全脳アーキテクチャイニシアティブ

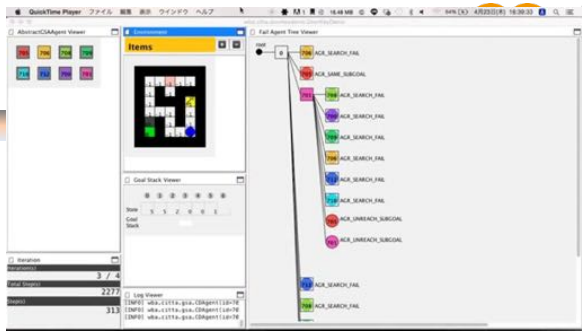
外部連携

理化学研究所
全脳アーキテクチャ・イニシアティブ

高橋恒一
@ktakahashi74

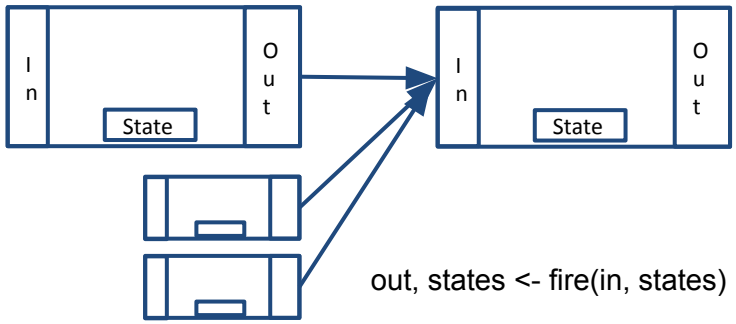
Whole Brain Architecture Initiative

BriCA: Brain-inspired Computing Architecture



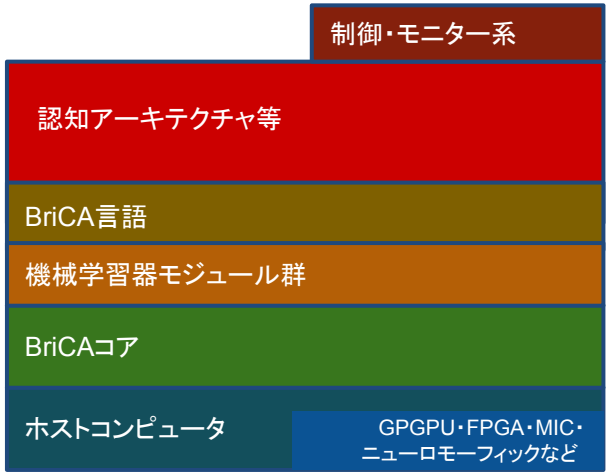
高性能プラットフォーム:
多数のモジュールを非同期で大規模に結合する
フレームワークを開発。

非同期並列計算モデル



- 関数型とオブジェクト指向のハイブリッドにより状態更新を非同期化。
- 大規模な計算モデルを局所的には同期、モジュール間は非同期で構築。
- 並列離散事象スケジューリングを行うことで効率よく並列計算。
- 実時間/仮想時間の両方をサポート。

モジュラーアーキテクチャ



- 多種の機械学習プログラムをモジュールライブラリ化し、それらの結合をDSL(BriCA言語)で記述、構築。
- ハードウェアを抽象化、リソース管理を行い、SoC、PCクラスターからスパコンまでをカバーの目論み。

- 現在主流の同期型フレームワークにおける、(1)限定的なスケーラビリティーおよび(2)リアルタイム性の低下を克服し全脳規模の認知アーキテクチャーを実現するのが目的。

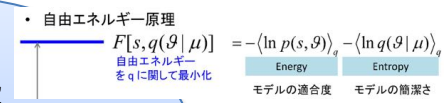
理研と全脳アーキテクチャ・イニシアティブで共同開発

(Takahashi *et. al.*, JSAI15, BICA15, Proc. Comput. Sci 2016, ICONIP16)

文部科学省　ポスト「京」研究開発萌芽領域的課題④
思考を実現する神経回路機構の解明と人工知能への応用
脳のビッグデータ解析、全脳シミュレーションと脳型人工知能アーキテクチャ

A: 脳の構造と活動の
大規模データ解析
大羽成征(京都大学)

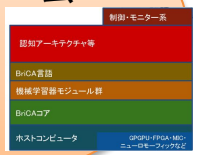
E: 脳型人工知能
アーキテクチャの開発
石井信(京都大学)



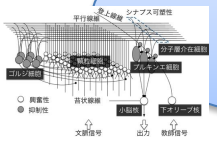
B: 大脳皮質神経回路の
データ駆動モデル構築
五十嵐潤
(理化学研究所)

D: 大脳皮質・基底核
・小脳モデル統合による全脳シ
ミュレーション
研究代表者:銅谷賢治
(代表機関:沖縄科学技術
大学院大学)

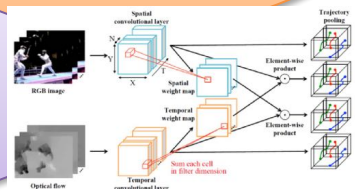
F: 脳型人工知能用大規模
高性能計算プラットフォーム
の開発
髙橋恒一
(理化学研究所)



