

IoT/AI時代の サービスプラットフォームを支える “新計算原理”

2017年3月7日

システム・情報科学技術ユニット

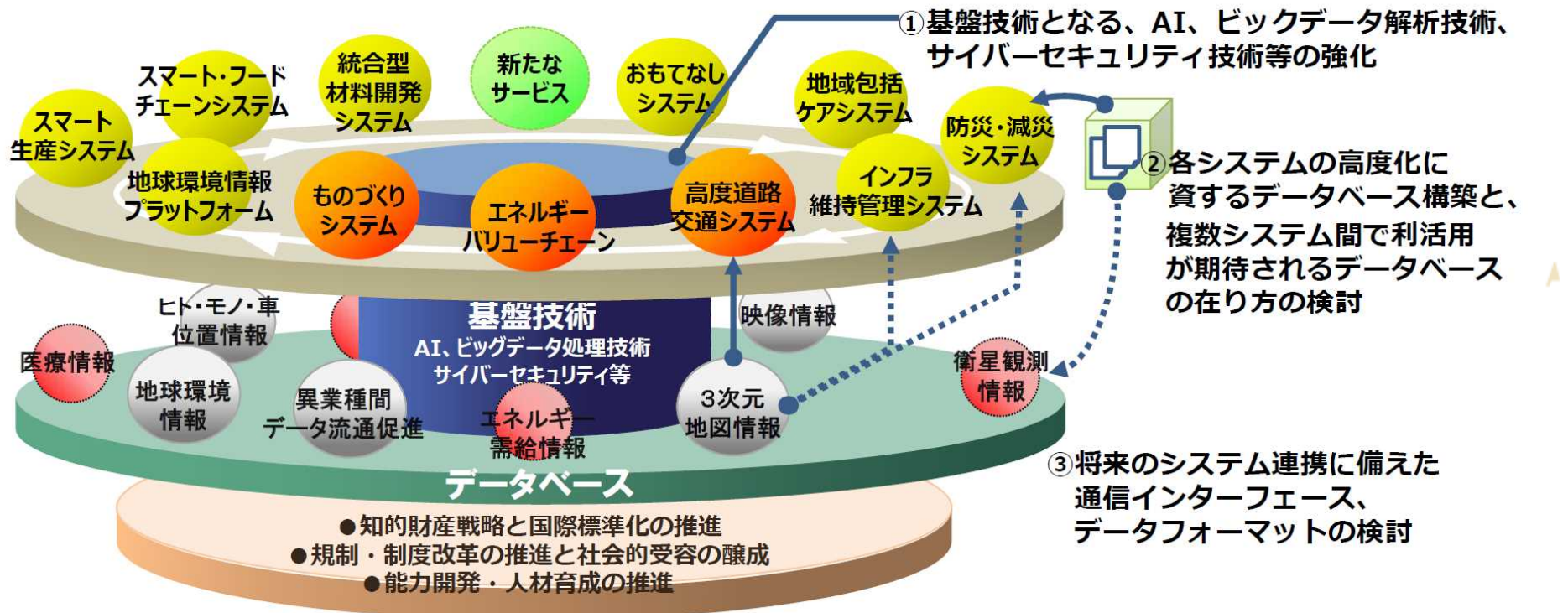
木村 康則



国立研究開発法人科学技術振興機構 研究開発戦略センター
Center for Research and Development Strategy Japan Science and Technology Agency

Society 5.0 サービスプラットフォーム

●「Society 5.0」プラットフォーム構築のイメージ



※今回取り上げたデータベースは参考例

1 1110 00

11

出典: 内閣府発表資料

<http://www8.cao.go.jp/cstp/tyousakai/juyoukadai/system/6kai/siryo2-1.pdf>

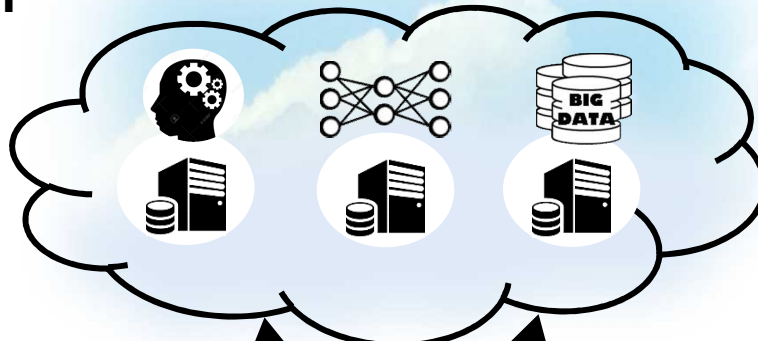
0011 1110 000

00 11 001010 1

11 1110 000

Society 5.0 実現アーキテクチャ

クラウド



機械学習/深層学習
ビッグデータ処理
高性能
大規模

インターネット



低レイテンシ
高バンド幅
データ処理
(Edge Computing)

