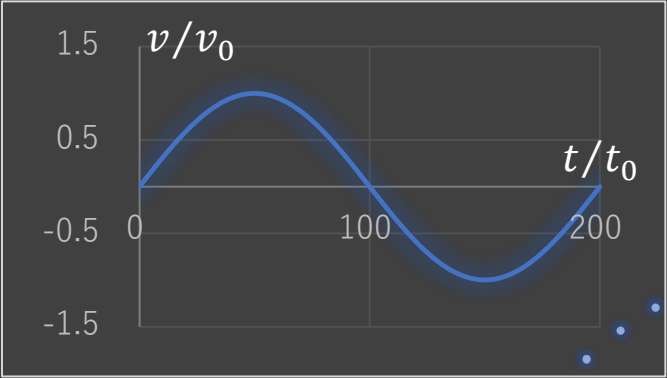
$$\frac{dx(t)}{dt} = \frac{1}{D} \frac{dw(t)}{dt} = \frac{\mu_d R_{\text{on}}}{D^2} \cdot i(t)$$

$$x(t) = \frac{\mu_d R_{\text{on}}}{D^2} \int i(t) dt = \frac{\mu_d R_{\text{on}}}{D^2} \cdot q(t)$$

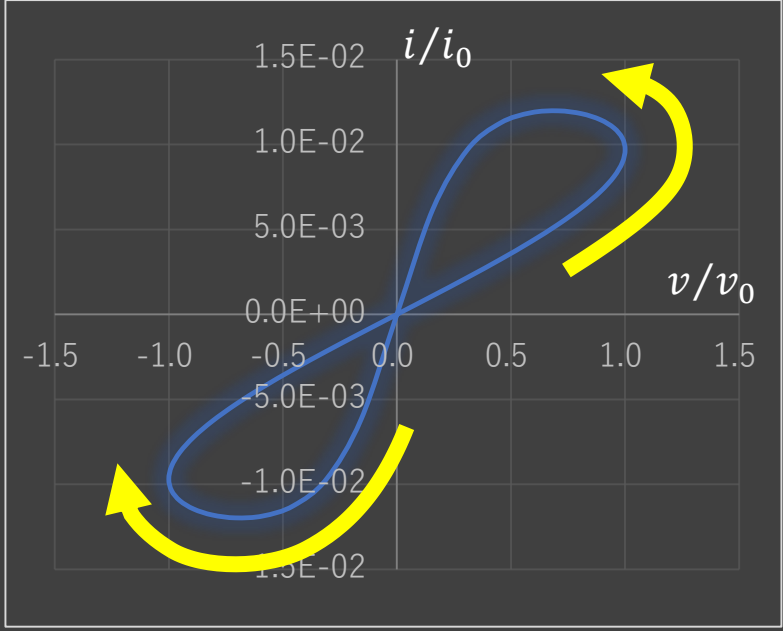
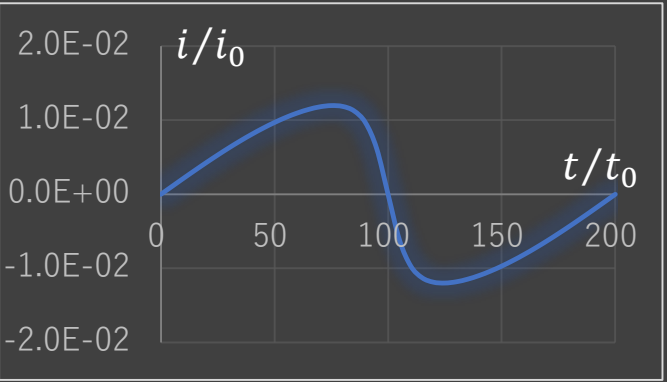
The "state" is proportional to charge q at t

メモリスタ特性の数値シミュレーション

Input : $v(t) = v_0 \sin \omega t$



Output: $i(t) = \frac{v_0 \sin \omega t}{R_{\text{on}} \sqrt{\left(\frac{R_{\text{off}}}{R_{\text{on}}}\right)^2 + 2 \left(\frac{R_{\text{off}}}{R_{\text{on}}} - 1\right) \left(\frac{\mu_d}{D^2} \frac{v_0 \cos \omega t}{\omega} + \text{Const}\right)}}$

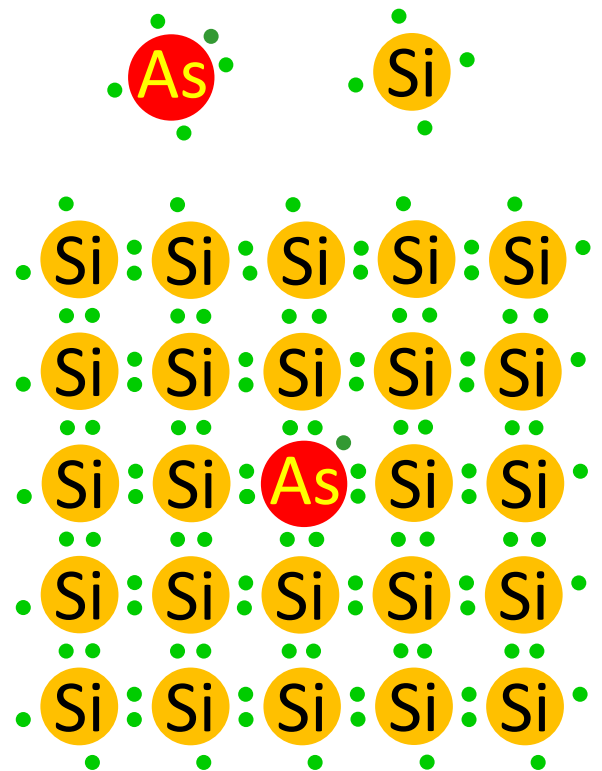


t [s]	R _{off} /R _{on}	v ₀ [V]	i ₀ [A]	D [m]	μ _v [m ² /Vs]	t ₀ [s]	ω	w ₀ /D
2.0000	160	1.00	0.010	1.0E-08	1.00E-14	0.010	3.1416	0.10

$t_0 = D^2 / v_0 \mu_v$

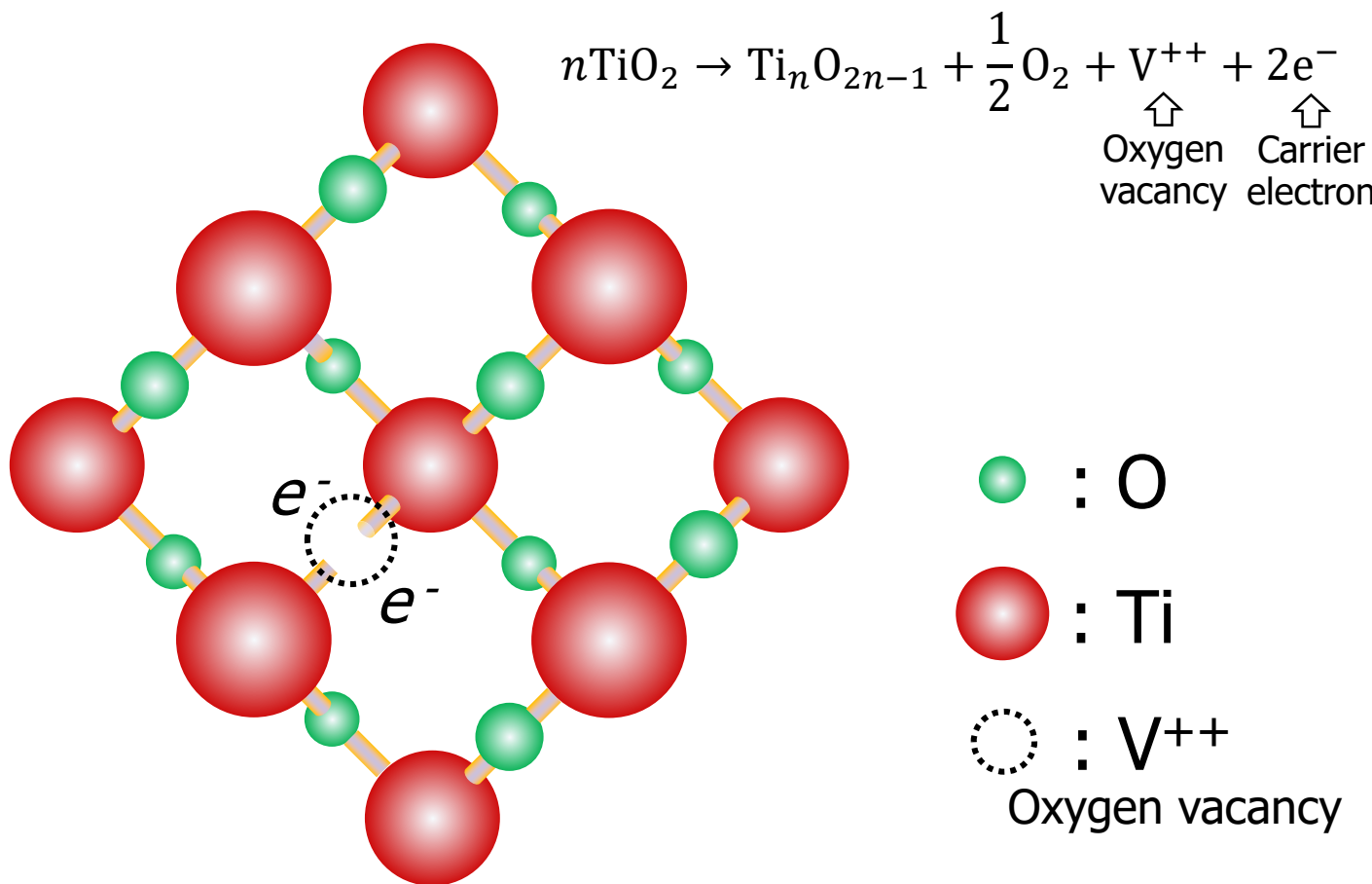
還元性金属酸化物はメモrista材料となる

n-type Si



As atom, a donor, **never** drift when applying voltage bias

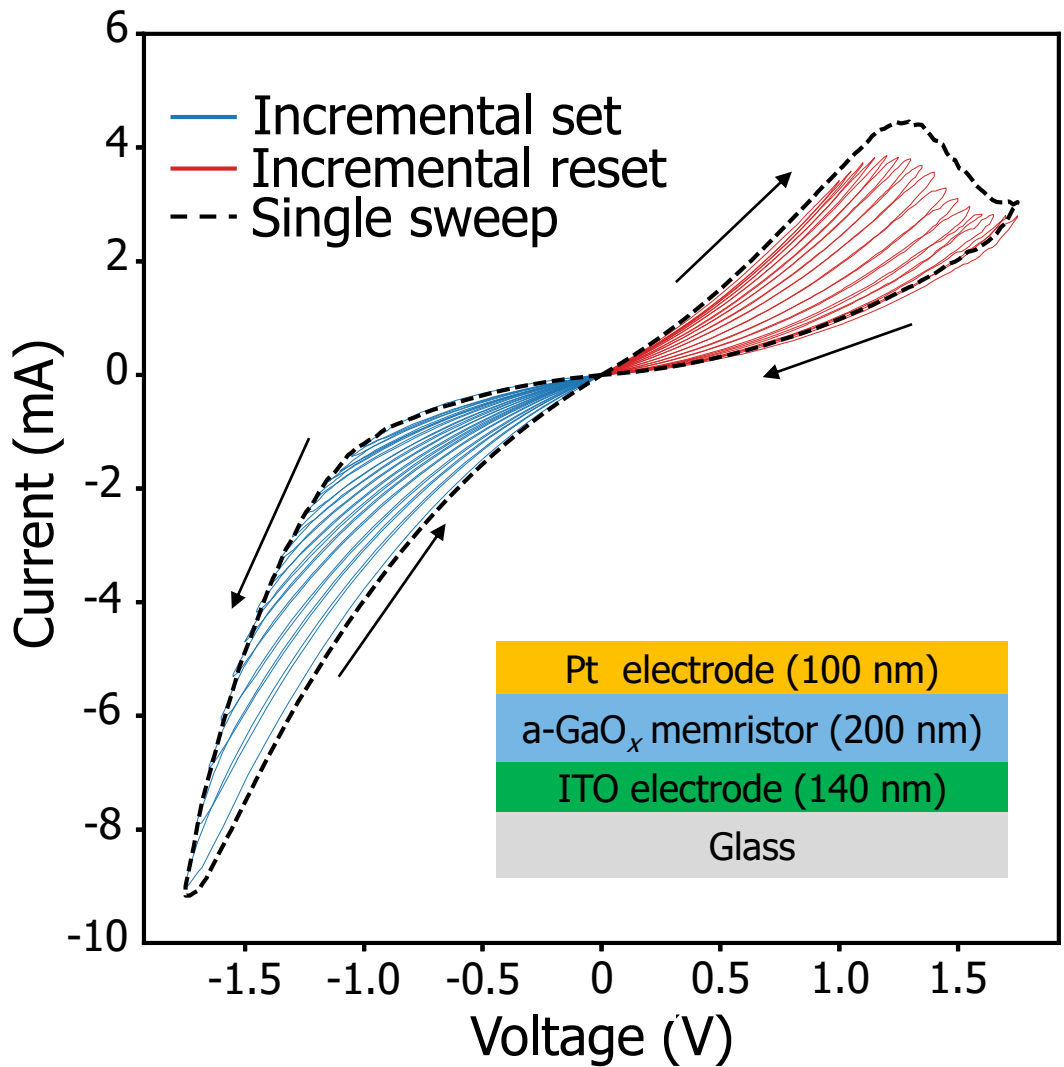
Reduced metal-oxide, e.g., TiO_{2-x}



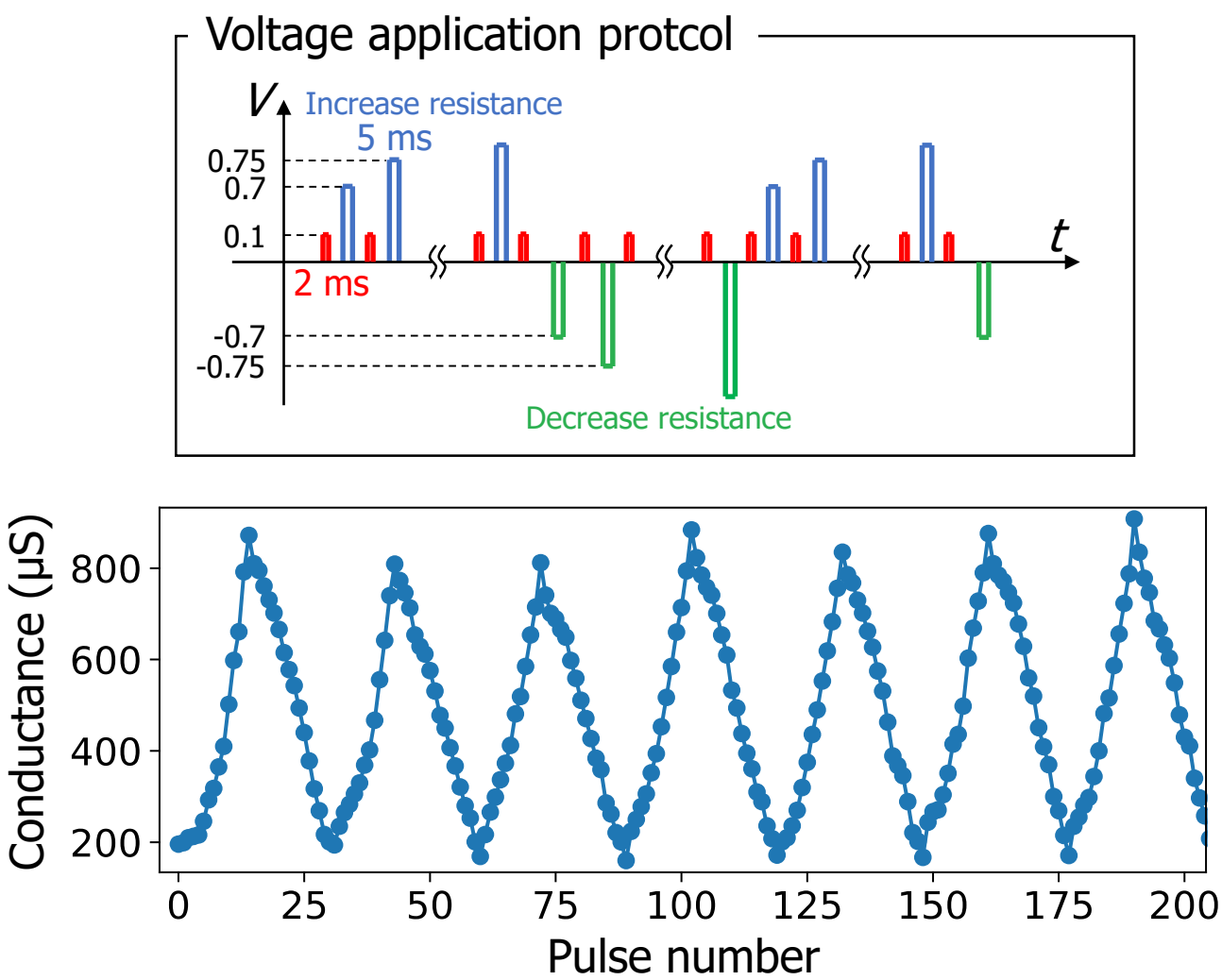
Oxygen vacancy, a donor, **can** drift when applying voltage bias

アモルファスGaO_xメモ리스タ

Current-Voltage characteristic

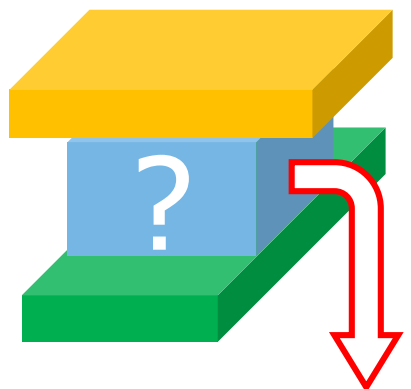
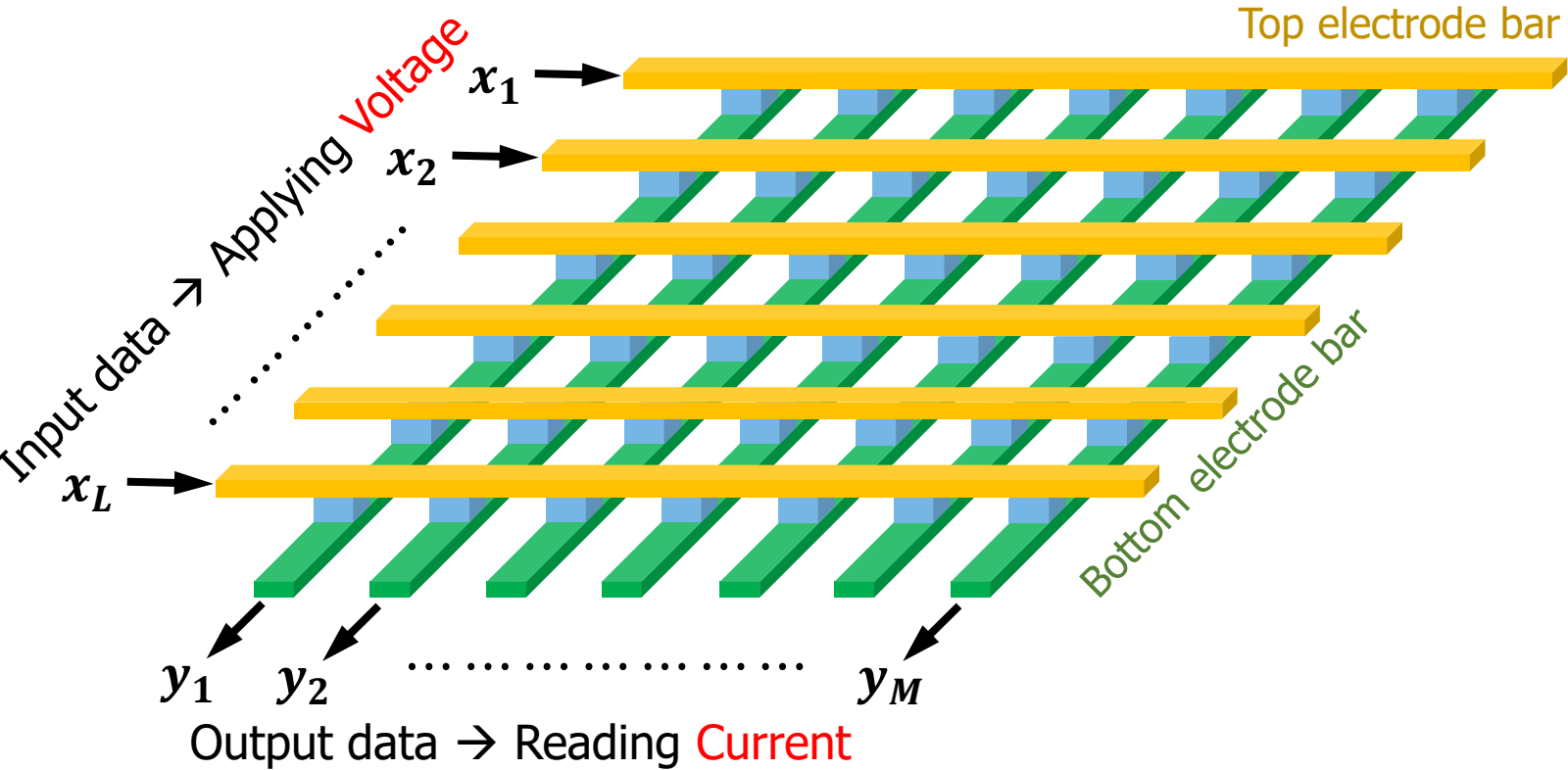


Conductance variation by voltage pulses



Multilevel conductance can be effectively used for synaptic weight change

クロスバーアーキテクチャは積和演算問題を解決する



The material that can vary the conductance (resistance) as an artificial synapse is needed !

