



# 脳型コンピュータの可能性

Makoto Taiji

Deputy Director  
RIKEN Quantitative Biology Center

Chief Scientist, RIKEN

taiji@riken.jp

# 自己紹介

## ■ 科学計算用専用計算機の開発(1986～)

- ▷ スピン系のモンテカルロ計算(m-TIS I, II)
- ▷ 重力多体問題(GRAPE-4,5)
- ▷ 分子動力学(MD-GRAPE, MDGRAPE-3,4)
- ▷ 密行列計算、密度汎関数法(MACE)

## ■ 超高速時間分解分光(1987-92)

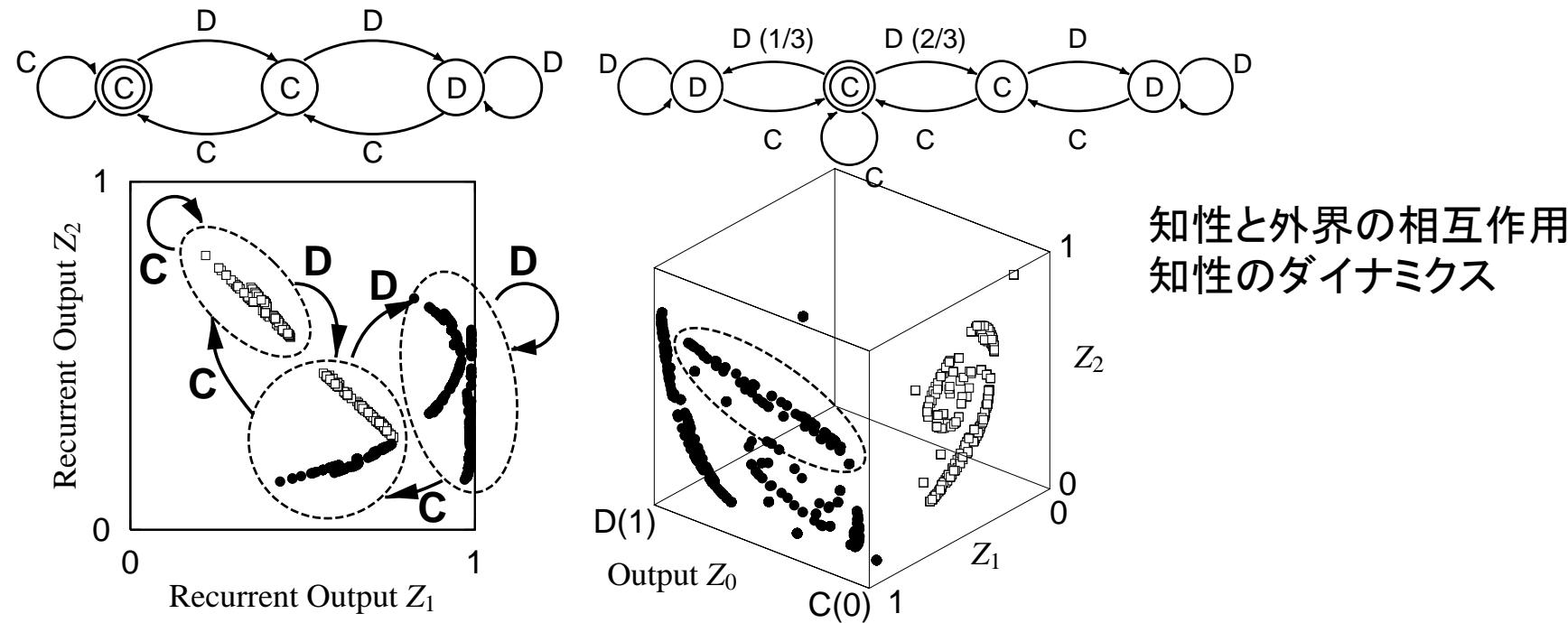
## ■ A-Life, Recurrent Neural Network (1996～)

- ▷ ゲーム理論 + 機械学習

## ■ 分子動力学計算の生命科学への応用 物理乱数

# Internal models in game players

- Iterated Prisoner's Dilemma Game
- Create opponent's models using recurrent neural networks



# 脳型コンピュータの背景

- 人工知能・機械学習  
の発展

- ▷ Deep Learning etc.