デバイス

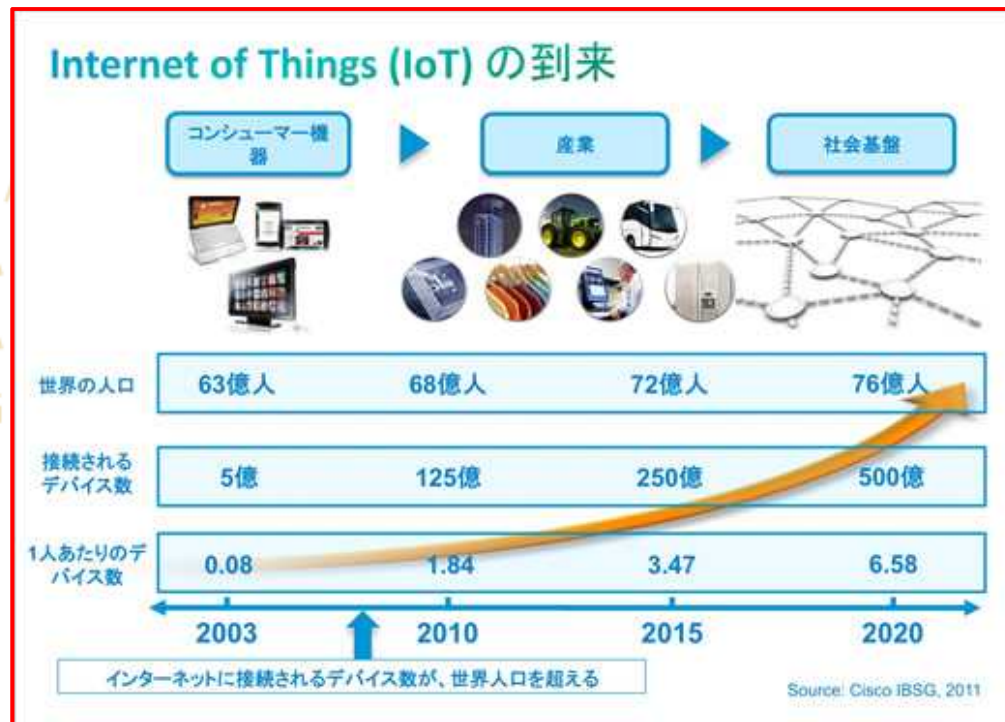


種々のデータ処理
リアルタイム
低消費電力
安価, 小型, 強靱

セキュリティ, プライバシー保護

IoT/AIデバイスの役割と数

- スマートデバイス中でローカルに処理できるもの
⇒ **組み込みAIを使ったエッジコンピューティングにより瞬時に高速処理。**
- より難しく複合的な問題
⇒ **クラウドに上げられて、クラウド側AIにより知的情報処理。**
- スマートデバイス、クラウド間のアーキテクチャは、目的、対象により変化。
⇒ **柔軟なシステム構成、短TATでのデバイスデザインが必要。**