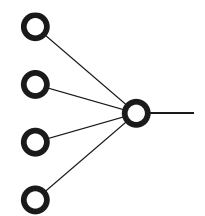
Memristor !

$$[x_1 \ x_2 \ \cdots \ x_L] * \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^M \\ w_2^1 & w_2^2 & \cdots & w_2^M \\ \vdots & \vdots & \ddots & \vdots \\ w_L^1 & w_L^2 & \cdots & w_L^M \end{bmatrix} = \begin{bmatrix} y_1 = w_1^1 x_1 + w_1^2 x_2 \cdots + w_1^L x_L \\ y_2 = w_2^1 x_1 + w_2^2 x_2 \cdots + w_2^L x_L \\ \vdots \\ y_M = w_M^1 x_1 + w_M^2 x_2 \cdots + w_M^L x_L \end{bmatrix}$$

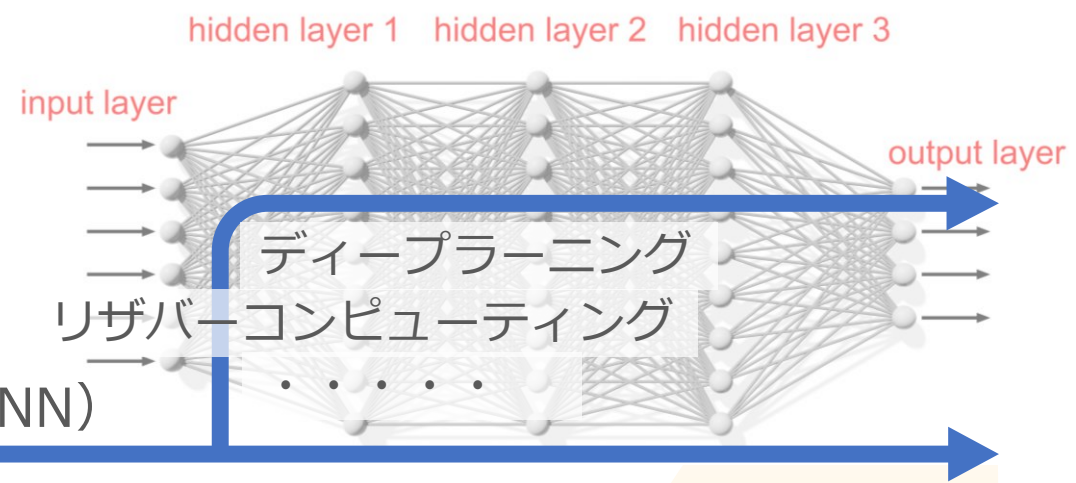
Voltage × Conductance = Current

MAC can be perform in a hardware with cross-bars and cross-points !

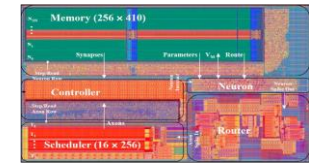
AI（人工知能）開発の系譜



人工ニューラルネットワーク
(Artificial Neural Network: ANN)



ニューロモルフィックチップ
TrueNorth

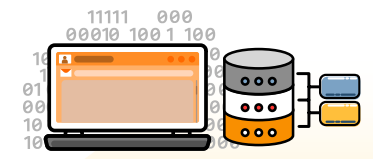


Loihi
Kirin970
A11 Bionic
.....

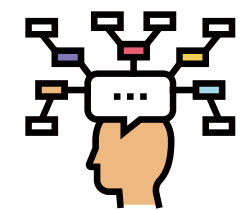


エキスパートシステム オントロジー

Mycin



Deep Blue



Watson



対話システム

Eliza

Siri



1960年代

1970年代

1980年代

1990年代

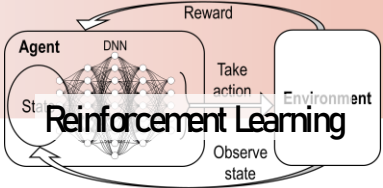
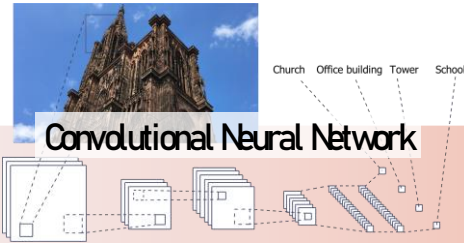
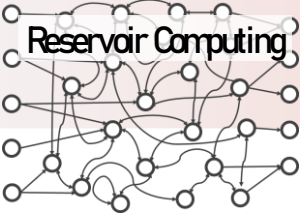
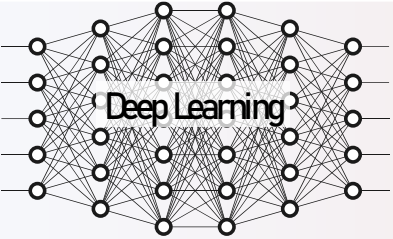
2000年代

2010年代

2020年代

ANNと神経科学

Information Science

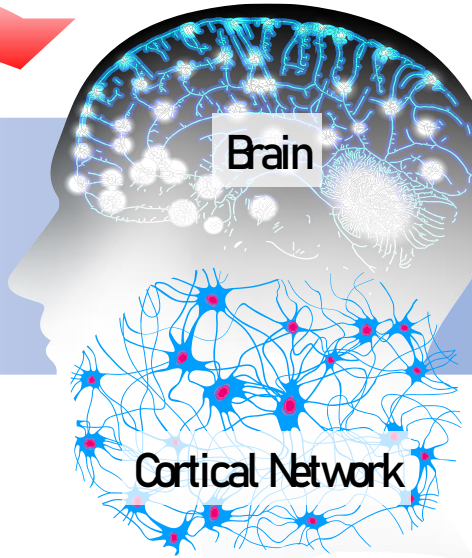
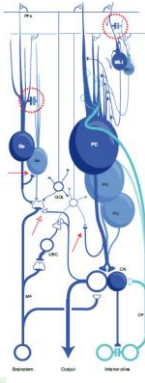
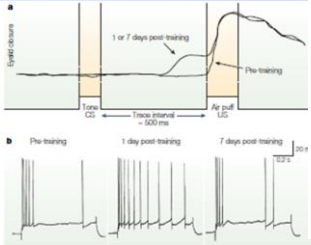
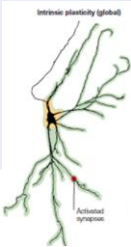
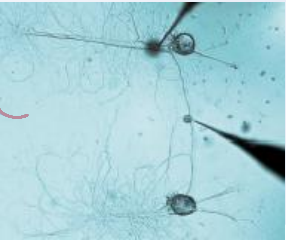
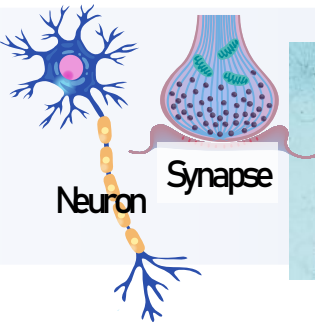


人工シナプス素子ハードウェアの
観点からANNをリブートする



Hyper
AI

AI Electronics 情報科学と神経科学の発展的な橋渡し



脳の理解
疾患治療

Neuroscience

人工シナプス素子

シナプス可塑性の模倣

+

変調性

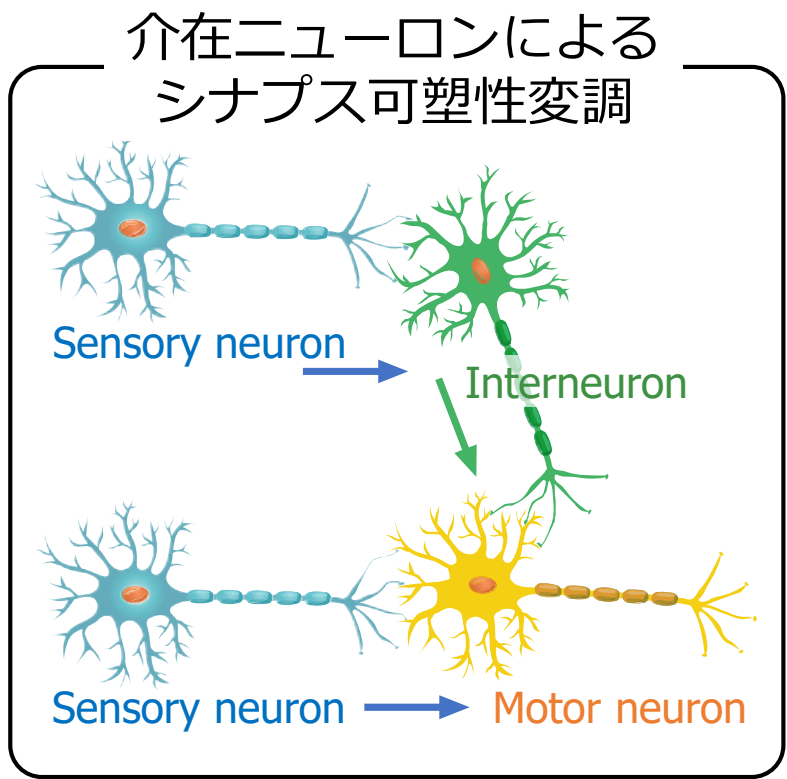
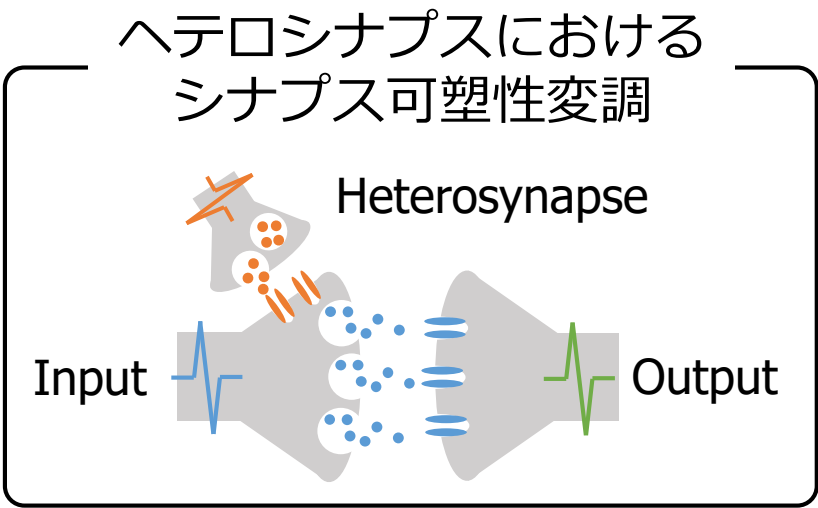
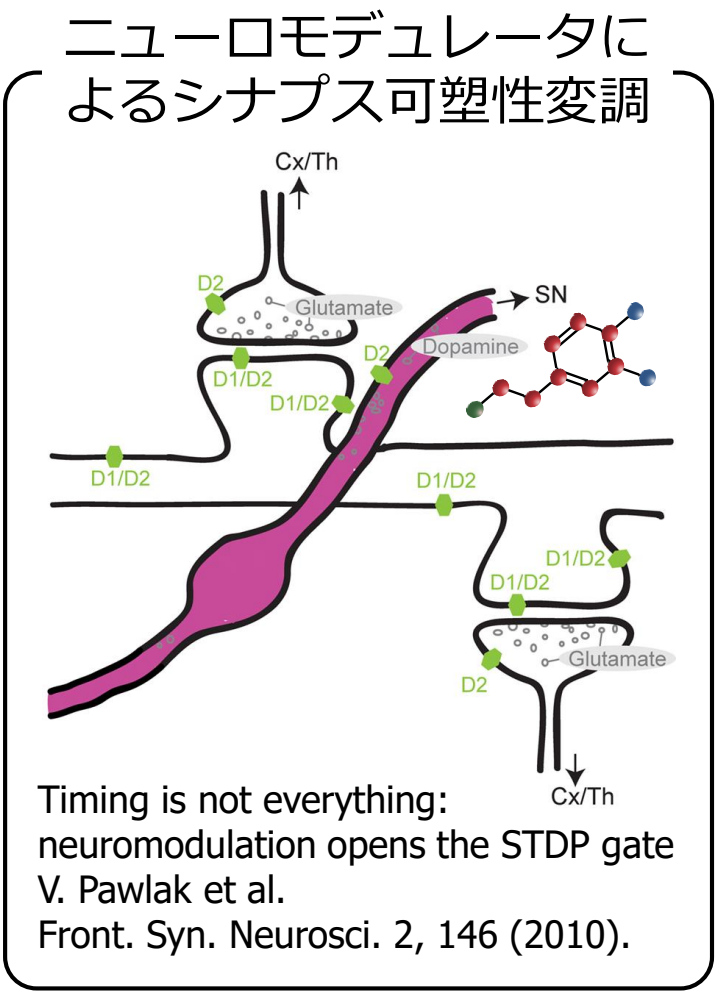
連合性

相関性

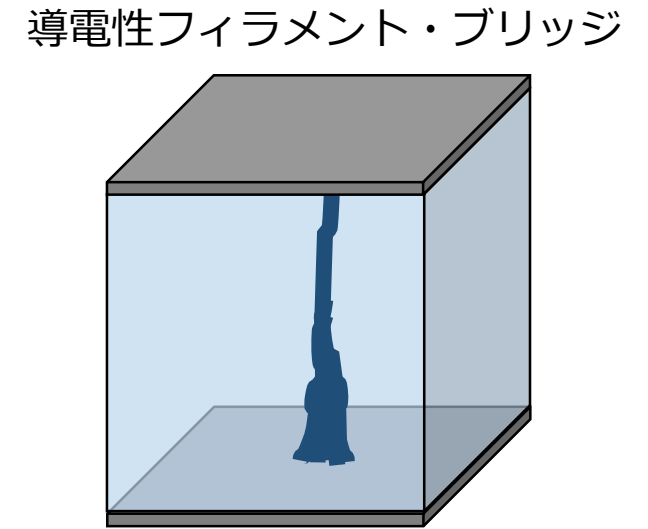
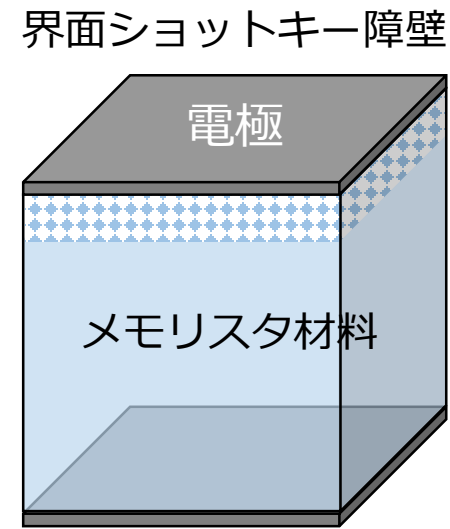
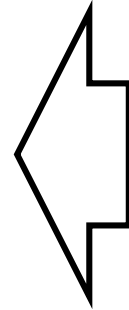
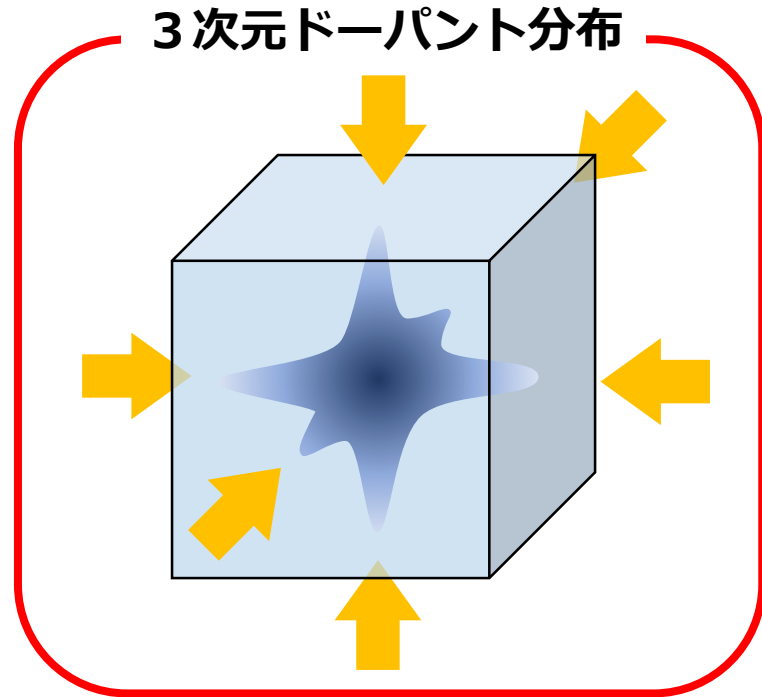
⋮



多端子メモリスタ素子の創製



酸素空孔分布制御型多端子メモ리스タ素子の機能原理



ドーパント分布のトポロジを電界で制御し
バルクの電気伝導に方位依存性を持たせて
種々の抵抗変化特性を発現させる

ドーパント = 酸素空孔

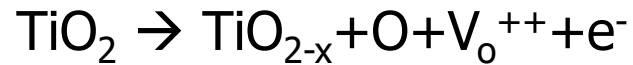
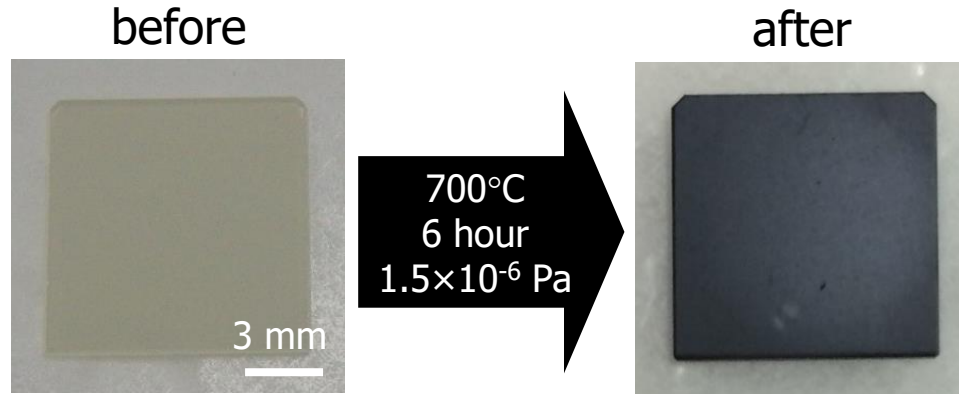
■ 単結晶 TiO_{2-x}
還元バルク基板
エピタキシャル薄膜

