



2024年10月28日
サイエンティフィックシステム研究会

高性能計算と人工知能で加速する脳シミュレーション

理化学研究所 計算科学研究センター
高性能人工知能システム研究チーム

五十嵐潤

略歴と日本のスパコンの性能推移

(FLOPS: FLoating-point OPeration per Second
1秒間当たりの浮動小数点演算の回数)

五十嵐 潤

- 九州工業大学 生命体工学研究科(2002–2007年): 嗅内皮質の研究、プレイステーション3による脳シミュレーション
- ホンダ・リサーチ・インスティチュート・ジャパン(2008年): GPUによる基底核のリアルタイムシミュレーション
- 理化学研究所 (2009–2013年): 京での脳のシミュレーション
- 沖縄科学技術大学院大学(2013年–2016): 京によるパーキンソン病状態の脳のシミュレーション
- 理化学研究所 (2016–現在): 富岳による全脳シミュレーションを目指し研究活動中

スパコン

地球シミュレータ:
41 Tera FLOPS

Tera: 10^{12}

> x100

京:
11 Peta FLOPS

Peta: 10^{15}

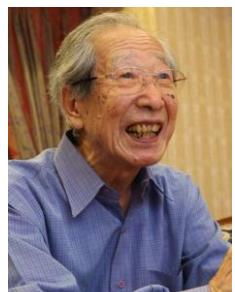
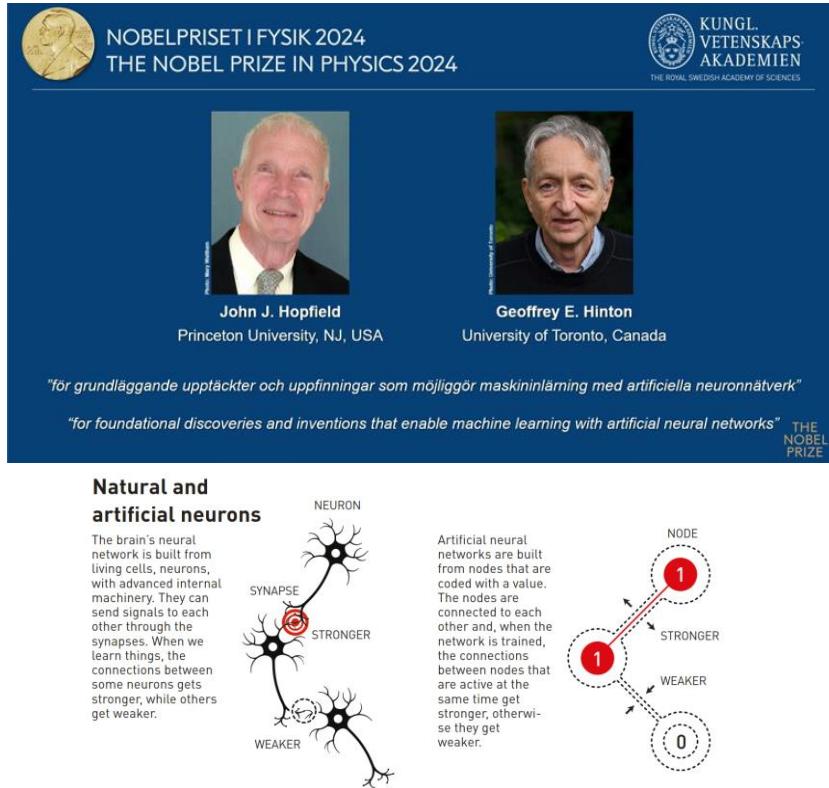
~x40

富岳:
~0.5 Exa FLOPS
Exa: 10^{18}

本日の内容

0. 自己紹介、本日の内容
1. 脳について
2. HPCを用いた脳シミュレーションの現在
3. 神経科学の技術動向とAI-for-Neuroscience

2024年ノーベル物理学賞にみられるAIと脳の関係



甘利俊一氏



福島邦彦氏

- 2024年のノーベル物理学賞:「人工ニューラルネットワークによる機械学習を可能にした基礎的発見と発明に対する業績」で、Geoffrey Hinton氏(ボルツマンマシン)、John Hopfield氏(連想記憶)が受賞

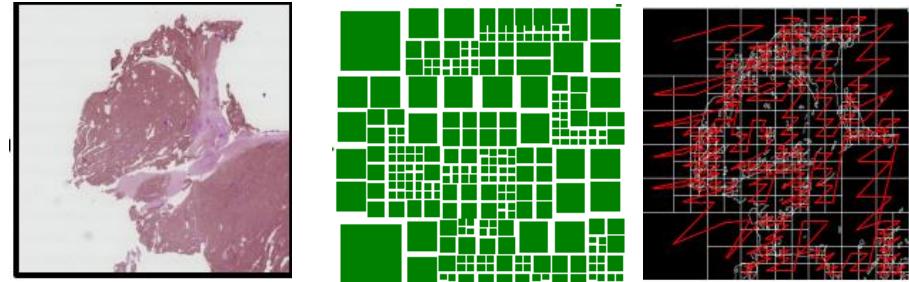
受賞の説明のペーパーで甘利氏、福島氏の研究を引用

- 甘利俊一氏:理化学研究所脳科学総合研究センター長。しきい値素子からなる自己組織化ネットワークの学習
- 福島邦彦氏:日本神経回路学会初代会長・名誉会員。視覚野からヒントを得たネオコグニトロン
- 現在のAIで用いられる人工ニューラルネットワークと脳・神経科学の密接な関係

脳・神経科学とAIの相互作用による発展

◎ AI for Neuroscience

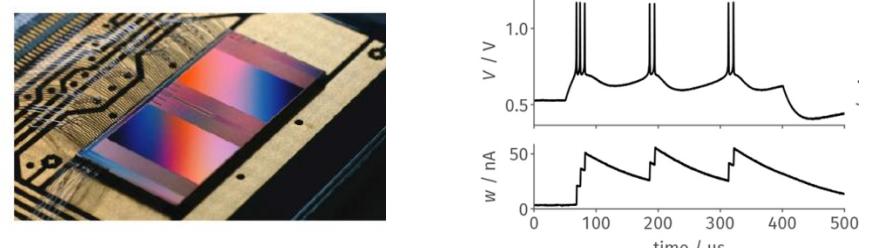
- ・脳のデータの解析、セグメンテーション、時空間活動、データ同化、コード生成



Zhang et al., 2024

◎ Neuroscience for AI

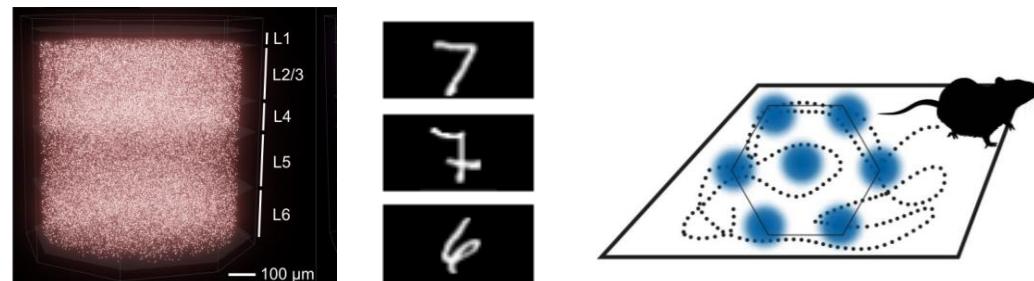
- ・脳の特徴を模倣、AIに導入。
- ・Neuromorphic computing (低消費電力)



Pehle et al., 2022

◎ Fusion of AI and Neuroscience

- ・視覚野脳モデル+バックプロパゲーションで画像域別
- ・LSTMによる空間学習で嗅内皮質のグリッド状の場所表現を獲得



Banino et al., 2018

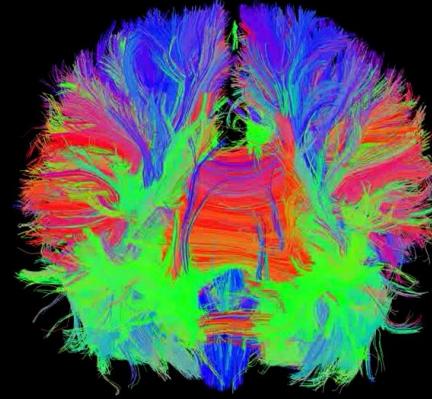
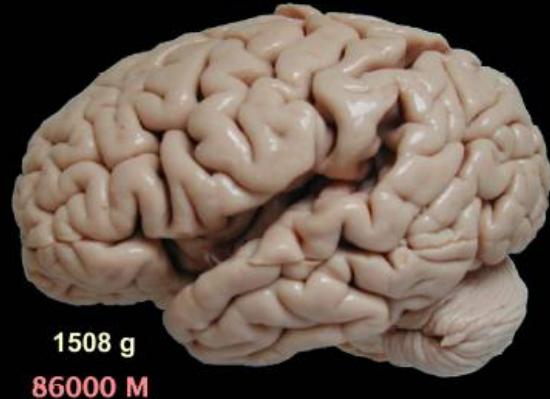
Scherr and Maass, 2022

本日の講演の趣旨

- ・2023年度末に、AIを用いた科学、“AI-for-Science”の技術動向に関して、HPCIコンソーシアムで調査を行い、脳・神経科学部門のロードマップの執筆を行った。
- ・また、最近、富岳成果創出加速課題、AMED脳統合プログラムなど、さまざまな異なる研究プロジェクトで、AIを用いた取り組みを提案することになり、その可能性について検討している。
- ・その観点から、AIとHPCを利用した脳シミュレーションの可能性について見えてきたことについてお話しする。

1. 脳について

膨大な要素からなる脳と大規模言語モデル(LLM)