C: ヒト全小脳モデル構築と
大脳小脳連関シミュレーション
山崎匡(電気通信大学)



G: 脳型人工知能の大規模
実問題への応用
原田達也(東京大学)
PFN



思考を実現する神経回路機構の解明と人工知能への応用
課題F：脳型人工知能用大規模高性能計算プラットフォームの開発 （理化学研究所・高橋恒一）

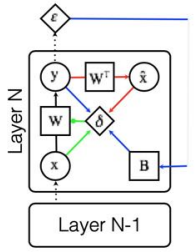
目標

全脳に匹敵する規模での非同期並列計算に耐えうる脳型人工知能用高性能計算基盤ソフトウエアを開発する

成果内容と科学的・社会的意義

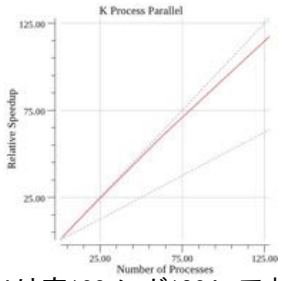
成果(1)　非同期並列基盤ソフトウエアBriCAを開発した。課題内連携のほかに課題外でも利用実績を積んだ。
成果(2)　新規の非同期学習手法を提案し、「京」を用いて世界で初めて1000コア規模のモデル並列学習に成功した。

(1)の成果により、非同期分散型の脳型認知アーキテクチャを高並列で実行する基盤が整備された 。
今日広く使われているTensorFlowなどの人工知能エンジンは、GPUでの同期的処理に最適化され、モデル並列実行時の並列性能は厳しく制限されている。 BriCAの開発により、人工神経回路の数千コアから数万コア規模での学習実行が可能になった。
BriCAを用いて、サブ課題Eで開発するMatcherNetをARMサーバー上で48並列で実行することに成功した 。また新学術領域「脳情報動態」でも大脳新皮質の6層構造を真似たマスターアルゴリズムフレームワーク（ MAF）の研究に用いられている。さらに、当課題のほか多数の研究機関・企業が協賛する全脳アーキテクチャ・ハッカソンでも採用され、新皮質＝大脳基底核ループや海馬モデルに基づいた 脳型人工知能の研究・開発・さらに教育にも展開 されている。
この成果は今後サブ課題ABCDの脳神経回路モデルを脳型認知モデルに応用する際の基盤ともなる。



Asynchronous Direct Feedback Alignment（ADFA）の一層

(2)の成果では分散メモリ環境において人工神経回路のモデル並列計算を高効率で実行可能なことを示した。
現在主流の誤差逆伝播に基づくend-to-end学習では、ネットワーク全体の同期処理が発生するため分散並列実行に根本的な足枷がある。本研究では、いくつかの生物学的に忠実な (biologically plausible)学習手法を提案あるいは拡張し、高並列実行に適したアルゴリズムを開発・実装し、性能の検証を行っている。
特に、Direct Feedback Alignment法(DFA)をベースに新規開発したAsynchronous DFA(ADFA)では「京」を用いて128ノード1024コアで良好なウイークスケーリングを示したが、この規模の性能は文献等で調査する限り世界初 である（論文投稿中）。このほかに、Decoupled Neural Interface(DNI)の高並列拡張などにも取り組んでいる。
この成果は、今後より大規模な人工神経回路を複合した自律性の高い人工知能アーキテクチャの実現の基礎となる。



ADFAは京128ノード1024コアまでの良好なスケーリングを確認

□ □ □ □ □ □ □ □ (BP)