■ アモルファス GaO_x 薄膜

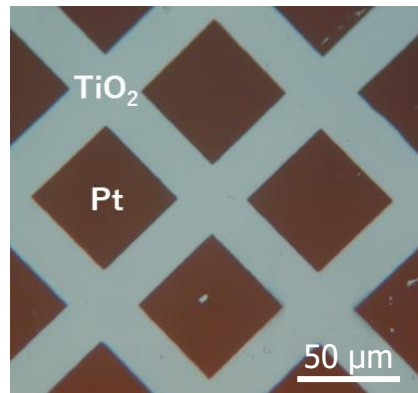
還元TiO₂ (TiO_{2-x}) 結晶・薄膜

■ ルチル型TiO₂単結晶の熱還元

真空熱処理



■ 4端子平面型素子



Optical microscope image

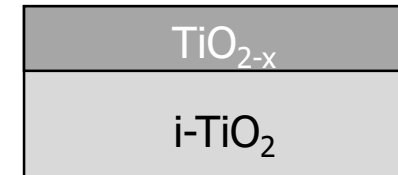
■ 数十 μmスケール素子

■ 電気着色現象による酸素空孔分布の可視化

Ti³⁺ color center

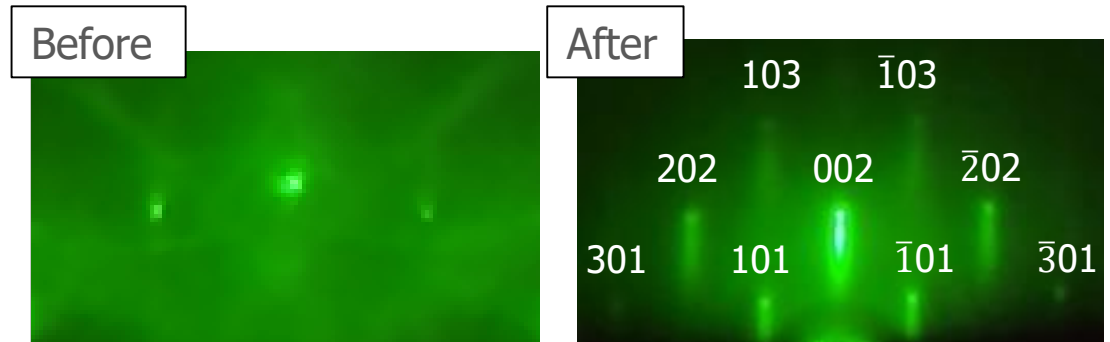
■ ルチル型還元TiO₂エピタキシャル薄膜成長

Pulsed Laser Deposition



Thickness: 50~100 nm

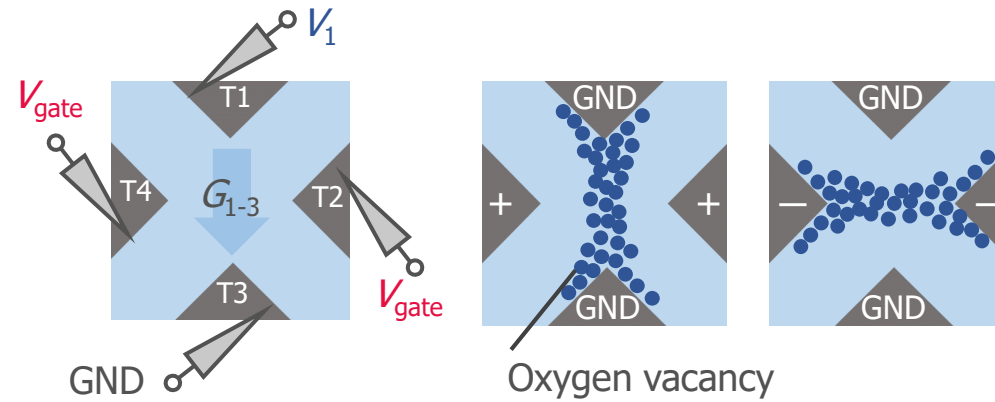
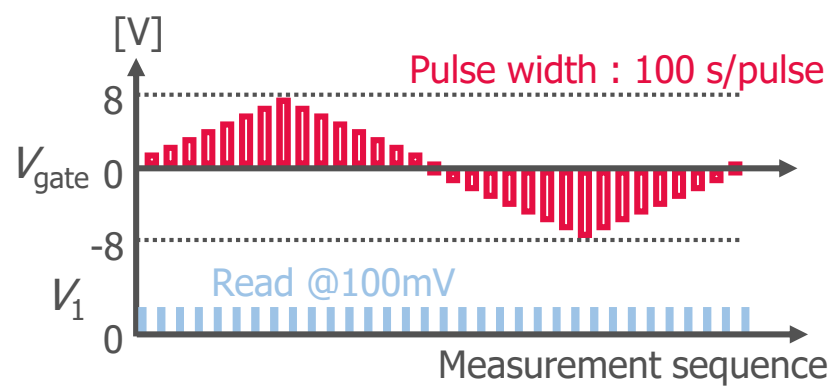
Resistivity: ~10⁻³ Ωcm



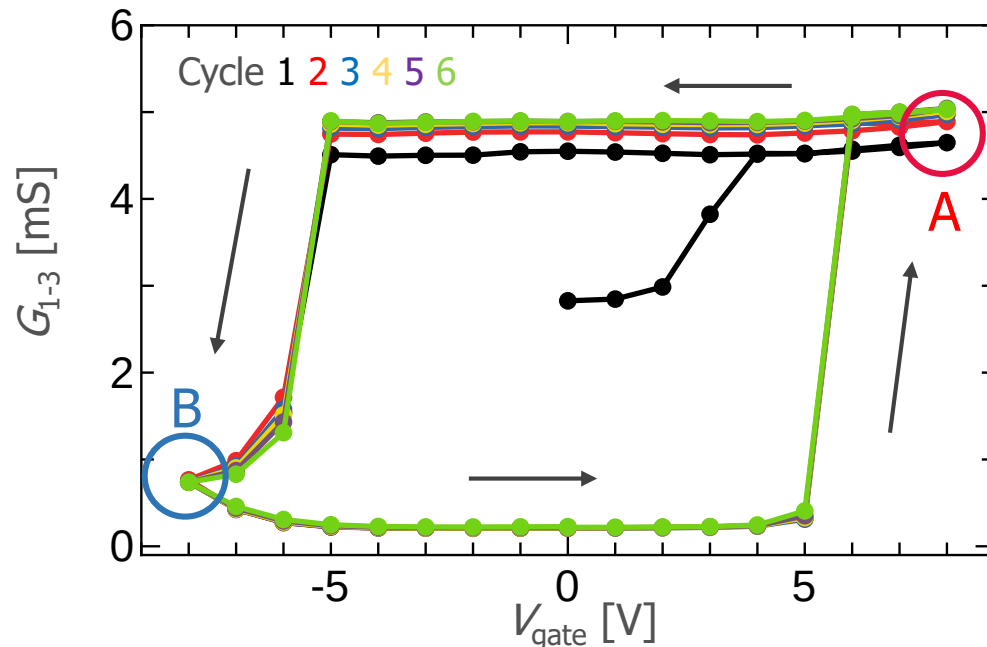
In situ RHEED pattern

4端子TiO_{2-x}メモリスタ

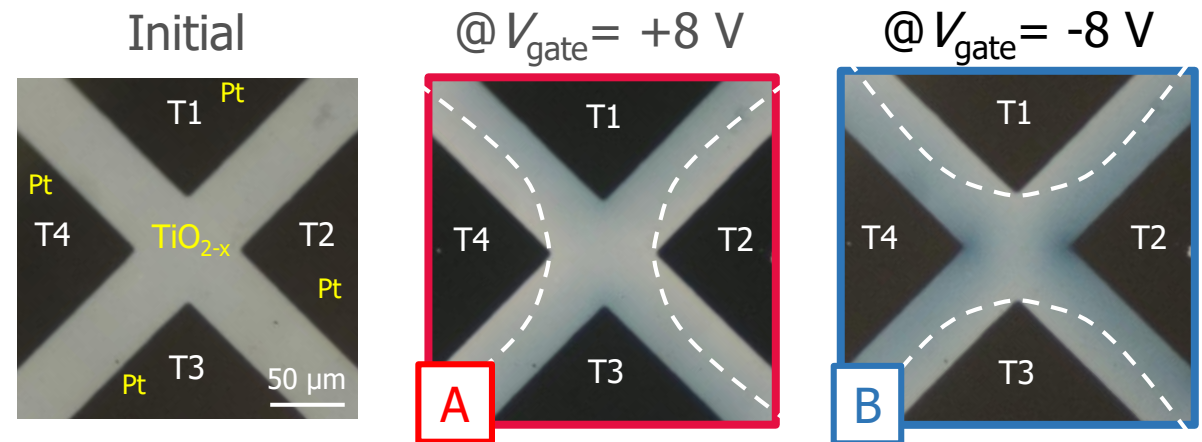
Voltage application protocol for measuring conductance G_{1-3} between T1 and T3



Variation of G_{1-3} and oxygen vacancy distribution depending on V_{gate}



Optical microscope images

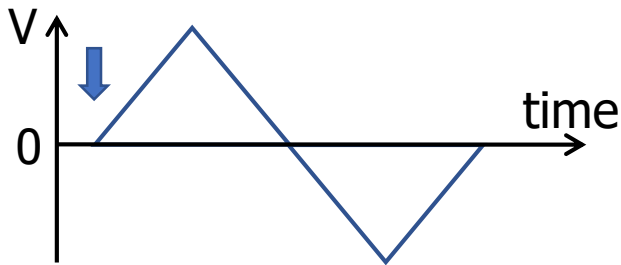


Electrocoloring clearly reveals V_{gate} -modulated oxygen vacancy distribution, consistent with G_{1-3} modulation

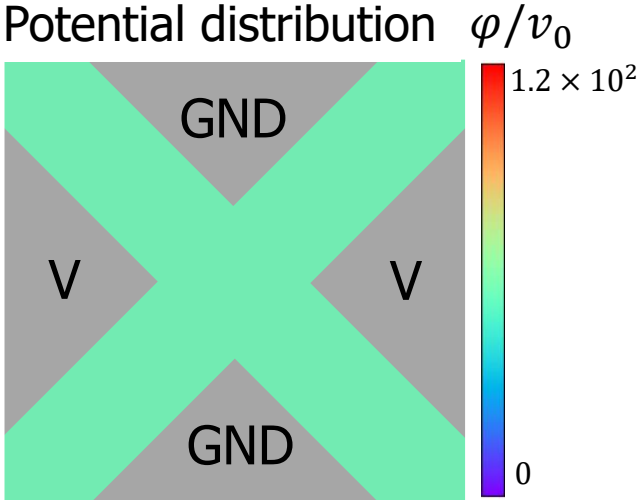
2次元ポテンシャル・酸素空孔分布シミュレーション

酸素空孔のドリフト・拡散モデル

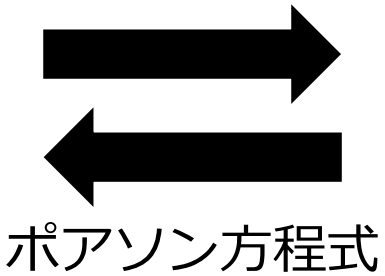
ポアソン方程式	$\nabla^2 \varphi = -\frac{q}{\epsilon}(p - n + N_d - N_a)$
キャリア連続の式	$\nabla \cdot (qp\mu_p \nabla \varphi + qD_p \nabla p) = 0 \quad \nabla \cdot (qn\mu_n \nabla \varphi - qD_n \nabla n) = 0$
ドナーイオン連続の式	$q \frac{dN_d}{dt} = \nabla (qn\mu_i \nabla \varphi - qD_i \nabla N_d) \quad \leftarrow \text{ドリフト・拡散近似}$



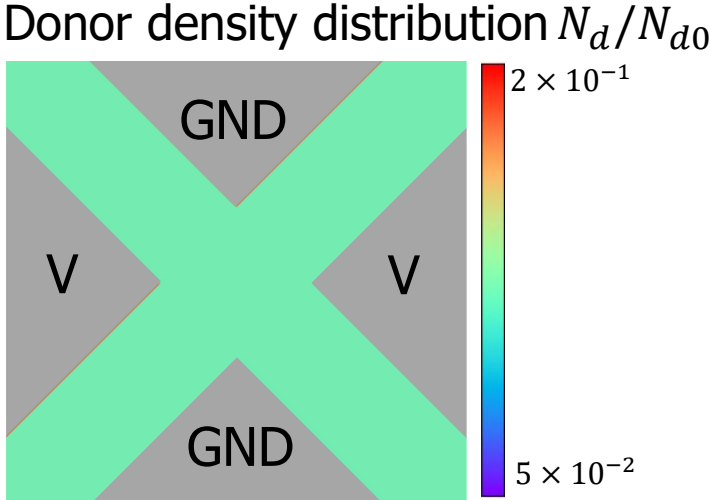
熱起電力 $v_0 = k_b T / e$
電極内 $N_{d0} = 5 \times 10^{26}$
初期値 $N_d^* = 5 \times 10^{25}$



キャリア連続の式
ドナーイオン連続の式



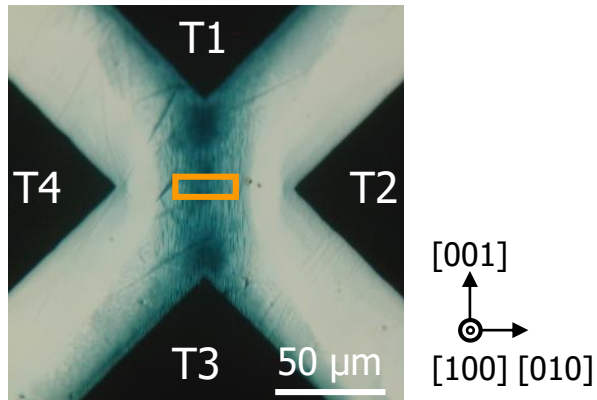
ポアソン方程式



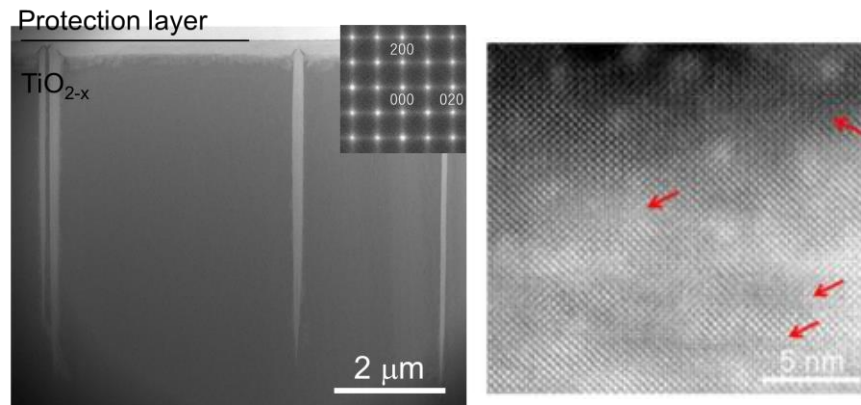
TiO_{2-x}(100)素子と(001)素子の比較：結晶構造

HAADF-STEM observations

(100) Oriented device



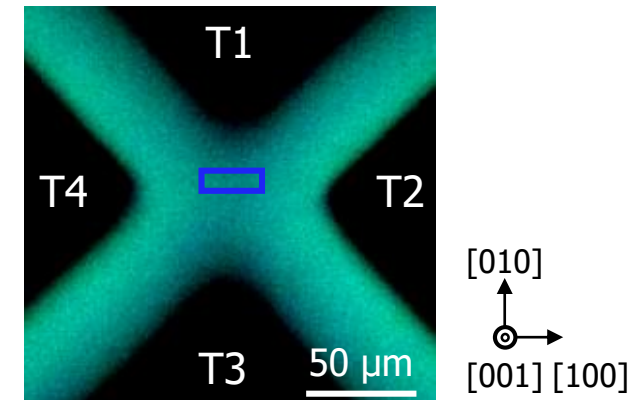
Crystal structure in colored region



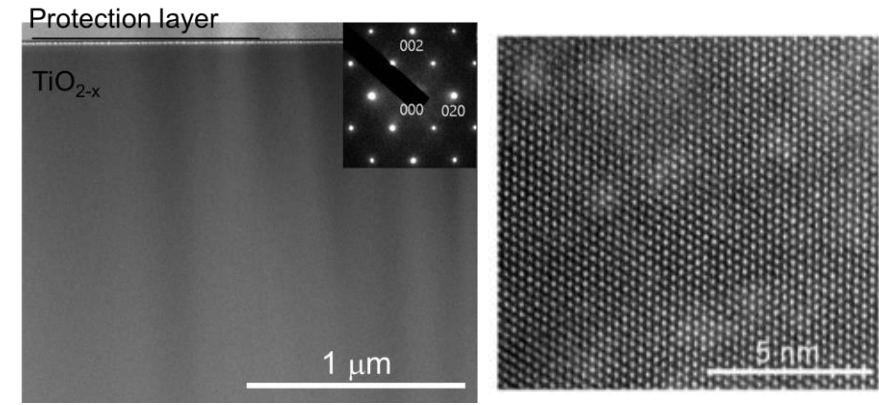
面状欠陥構造（筋状明コントラスト）
局所格子欠陥

S. Takeuchi, AS et al., Sci. Reports 9, 2601 (2019).

(001) Oriented device



Crystal structure in colored region



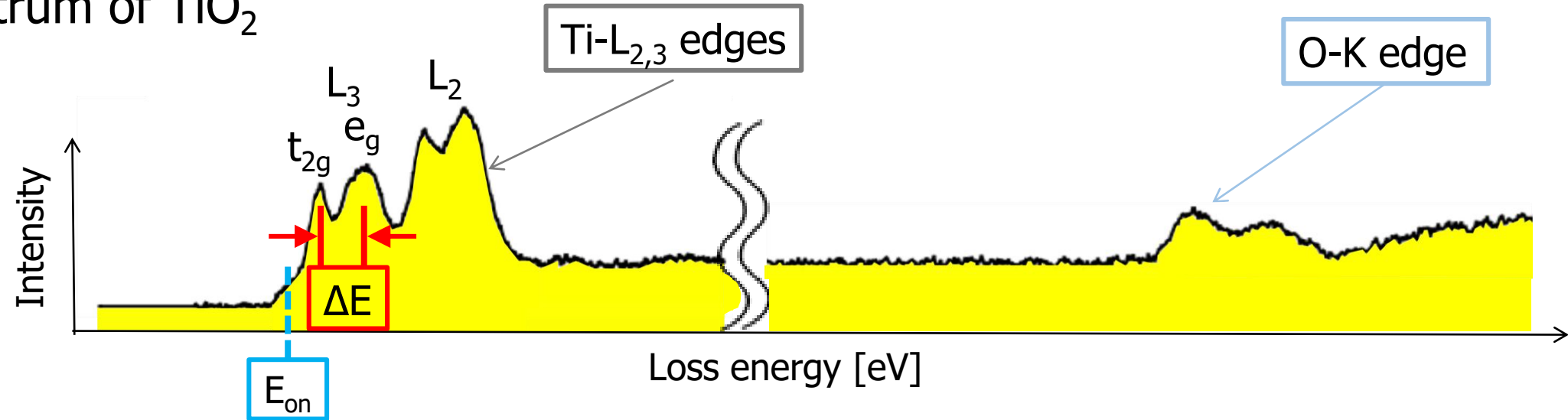
結晶構造変化が少ない
均一な格子構造

Z. Nagata, AS et al., Sci. Reports 9, 10013 (2019).

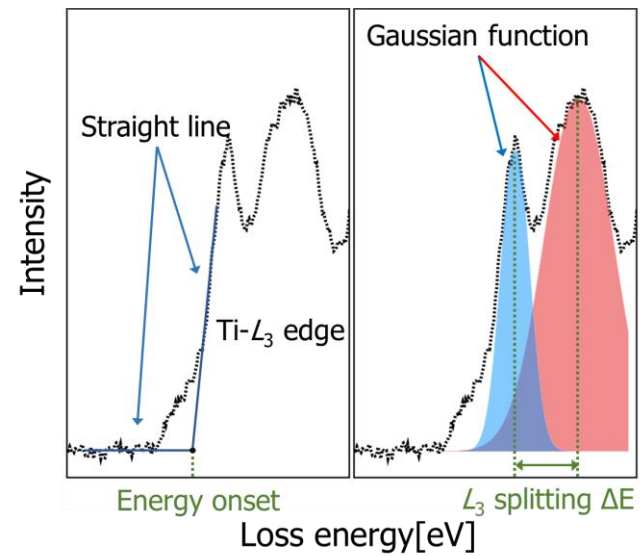
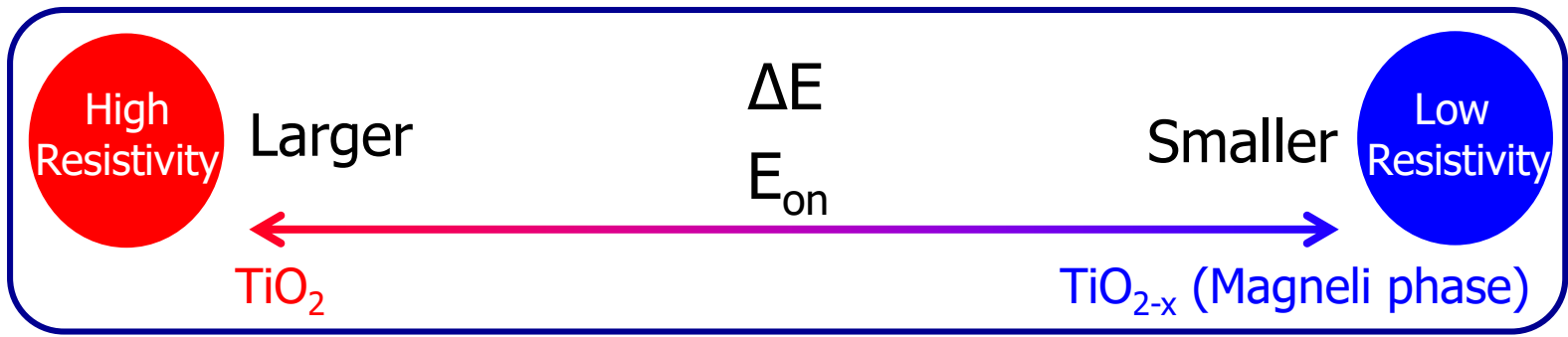
STEM-EELS 分析