人間の脳にある神経細胞と結合数
約860億個、約100兆本

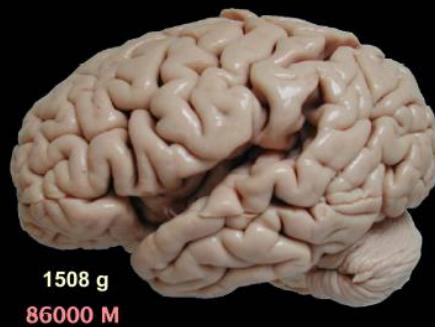
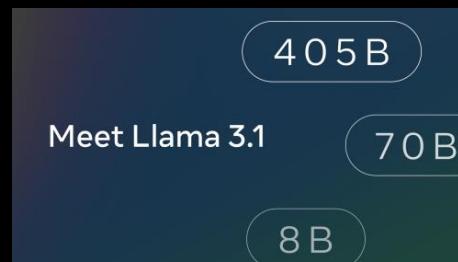
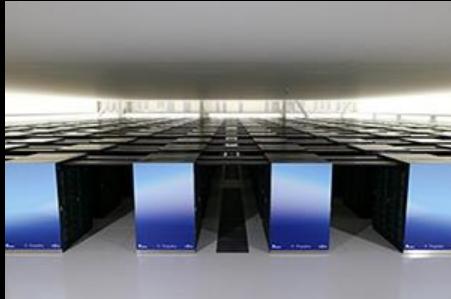
出典 脳の写真 Herculano-Houzel, 2009, Dynamic connectome Lab



LLMの推定パラメータ数
OpenAI GPT-4: 1.8兆 (推定)
Meta Llama-3.1-405B: 4050億

脳はこの世で一番複雑な構造物といわれているが、LLMのパラメータ数
(神経細胞数や結合数)が近づきつつある。

脳は超低消費電力



富岳 (TOP500 4位): 28.3メガワット

Meta Llama 3.1-405B (Llama team, 2024)

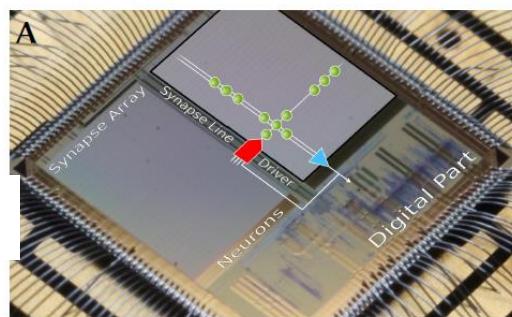
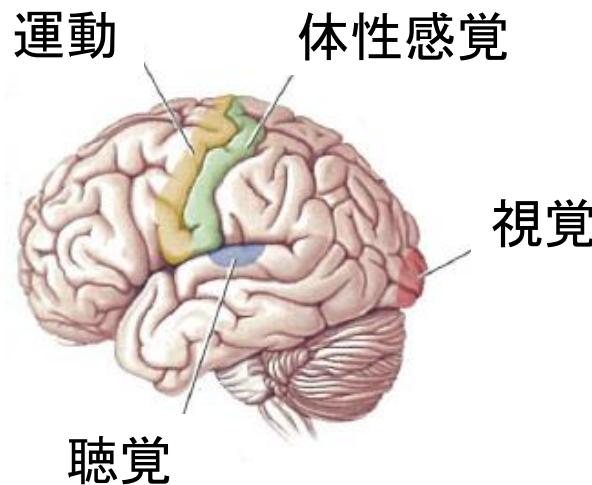
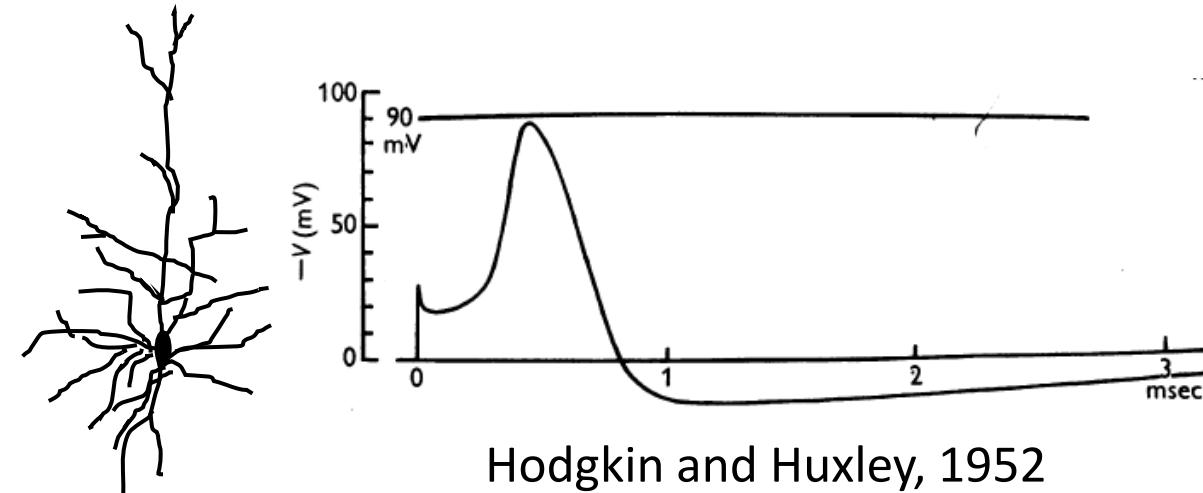
訓練 (16,000 H100): 15メガワット程度

推論 (16 H100): 15キロワット程度

人間の脳 : 約10ワット

スペコンやLLMに比べ、脳の消費エネルギーは 10^3 - 10^6 分の1で、エネルギー効率は非常に優れている。

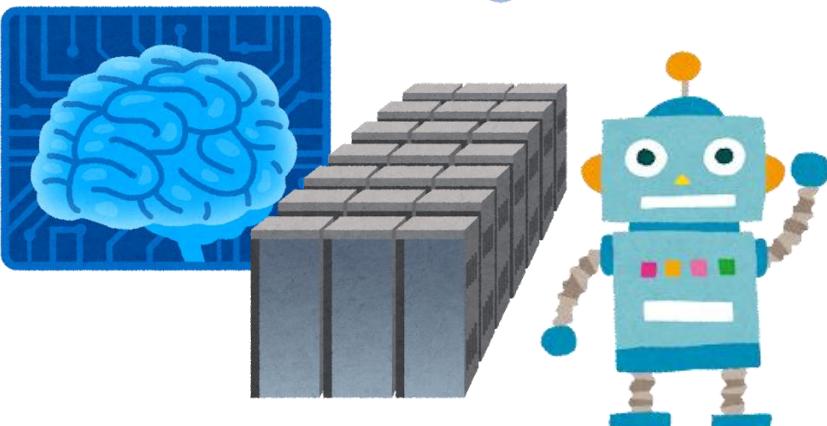
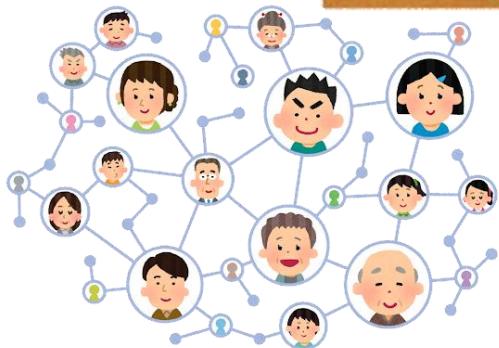
なぜ脳が省エネなのか



Petrovici et al., 2017

- 低電圧(100mV)、低周波数駆動(>1KHz)
- イベント駆動型の信号発生
- 発火の2値表現
- 演算部とメモリ部の混載構造
- 3次元実装
- 機能別専用回路
- 非同期処理
- 確率的神経素子を用いる確率的情報処理
- 液体によるエネルギー供給

脳の情報を処理する仕組みはまだよくわかっていない



Q. 脳の仕組みを理解する必要があるのか？

A. 医療、教育、社会学、AI、計算機、ロボットの発展、我々自身の理解という哲学的追求のため必要

Q. なにがわかっていない？なぜわからない？

A. 神経細胞がどのように相互作用して情報処理するかが未解明。複雑かつ数が多くて観測しきれない(LLMも内部機構は未解明で、観測できるが内部解析が不十分)

★シミュレーションは、コンピュータ上で膨大な数の神経細胞を造り出し、どんな状態でも、何度でも、試せる!