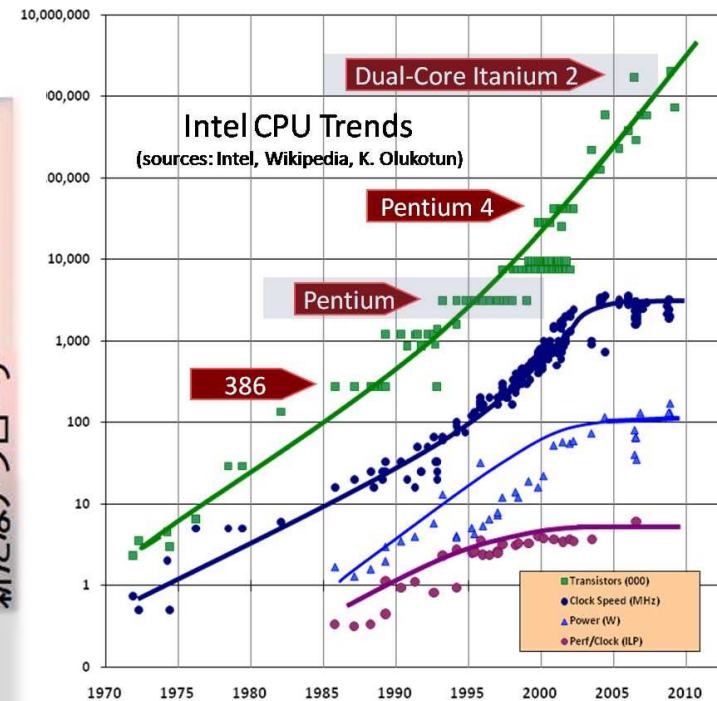
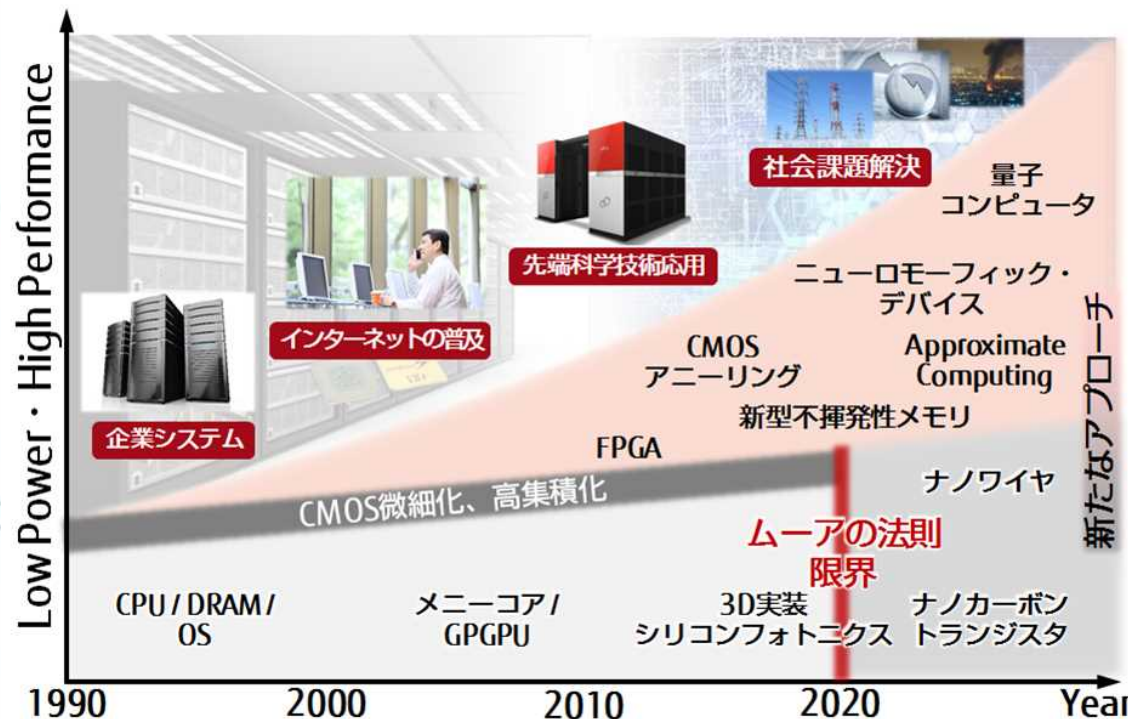
出典: Cisco 2011

高機能計算のニーズ

- ・ ムーアの法則の限界(微細化の限界, 電力消費の問題)
- ・ 製造コストの問題
- ・ 応用範囲(例. A.I.)と計算ニーズの爆発的拡大
- ・ 収集したデータ量の爆発的増大

⇒ **ノイマン型に代わる新たなアーキテクチャ
および数値計算方法の探求**

- ・ 2005年頃には既にプロセッサの周波数は頭打ちだった
- ・ 熱の限界: 1 W/mm^2



(左図の引用元) 富士通研究所プレスリリース(2016年10月20日) <http://pr.fujitsu.com/jp/news/2016/10/20-1.html>

(右図の引用元) <http://www.gotw.ca/publications/concurrency-ddj.htm>

実世界の学習に要する計算リソース

今後の機械学習/深層学習が必要とする計算リソース

学習を1日で終わらせるのに必要な計算リソース

画像/

バイオ・ヘルスケア

P:Peta
E:Exa

AlphaGOの例

DeepMind
AlphaGO
(2,000CPU, 300GPU)

(4勝1敗)

世界最高クラスの棋士
イ・セドル九段

250,000W

思考時:21W

休息時:20W

人工的に生成された10万時間の
音声データを基に学習 [Baidu 2015]