- 半導体技術の  
「成熟」

- ▷ 消費電力などの限界

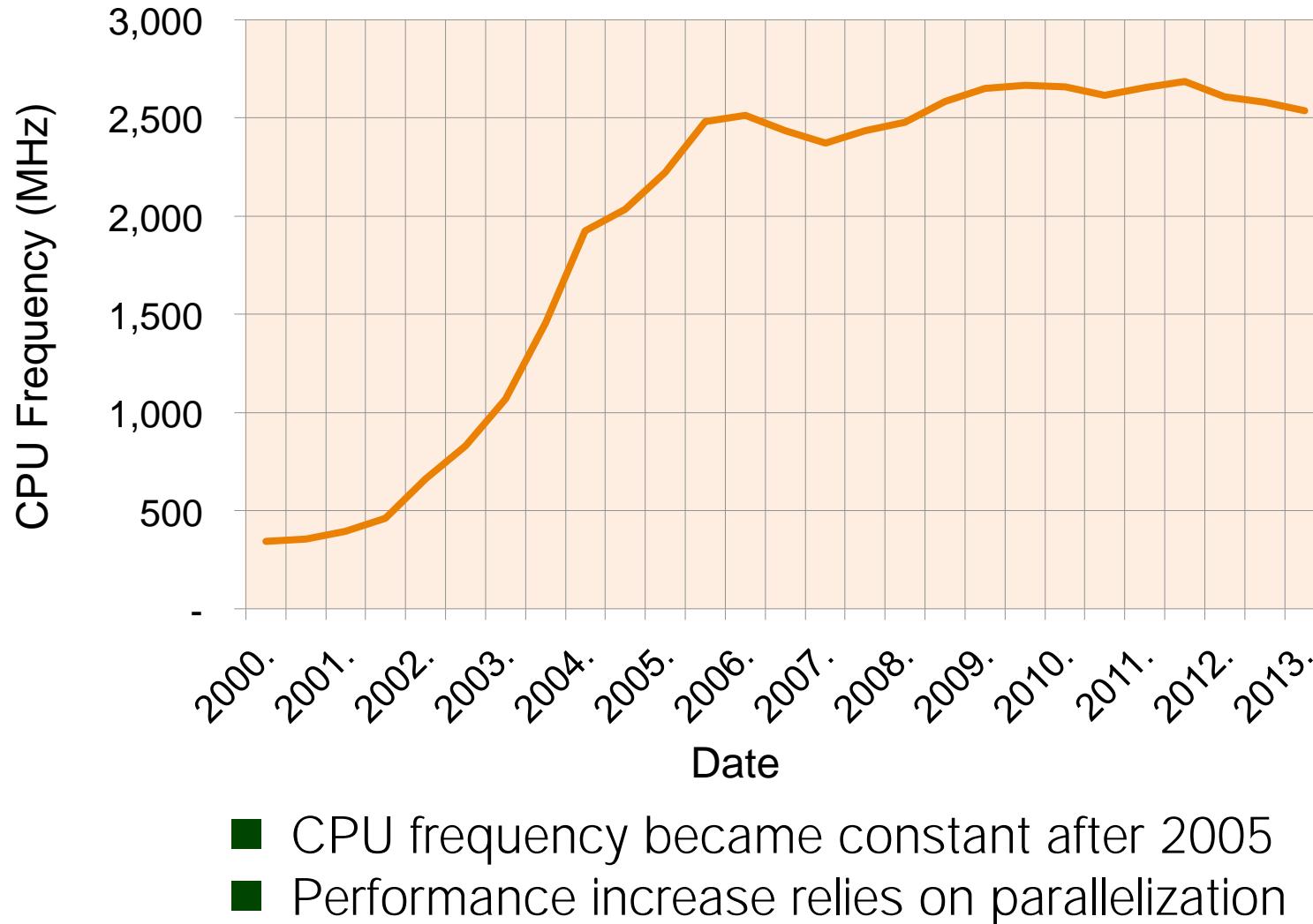
新しい情報処理の  
可能性

新しい計算原理の  
必要性

脳型コンピュータ

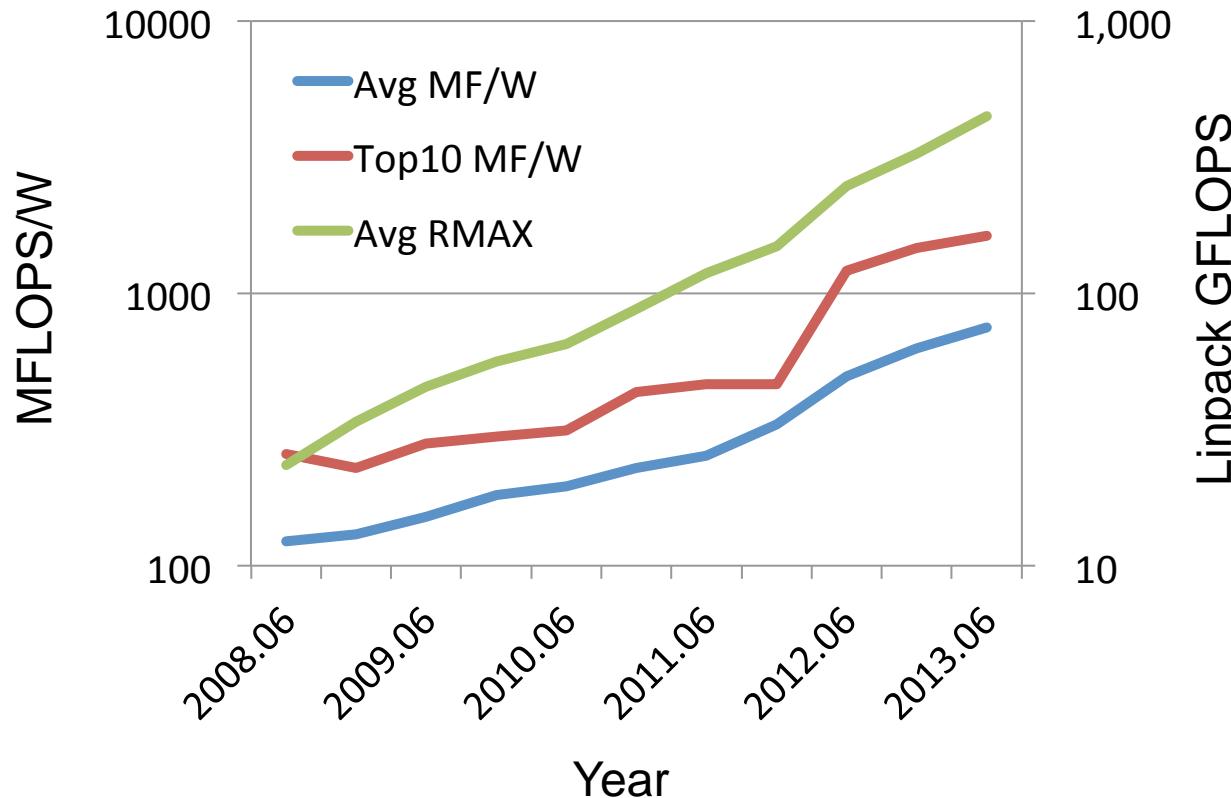
# Trends in High Performance Computing (1)

## TOP500 Average Frequency



# Trends in High Performance Computing (2)

## TOP500 Average Power Efficiency



- Pro : Exponential Growth
- Con: Growth speed slower than performance growth

# 半導体技術の成熟と課題

- 半導体技術は2020年頃に飽和
- 消費電力の増大
  - ▷ 電圧はこれ以上下げられない
  - ▷ ゲートあたりの消費電力は減る
  - ▷ 面積あたりの消費電力は増大
  - ▷ 増やさないように
    - 速度を落とす
    - 「ダークシリコン」… 同時に動かせない部分
- 汎用アーキテクチャでは解決困難
  - ▷ ヘテロジニアスプロセッサ