2. HPCを用いた脳シミュレーションの現在

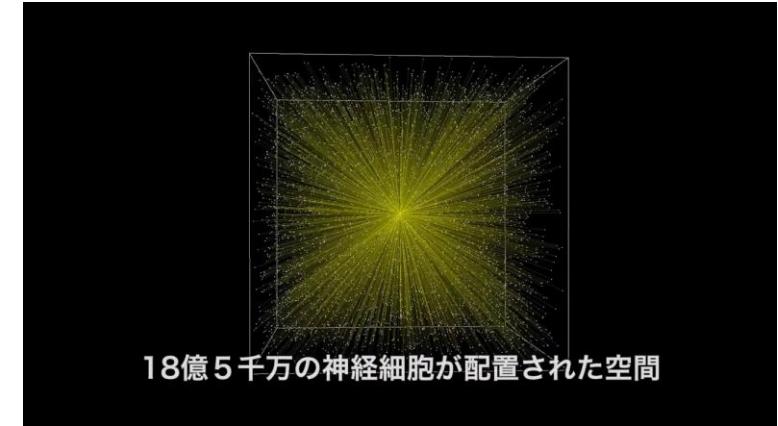
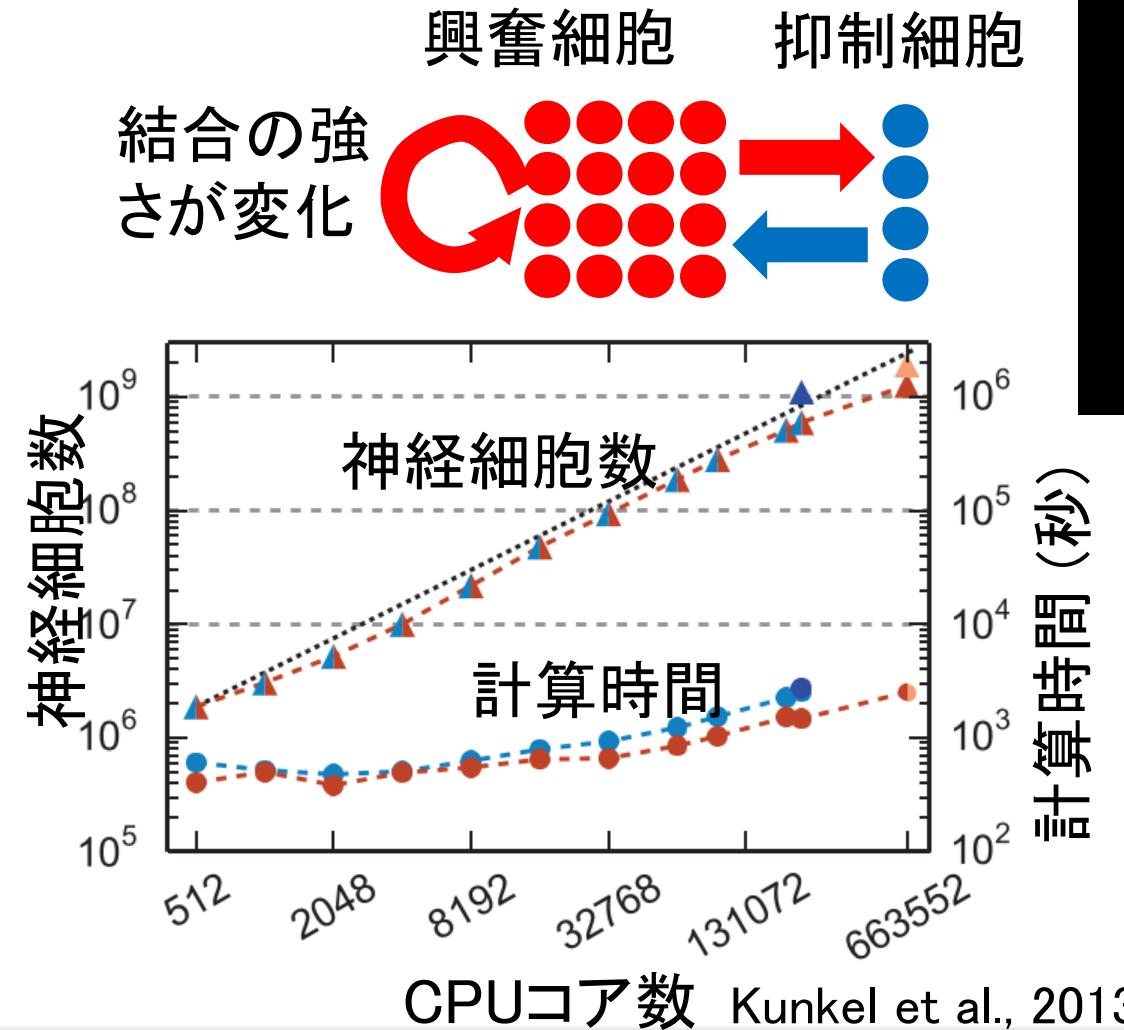
京による小型サル規模大脳皮質シミュレーション(2013年)



+

nest::

(ソフトウェア)



ヨザル 約14億個

18.5億個の神経細胞と11.1兆本の大脳皮質モデル1秒の活動を40分で計算

Leaky Integrate-and-Fire neuron model and conductance-based synaptic model with matrix exponential method (Rotter and Diesmann, 1999)

$$u_{i,k+1} = u_{i,k} + \frac{\Delta}{\tau}(-u_{i,k} + u_{rest,x} + R_x(I_{syn,i,k} + I_{ext,i,k}))$$

$$\text{if } u_{i,k+1} = \theta \Rightarrow u_{i,k+1} \rightarrow u_r, (k \leq k + k_{RP})$$

$$I_{syn,i,k} = \sum_y (v_{0i,y,k+1} + v_{1,i,y,k+1})(u_{i,t} - E_{syn,x,y})$$

$$v_{0i,y,k+1} = \left(c_{0,x,y} (v_{0i,y,t} + Z_{i,y}) \right)$$

$$v_{1,i,y,k+1} = \left((c_{1,x,y} (v_{1,i,y,t} + Z_{i,y}) + c_{0,x,y} v_{1,i,y,t}) \right)$$

$$Z_{i,y} = \sum_{j(y)} W_{i,j} S_{i,j,(k)}$$

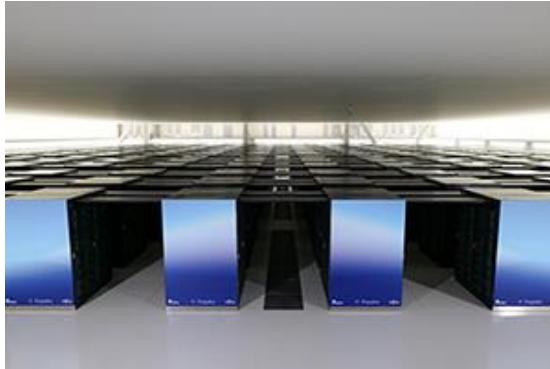
$$\text{if } t = t_{j,y}^f \Rightarrow S_{i,j,(k)} = S_0$$

i : postsynaptic neuron, j : presynaptic neuron, x : postsynaptic neuron type, y : presynaptic neuron

type, u : membrane potential, I_{syn} : synaptic current, I_{ext} : external input, R_x : membrane resistance, t_{RP} :

refractory period, W : connection weight, $g_{syn,i,y}$: synaptic conductance, E_{syn} : reversal potential, $t_{j,y}^f$: j , y spike time, c_0, c_1 coefficients for matrix exponential method, v_0 : $a_{x,y}v + \dot{v}$, v_1 : synaptic conductance, $Z_{i,y}$: postsynaptic conductance at initial step

v_0 : $a_{x,y}v + \dot{v}$, v_1 : synaptic conductance, $Z_{i,y}$: postsynaptic conductance at initial step

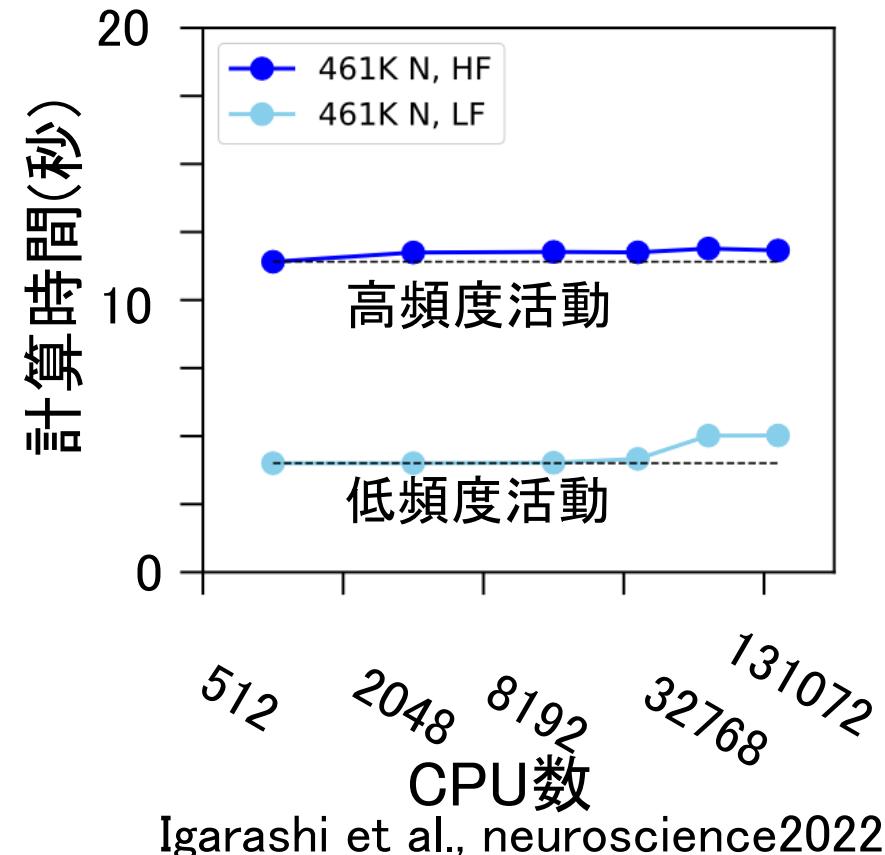


+

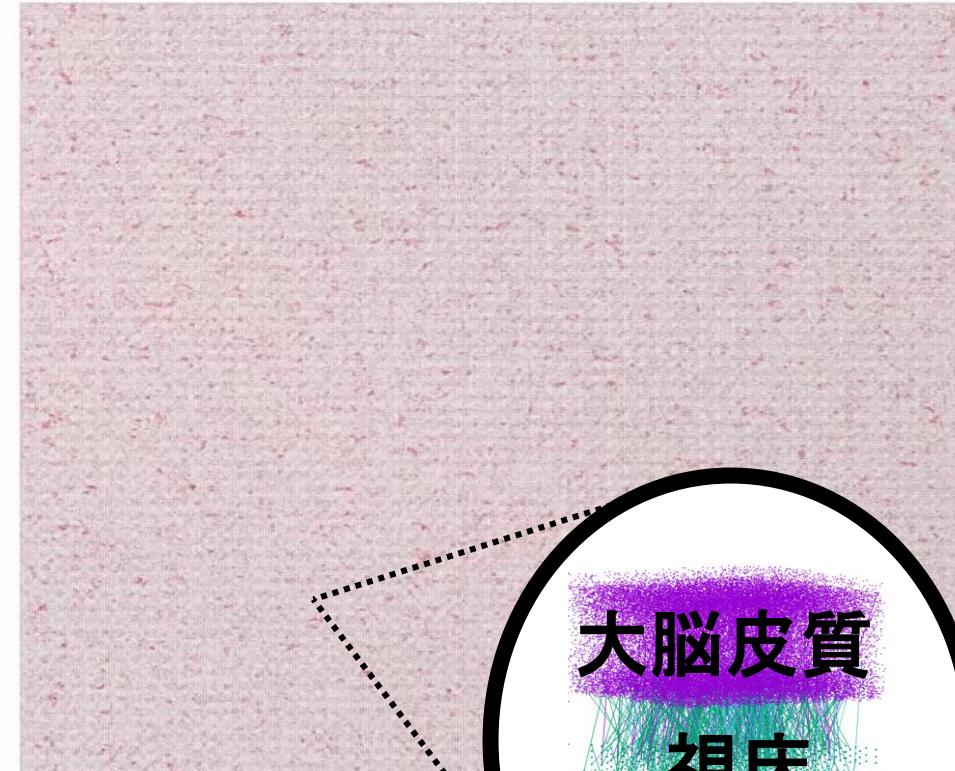
MONET
(ソフトウェア)