Forward : $f_i(x) = W_i x$

Backward : $b_i(x) = W_i^T x$

Real-time predictive coding -> 'followdictive' coding

Itaya, Sakai, Takahashi, Yamakawa JSAI19

文部科学省科学 新学術領域研究(研究領域提案型)　領域番号：4905　略称：脳情報動態

脳情報動態を規定する多領野連関と並列処理

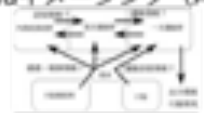
領域代表：　東大医学部　尾藤春彦

A03:脳情報動態に学んだ非同期並列情報処理アーキテクチャの提案と実証

研究代表者
高橋恒一・理化学研究所・チームリーダー
WEB http://lbcs.riken.jp

連携研究者
山川　宏・ドワンゴ人工知能研究所・所長／東京大学医学部・客員研究員

2光子Caイメージング（A01松崎）



fMRI/MEG計測
（A03春野）

多領野
イメージング
（A02尾藤）

大脳＝小脳動態計測
（A02喜多村）

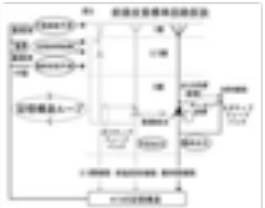
認知機能動態

比較
対照
検証

回路解析・探索

本研究