*EELS : Electron Energy Loss Spectroscopy

Spectrum of TiO_2



組成情報の抽出 Energy splitting of e_g and t_{2g} : ΔE [1], [2]
Energy onset of Ti- $L_{2,3}$ edges : E_{on}

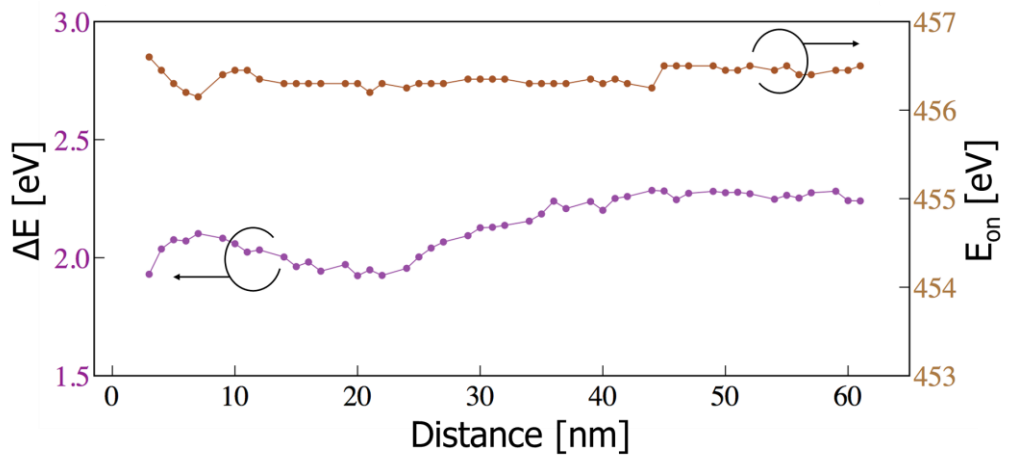
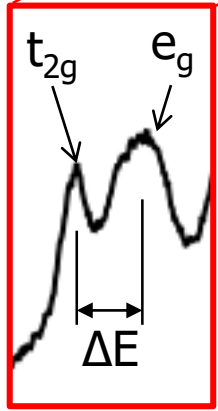
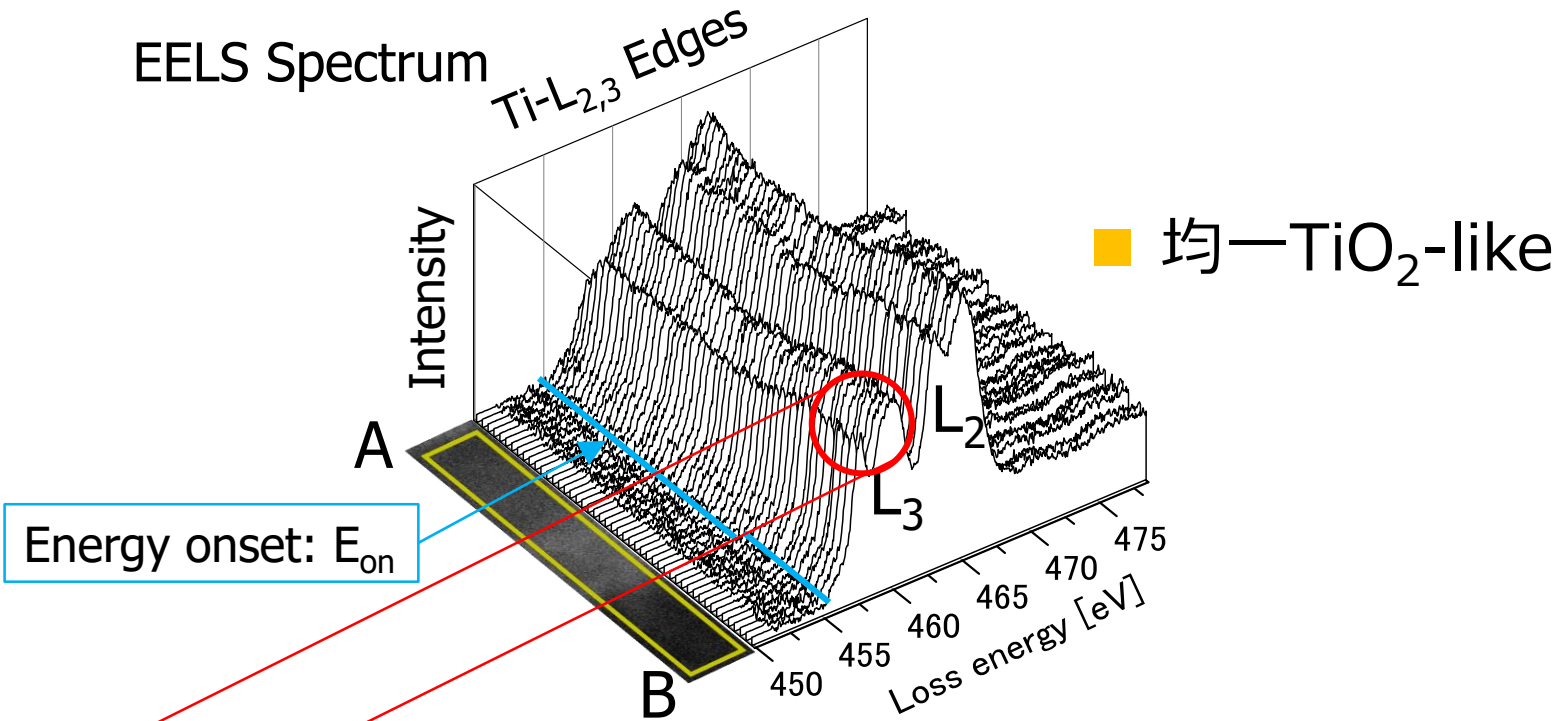
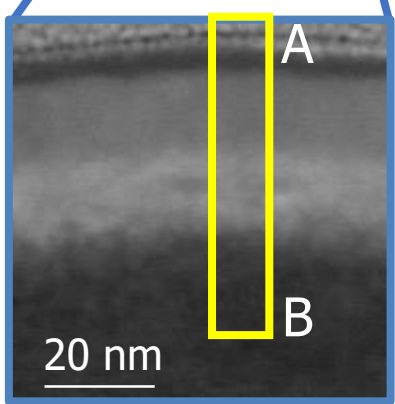
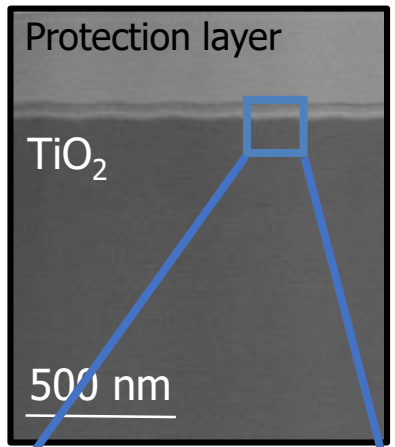


[1] A. Moballegh et al., Acta Materialia 86, 352 (2015). [2] E. Stoyanov et al., American Mineralogist 92, 577 (2007).

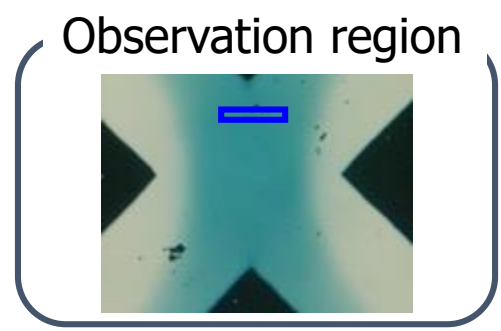
STEM-EELS 分析 : Colorless領域



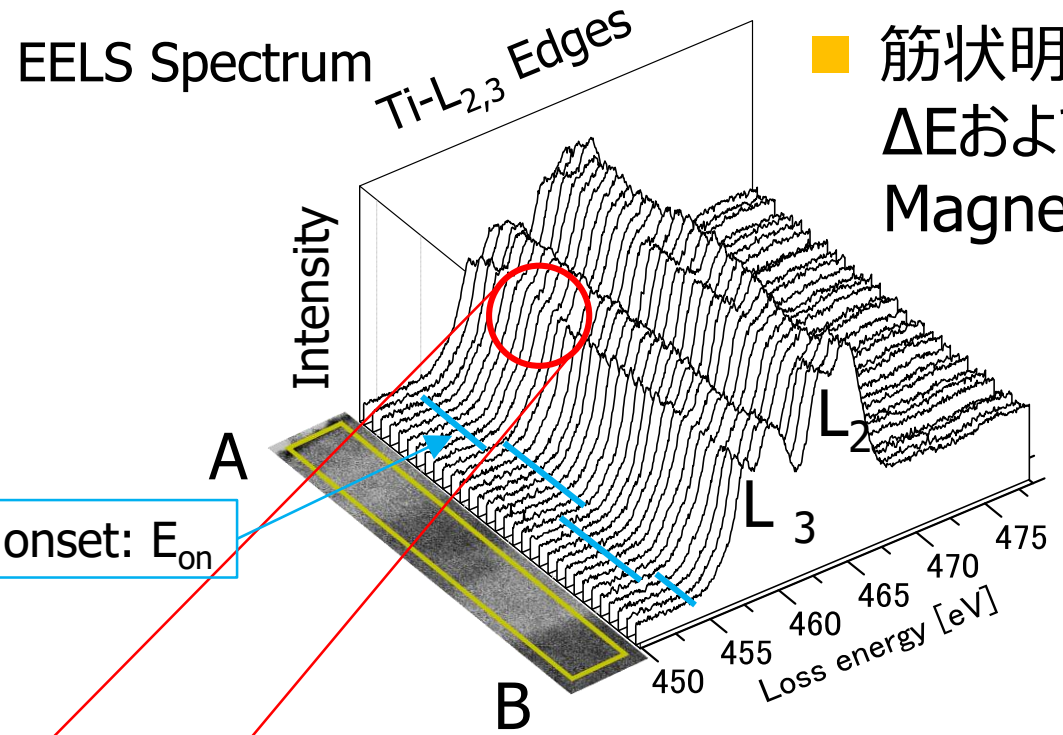
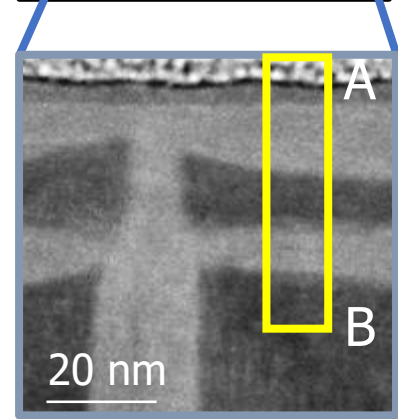
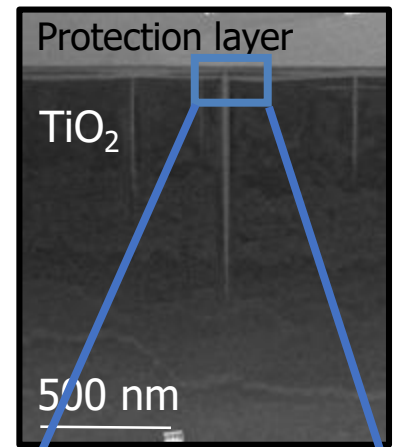
HAADF-STEM Image



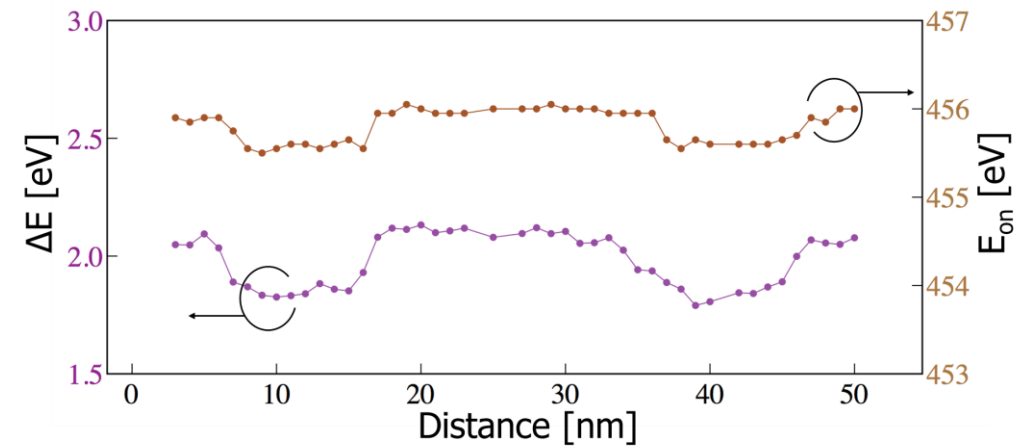
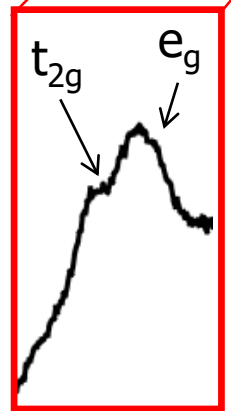
STEM-EELS 分析 : Colored 領域



HAADF-STEM Image

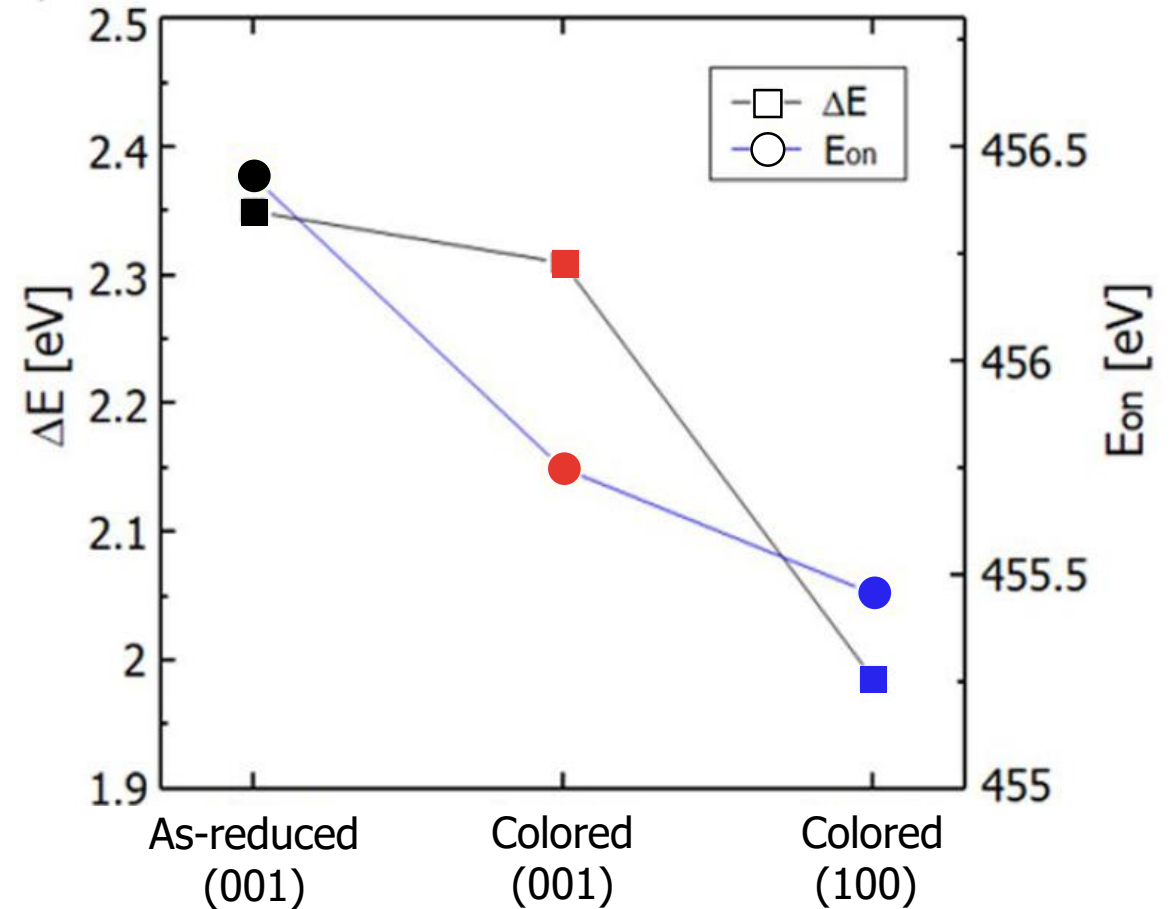
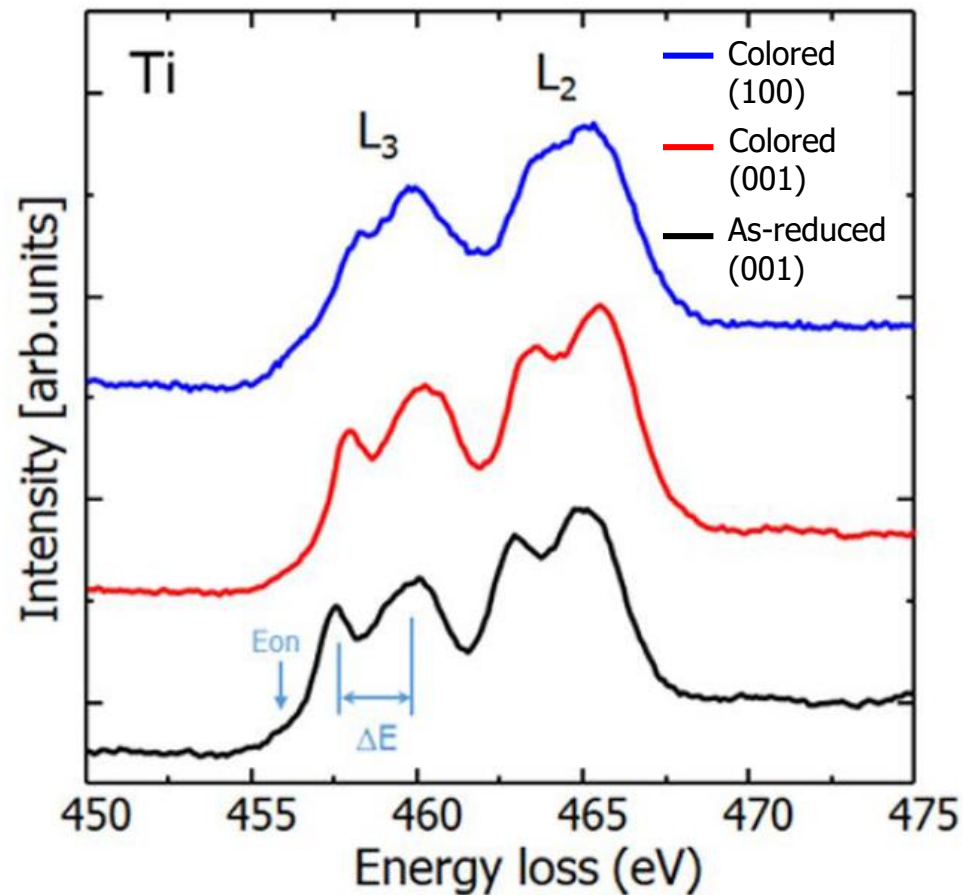