10台~1000台, 100日分の走行データの学習

100万台~1億台から得られた
データで学習する場合

機械学習、深層学習は学習データが大きいほど高精度になる
現在は人が生み出したデータが対象だが、今後は機械が生み出すデータが対象となる

各種推定値は1GBの学習データに対して1日で学習するためには
1TFlops必要だとして計算

10PF

100PF

1EF

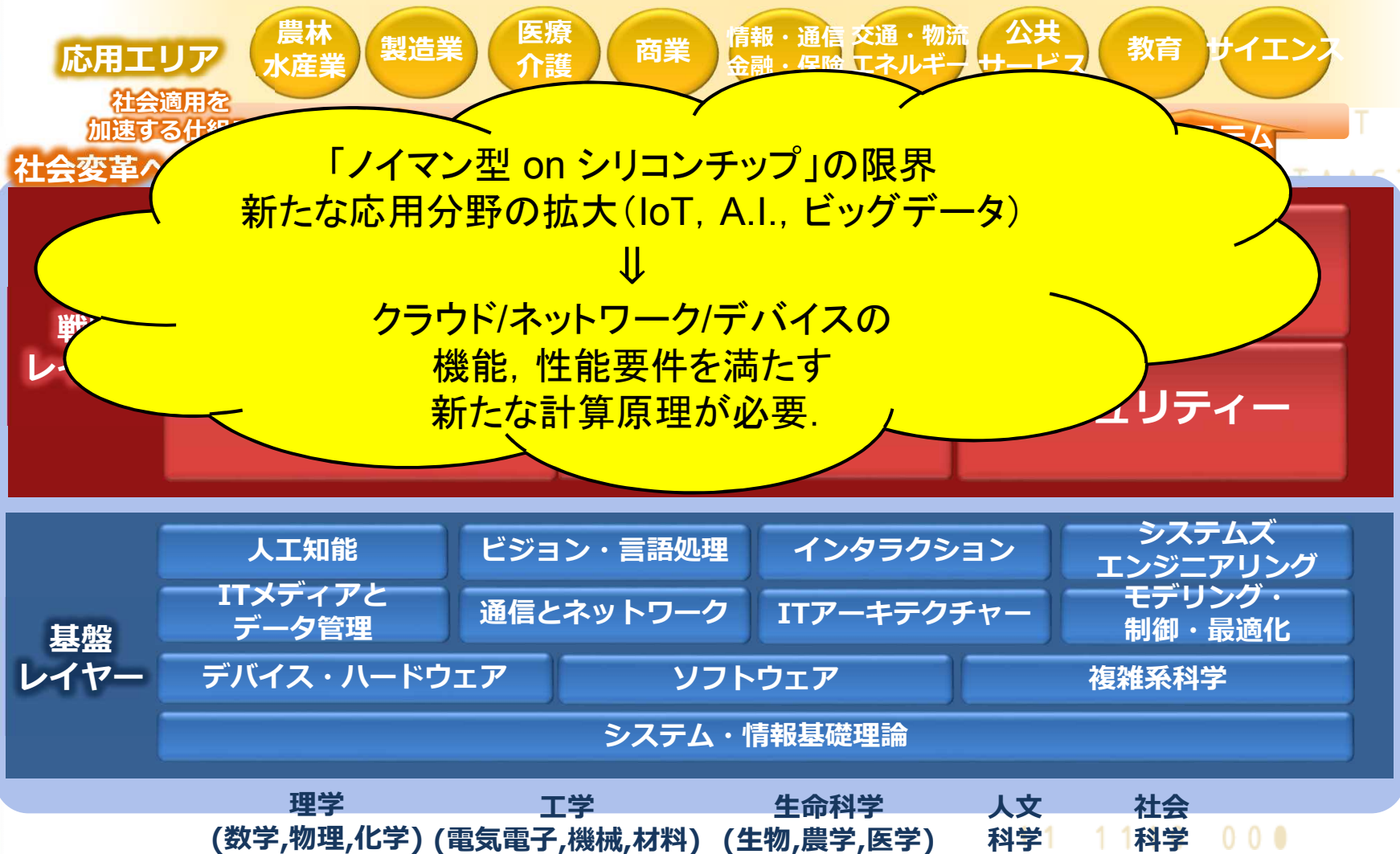
10EF

100EF

出典: 岡野原大輔(Preferred Networks), 実世界の人工知能 ~ 交通, 製造業,
バイオヘルスケア~, Interop 2016 (2016/6/8) 講演資料