富岳のCPU15万個と4ペタバイトのメモリを使って、450億神経細胞の脳モデル1秒分の活動を約12秒で計算実行

富岳によるヒト規模大脳皮質-視床-小脳シミュレーション(Igarashi et al., 2022, neuroscience2022)



450億神経細胞を持つ大脳皮質-視床-小脳モデル



大脳皮質
視床
小脳

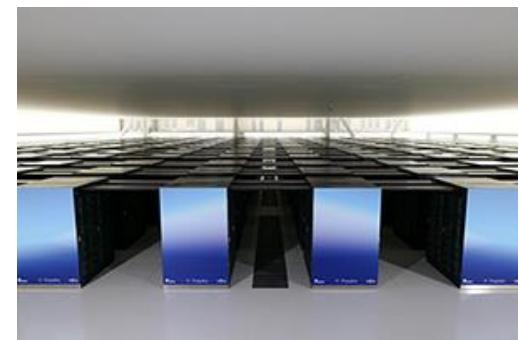
大脑皮質モデル

京(2011–2020)
 10^{16} 演算/秒



$\times 100$ 倍=

富岳(2021–現在)
 10^{18} 演算/秒



350

4

富岳で計算が速くなるとなにができるようになる？

- ヒトの脳をシミュレーション
- マウスや小型のサルの脳を数百体同時にシミュレーション
- 長時間活動シミュレーションで運動、認知、学習、病気などを調べられる

1000

100000

512

131072

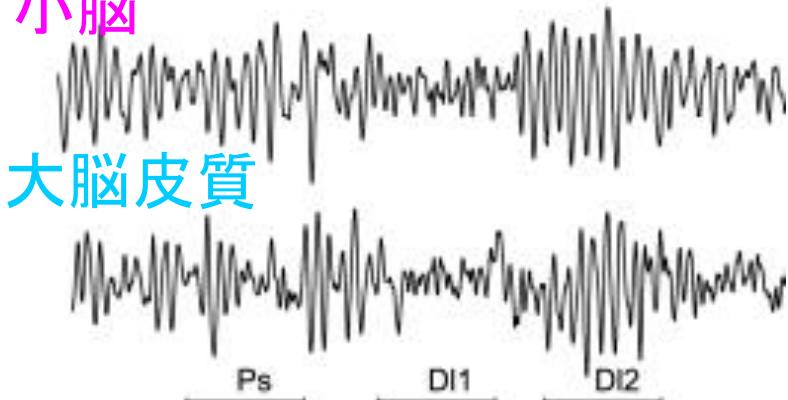
CPUの数

CPUの数

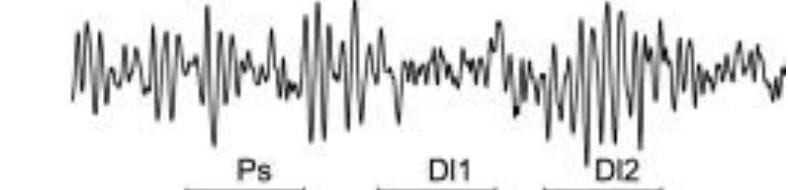
1秒分のシミュレーションが300秒から2秒に高速化 (150倍!)