データ統合モデリング（A01石井）



前頭前野標準回路（A01川口）

モデル推定・
標準回路仮説

アーキテクチャ記述言語
（BriCA言語）

マスターアルゴリズム
フレームワーク

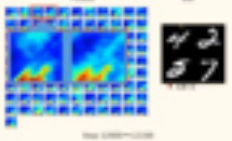
1. 回路セマンティクス・アーキテクチャ定義

非同期並列計算モデル

拡張PredNet、Active
Inference Network等・・

2. 脳の領野間連携に学んだ並列計算手法群の開発

能動推論タスク

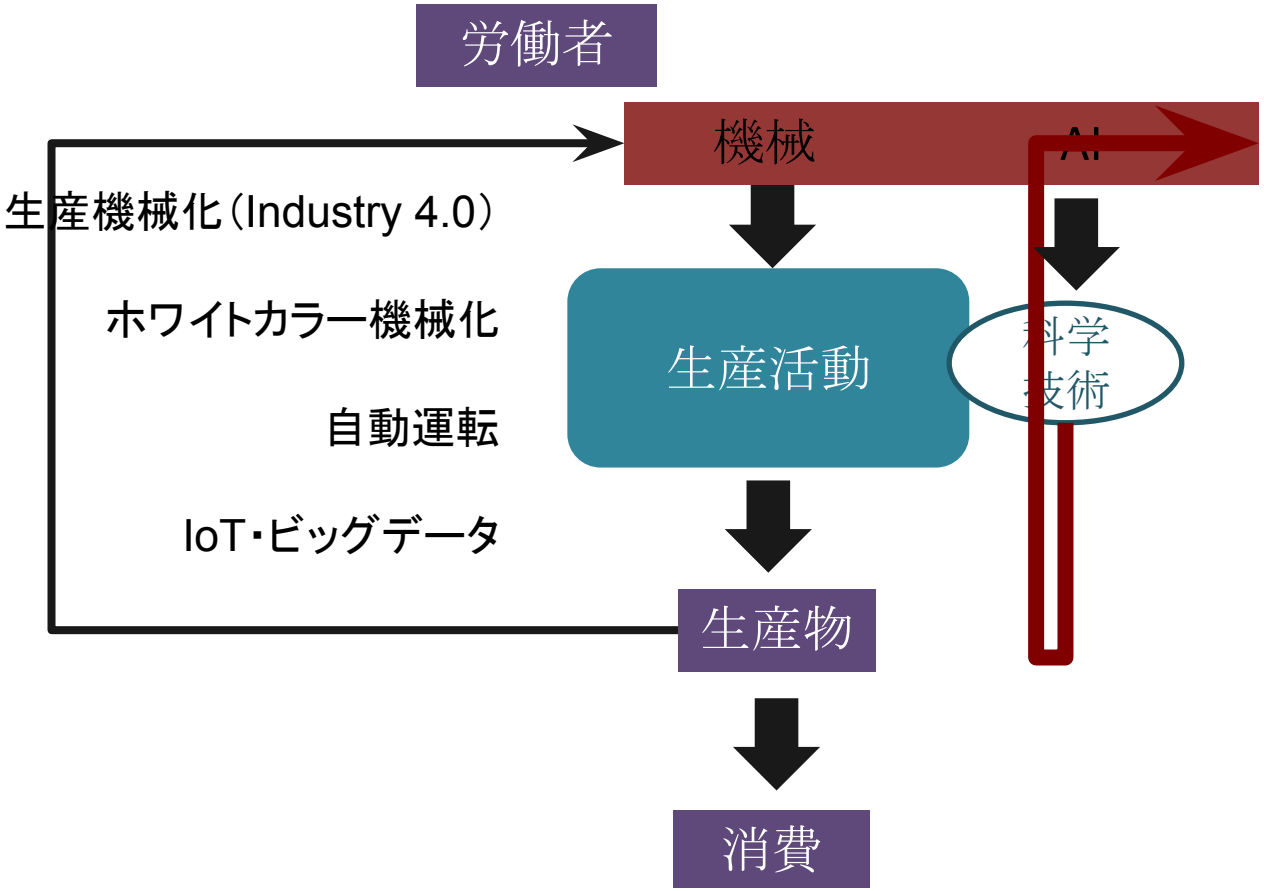


既存NN手法の並列化

3. 並列実装・実証

AI革命後のマクロ経済モデル

AI駆動型科学の重要性



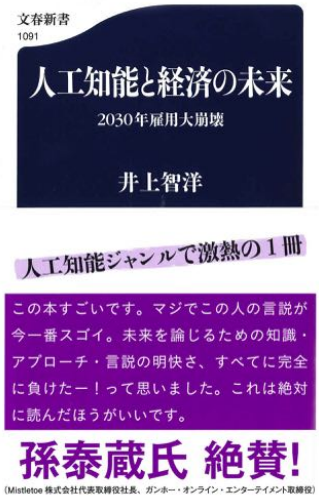
研究者

高度な情報技術
による科学・技術
イノベーション
の効率化・加速

最終的には知的労
働全般の自動化

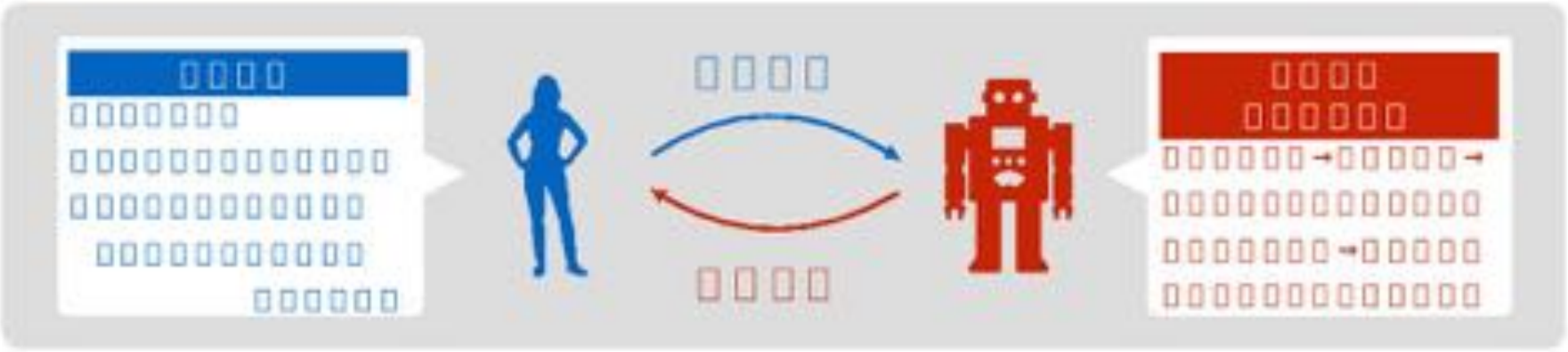
$$Y = AK \quad \Rightarrow \quad \frac{\dot{Y}}{Y} = sA(0)e^{gt} - \delta + g$$