# 専用計算機の歴史

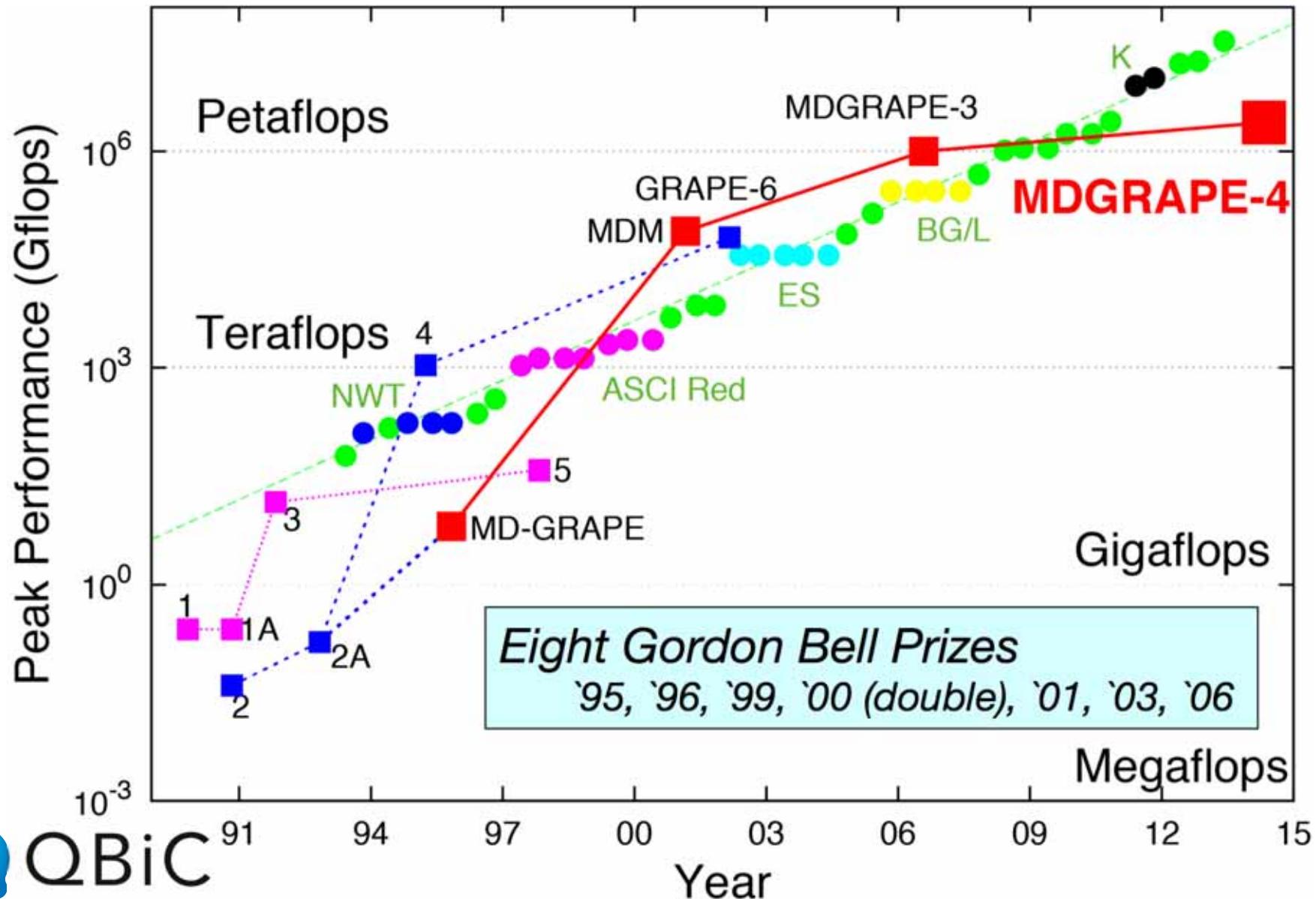
- 第一世代：初期の電子計算機
  - ▷ Atanasoff-Berry Computer
- 第二世代：「丸ごと」専用計算機
  - ▷ デルフト分子動力学計算機  
(Delft Molecular Dynamics Processor)など
  - ▷ 開発が大変で、時間かかった
- 第三世代：m-TIS/GRAPE
  - ▷ パソコンやワークステーションの加速装置
- 第四世代：Anton, MDGRAPE-4
  - ▷ System-on-Chip専用計算機・アクセラレータ

# GRAPE計画

- 1989～
- GRAPE (GRAvity PipE) 計画
- 銀河・銀河団、球状星団、惑星形成、宇宙論など
- 重力多体問題のための専用計算機
- 粒子系一般、特に分子動力学への拡張



# History of GRAPE computers



# Anton (D.E.Shaw研究所)

Forbes 400 Richest Americans      World's Billionaires      World's Most Powerful People      America's Best Small Comp

◀ #164 A. Jerrold Perenchio      Browse list ▾



**David Shaw**

Net Worth **\$2.2 B**  
As of March 2011

+ Follow David Shaw 6

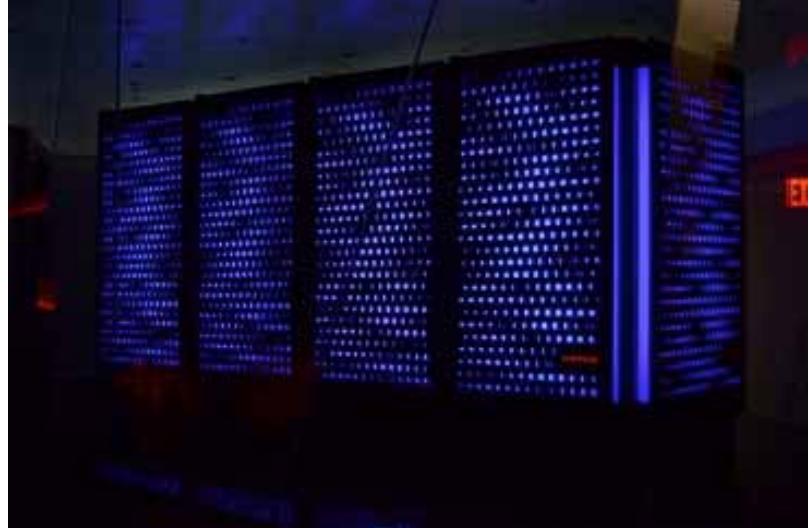
**At a Glance**

Age: 60  
Source: hedge funds , self-made  
Residence: New York, NY  
Country of citizenship: United States  
Education: PHD, Stanford University; BA/BS, University of California, San Diego  
Marital Status: Married

**Forbes Lists**

#540 Forbes Billionaires  
#179 in United States  
#164 Forbes 400

Like 4 likes. Sign Up to see what your friends like.



分子シミュレーション  
専用計算機  
高度な専用化を達成



Bill Gates visit at Schroedinger Inc.