マウス大脑皮質一小脳における振動の伝搬のシミュレーション

小脳



大脑皮質



Courtemanche and Lamarre, 2004

シミュレーション

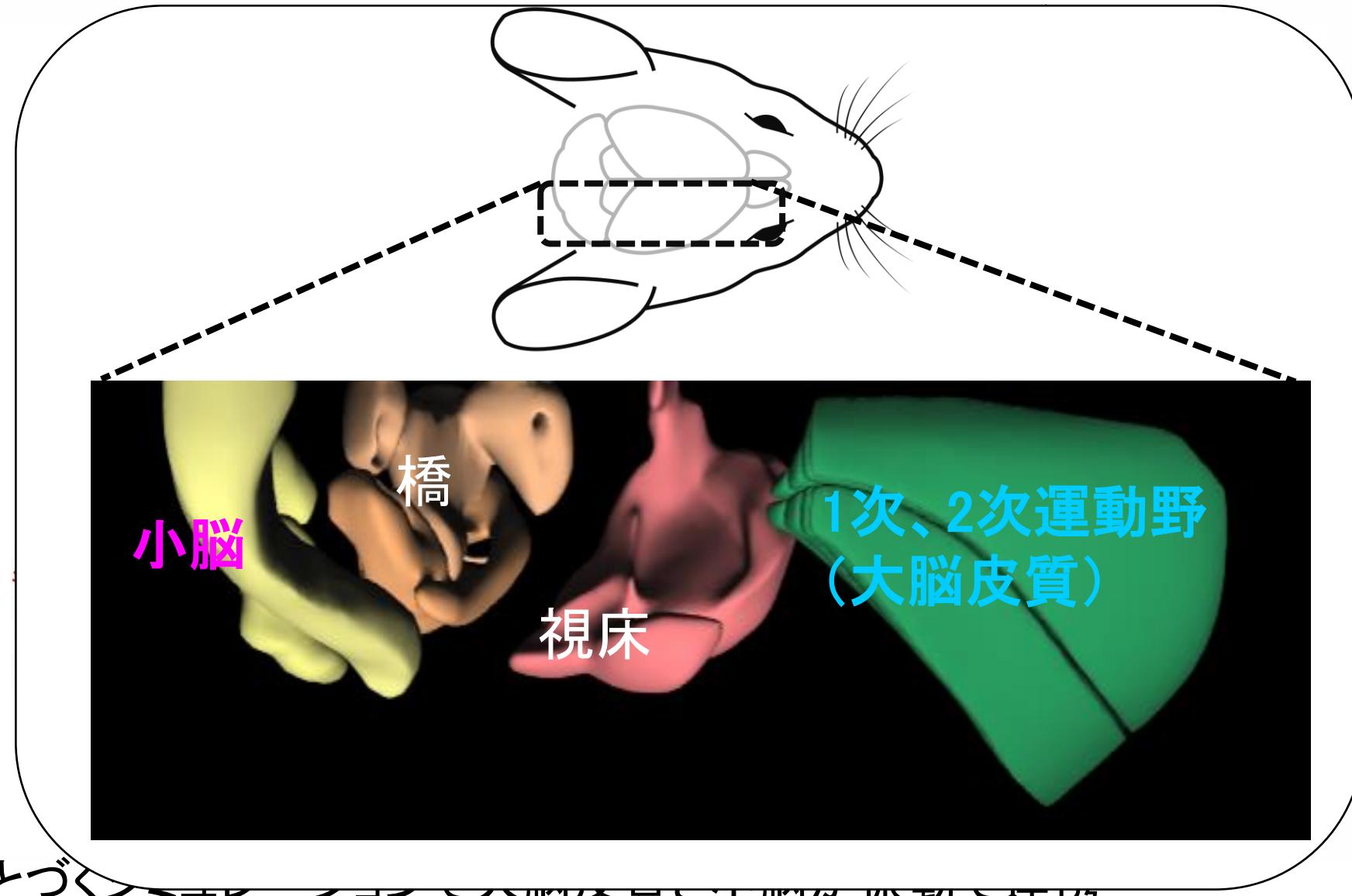
小脳



大脑皮質



脳構造・結合データにもとづく



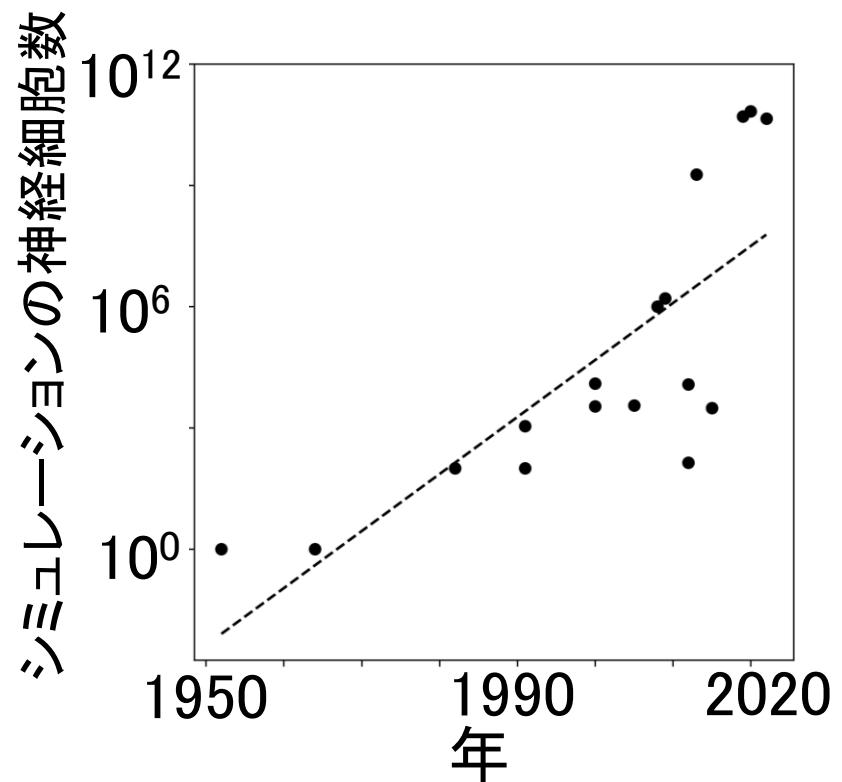
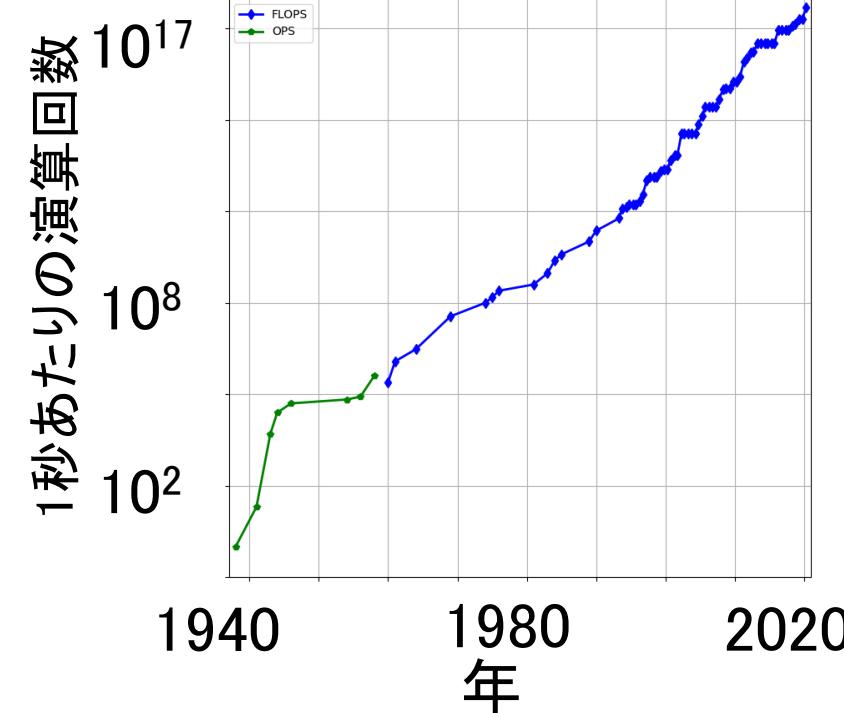
3. 神経科学の技術動向とAI-for-Neuroscience

脳・神経科学におけるAI-for-science

- ・ 脳データの解析(脳領域のセグメンテーション、脳活動の時空間構造解析)
- ・ 脳シミュレーションにおけるデータ同化の適用
- ・ 複雑な神経細胞モデルのサロゲートモデルによる置き換え
- ・ LLMによる神経科学の知見の収集、解析やシミュレーションのコード生成

これらの背景として、脳データの爆発的増加、計算機性能の指数関数的上昇、LLMの急激な性能向上がある。

スパコン性能の上昇と脳モデルサイズの増加

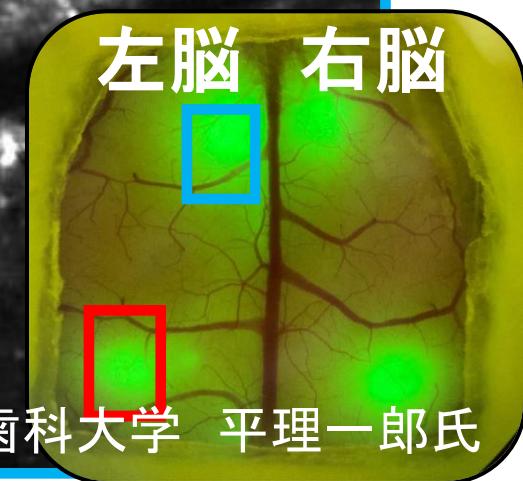
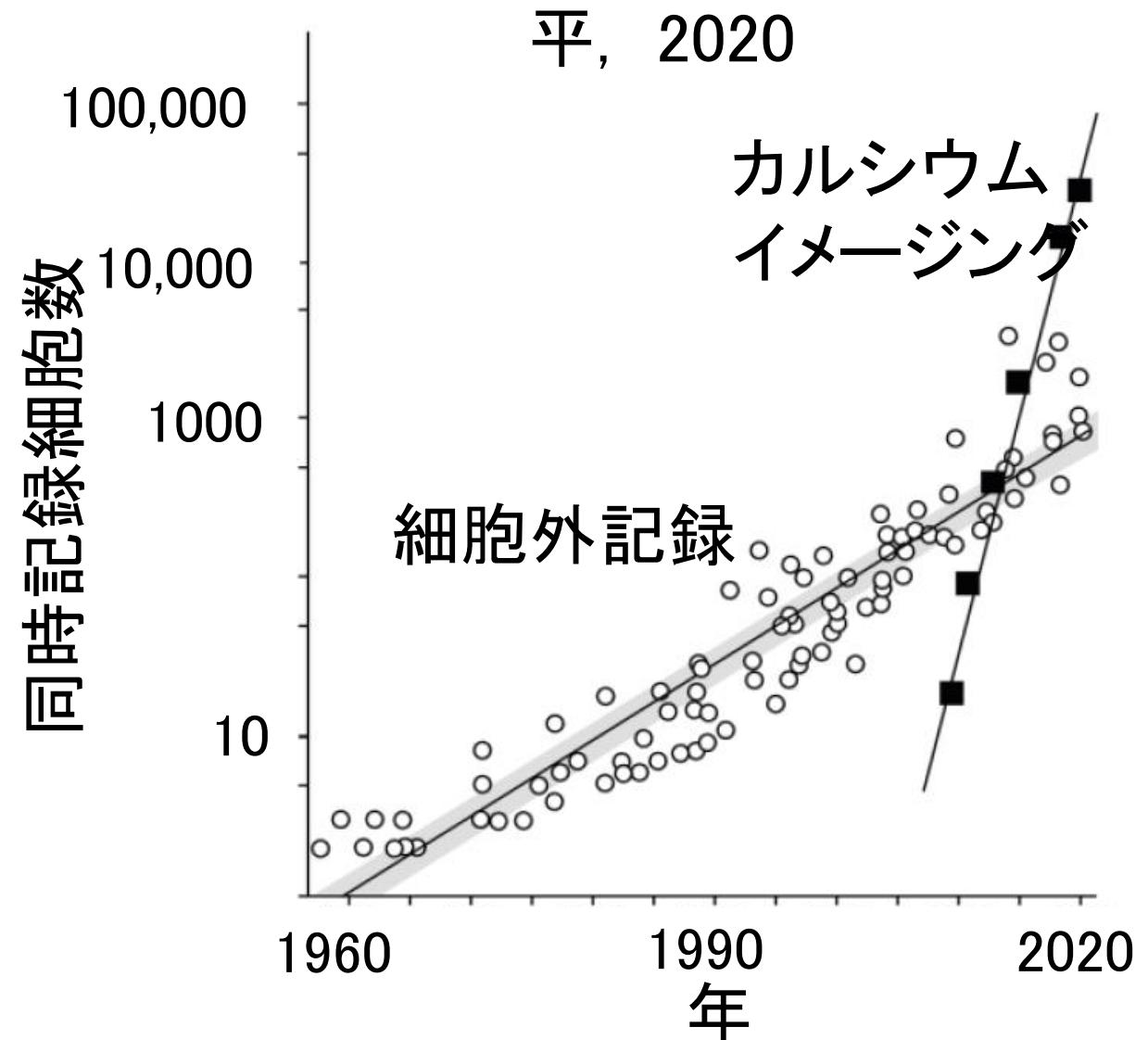


[ScienceNews2013]「京」でつなぐ神経回路 世界最大の脳シミュレーション

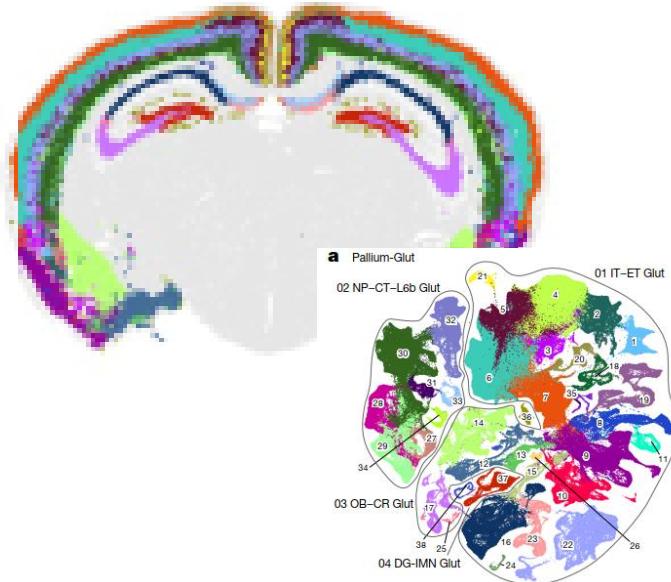
<https://www.youtube.com/watch?v=pbeCxQM3n0c>

