

技術戦略研究センターレポート

TSC Foresight

Vol. 30

人工知能を支えるハードウェア分野の 技術戦略策定に向けて

2018年10月

1		
章	人工知能を支えるハードウェア分野の概要	2
	1-1 AI技術の開発経緯·····	2
	1-2 コンピュータの分類	3
	1-3 AIを支えるハードウェア ·····	5
人 章	人工知能を支えるハードウェア分野の置かれた状況	7
	2-1 市場規模	7
	2-2 技術開発の動向	7
	2-3 国内外の研究開発及び政策の動向	14
2		
う 章	人工知能を支えるハードウェア分野の技術課題	21
/		
4 ÷	+>+>h)?	94

TSC とは Technology Strategy Center (技術戦略研究センター)の略称です。

1章

人工知能を支える ハードウェア分野の概要

1

-1 AI技術の開発経緯

あらゆる物事がデジタルデータとして情報化されインターネットによって結びついてゆく「IoT (Internet of Things)」や、身の回りの実空間とこれに対応するデータによるサイバー空間とが連結したシステム「CPS (Cyber-Physical Systems)」が進展しつつあり、膨大なデータに裏付けられた情報化社会が現実化している。そして、この情報化社会の現実化には、半導体とコンピュータの進展が大きく貢献していることは言うまでもない。

1940年代にペンシルベニア大学で電子式のコンピュータENIACが開発され、続いてプログラムを内蔵したコンピュータ (ノイマン型コンピュータ) の原型となる EDVAC 等が開発された。その後、商用のコンピュータが発売され、コ

ンピュータの大型化が進展した。また、1950年代の集積回路の誕生から電子回路技術が進展し、1970年代にマイクロプロセッサが開発され、これによりコンピュータの進化が加速し、現在に至っている。

一方、コンピュータが発明された1940年代は、脳科学も大きく進展し、脳神経回路のモデルが作られた。その構造は、コンピュータとは大きく異なるものであったが、人間に匹敵する情報処理能力を得るべく、コンピュータによって脳機能をシミュレーションする、あるいは脳に近いアーキテクチャを持つコンピュータの研究が行われた(図1)。

最初のモデルは、Rosenblattが1958年に発表したパーセプトロンである。パーセプトロンの能力は限定的であったが、1980年代後半には、入力層と出力層の間の中間層である隠れ層を備えることによって、認識能力、学習能力を飛躍的に向上させたニューラルネットワークが登場した。そして、ニューラルネットワークによって知能を実現しようとするコネクショニズム研究が盛んに行われ、ニューラルネットワークを高速に計算するアクセラレータが開発された。

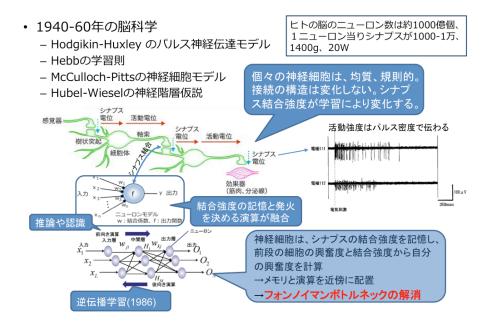


図1 脳科学とニューラルネットワーク

その後、マシンラーニング (Machine Learning:機械学習)として、数学的根拠が明確なSVM (Support Vector Machine) や確率モデルを使った隠れマルコフモデル、また幅広い探索を学習的に行う遺伝的アルゴリズム、あいまいな状況で的確な判断を下すファジー理論など、多数の提案がなされ、ニューラルネットワークはマシンラーニングの1つの事例 (インスタンス) になった。

2010年頃、隠れ層をさらに増やし、ニューロン数を数百万以上にまで高めることで、高い能力を発揮するディープニューラルネットワークが開発された。このネットワークによるマシンラーニング(ディープラーニング)は、他のマシンラーニングを凌駕し、AlphaGOのように最優秀な人の能力を超える能力を発揮するに至った。この能力拡大は、インターネット等から採集される学習用の膨大なデータと、並列処理に優れる画像処理チップGPU(Graphics Processing Unit)を活用して、膨大な繰り返し計算を行うことにより得られている。GPUは、元々グラフィックス用の演算チップとして、3次元の座標変換や画像を構成する多角形ポリゴンの高並列の演算向けに開発されたが、高並列演算機能を汎用計算にも適用できるように開発が進み、ディープラーニングの学習や推論の処理に適用されている。