■ 筋状明コントラスト
ΔEおよびE_{on}の局所的減少
Magneli phase-like



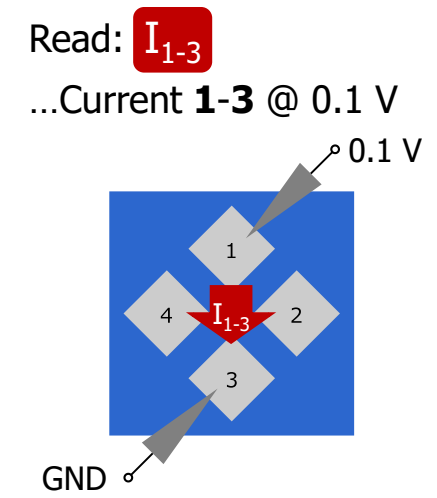
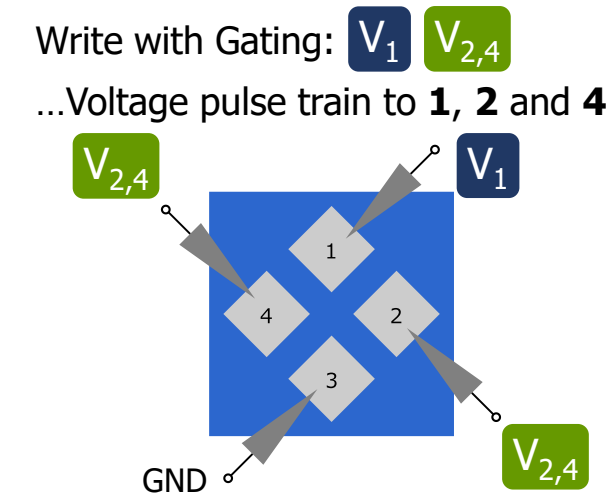
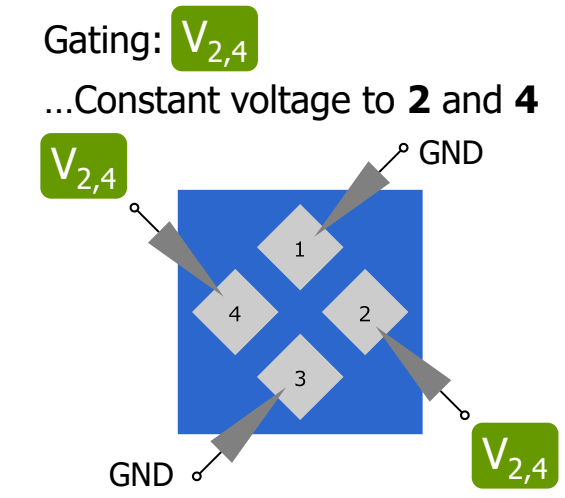
TiO_{2-x}(100)素子と(001)素子の比較：電子構造

STEM-EELS measurements

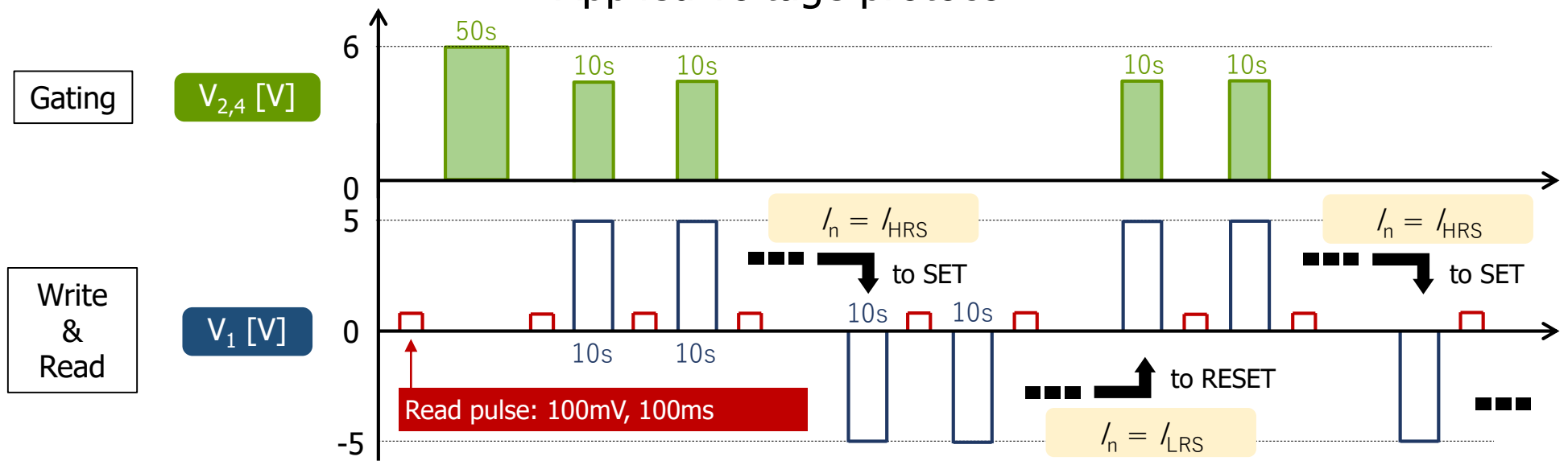


- Ti 価電子状態に明確な違い
- 過度の結晶変形を伴わないmoderateな電子状態変化

4端子メモ리스タ素子：ゲート変調の電圧印加プロトコル

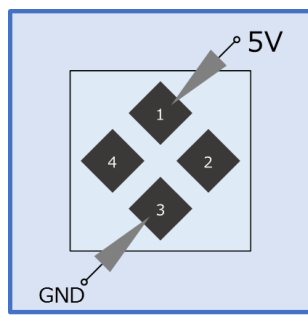
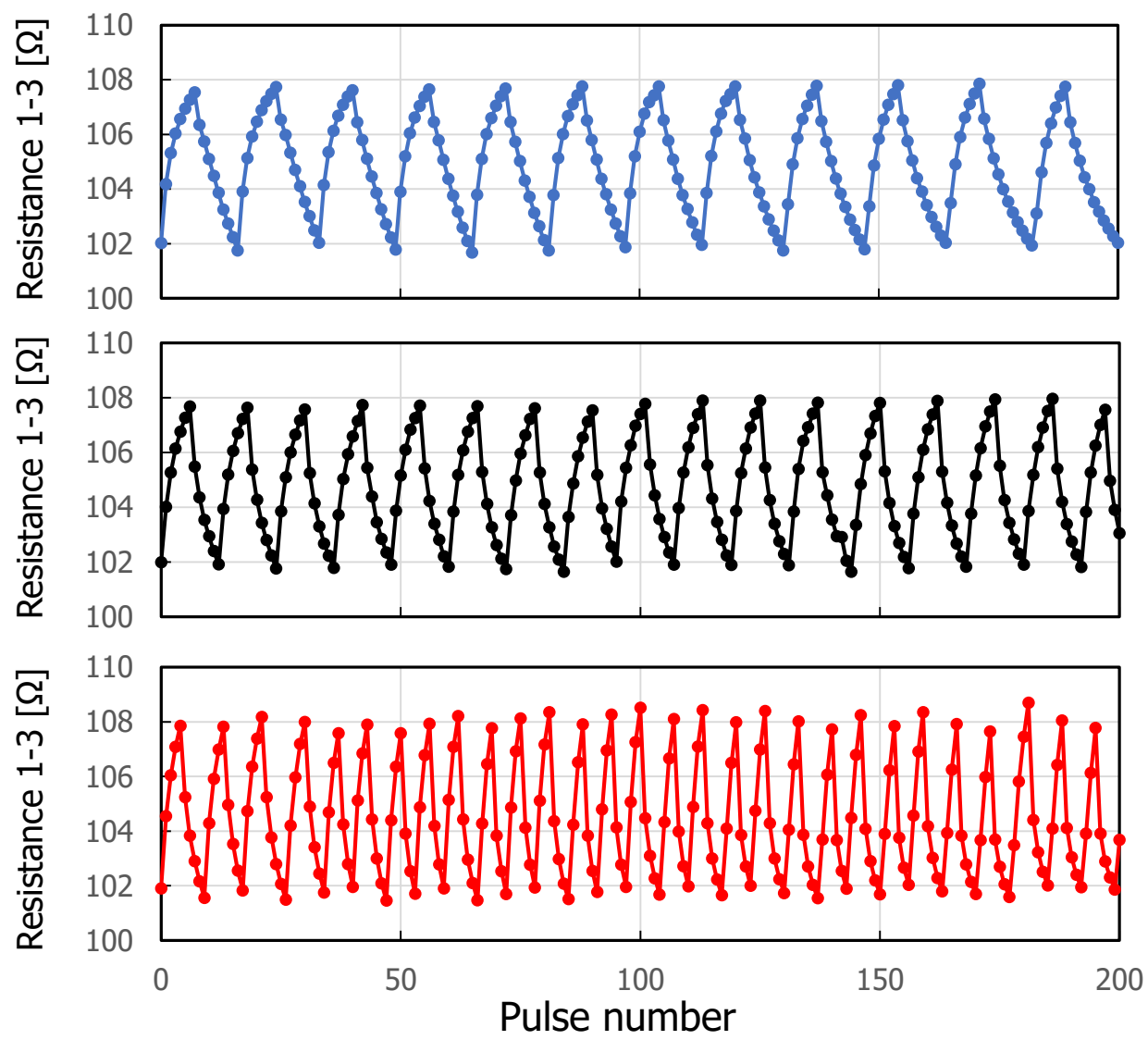


Applied voltage protocol

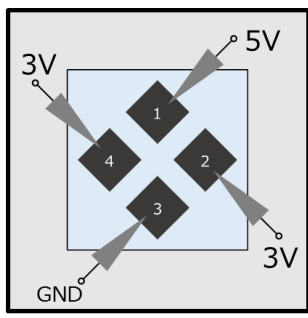


Depletion/Potential 特性のゲート変調

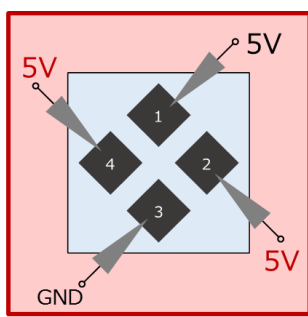
Gate voltage



n/a



3 V



5 V

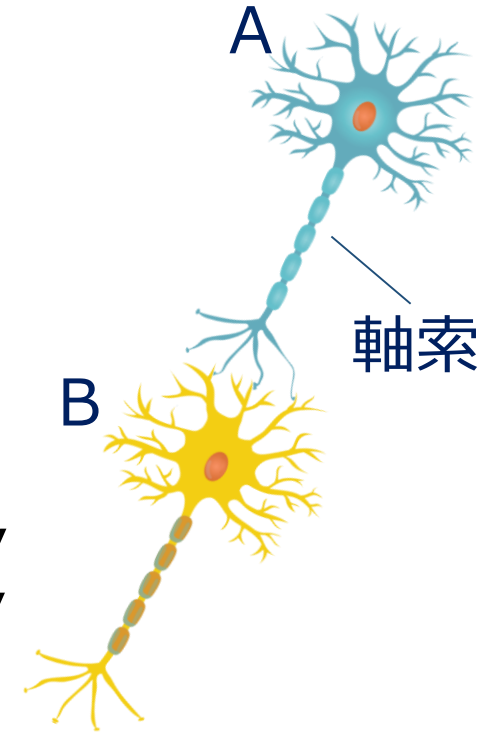
■ ゲート変調によるシナプスの学習／忘却効率の制御が可能

Hebbの法則

『The Organization of Behavior』 D. O. Hebb 1949

"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

「細胞Aの軸索が細胞Bを興奮させるのに十分に近く、かつそれが繰り返しあるいは持続的にBの発火に関与する場合、AまたはB、あるいはその両方において、何らかの成長過程や代謝変化が生じ、AがBを発火させる効率が向上する。」

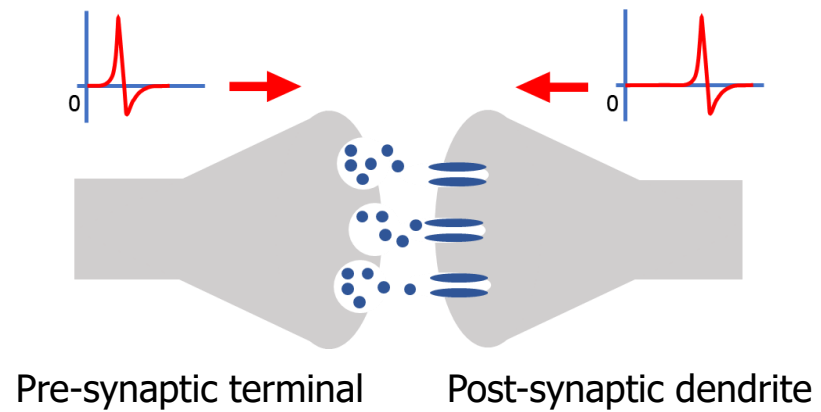


「一緒に発火するニューロンは結びつく」というシナプス可塑性の基本概念

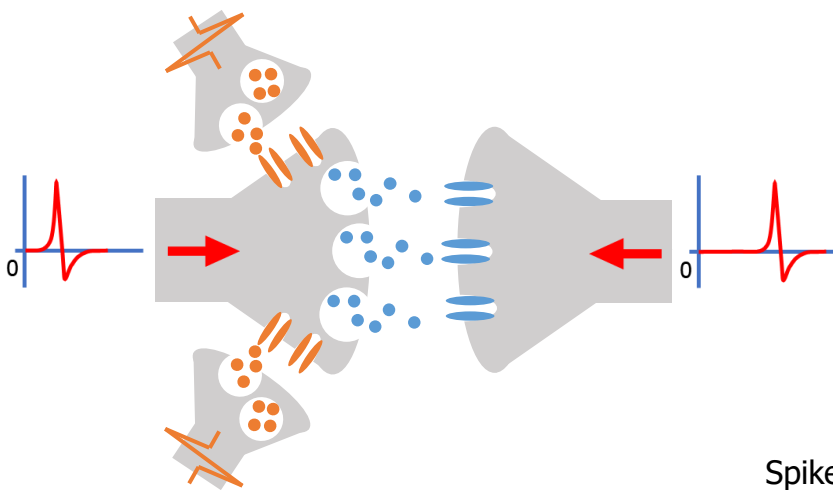
シナプス可塑性：経験や学習によって神経細胞間の結合（シナプス）強度が変化する現象

Spike-timing-dependent plasticity (STDP)

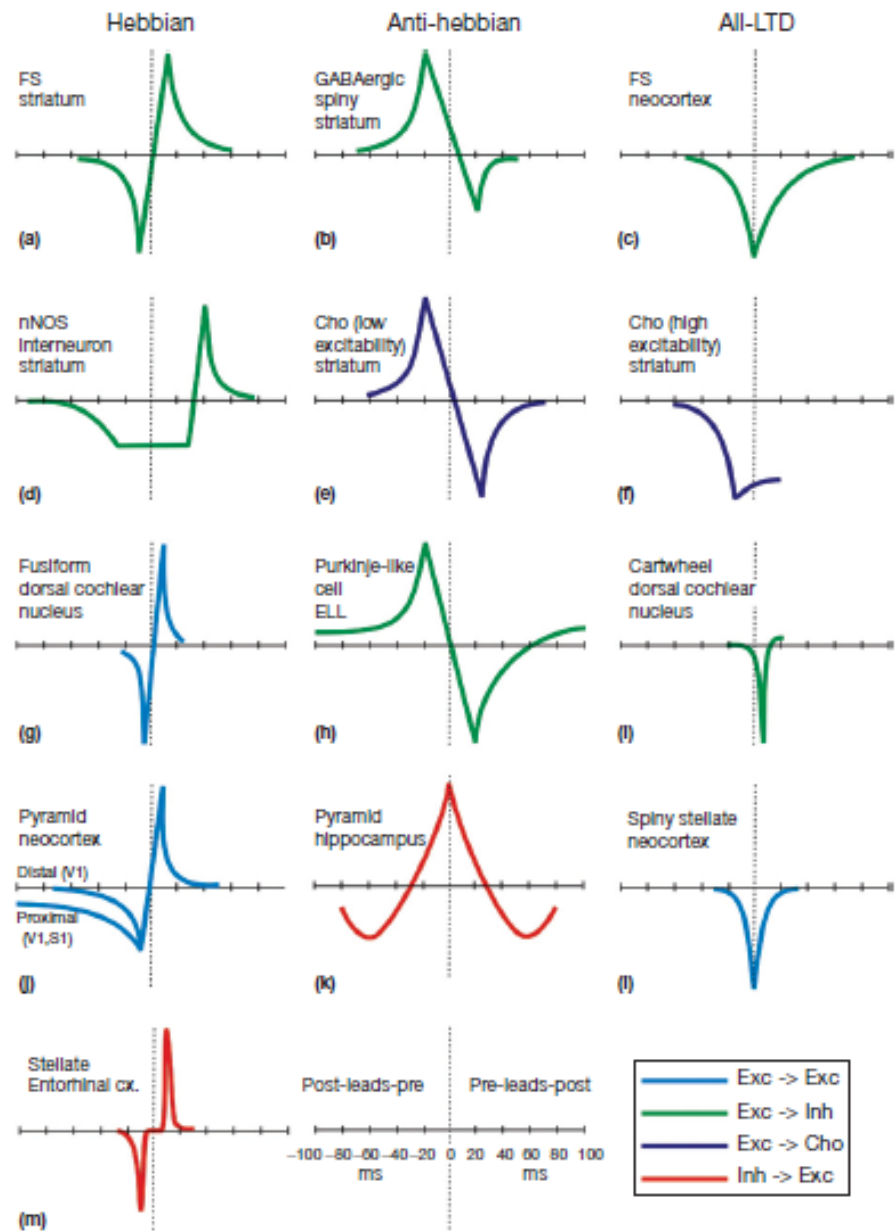
信号入力時刻差に依存するシナプス重み変化



ヘテロシナプス可塑性によるSTDP特性の変調

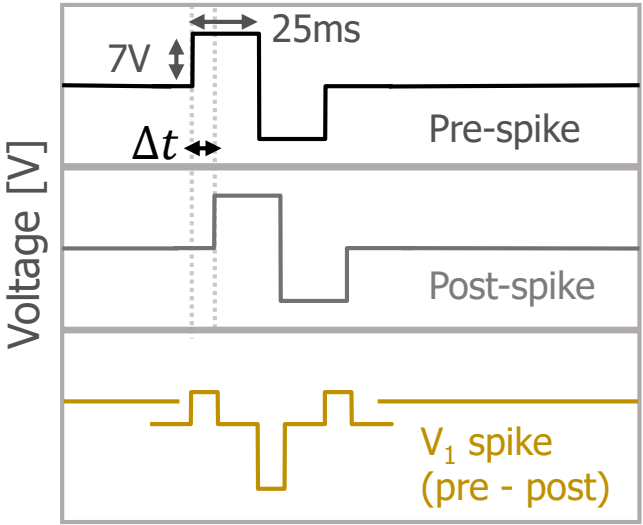


Spike Timing-Dependent Plasticity
D. E. Shulz and D. E. Feldman
From *Comprehensive Developmental Neuroscience Neural Circuit Development and Function in the Healthy and Diseased Brain*