10年前の予測と一致

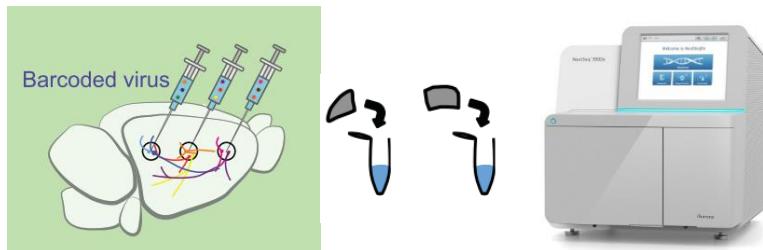
大規模化する脳の活動計測



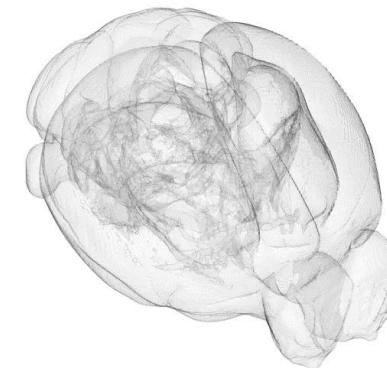
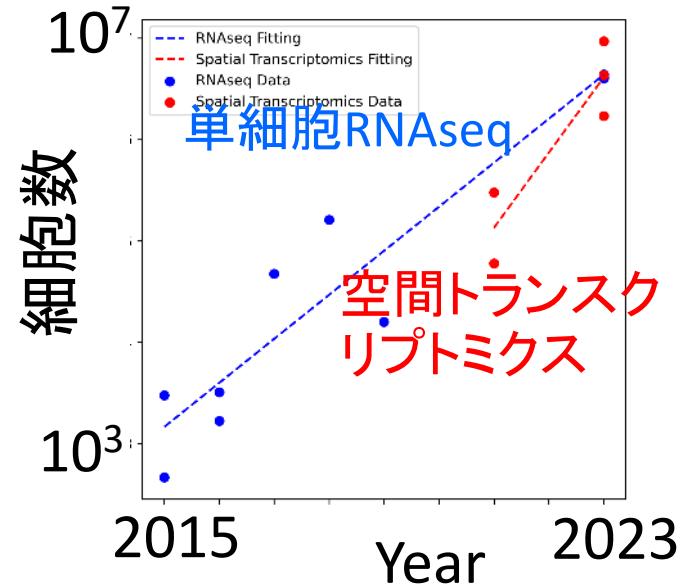
ranscriptomicsによる細胞種分類と結合計測



Yao et al., 2023



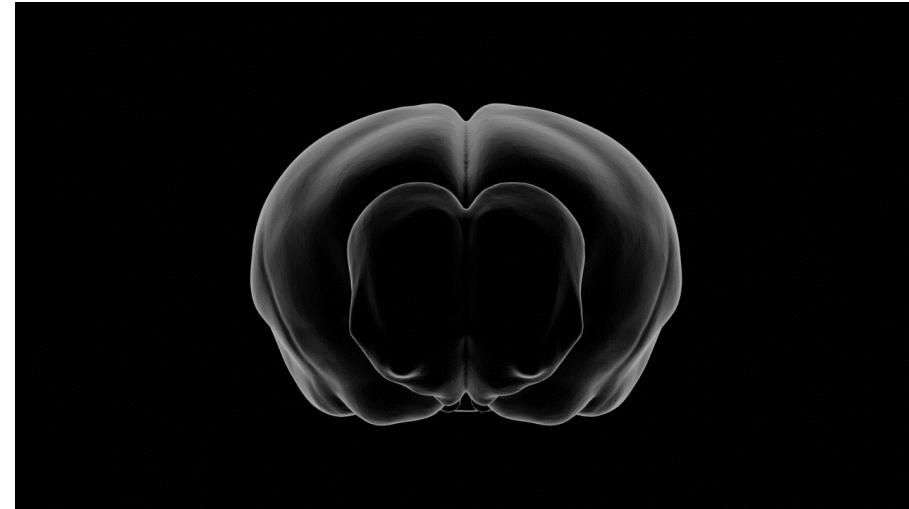
Huang et al., 2020



- 急速な遺伝子技術の発達(次世代シーケンサ、空間トランスク립トミクス)
- 最大で数百万神経細胞のRNA発現情報を入手可能に
- RNA発現レベルでの細胞種分類
- ウイルス感染で遺伝子を神経細胞に導入し、結合を介してRNAが伝播した後に、遺伝子情報を読み出し、結合情報を入手

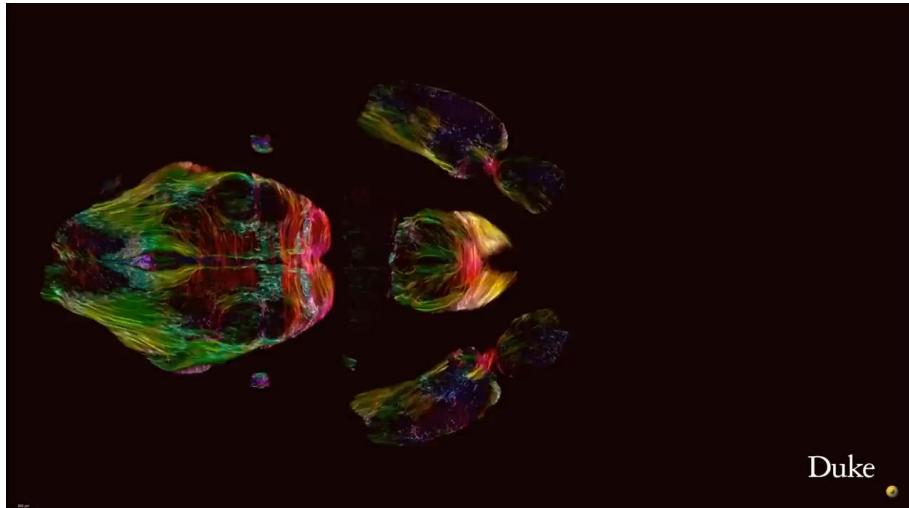
色素注入

(Winnubst et al., 2019)

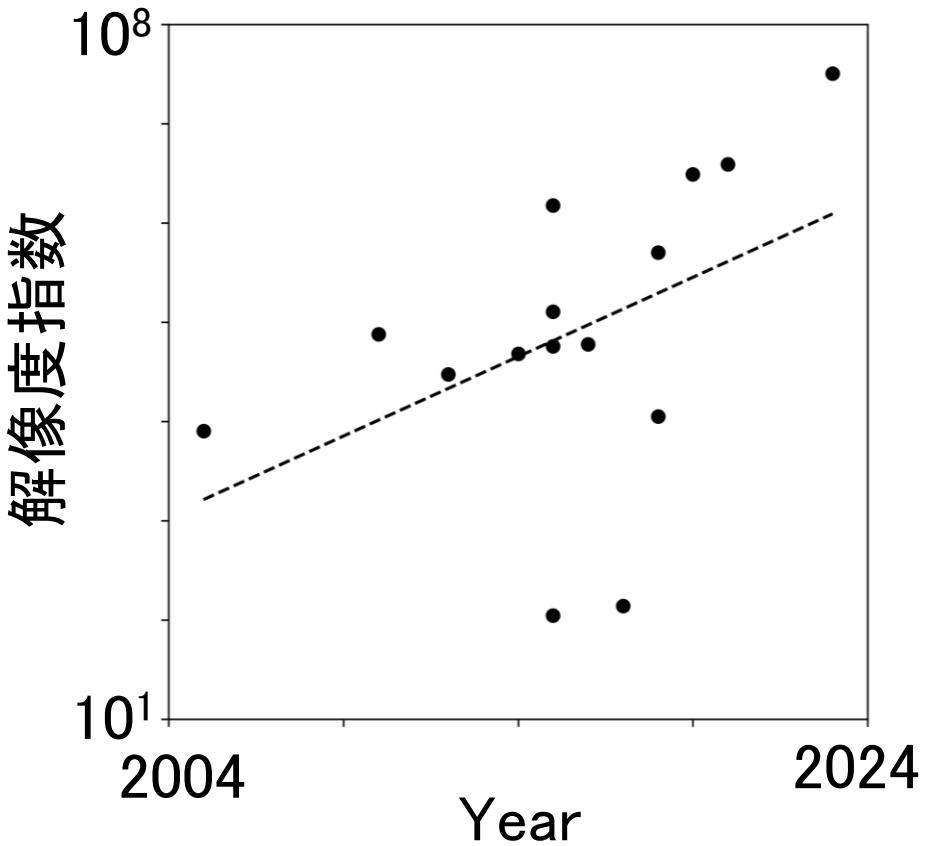


高解像度拡散磁気共鳴画像

(Jonson et al., 2023)