近年、社会実装が進み、認識アルゴリズムを使った情報 処理によって膨大なデータから必要な情報を得ることができ るようになってきた人工知能 (AI: Artificial Intelligence) は、これらの技術成果の1つと言える。

1 -2 コンピュータの分類

一般に、現在のコンピュータは、ノイマン型コンピュータと言われ、メモリ中にプログラムとデータを同居させ、プロセッサは、メモリからプログラムを逐次読み出し、その命令コードにしたがって、メモリ中のデータを処理するという動作を繰り返す。そのため、メモリとプロセッサは分離して、伝達経路である「バス」によって結合されているが、このバス結合が情報の隘路となってコンピュータの実行速度を律速しやすいという欠点がある(ノイマンボトルネック)。そこで、プロセッサと専用のキャッシュメモリを組みにしたコアを多数用意するマルチコア、メニーコアといったアーキテクチャが開発されている。現時点では、ハイエンドPC向けに18コアのCPUが発売されており、また、サーバー向けとして28コアの技術が国際会議で発表されている。

これらのコアをさらに多数集めて太いインターコネクトで接続したのが、現在のスーパーコンピュータであり、CPU (Central Processing Unit) や GPUを使ったディープラーニングによる AIコンピュータである。これらのコンピュータは CMOS (Complementary Metal Oxide Semiconductor) 半導体により実現し、性能やコストが決まっている。現在は CMOS 半導体の性能向上を支えてきた微細化によるスケーリング則 (ムーア則) が限界に近づいており、将来のコンピュータ性能の向上を支える代替技術の必要性が高まっている。

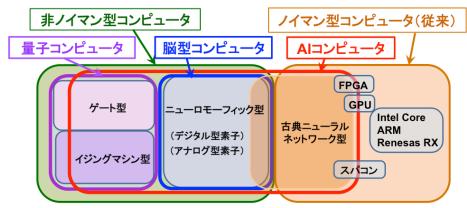
一方、ノイマン型とは異なる演算手法やアーキテクチャを特徴とする様々なコンピュータ(非ノイマン型コンピュータ)も研究開発されている。プログラムの逐次実行ではなく、必要データを揃えて演算を行うデータフローコンピュータや、動作時のハードウェア再構成が可能なリコンフィギャラブルコンピュータでは、CMOS半導体による研究開発が進んでいる。CMOS半導体を越えた取組として、脳の神経回路網の働きを模倣した脳型(ニューロモーフィック)コンピュータ

と、量子力学的な特性を応用して現状のスパコンを大幅に 上回る大規模な超並列の情報処理を可能にする量子コン ピュータがある。さらに、光コンピュータ、分子コンピュータ、 バイオコンピュータなどが研究されているが、これらにおい ては基本素子レベルの研究段階であり、複雑な演算を実 行するには至っていない。

図2に、現在のAI技術の中核であるディープラーニング

をサポートするコンピュータをAIコンピュータと定義し、AI 支援という視点で分類した結果を示す。ただし、基本素子 研究段階にあるコンピュータは省いた。

AIコンピュータは、その構成要素からノイマン型と非ノイ マン型のどちらも含んでいる。また、非ノイマン型は脳型コン ピュータと量子コンピュータの2種類から構成されるが、量 子コンピュータは全てが AI用ではない。



【語句定義】

ノイマン型コンピュータ:メモリにデータとプログラムを内蔵、メモリの命令を逐次取り出しプロセッサで実行 非ノイマン型コンピュータ:ノイマン型以外 AIコンピュータ:機械学習、組合せ最適化等の演算処理を行う(プロセッサ、メモリ等の集合体)

脳型コンピュータ(ニューロモーフィック): ニューロン・シナブスのような脳機能を模倣するデジタル素子やアナログ素子で演算処理を行う

量子コンピュータ(ゲート型): ロジック演算を行う

量子コンピュータ(イジングマシン型):アニーリング等、組合せ最適化問題に特化した演算を行う

図2 AIを支えるコンピュータの分類

出所:「AI白書2017」を基にNEDO技術戦略研究センター作成(2018)

1

-3 AIを支えるハードウェア

今後の情報化社会では、AI活用が重要な技術となるが、 その実現には、膨大なデータの情報処理 (コンピューティング) が必要である。しかし、従来手法でこの処理を実現できない状況が想定され、その限界を超えなければならない。