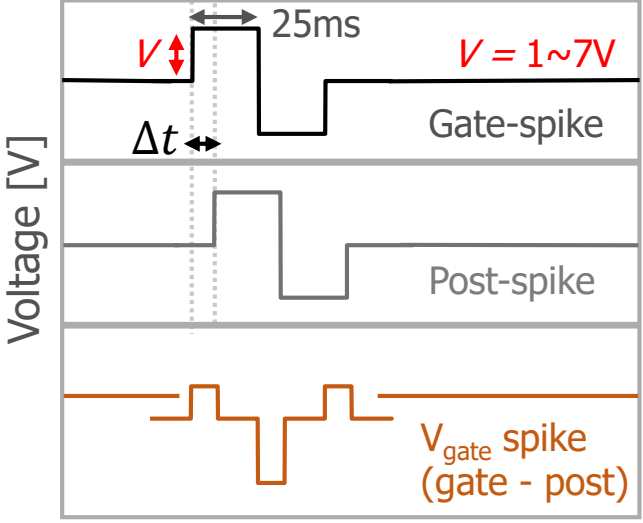


ゲート電圧印加によるSTDP特性の変調

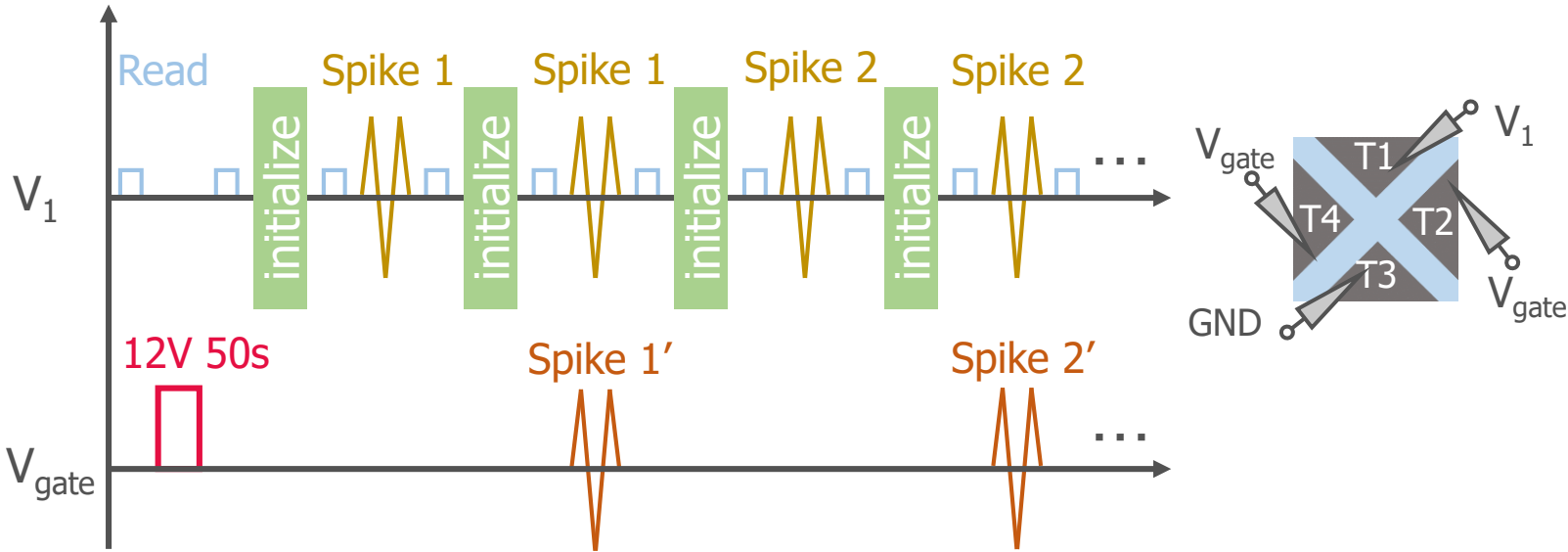
V_1 Spike Pulse



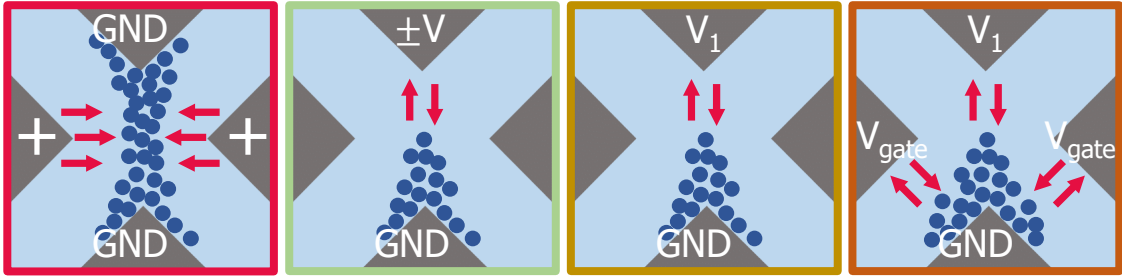
V_{gate} Spike Pulse



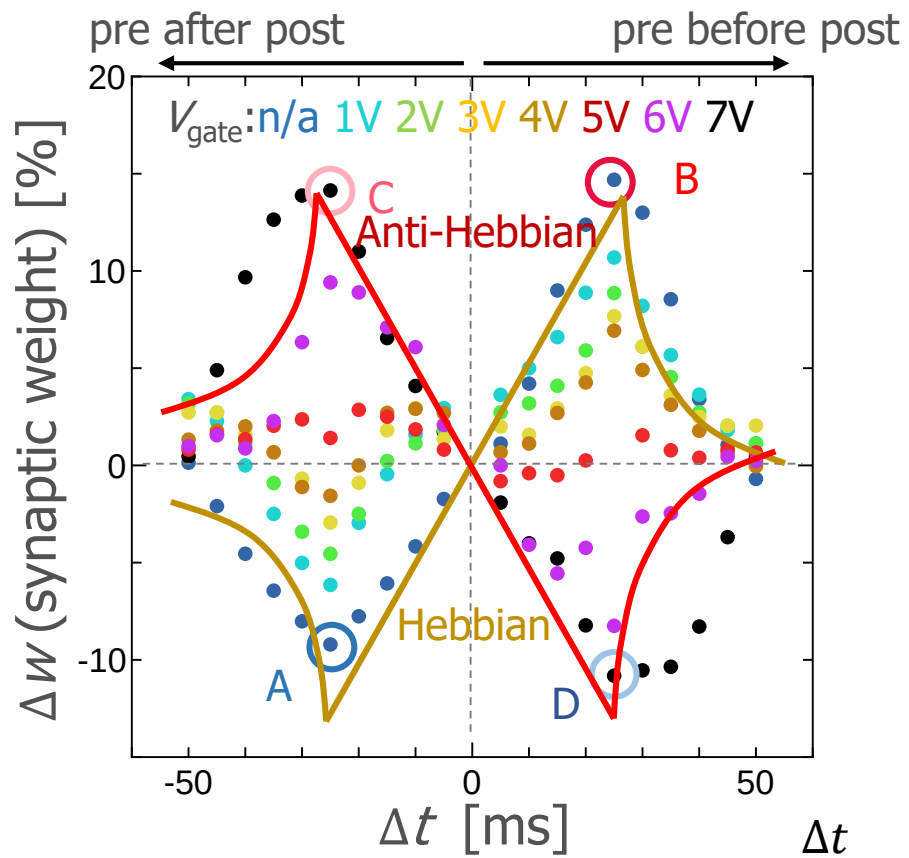
電圧印加プロトコル



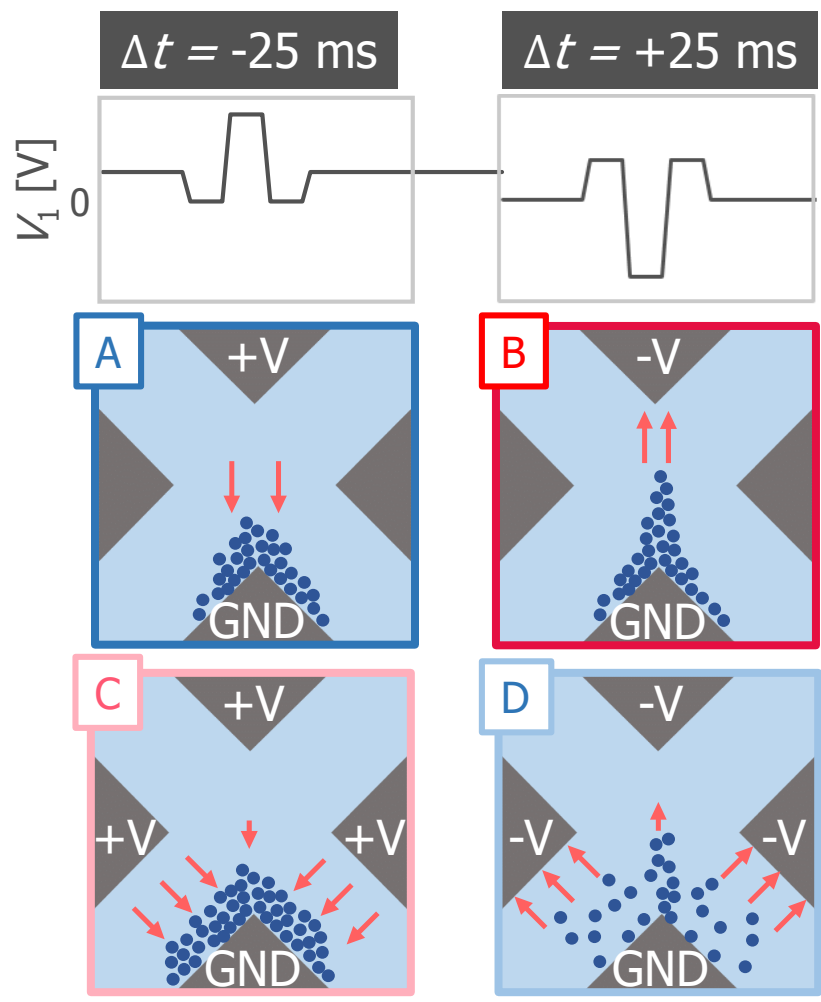
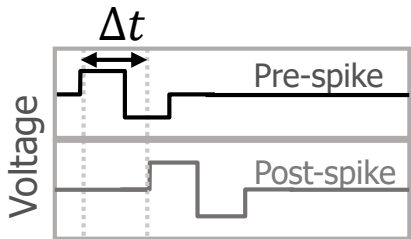
酸素空孔分布



ゲート電圧印加によるSTDP特性の変調



$$\Delta w = \frac{G_{after} - G_{before}}{G_{before}} \times 100 [\%]$$



- 2端子モードでのSTDP特性
- ゲート電圧に依存するSTDP特性

- Silent synapseの実現
- HebbianからAnti-Hebbianモード

日常的に起こる馴化と鋭敏化（感作）

馴化



当初に気になっていた騒音も慣れると
気にならなくなる

鋭敏化



突然大きな音を聞くとこれまで気になら
なかった小さな音でも驚くようになる

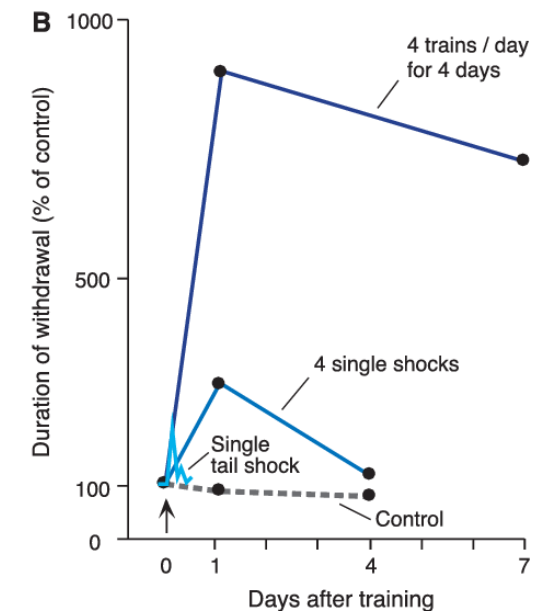
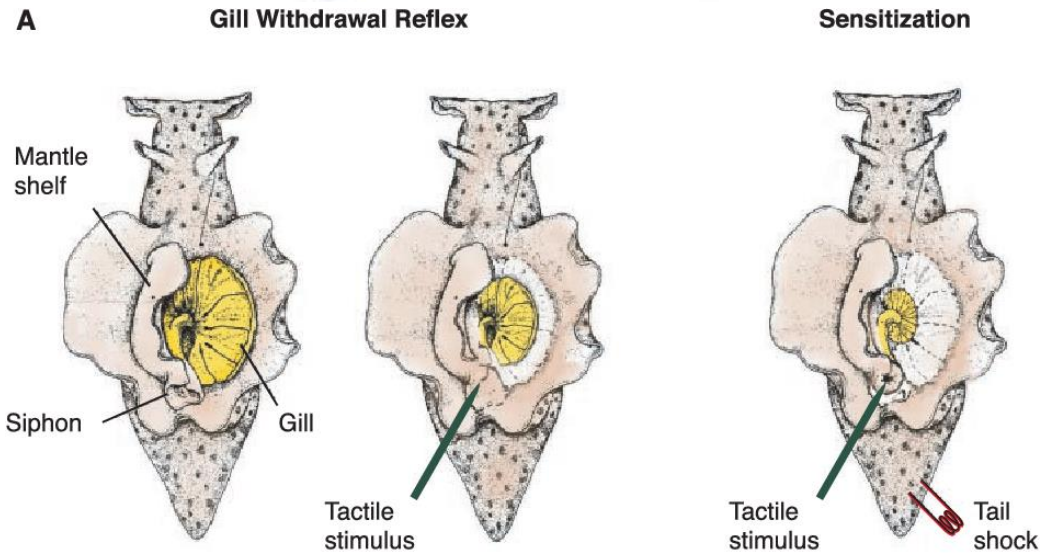
連合／非連合学習の神経科学的アプローチ

Aplysia



Eric Kandel

<https://jp.weforum.org/people/eric-kandel/>



The Molecular Biology of Memory Storage: A Dialog Between Genes and Synapses

E. Kandel, Bioscience Reports, Published 10 August 2004

DOI:10.1007/S10540-005-2742-7

馴化／鋭敏化の神経メカニズム

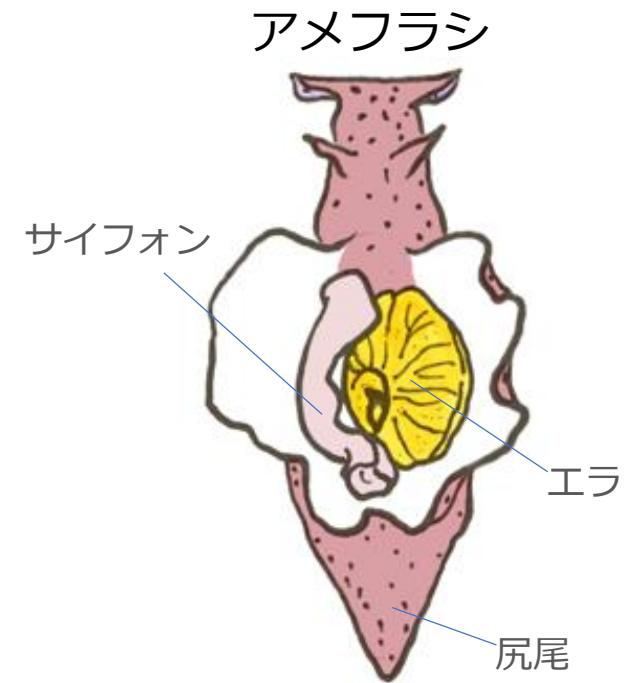
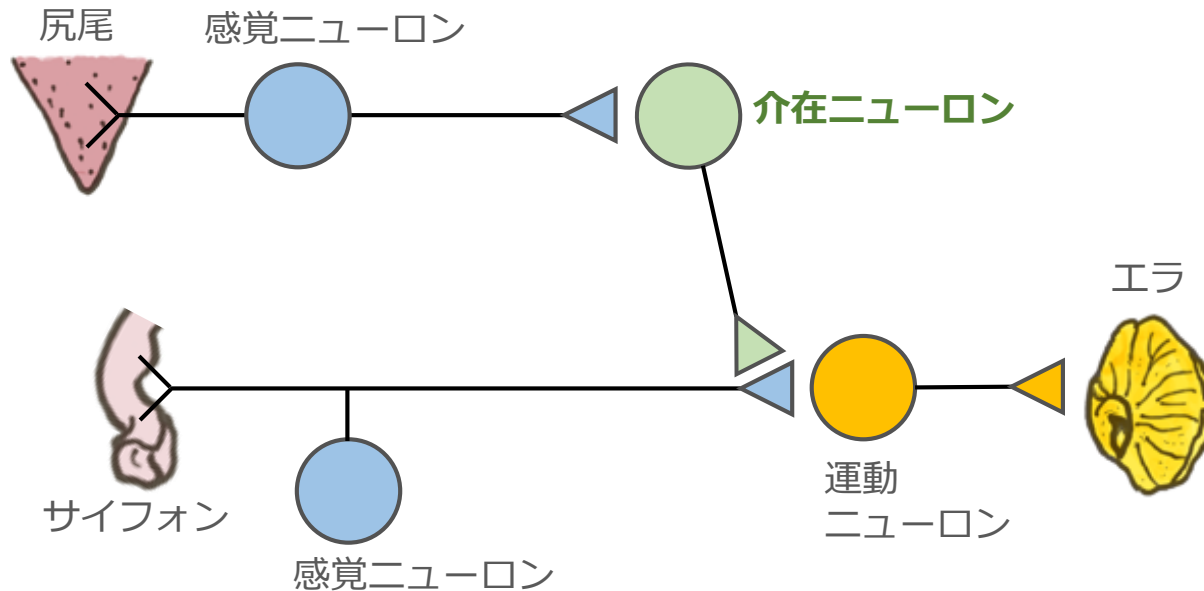
シナプス特性における馴化／鋭敏化

特定の刺激印加・印加プロトコルに対応してシナプス重みが変わる

馴化：不必要な（無害な）刺激を無視する非連合学習

鋭敏化：当初弱い反応の刺激に対して感応度が増加する連合学習

馴化



馴化／鋭敏化の4端子メモリスタ素子への実装

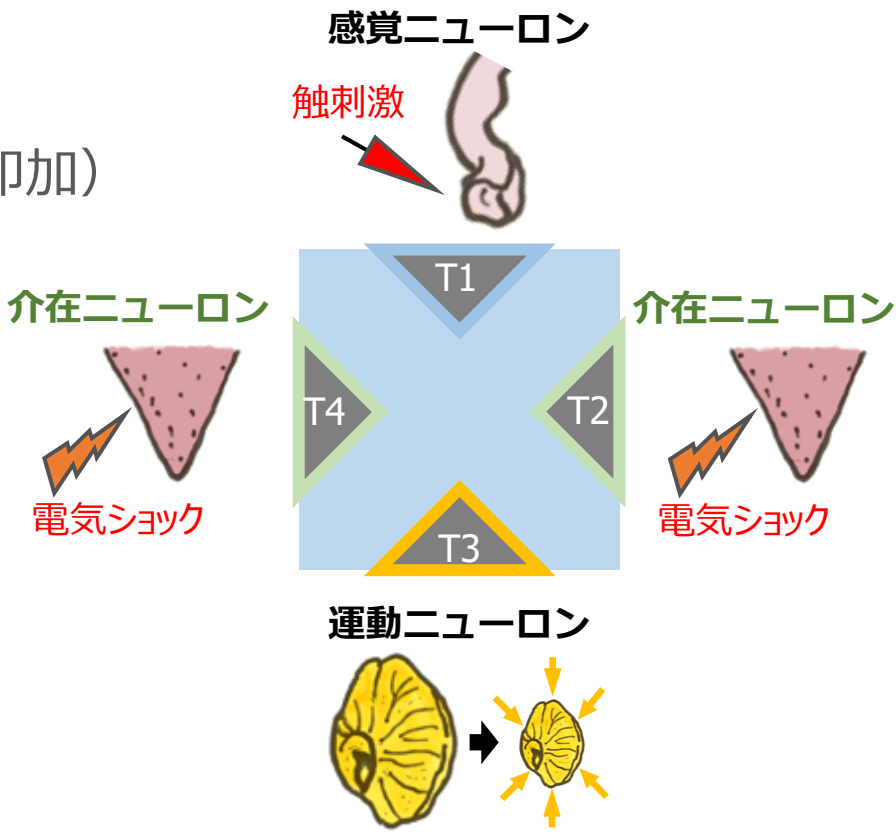
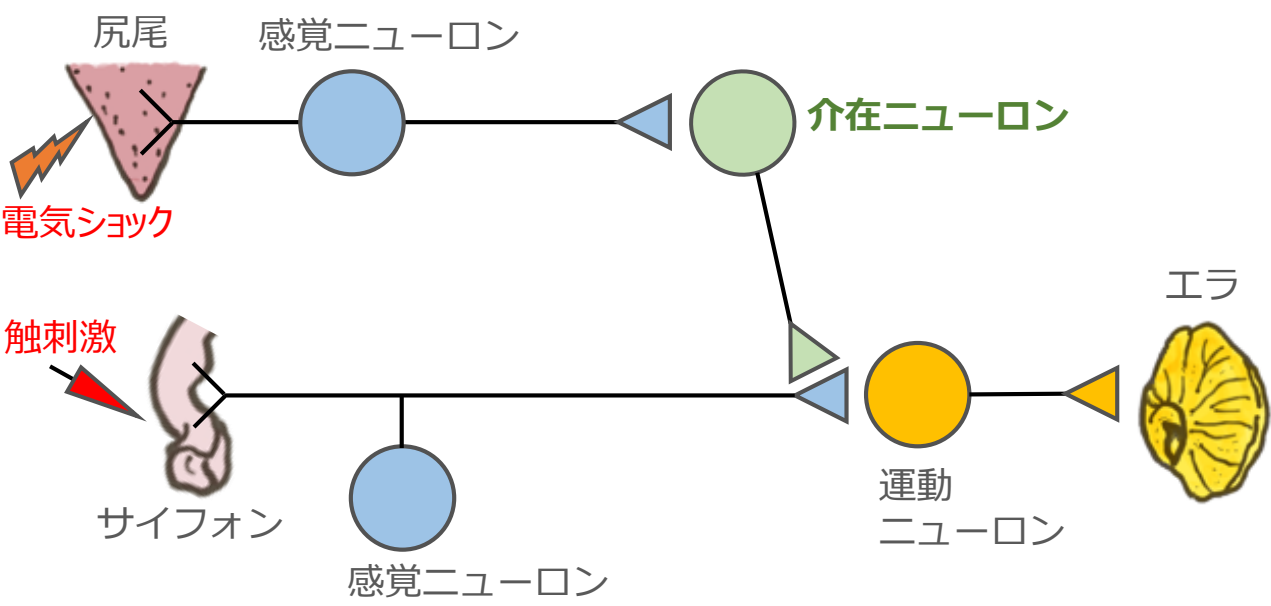
シナプス特性における馴化／鋭敏化