# MDGRAPE-4

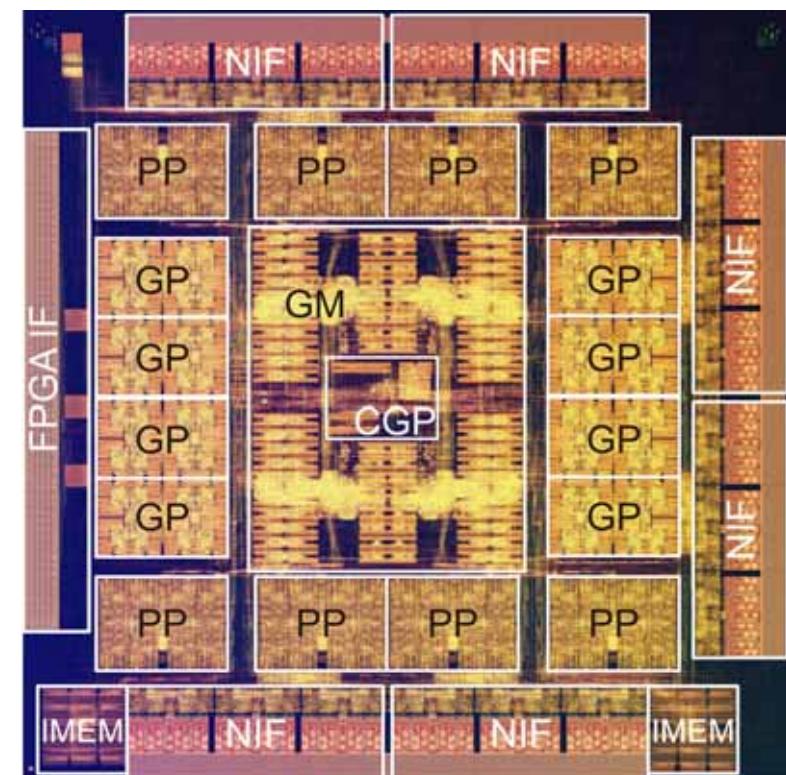
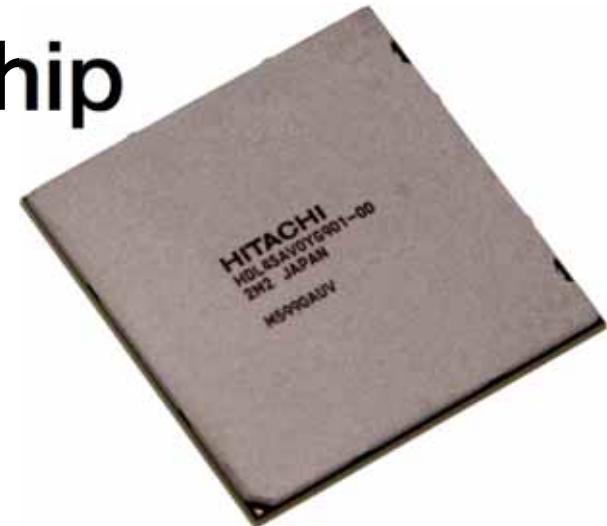
- Special-purpose computer for MD simulation
- Target performance
  - ▷ 50μsec/step for 50K atom system
  - ▷ 3μsec/day (2fsec/step)
- Target application : GROMACS
- Completion: 2014
- Enhancement from MDGRAPE-3
  - ▷ 130nm → 40nm process
  - ▷ SoC Integration of Network / CPU

# MDGRAPE-4 System-on-Chip

## ■ Fully Designed in QBiC

- ▷ Largest LSI designed in academia
- ▷ Such developments will be possible only by us and D.E.Shaw Research

- 40 nm (Hitachi), ~ 230mm<sup>2</sup>
- 64 force calculation pipelines @ 0.8GHz **2.5TFLOPS**
- 65 general-purpose processors  
Tensilica Extensa LX4  
@0.6GHz
- 3-dimensional torus network  
@ 7.2GB/sec/direction



# 専用計算機の現代的意義

## ■ 演算の最適化

- ▷ 精度の最適化
- ▷ パイプライン化

MDGRAPE-4では...