システム・情報科学分野の俯瞰と戦略的研究領域

技術や社会・経済のトレンドを的確にとらえ、システム・情報科学技術を俯瞰し、
あるべき姿(Society 5.0)の実現に寄与する



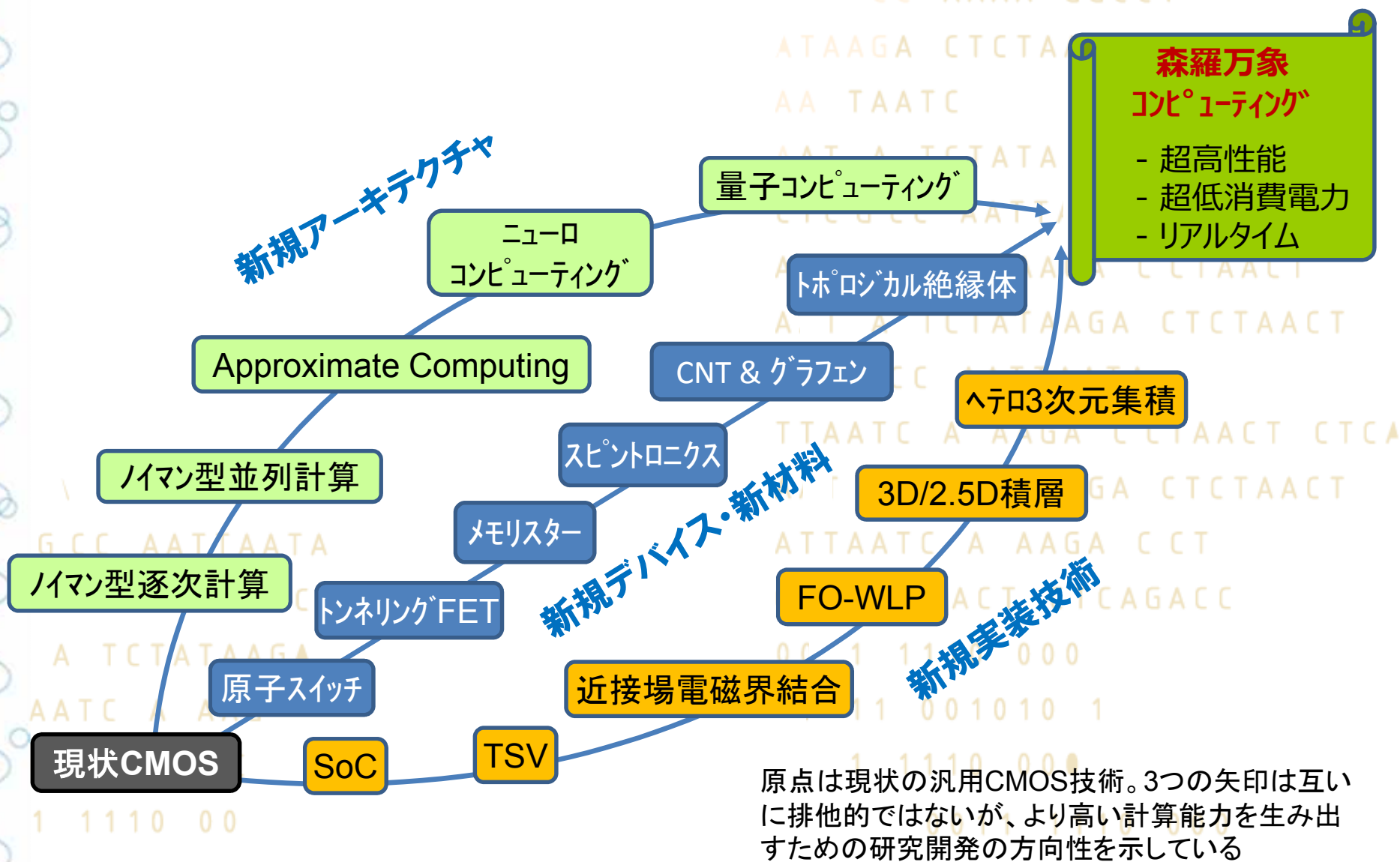
「新計算原理」の内容

森羅万象のデジタル化が劇的に進む超ビッグデータ時代に資する「新計算原理」を探求し、**既存の計算原理の性能限界を突破**するための基盤技術に関する戦略的研究開発を実施する。

具体的には、以下の3つの研究開発を主軸とする。

- ① **近似計算 (Approximate Computing)**を積極的に導入した「新原理計算」の探求と**高効率低エネルギー情報処理**デバイスシステムの**SW/HW協調によるコ・デザイン**
- ② **脳型コンピュータ・アーキテクチャの深化**：論理的・計算的な情報処理（左脳）モジュールおよび直観的・空間的な情報処理（右脳）モジュールの**SW/HW協調によるコ・デザイン**、ブレインモルフィックAI技術のための「新原理計算」の探求
- ③ **森羅万象コンピューティングのための新計算原理・設計指針**の確立および計算限界の解明による**森羅万象解析手法**の開発
(例)データフロー/データセントリック・コンピューティング

IoT/AIチップの革新、将来のコンピューティングへの技術潮流



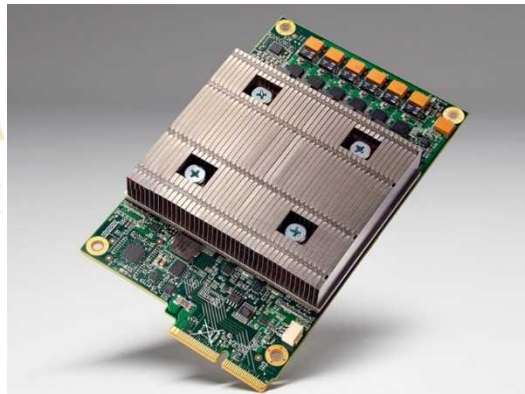
FET: Field Effect Transistor, SoC: System on Chip,
TSV: Through Silicon Via, FO-WLP: Fan-Out Wafer Level Package

Approximate Computing (積極的な近似計算)

AI では厳密な計算が必ずしも必要ない
計算の精度を落とす(エラーを許容する)ことで高速化・低電力化

米国Googleが深層学習専用プロセッサTPU(Tensor Processing Unit)公表
「消費電力当たりの性能はGPUの10倍」と主張 (2016年5月)