馴化

繰り返し電圧印加に伴うコンダクタンス変化度の減少

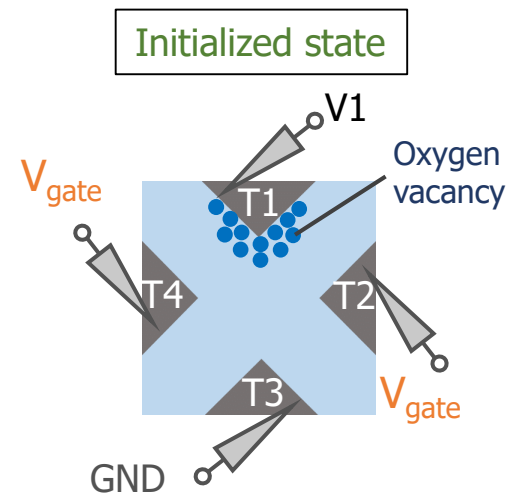
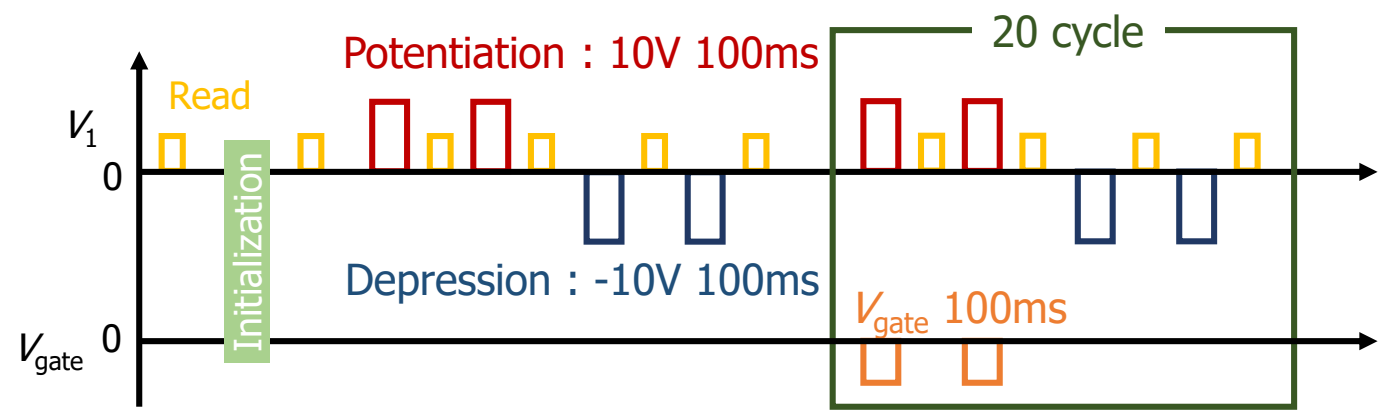
鋭敏化

介在ニューロンに見立てたゲート電極への強刺激（電圧印加）によるコンダクタンス変化度の増大

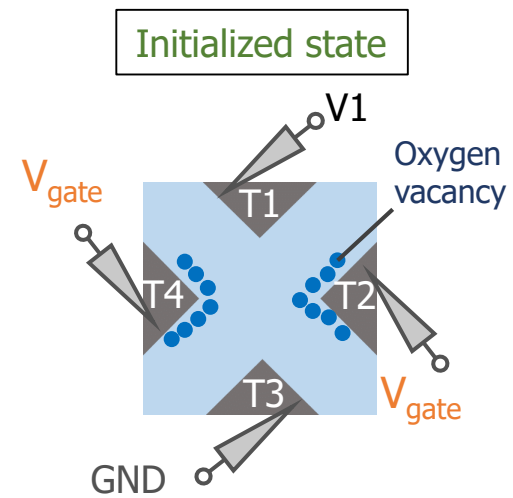
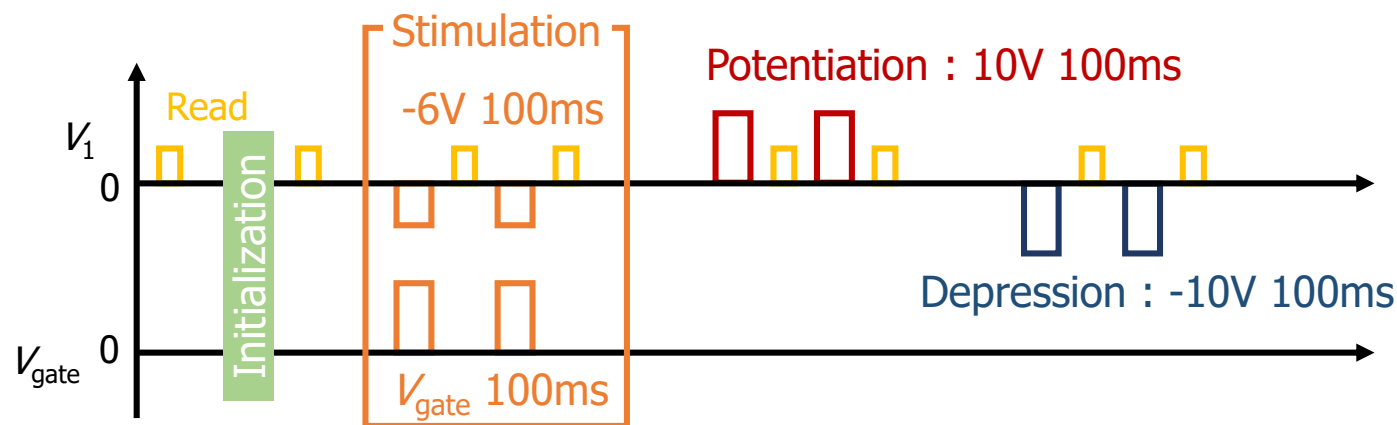


馴化／鋭敏化の電圧印加プロトコル

馴化

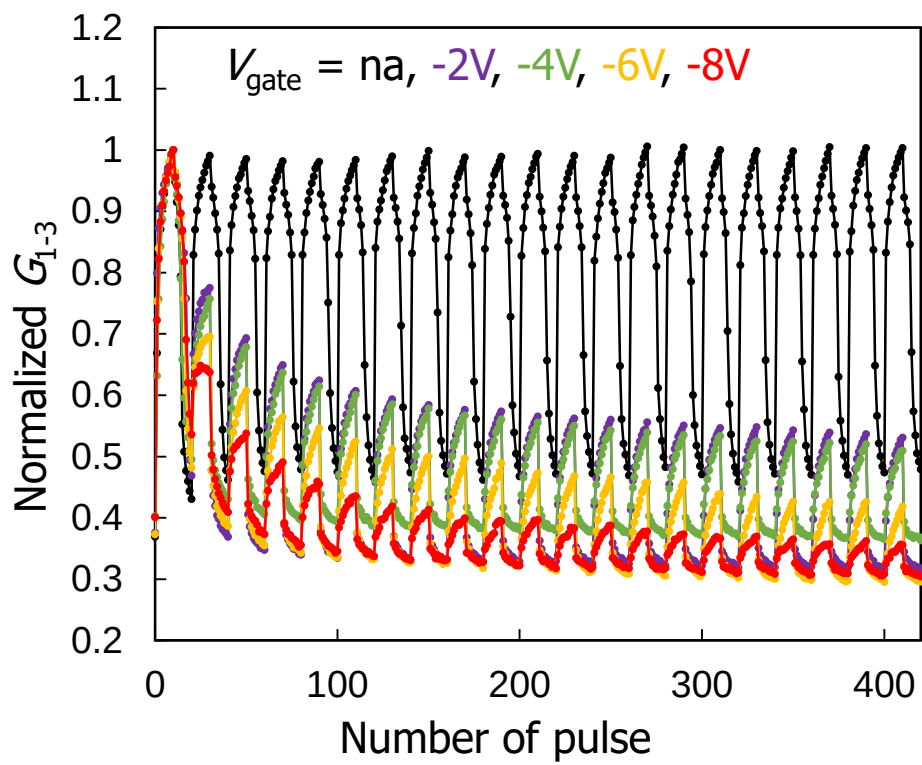


鋭敏化

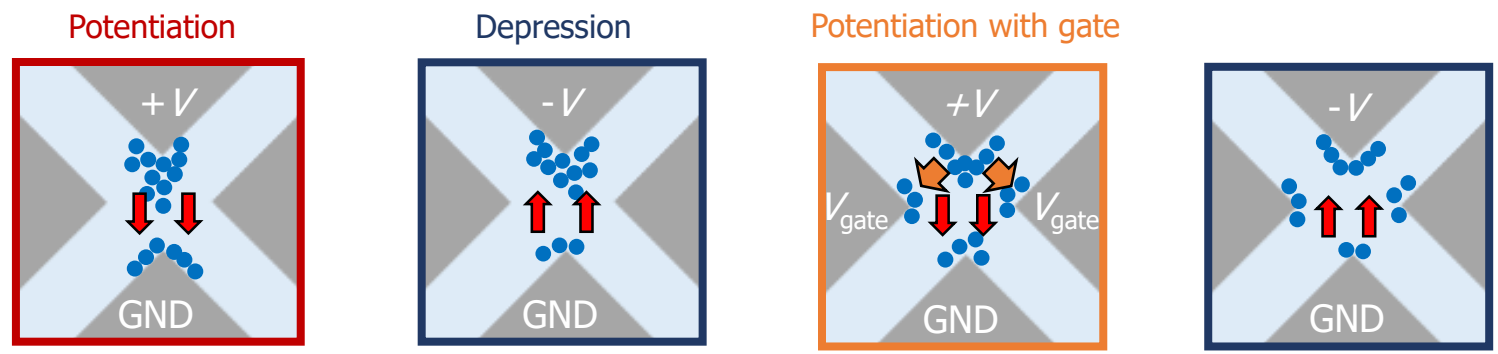
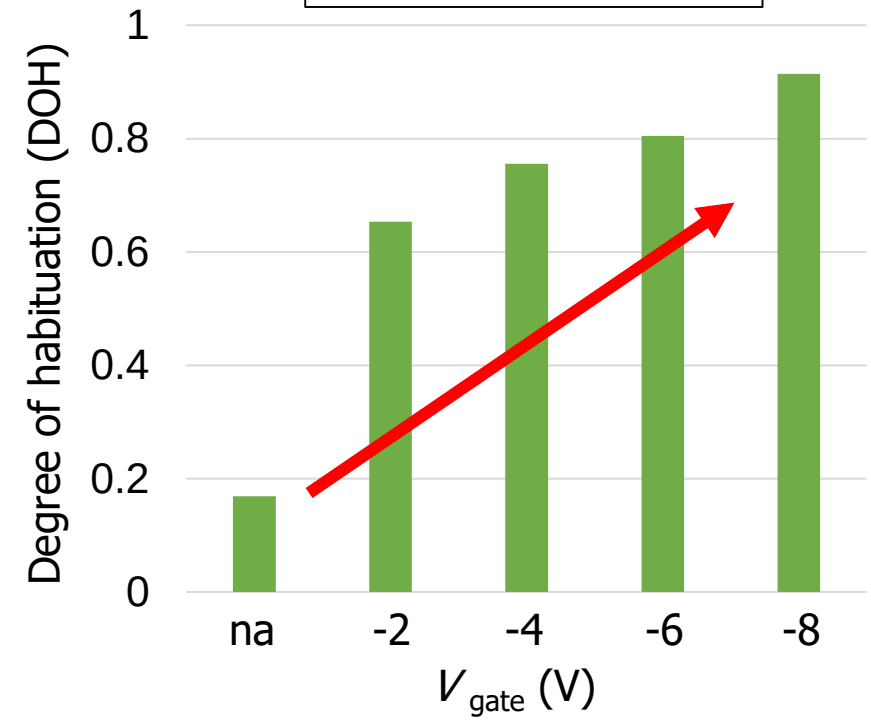


馴化を模倣するコンダクタンス変化

Modulation of conductance G_{1-3}

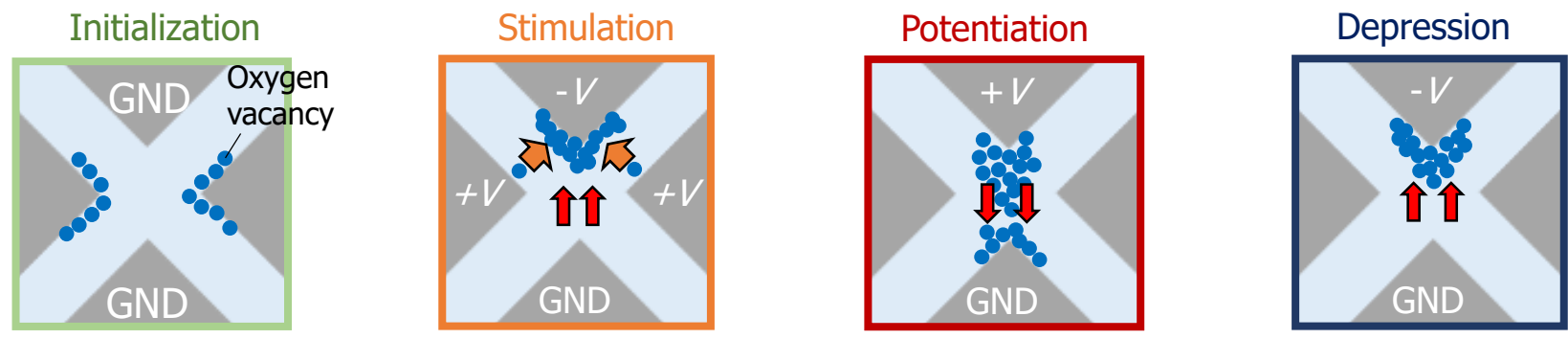
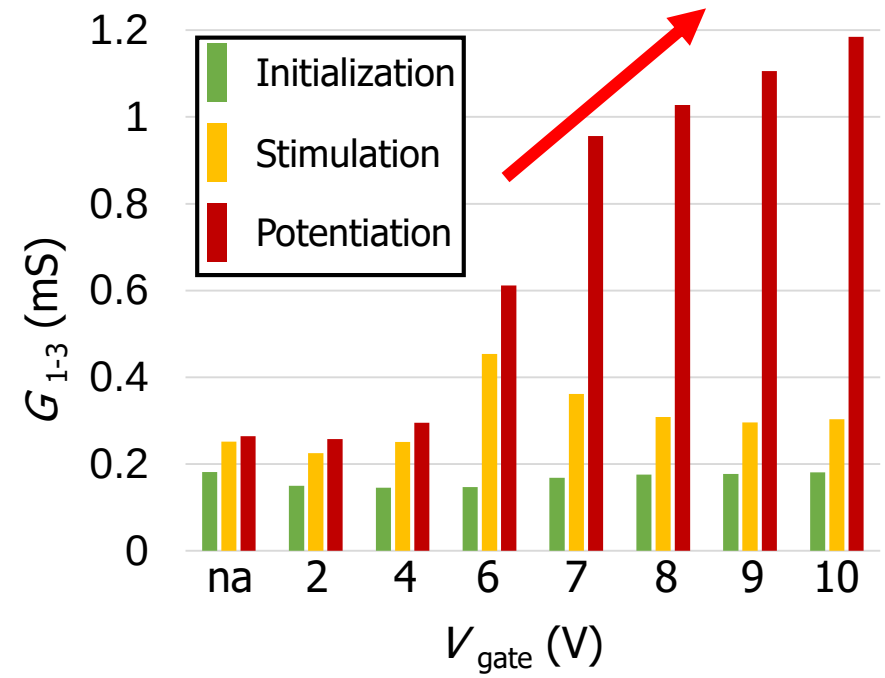
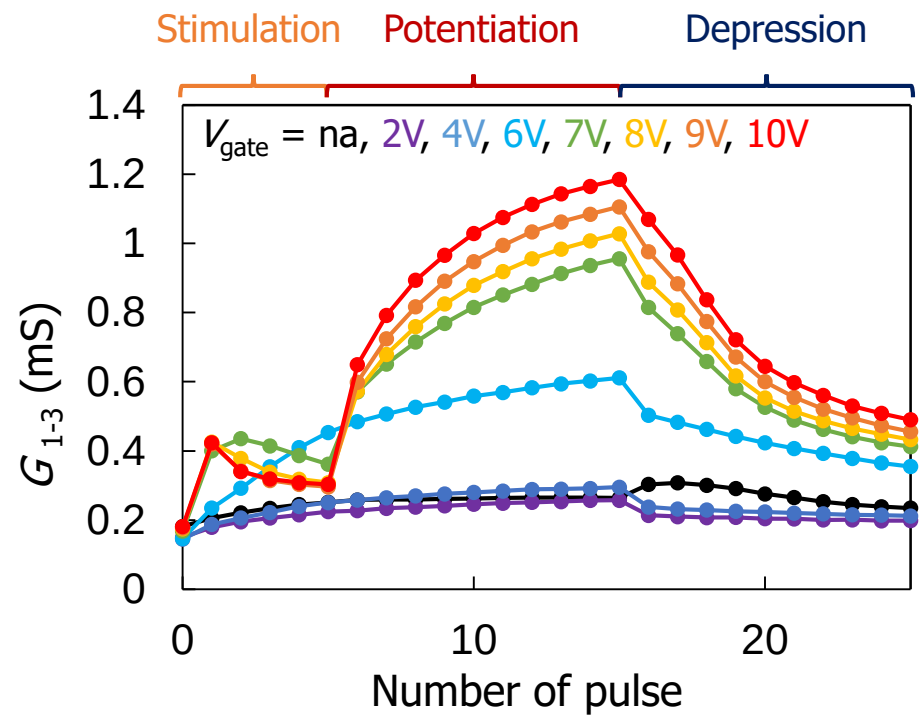