単精度  
28段・50演算  
専用パイプライン

## ■ データ通信の最適化

- ▷ 遅延の短縮
- ▷ メモリ構成の最適化
- ▷ 同期の高速化

400-600nsec

1.8MBエンベッド共有  
積算機能つきメモリ

# 発想の背景

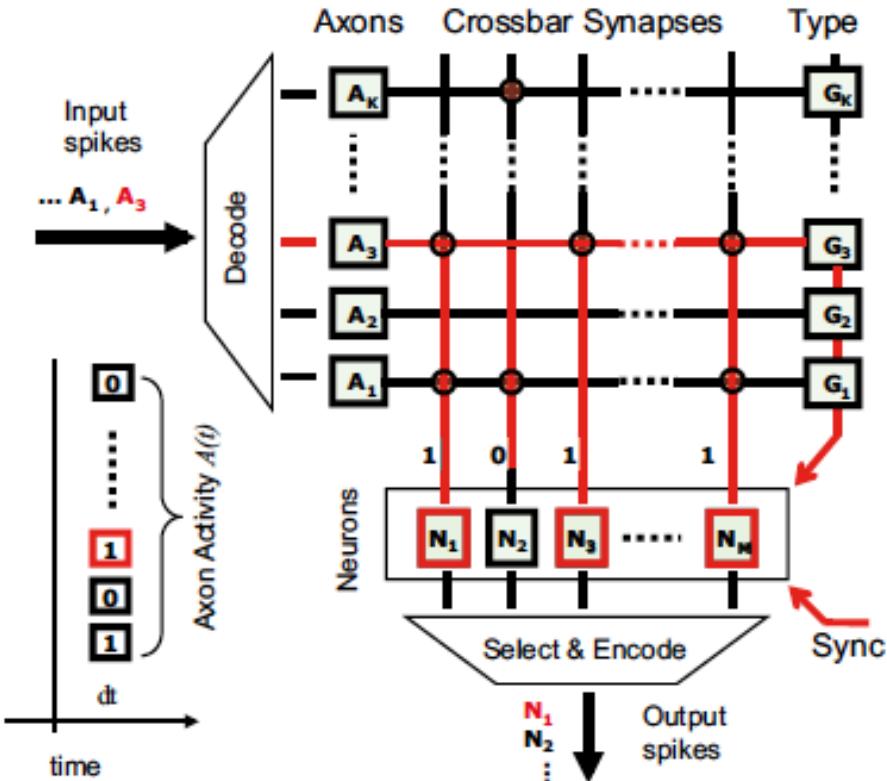
- COI-T「ライフコース・データに基づく健康医療情報プラットフォームの構築と新しいパブリックヘルスの実現」
  - ▷ 医療データの機械学習による解析(京大奥野先生)
- ビッグデータCREST「医薬品創薬から製造までのビッグデータからの知識創出基盤の確立」(東大工・船津公人代表)
  - ▷ 大規模化合物ライブラリの生成と統計解析
- 大規模ANNが使えそうになっている
- 低電圧素子の応用に向いているのでは

# 人工神経回路用LSI・ハードウェア

- SyNAPSE (IBM)
- FACETS (U. Heidelberg)
- Neurogrid (Stanford)
- Many FPGA-based works
  
- どちらかというと、ANNでなく「神経回路のシミュレーション」がターゲット
  - ▷ FPGAではANNが中心

# SyNAPSE

- IBM, Digital
- 45pJ/synapse
- 250K synapse
- RBMをマップ  
但し、学習・  
Classifierはホスト

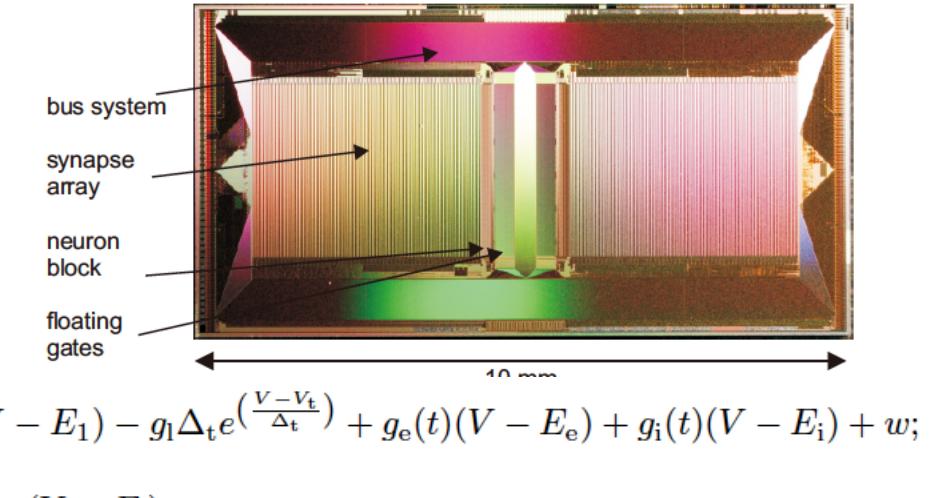


$$V_i(t+1) = V_i(t) + L_i + \sum_{j=1}^K \left[ A_j(t) \times W_{ji} \times S_i^{G_j} \right].$$

Merolla, Paul, et al. "A digital neurosynaptic core using embedded crossbar memory with 45pJ per spike in 45nm." Custom Integrated Circuits Conference (CICC), 2011 IEEE. IEEE, 2011.