- ※ Google TPU (8bit 整数形式): 深層学習のためにGoogleが独自開発した **ASIC** (Application Specific Integrated Circuit)
- ※ Googleが2015年11月にオープンソースソフトウェア(OSS)として公開した機械学習ソフトの「**TensorFlow**」に対応
- ※ Googleが2016年3月にサービスを開始した機械学習のクラウドサービス「**Google Cloud Machine Learning**」に使用



- ※ 囲碁AIの「**AlphaGo**」にも使用
- ※ 浮動小数点数の計算の精度を抑えた設計 (8bit 整数形式)

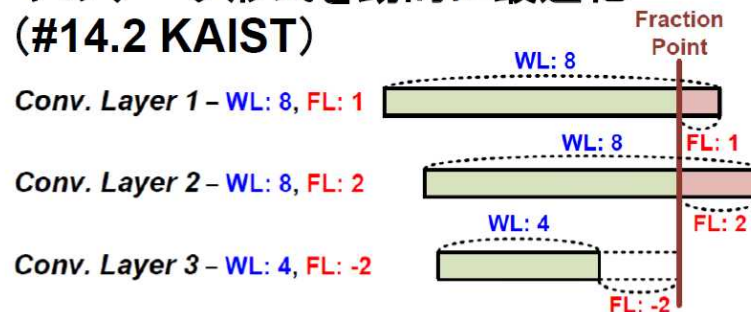
⇒ 消費電力当たりの性能が高い!



ISSCC2017のテーマ:スマート社会に向けたインテリジェントなチップ

ISSCC (International Solid-State Circuits Conference) 2017 Deep-Learning Processor セッション Approximate Computingの例

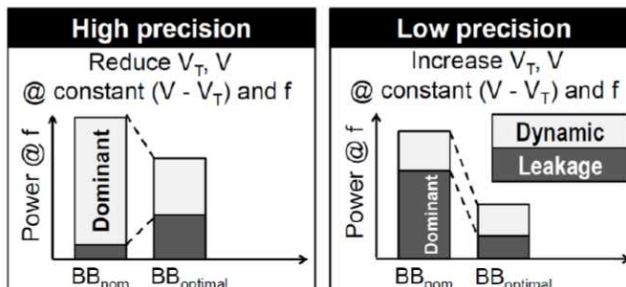
- 計算量の大半を占める畳み込みレイヤのデータ形式を動的に最適化 (#14.2 KAIST)



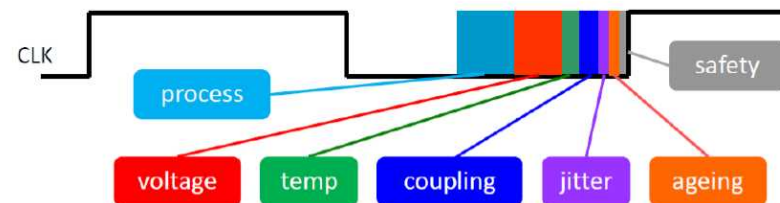
- 基板バイアスで V_{th} を増加。精度を犠牲にして低電力化。 (#14.5 KU Leuven)

Precision Scaling – BB in FDSOI

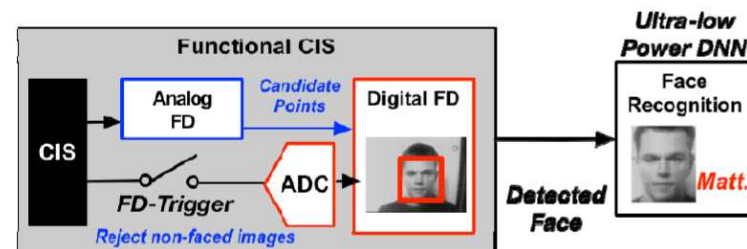
- DVAFS modulates leakage-vs-dynamic balance
- Body-Bias tuning allows minimizing energy



- 時々クリティカルパスが不良することを許容 (#14.3 U. Michigan)



- 顔認識で低精度・低電力のアナログフィルタ(予選) + 高精度のデジタルフィルタ(決勝)を組み合わせ 40%低電力化 (#14.6 KAIST)