AK型経済： 労働項Lが消える。
時間発展に注目すると資本項Kも消え、技術進歩率Aのみが残る。

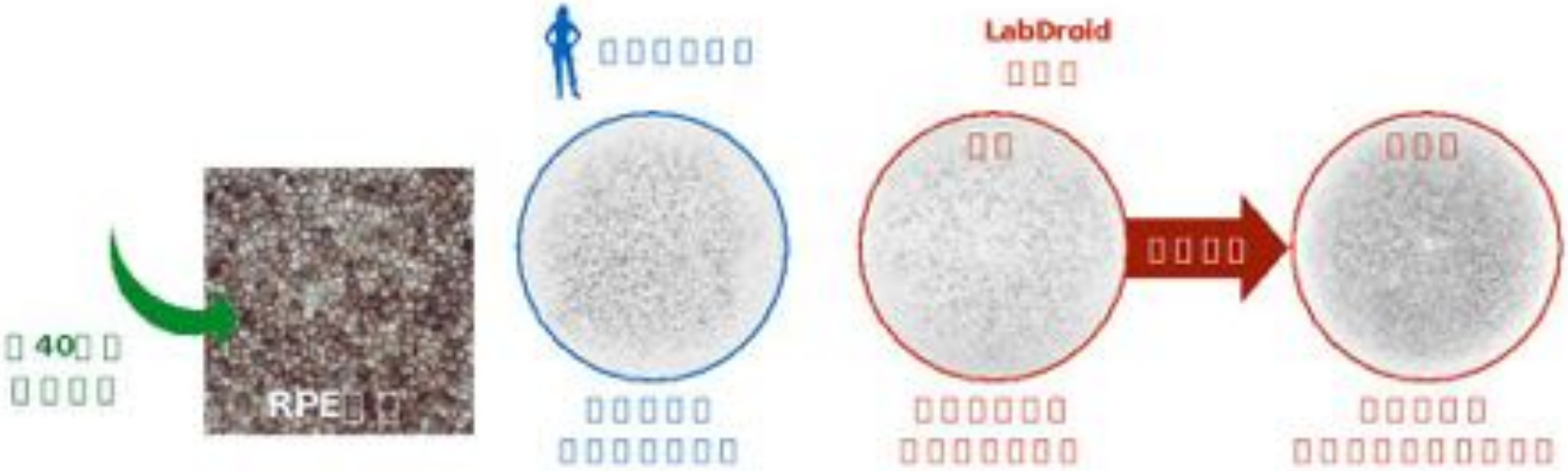


Prototyping Lab

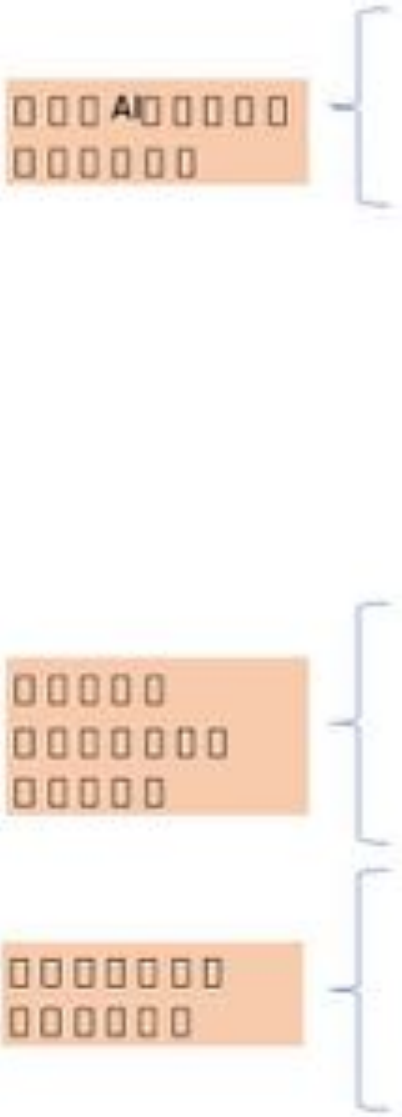
AI



RPE



AI



第一の科学	経験記述（実験）
第二の科学	理論
第三の科学	シミュレーション
第四の科学	データ

第五の科学　　自動化

AI駆動型科学

1 自動**実験・データ処理**

2 自動**実験計画**

(*e.g.* 仮説演繹法、実験計画法)

3 自動**仮説生成**

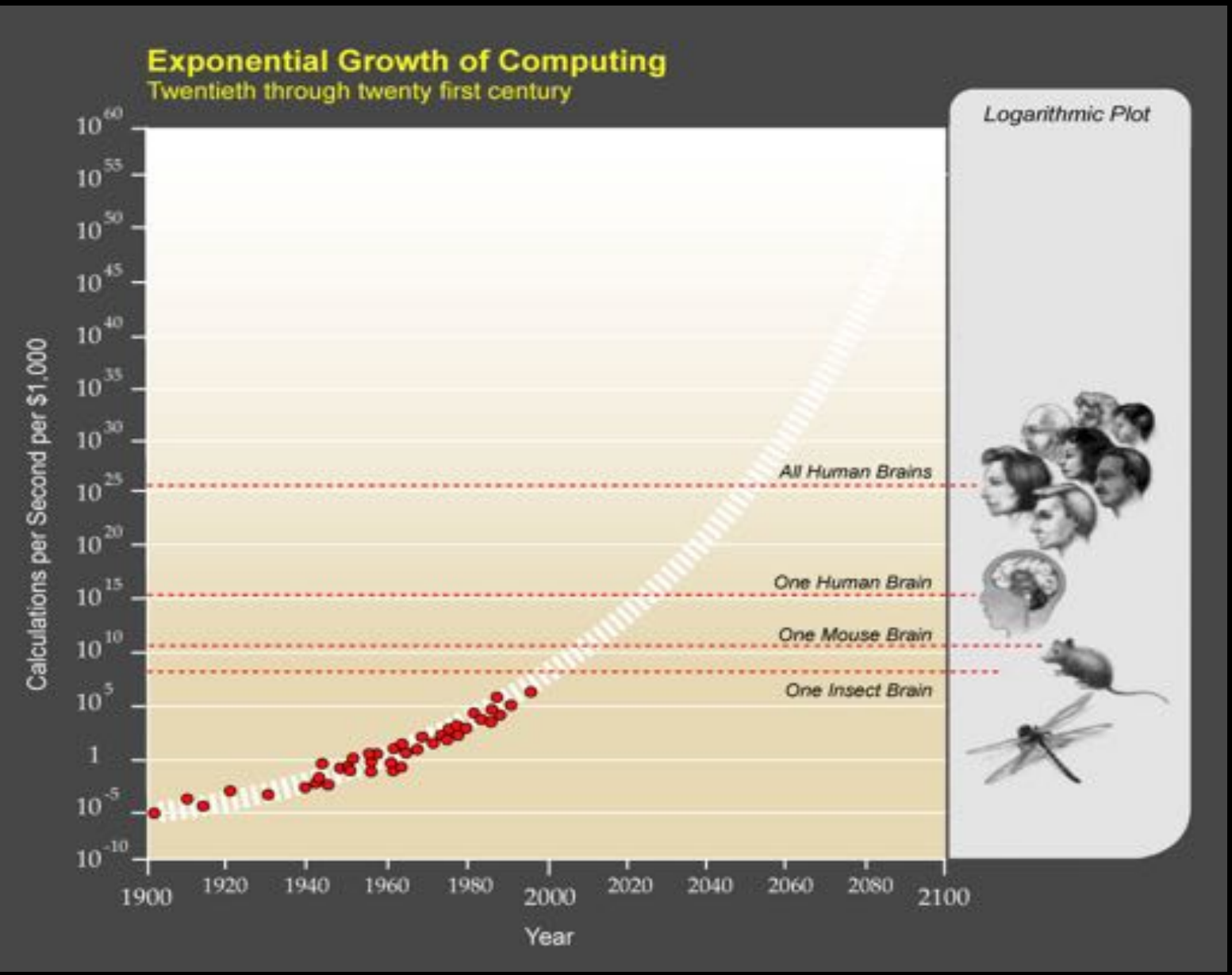
(*e.g.* アブダクション、能動推論)