# FACETS Neuromorphic hardware

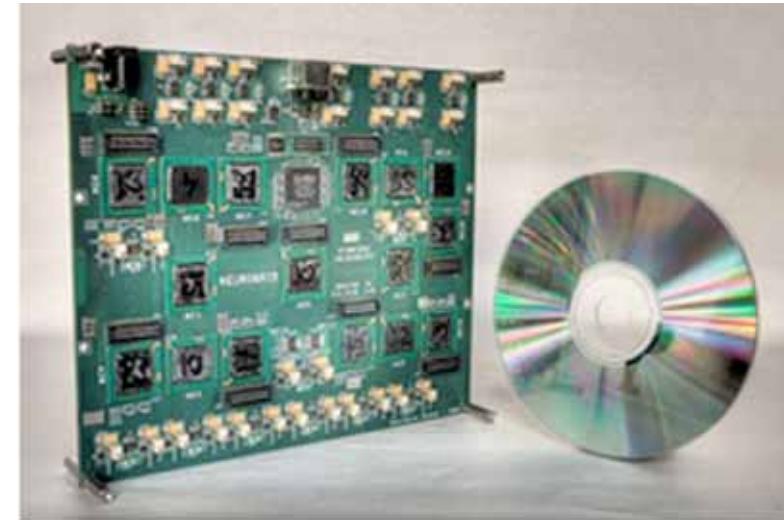
- U. Heidelberg
- Analog/Digital mixed
- HICANN chip
  - ▷ 180nm
  - ▷ 512 neurons, 224 synapses/neurons
- Wafer-scale integration
  - ▷ 384 HICANN/wafer
  - ▷ 200k synapses



$$-C_m \frac{dV}{dt} = g_l(V - E_l) - g_l \Delta_t e^{(\frac{V - V_t}{\Delta_t})} + g_e(t)(V - E_e) + g_i(t)(V - E_i) + w;$$
$$-\tau_w \frac{dw}{dt} = w - a(V - E_l).$$

# Neurogrid

- Stanford
- Analog/Digital mixed
- 180nm
- $10^6$  Neurons with 2 compartments,  $10^9$  synapsesを“Real-time”で計算、5W
- GRAPE-6に触発されたと書いてある…



Choudhary, Swadesh, et al. "Silicon neurons that compute." Artificial Neural Networks and Machine Learning—ICANN 2012. Springer Berlin Heidelberg, 2012. 121-128.