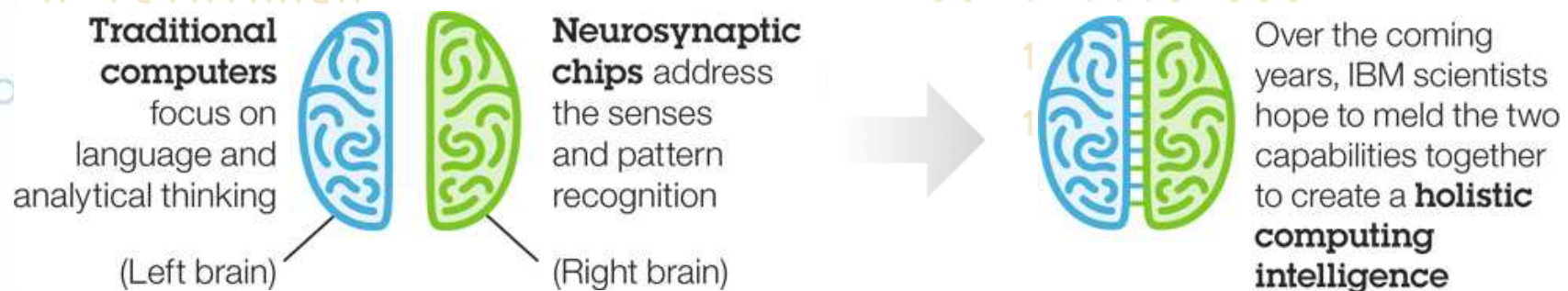


脳型コンピュータ・アーキテクチャの深化

IBMは、脳型チップ「SyNAPSE」(開発コード名はTrueNorth; ネズミ並みの数のニューロンとシナプスを備えた非ノイマン型プロセッサ)を発表 (2014 年夏)

- SyNAPSEの最大の強み: 画像認識や音声認識といった高度なパターン認識を極めて低い消費電力(約0.07W/chip)で実行可能
- IoT等のセンサーネットワーク末端におけるエッジコンピューティングに向けた用途開発の加速が目的! ?
- 2016年3月末, 米国ローレンスリバモア国立研究所はIBM リサーチと協力し, IBMのニューロシナプティックコンピュータチップSyNAPSEを使って, 脳からヒントを得たディープラーニングのための極低エネルギー効率スーパーコンピューティングプラットフォームの開発を開始

【Holistic Computing Intelligence (IBM)】 論理的・計算的な情報処理(左脳)モジュールと直観的・空間的な情報処理(右脳)モジュールの統合



(引用元)【Brain Power】 Scientists at IBM research unveil a brain-inspired computer and ecosystem
<http://www.research.ibm.com/cognitive-computing/brainpower/>

「新計算原理」の日本人先駆者

•量子アニーリング

- ✓ 1998年発表
- ✓ 西森秀稔教授, 門脇正史氏(東京工業大学)
- ✓ 1998年の門脇・西森の論文で、最適化問題のための汎用的な基本アルゴリズム「量子アニーリング」を提案した.
- ✓ D-Waveの量子コンピュータで実現.



西森秀稔教授 (Wikipediaより)



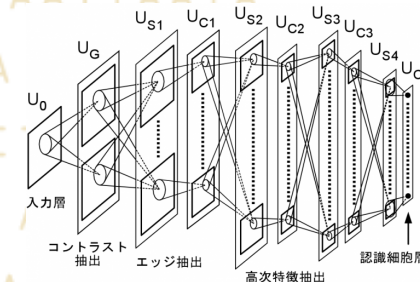
「NASA Ames Research Center」に設置された「D-Wave 2X」
<http://itpro.nikkeibp.co.jp/atci/news/15/120904017/>

•神経回路モデル: ネオコグニトロン

- ✓ 1979年発表
- ✓ 福島邦彦氏 (NHK技研, 大阪大学)
- ✓ パターン認識ネットワークを使った文字認識に使用し, 後にこれが畳み込みニューラルネットワークへと発展した. そして, 2006年, ジェフリー・ヒントンらにより, 現在のディープラーニングへと発展した.



福島邦彦教授 (Wikipediaより)



ネオコグニトロンの回路構造

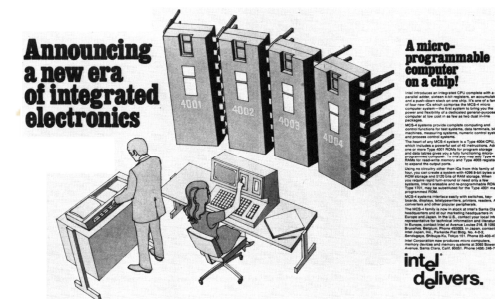
Copyright © 福島邦彦 2006, All Rights Reserved.

•世界初のマイクロプロセッサ Intel 4004

- ✓ 開発: 1971年3月
- ✓ 嶋正利氏 (ビジコン社)
- ✓ 動作周波数: 750KHz, チップサイズ: 3*4 mm
- ✓ 性能: 0.06MIPS, 1946年開発のENIACと同等性能
- ✓ インテルと共同開発し, 世界のコンピュータ産業に多大な影響を与えた.