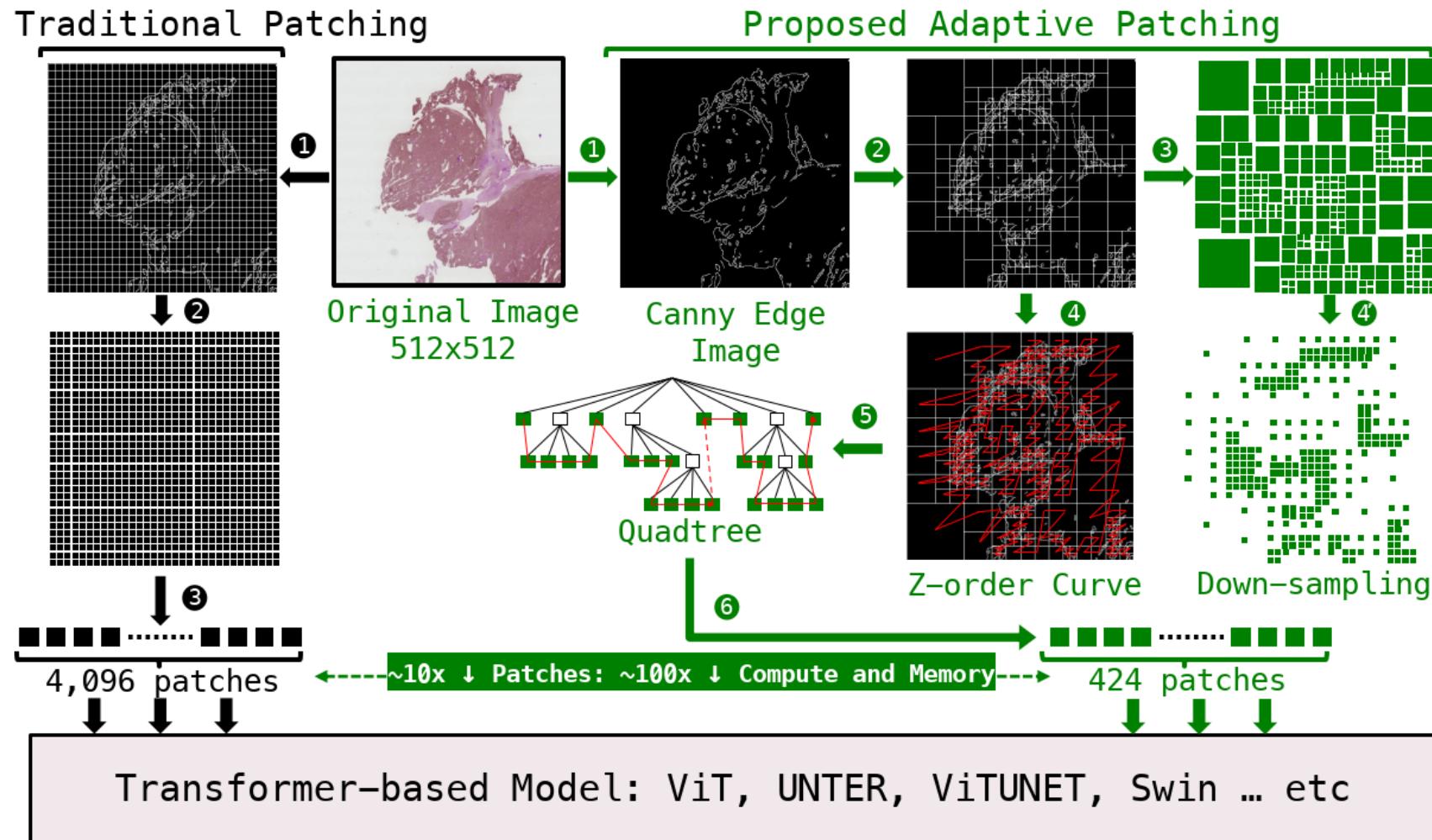


拡散磁気共鳴画像法による 脳結合計測の高解像度化



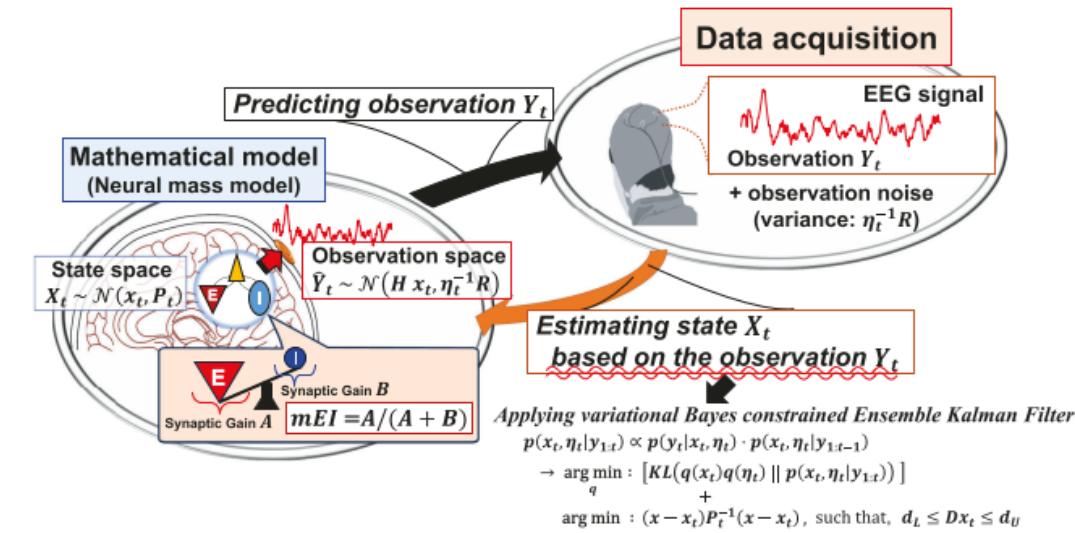
拡散磁気共鳴画像法の空間解像度は向上が
続き、細胞レベルの解像度に近づきつつある。

LLMによる脳のセグメンテーション (Zhang et al., 2024, SC24)



脳の数理モデルへのデータ同化の適用

- 脳の数理モデルのデータ同化は2000年代から (Voss et al., 2004)
- 大脳皮質のデータにUnsensed Kalman Filterを適用 (Schiff and Sauer, 2008)
- 脳波計データにUnsensed Kalman Filterを適用 (Escuain-Poole et al., 2018)
- 脳波データとneural massモデルにアンサンブルカルマンフィルタの適用 (Yokoyama and Kitajo, 2023)
- Bold信号にアンサンブルカルマンフィルタ、パーティクルフィルタ、マルコフ連鎖モンテカルロ法を組み合わせた階層的データ同化 (Zhang et al., 2024)
- 包括的な計測データ不足、数理脳モデルの計算リソース不足で、細胞解像度の数理脳モデルのデータ同化はまだ行われていない。

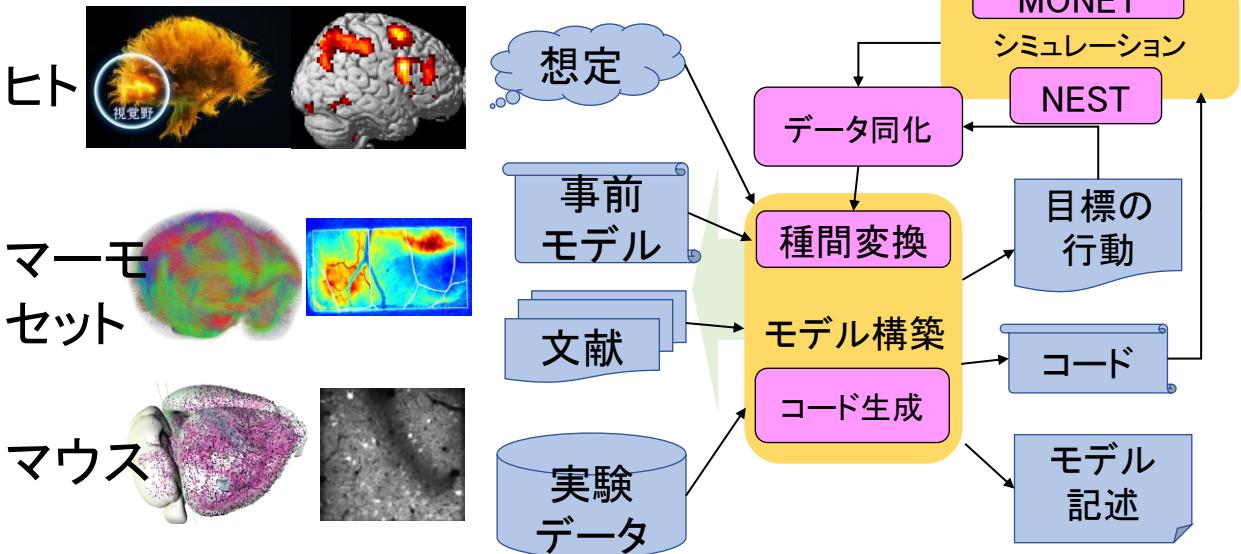


Yokoyama and Kitajo, 2023

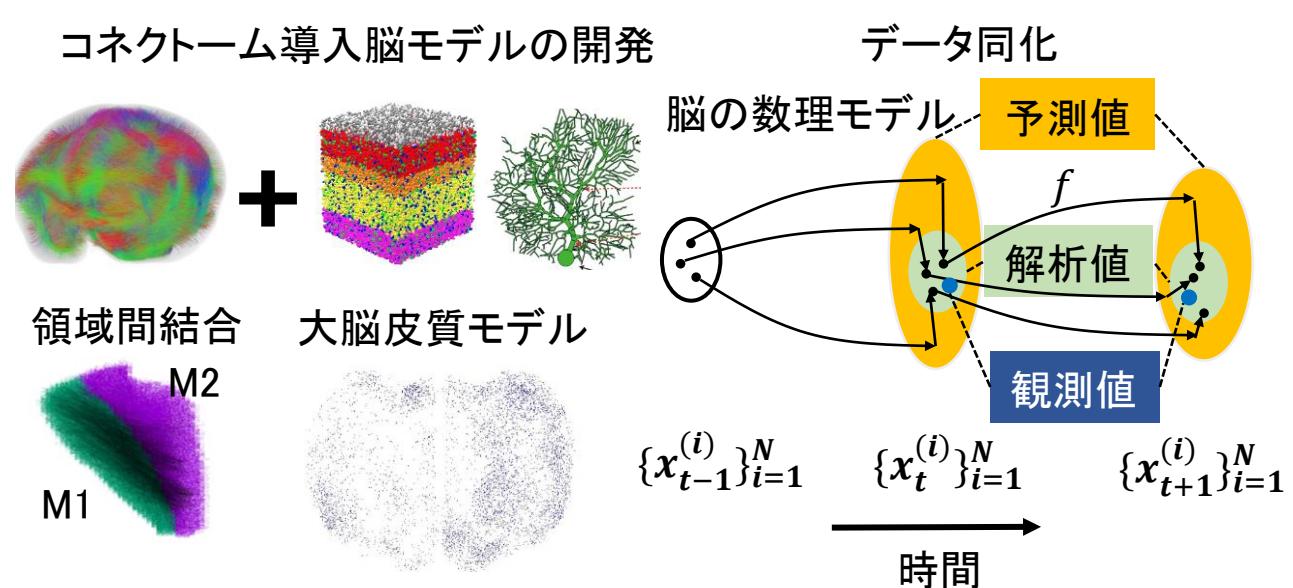
スーパーコンピュータ「富岳」成果創出加速プログラム: 包括的計測情報による多種全脳データ同化と特異的振動活動の探求