$$DOH = 1 - \frac{\Delta G_{1-3}^{last\ pot.}}{\Delta G_{1-3}^{first\ pot.}}$$



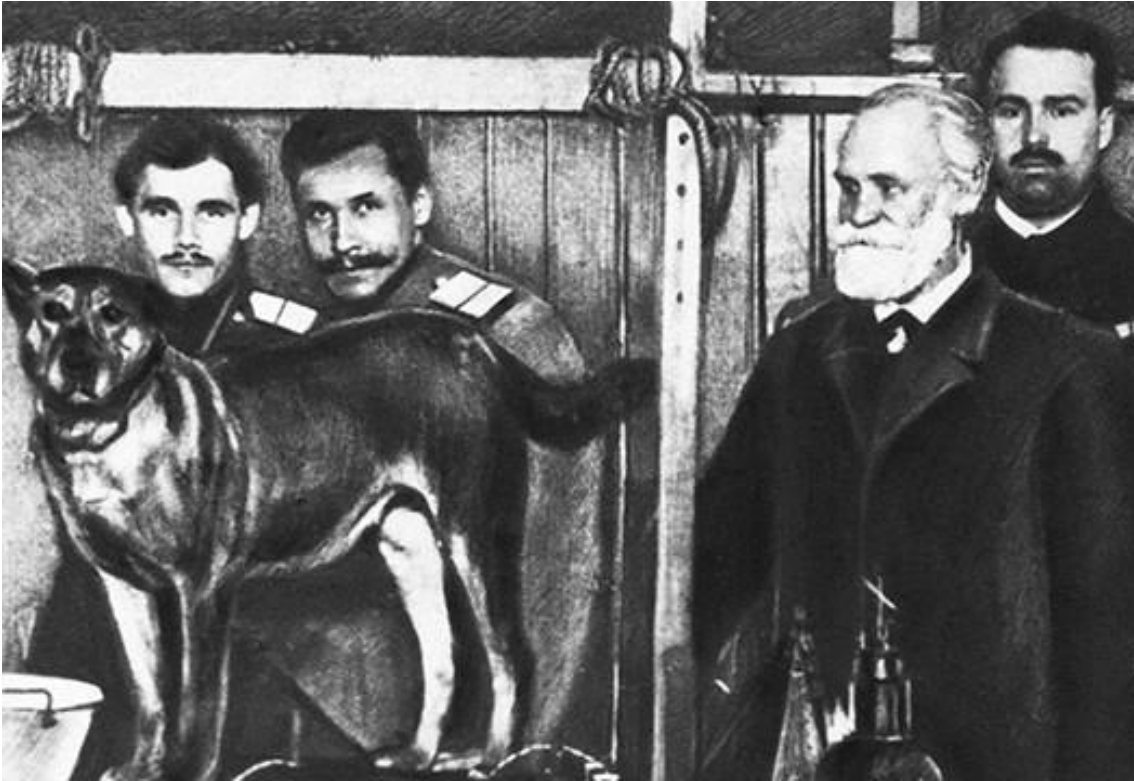
鋭敏化を模倣するコンダクタンス変化

Modulation of conductance G_{1-3}



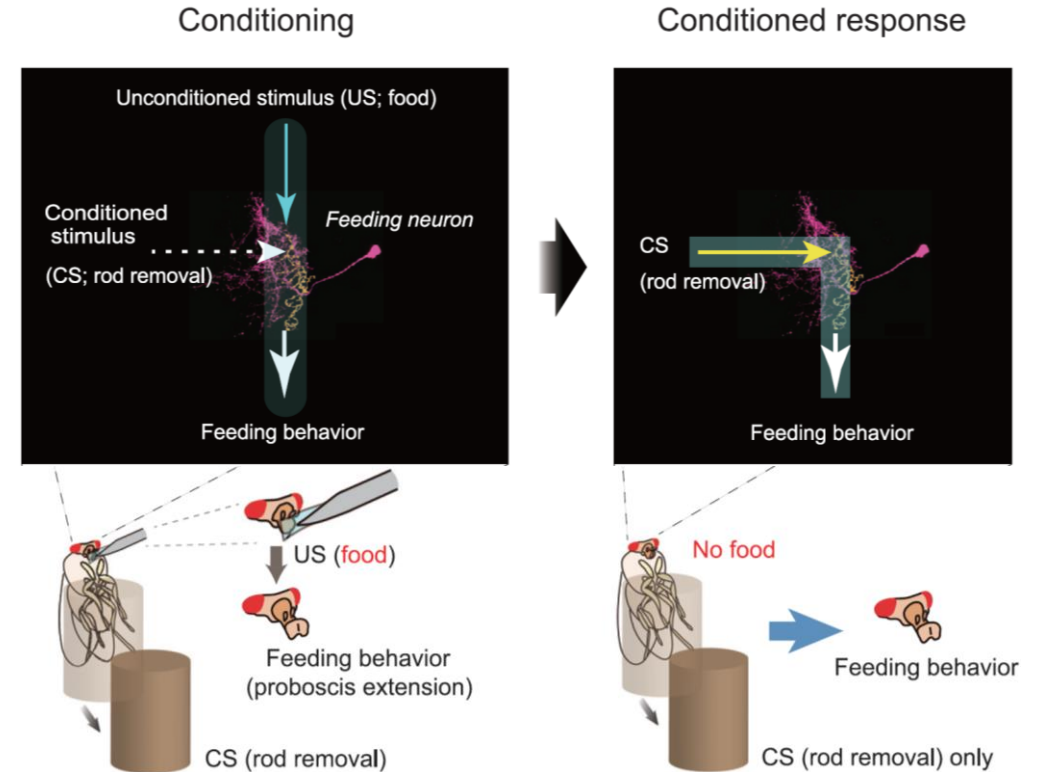
パブロフ型条件付け：連合学習

Ivan Pavlov : パブロフの犬



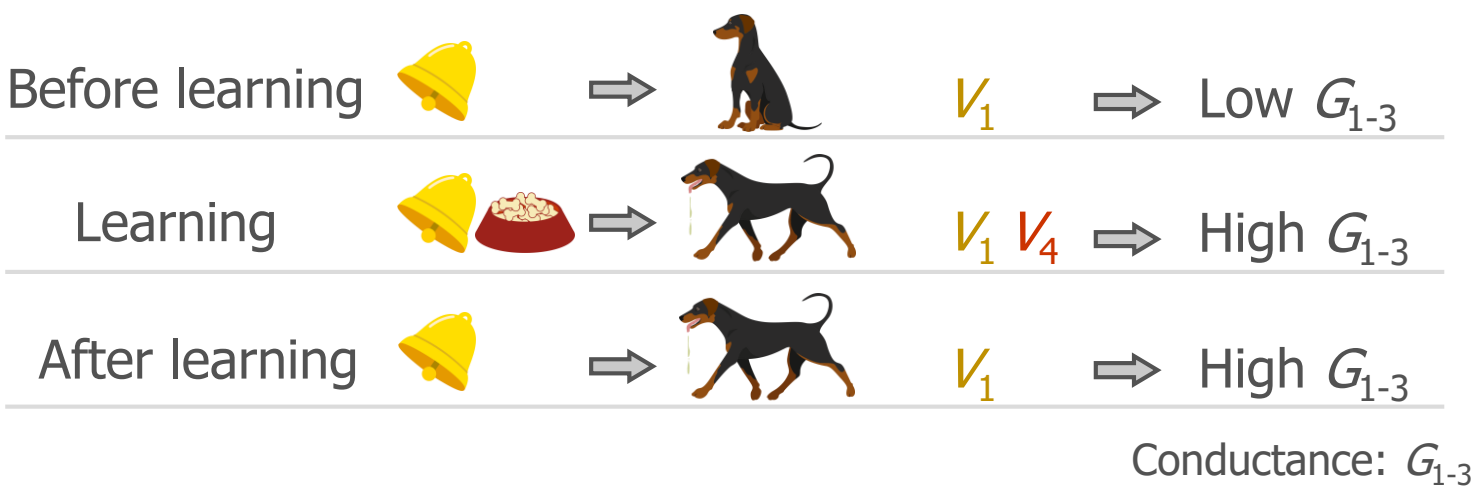
<https://historyofyesterday.com/pavlovs-dog-the-experiment-that-revolutionized-psychology-d4bfdc343c73>

ショウジョウバエの脳における摂食命令
ニューロンのパブロフ型条件付け



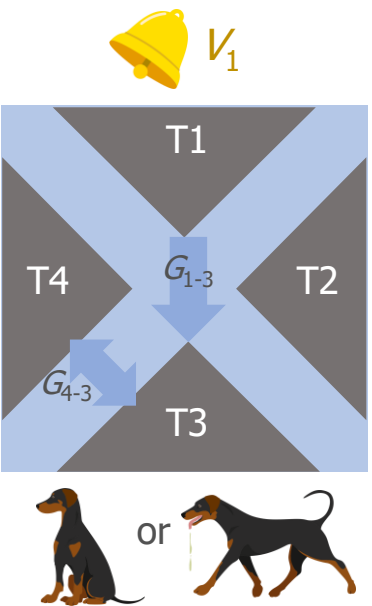
Alteration in information flow through a pair of feeding command neurons underlies a form of Pavlovian conditioning in the *Drosophila* brain
A. Sakurai et al.
Current Biology 31, 4163 (2021).

パブロフ型条件付けの4端子メモ리스タへの実装

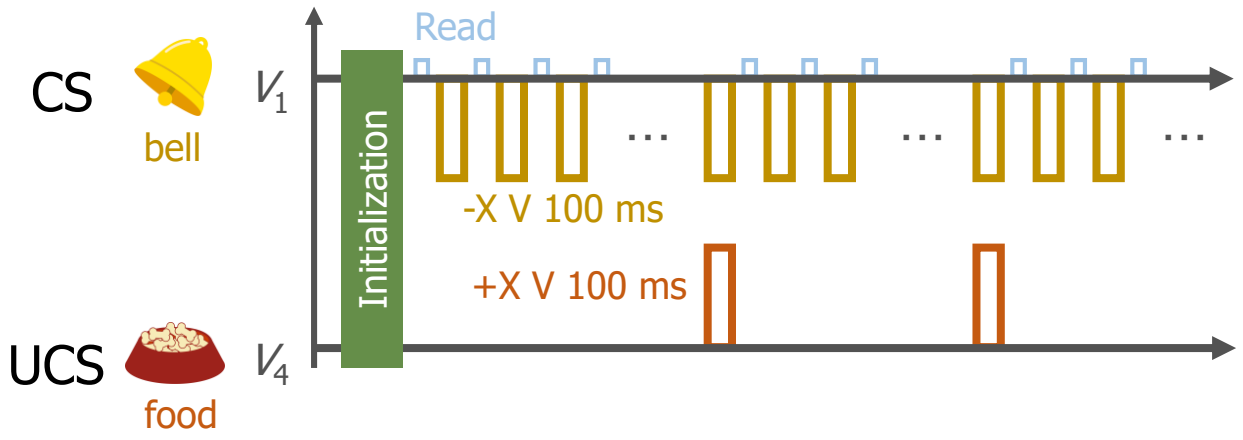
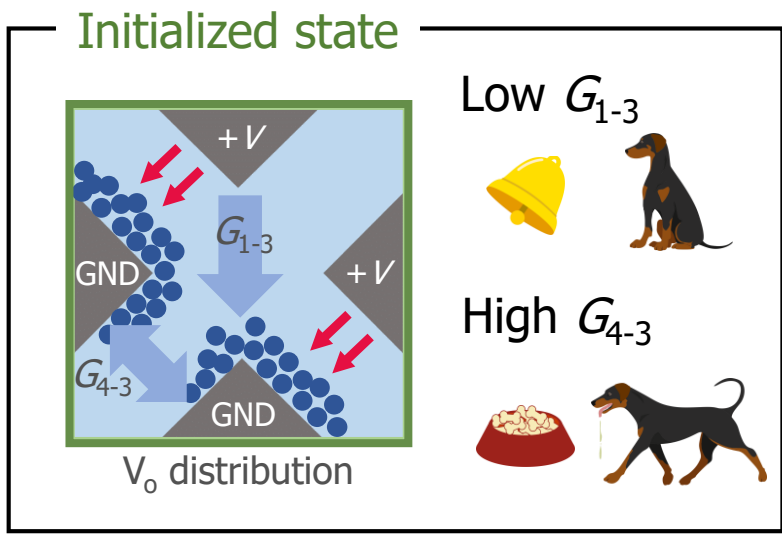


Conditional stimulus (CS)

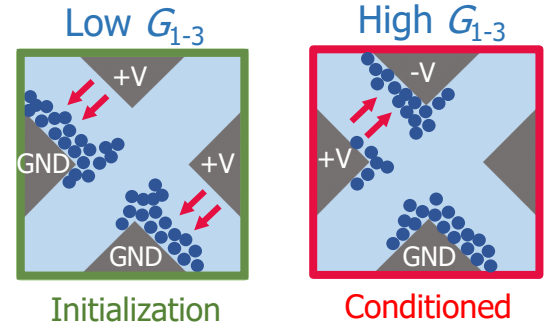
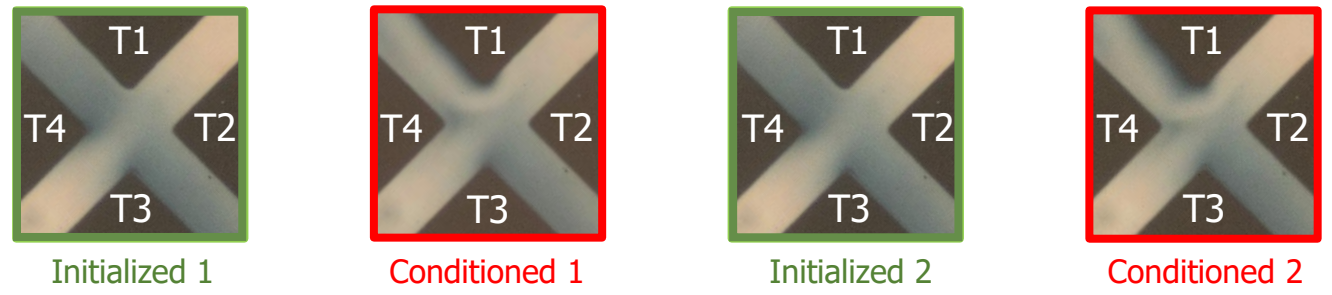
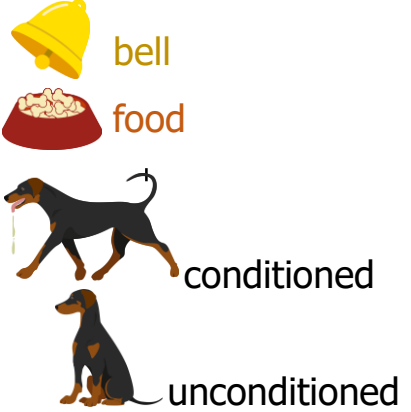
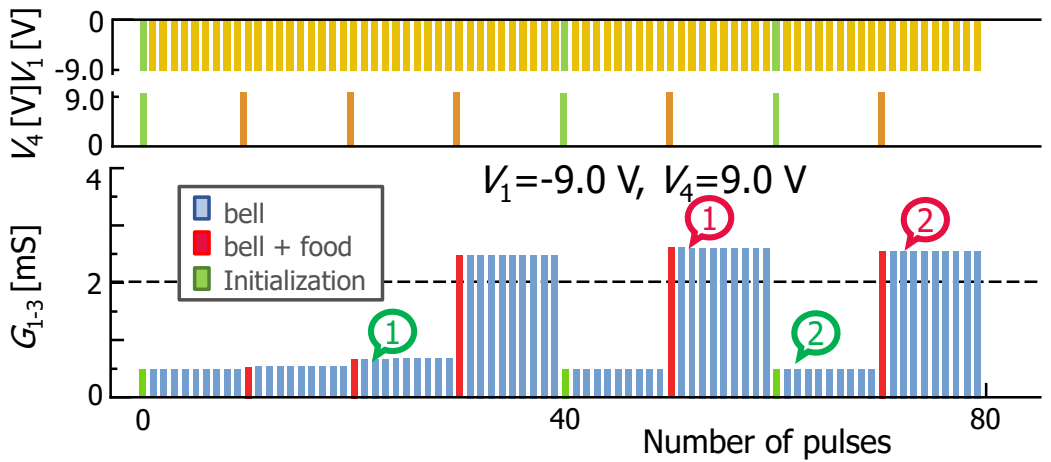
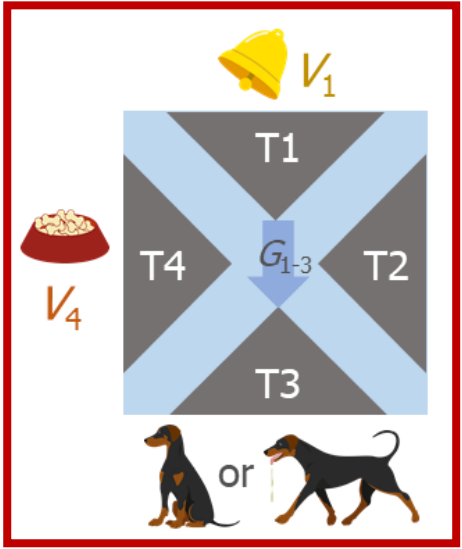
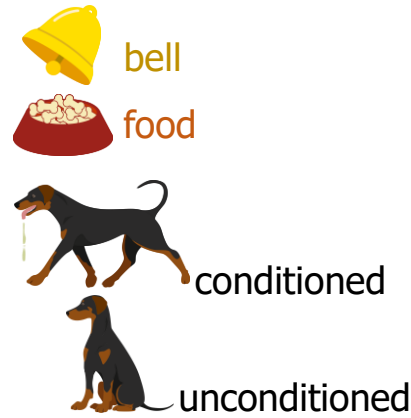
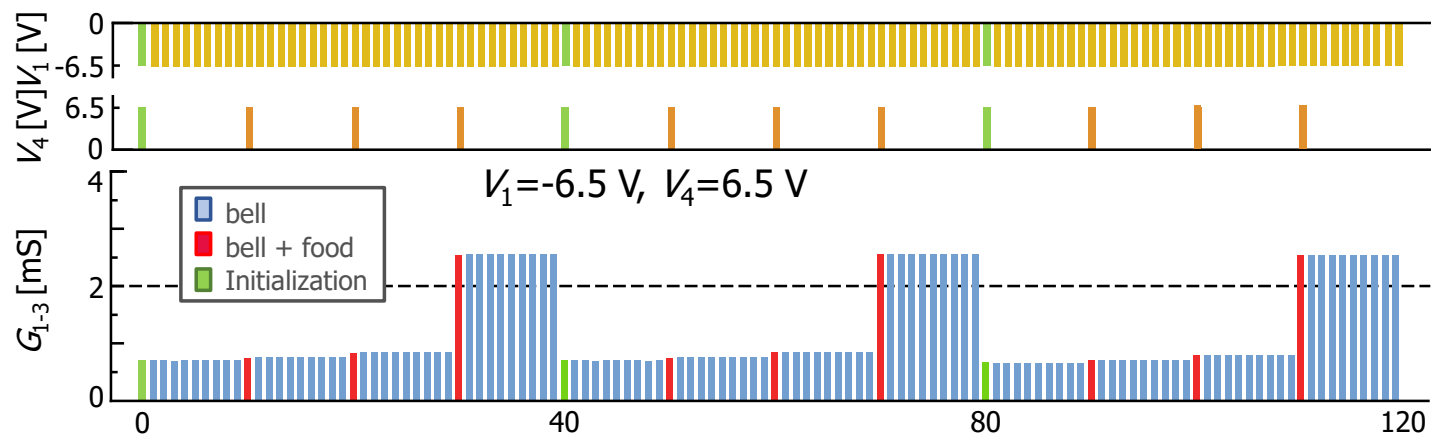
Unconditional stimulus (UCS)  V_4



Voltage application protocol



パブロフ型条件付けの4端子メモ리스タへの実装



Optical microscope images in the device active region

まとめ

- 半導体メモリデバイスは、IoT, AI, ロボティクス等の技術の中核に位置し、未来社会の構築に不可欠である
- データ処理数の増大とともにフォン・ノイマンボトルネックが問題となる
- 深層ニューラルネットワークは現在のAIの核である
- イン・メモリコンピューティングを実現するクロスバーアーキテクチャは積和演算に関する諸問題を解決する
- メモリスタはANNにおけるシナプス機能を発揮する
- ANNを発展的に再構築するうえで、神経系の高次機能（変調・連合・相関）を発現する人工シナプス素子が不可欠となる

謝辞 本研究はJSPS科研費（JP17K18881, JP20H00248, JP21K18723）の助成を得て行われた