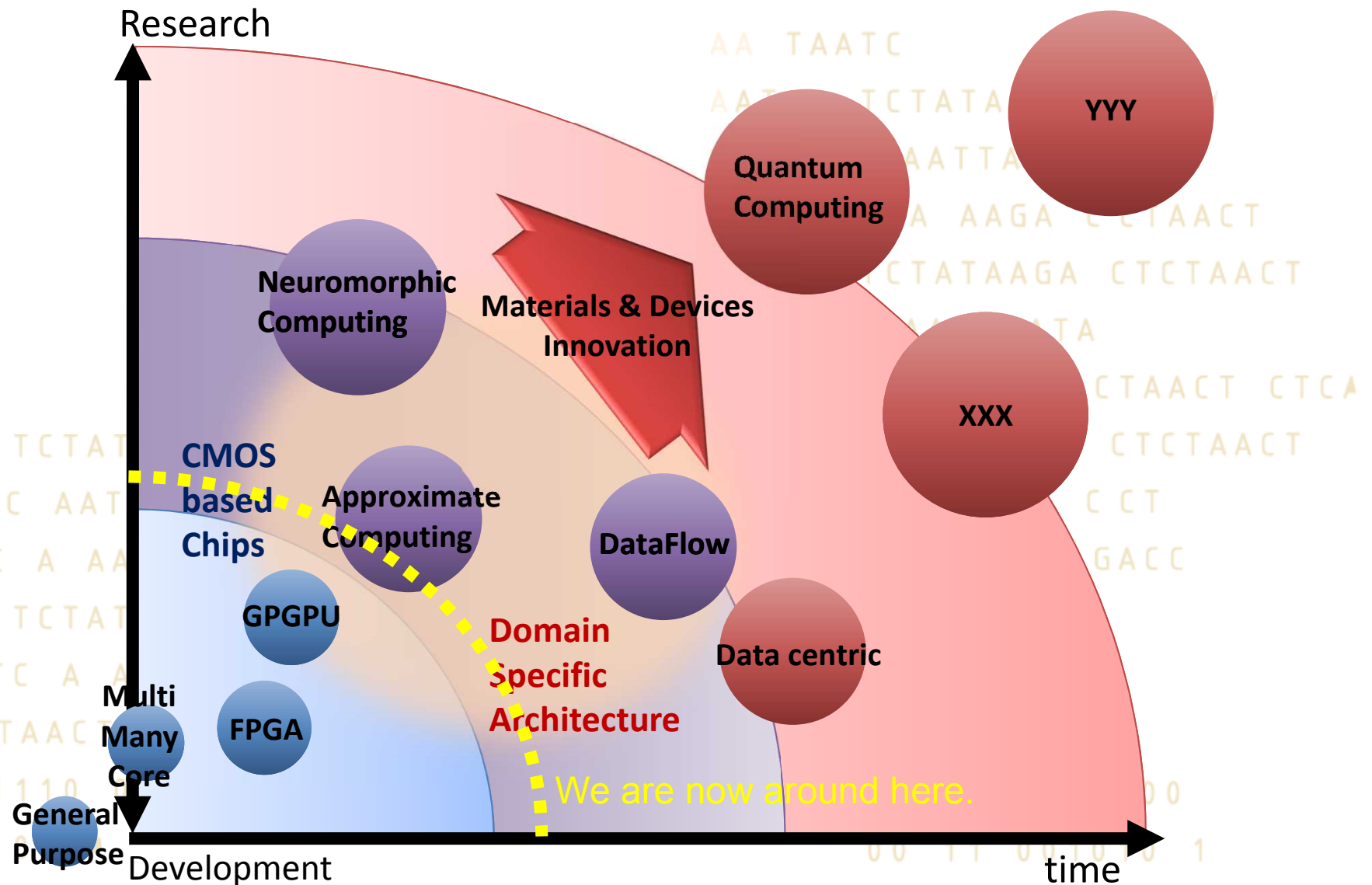
嶋正利氏 (Wikipediaより)



4004の発表当時の広告(1971)

[原出所] Electronic News, November 15, 1971

「新計算原理」R&D ポートフォリオ



まとめ

- Society 5.0実現に向けて新計算原理が必要な時代に
 - ✓ ムーアの法則, アムダールの法則による限界
 - プロセッサの性能向上の源泉
 - ✓ 新しい応用分野の出現
 - IoT, A.I.(3回目), ビッグデータ, etc.
- システム設計, デバイス設計の両方わかる人材の育成



ENIAC
真空管17,468本
167m²



System/360
1964年



Pezy
2015年



Sun WS
1982年



Multi Core Chip
Intel



Nvidia
2015年

新たな地平の開拓

写真はWikipediaから

AT A TCTATAAGA CTCTAACT

GA CCC

CC AAAA GGCCI

ATAAGA CTCTAACT CI

AA TAATC

AAT A TCTATAAGA CTCT/

CTCGCC AATTAATA

ATTAATC A AAGA CCTAACT

AAT A TCTATAAGA CTCTAACT

CTCGCC AATTAATA

TTAATC A AAGA CCTAACT CTCA

AAT A TCTATAAGA CTCTAACT

ATTAATC A AAGA CCT

GA CCTAACT CTCAGACC

00 1 1110 000

11 001010 1

1 1110 000

0011 1110 000

00 11 001010 1

11 1110 000

ご清聴ありがとうございました。

TCTATA

GCC AATTAATA

ATC A AAGA CC

A TCTATAAGA

AATC A AAG

CCTAACT C

1 1110 00

11 0010