<http://www.stanford.edu/group/brainsinsilicon/neurogrid.html>

# 脳型コンピュータのねらい

## ■ 脳型コンピュータで、普通の計算機でできないことができる？

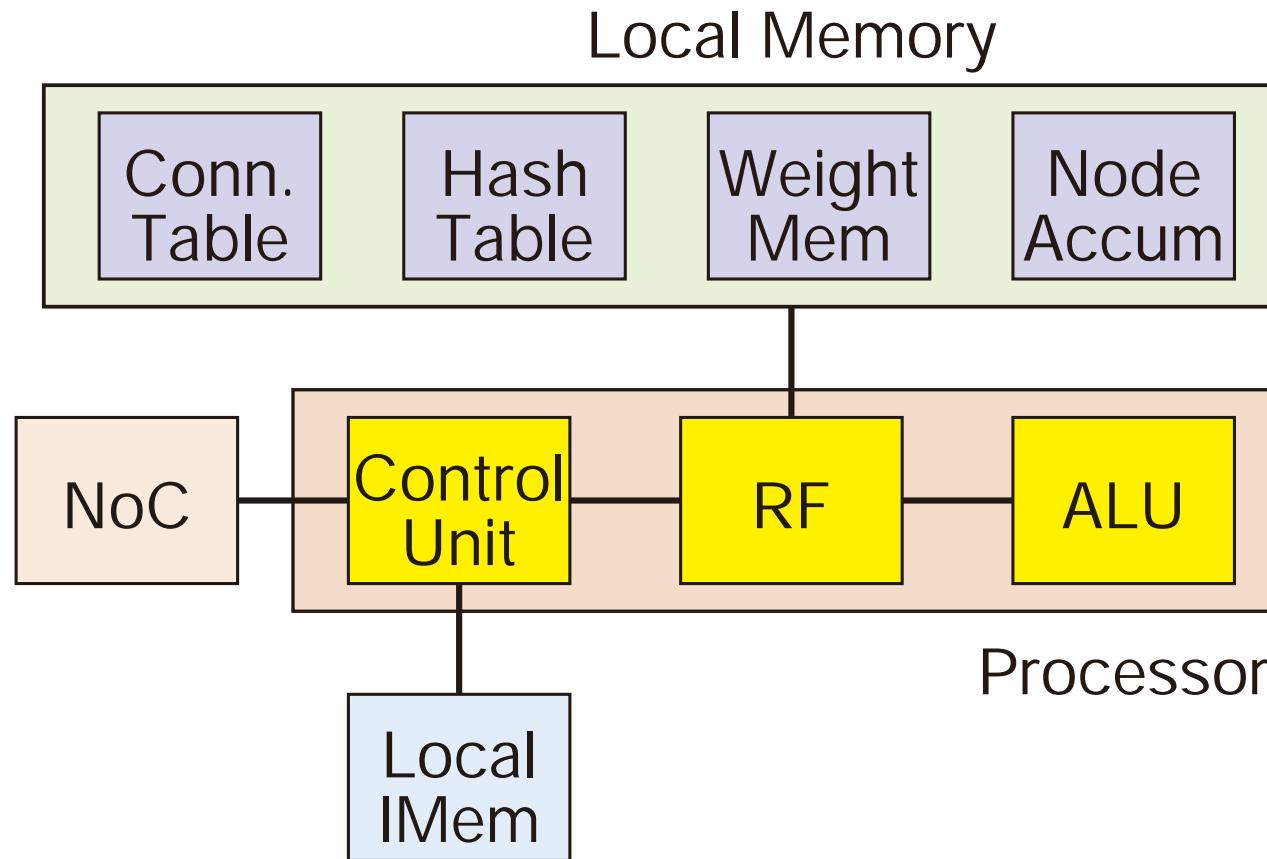
- ▷ というわけではない。×
- ▷ 大規模化・Embeddedの可能性は○

## ■ ポイント

- ▷ ANNにより注力した最適化
  - ▷ 精度の最適化
  - ▷ ネットワークの最適化
- により、高い性能を目指す



# ANN向けプロセッサ



- ・メモリ側の専用化
- ・特殊演算
- ・NW等からのイベント駆動の高速化

# 計算精度の問題

- FPGA実装では、様々な精度での実装報告あり
  - ▷ 16bitで問題ない報告
- A/D Mixed, STDP modelの場合、weight 4bitで十分な報告
- 学習方法にも依存すると考えられるので、今後さらに調べる

Moussa M, Areibi S, and Nichols K. "Arithmetic precision for implementing BP networks on FPGA: A case study." In "*FPGA Implementations of Neural Networks*" Ed. Omondi AR and Rajapakse JC. Springer-Verlag (2006).

Pfeil T, Potjans TC, Schrader S, Potjans W, Schemmel J, Diesmann M, and Meier K. "Is a 4-bit synaptic weight resolution enough? – Constraints on enabling spike-timing dependent plasticity in neuromorphic hardware." *Frontiers in Neuroscience* 6 (2012).

# ネットワークの課題

- LSI上のオンチップネットワーク
  - ▷ 2次元的な実装の制約を受ける
  - ▷ 2D mesh or 2D Ring bus
- 小パケットに特化
- 画像認識などregular topologyであれば比較的簡単
- 3次元実装・3次元ネットワークも将来的にあり得る

# 電力削減率

- 8bit程度まで精度を落とせるとすると、通常の32bitコアの1/5～1/10程度と予想される
- Networkの消費電力が課題
  - ▷ wire 1ch/mmあたり～0.2pJ/bit
  - ▷ 1um pitchで縦横に敷き詰めるとすると、2k wire/mm<sup>2</sup>
  - ▷ 160mm<sup>2</sup> で64nJ/bit, 1GHzとして 64W
  - ▷ 8k coreとするとCore当たり面積0.14x0.14mm<sup>2</sup>, 帯域35GB/s

# さらに消費電力を減らすには

- より電圧を下げる

- ▷ near-threshold, sub-threshold

- エラーの増大

- ▷ エラーを許容
  - ▷ 学習過程で自動的に回路の特性を考慮
  - ▷ 「ゆらぎ」の存在下で安定に動作する回路
  - ▷ 速度調整で非同期に動作なども考えられる

- エラーの存在前提で動く回路ターゲットとして、神経回路は向いている？

# GPUに勝てるか?

- Nvidia Echelon (10nm)
  - ▷ 20TF/chip, 50GF/W (System)
  - ▷ 1k core, 2.5GHz, 8 flops / core
- Target
  - ▷ 8k core, 1GHz, 8 ops/core, 64 TO/chip
  - ▷ Chip size ~ 200 mm<sup>2</sup>, 130W
  - ▷ ~ 500 GO/W
- 問題は実効性能:専用化、特にネットワークの最適化で高効率を目指す
- 中期的には、エラー許容などでさらに効率を上げられる可能性

# 脳型コンピュータの課題

- 動作アルゴリズムの確定
- 素子レベルでの電力削減(低電圧化)
- 誤差の存在下での動作解析・誤差の許容範囲の確定
- 回路レベルでの電力削減
- プログラミングモデルの作成
  
- 脳レベルの情報処理を同等エネルギーで実現