2023年4月-2026年3月、代表:理化学研究所 五十嵐潤

サブ課題1 データ駆動脳モデル構築フレームワーク

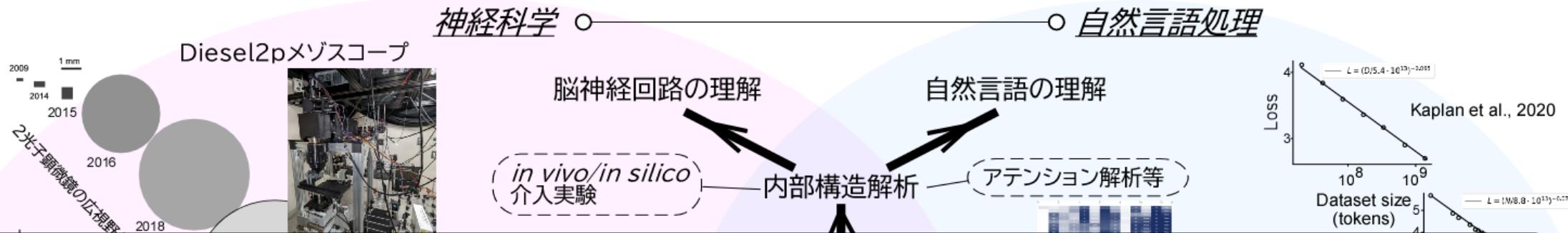


サブ課題2 全脳シミュレーションと振動現象の探求



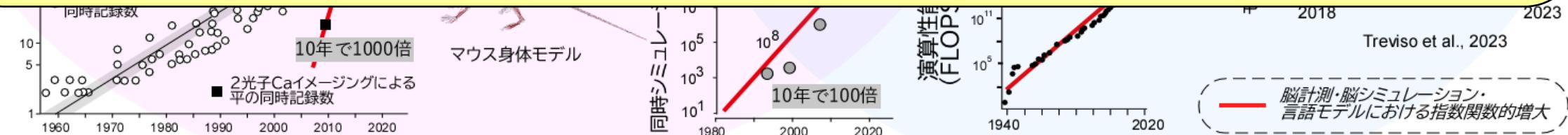
- 増大する脳計測データと計算コストに対して、富岳の計算性能を駆使し、データ同化の適用を目指している。
- マウスやマーモセットの脳モデルに対して、アンサンブルカルマンフィルタの適用を想定し、富岳で数十～数百体のアンサンブルメンバ一分の計算を実施。

AMED 脳神経科学統合プログラム 個別重点研究課題
LLMを介した大規模計算神経科学の加速
2024年10月–2030年3月、代表 東京科学大 平理一郎准教授

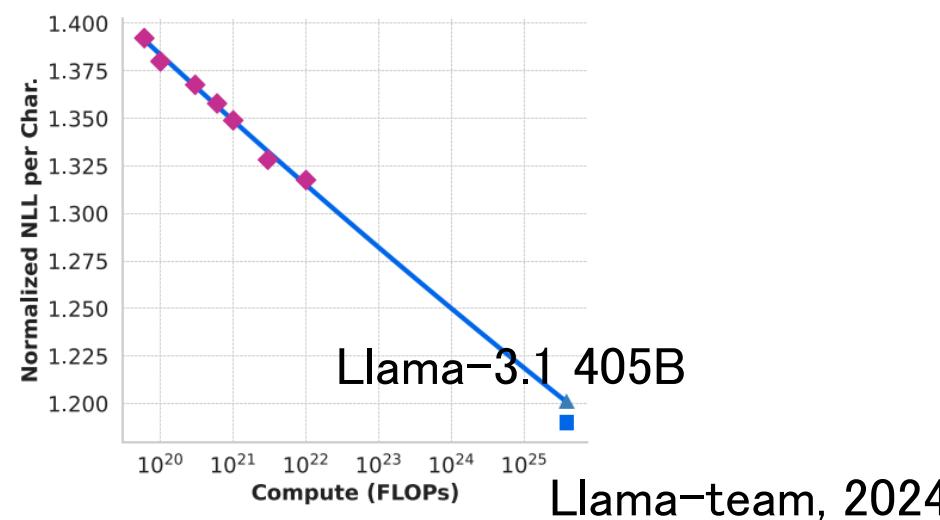
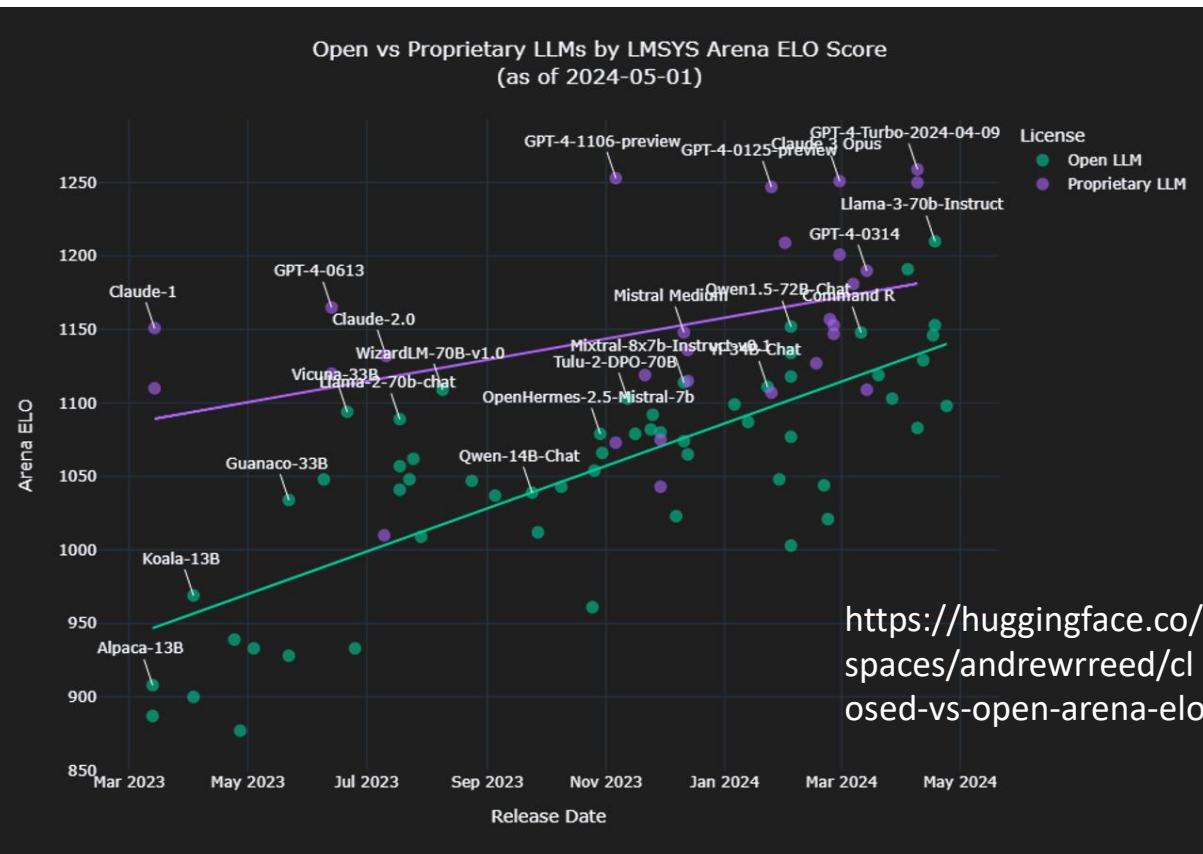


4つの目標:

1. 大規模脳計測・シミュレーションデータのための標準Transformerモデルの構築
2. LLM・LMMの内部構造理解による言語処理の可視化
3. LLMの内部構造解析手法の脳Transformerへの適用
4. コード生成LLMによるシミュレーションと解析の加速



Open vs Proprietary LLMs by LMSYS Arena ELO Score
(as of 2024-05-01)



LLMの急激な性能向上

- ・訓練時計算量増加による性能向上(Llama-team, 2024)
- ・ハードウェアの改善: 混合精度計算、ロジックとメモリ積層、専用ハードウェア化
- ・推論時の計算量スケーリング(OpenAI O1)
- ・AIの電力、データ量などの分析で、2030年までに1万倍計算量拡大の予想(Epoch.ai)

今後、数年間、数世代間にわたりLLMはさらに進歩し、科学を加速する可能性がある。
→HPC利用加速に向け、脳シミュレーションのコード開発の性能を調査

まとめ

- ・神経科学とAIは相互に影響を与え、発展してきている。
- ・脳のシミュレーションは、HPCの恩恵を受け、哺乳類脳全体のシミュレーションを扱えるようになりつつある。
- ・今後大幅な進歩が予想されるAIを利用し、HPCを用いた爆発的に増加する脳データの解析、HPCを利用した高精度な脳モデルのデータ同化、LLMのスケーラビリティを利用した脳活動モデル構築と内部解釈、HPCの並列化コード生成など、多方面から脳・神経科学の加速が期待できる。

Acknowledgements



RIKEN

Wahib Mohamed

Satoshi Matsuoka

Toshikazu Ebisuzaki



Juntendo University

Sun Zhe

Ryutaro Himeno



UEC

Hiroshi Yamaura

Tadashi Yamazaki



OIST

Carlos Enrique Gutierrez

Janne Moren

Kenji Doya



Kobe Univ.

Kentaro Nomura

Junichiro Makino



Tokyo Medical and dental Univ.: Riichiro Hira
RIKEN CBS: Shinichiro Tsutsumi
Waseda Univ.: Yuuki Mori



Brain/MINDS 2.0

This work was supported by JSPS KAKENHI Grant Number JP 21H05137 AMED BRAIN MINDS 2 (Grant Number JP23wm0625001). This work used computational resources of supercomputer Fugaku provided by MEXT as “Program for Promoting Researches on the Supercomputer Fugaku” (Project ID: hp210169, hp200139, hp230206).