現在の機械学習牽引型のAI技術のペイン

- アーキテクチャレベルの研究が比較的手薄
  - 大規模化に向けて不可欠な**参照アーキ**の定義、**モジュールの標準化**が不十分
  - Compositionalなアーキで直面する問題の洗い出しが不十分
    (Sculley *et al,* NIPS14)

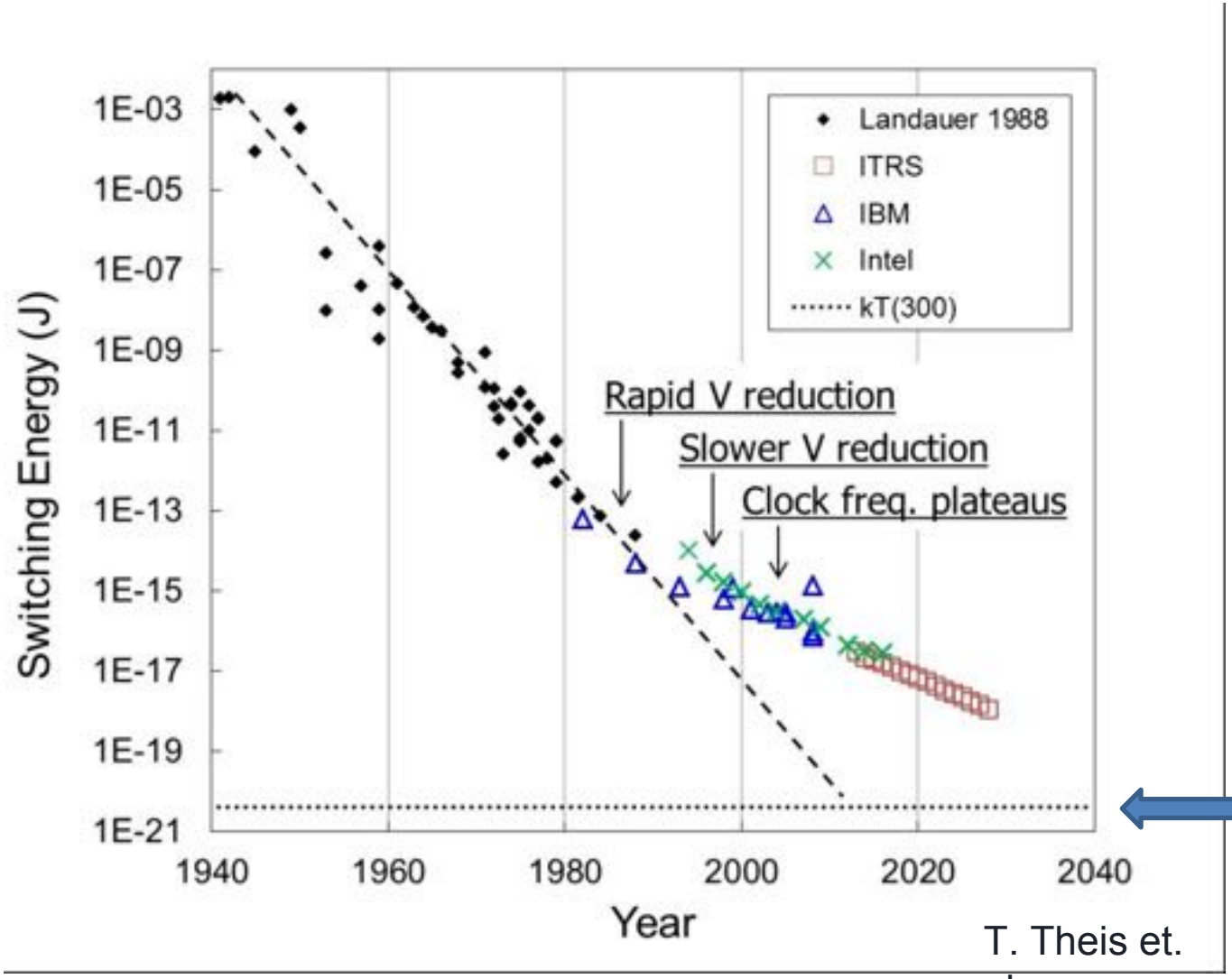
- **時間**を扱わない、もしくは扱いがナイーブ
  - 分類、生成は時間と無関係。モデルベース予測の方面を深化の必要。
  - ロボティクス方面（ROS等）は進んでいるが、理論的基盤は脆弱。

- 計算的基盤が未整備な結果、低レイテンシでスケールアウト可能な**アーキテクチャレベルの**並列計算モデルも未整備。



収益逓増の法則　Law of decreasing relative cost

Ray Kurzweil

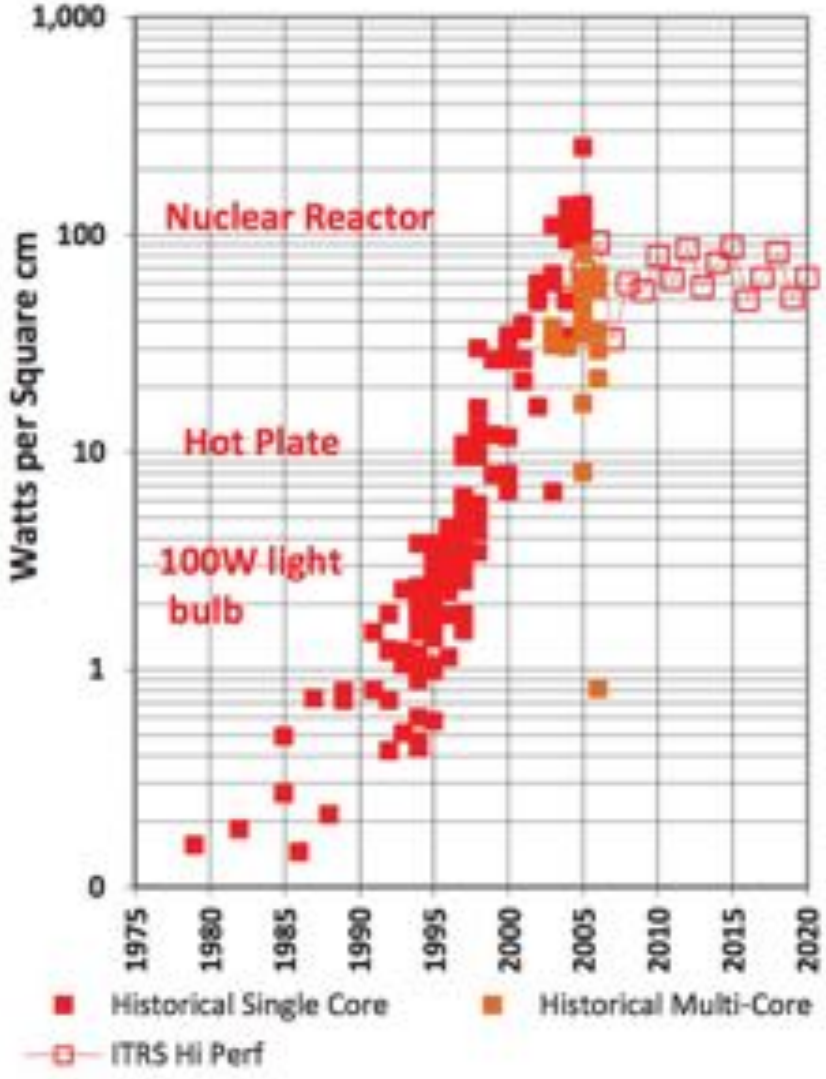


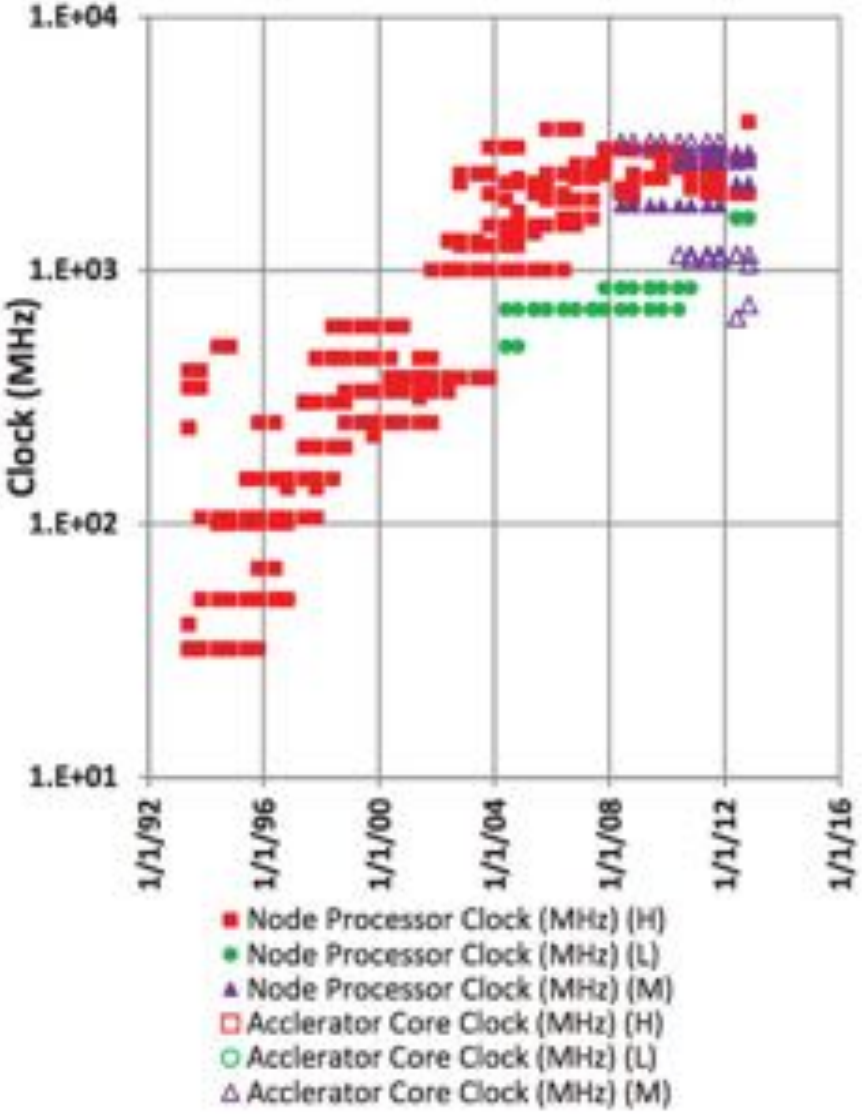
T. Theis et. al.

ランダウアー限界:

$k_B T \ln 2 = 2.87 \times 10^{-21}$ J (300K)

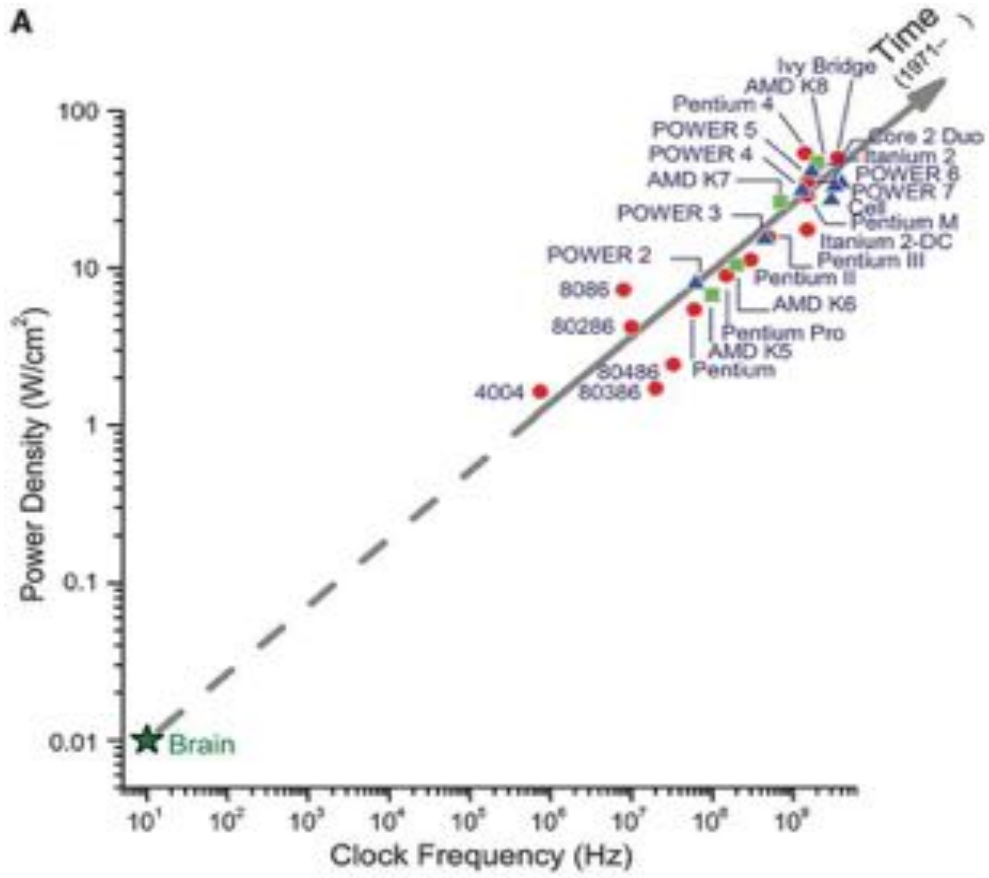
情報の消去などの論理的に非可逆な操作により散逸する1ビットあたりのエントロピーの下限（熱力学第二法則）





Source: Kogge and Shalf, IEEE CISE

脳型がムーアの法則の終わり(2020年前後)を超えて性能を向上させる鍵



P A Merolla et al. Science 2014;345:668-673

- **電力＝電圧の二乗　x 周波数**
- ノイマン型は局所のビット反転が全てを破綻させるため電圧を下げられない。
- ノイマン型は周波数と性能が比例するため電力と性能が比例。
- ニューラルネットは局所エラーが伝播しないためエラー許容率を緩和し電圧を大幅に下げられる可能性。

脳はとても効率的

- 脳の計算量はいくつかの独立な見積もりが0.1-10 PFLOPSに集中
- 器官としてのエネルギー消費は20W程度。純粋な計算自体には1W程度？
- 精度は10bit前後？　（ニューロンの不応期1ms、認知応答速度100ms、層数10^1とすると、10ms幅。この場合S/Nは20-30dBくらい）

一方、

- デジタル回路の情報損失
  加算 nbit + nbit -> nbit の場合、n bit
  乗算でもn bit（可逆論理ゲート）からn log n bit（現実的に）くらい
- 10PFLOPS, 1W, 損失100bit/単位演算だと、演算あたり 10^{-18} J

- ➡ ランダウアー限界から3–4桁
  半導体技術のプロジェクションと大幅には異ならない。

→ 2030年ころに1H普及とする予測と整合。
→ 1KHとか1MHとかになると、大量の電力が必要なことを示唆。

詳しくは「将来の機械知性に関するシナリオと分岐点」で検索