(1) 脳シミュレーション専用マシン

対応策の1つは、既存のコンピュータの高機能化である。 現在のデジタルコンピュータは、万能機械であり、神経細胞や脳の機能が解明されれば、その働きを忠実にシミュレーションすることができる。図3は、理化学研究所で行われたスーパーコンピュータを使った脳シミュレーションの規模拡大の過程を示している。しかし、ニューロン数でも、計算時間でも、まだ脳には追いついていない。脳シミュレーションの専用マシンを作ることにより、この限界を突破できると期待されている。

(2) AI技術の高度化

もう1つの対応策としては、AI技術の高度化が考えられる。 特に、脳型(ニューロモーフィック)コンピュータや量子コンピュー タは、そのポテンシャルから期待度が高い。また機械学習を高 度化する確率モデルコンピューティングも注目されている。

①脳型コンピュータ

脳モデルに近い計算を行う脳型 (ニューロモーフィック) コンピュータは、脳のニューロンの模擬を、デジタル型素子で行うか、アナログ型素子で行うかの2つに大別される(表1)。

デジタル型素子では、ニューロンへの入力の積算が閾値を超える(発火)と、次のニューロンへと情報が伝達されて行き、結果は学習情報としてメモリに格納される。一方、アナログ型素子では、ニューロンへの入力となるアナログ電圧の比較により発火に至り、その学習情報は、例えばプログラマブルな抵抗素子特性で表現される。

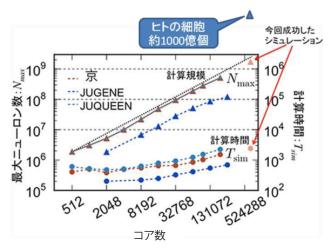


図3 スーパーコンピュータによる脳シミュレーションの規模

出所: 理化学研究所資料 (http://www.riken.jp/pr/topics/2013/20130802_2/) に NEDO 技術戦略研究センター追記 (2018)

表1 脳型 (ニューロモーフィック) コンピュータの特徴比較

	古典ニューラルネット	脳型コンビュータ	
演算処理	デジタル	デジタル型素子	アナログ型素子
演算素子	CMOS	CMOSベース	メモリスタ, ReRAM, 相変化メ モリ, スピントロニクス, etc.
消費電力 (相対比(例))	高(100000)	小 (10)	極小(1)
問題記述	プログラム	結合強度 (オフチップ学習)	結合強度、しきい値など (オンチップ学習)
開発フェーズ	商用チップ〜システム開発	素子〜システム開発	素子〜チップ開発
原理イメージ	論理回路 AND OR	ニューロン-シナブス模 ×: ************************************	шл (ca)

②量子コンピュータ

量子コンピュータは、ゲート型とイジングマシン型に大別できる(表2)。現在、その適用範囲を含めて各国で研究開発が進んでおり、将来、AIコンピュータの一端を担う可能性がある。

量子コンピュータは、計算単位として従来の0または1のビットではなく、0または1または量子力学的な重ね合せ状態を確率として持つ量子ビット(Qubit)を用いる。量子ビットの実現には磁束の向きによる超伝導量子ビットなどが開発されてきており、現在、コンピュータとして開発が進んでいる段階である。量子ゲート型は現行のコンピュータの上位互換に対応して高速処理を可能とし、イジングマシン型はイジングモデルを使った組合せ最適化問題に適している。

なお、量子コンピュータの定義については現時点で明確になっていない部分がある。狭義には、量子ビットを使い、量子ビットが量子もつれといった量子的相関関係を有した上で演

算処理を行うゲート型のようなものを指すが、広義には量子 ビットを使っているものを量子コンピュータに含める場合もある など、IEEE (The Institute of Electrical and Electronic Engineers:米国電気電子学会)においても議論が行われ ている。ここでは、量子コンピュータは量子ビットを用いて量 子的な振る舞いを演算処理に応用したコンピュータとして捉 え、量子的な挙動を模したものも便宜的に含めている。

③確率モデルコンピューティング

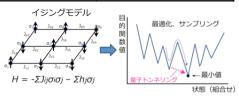
確率モデルを使ったコンピューティングは、複数のデータ群に対し、平均値等をその代表値として扱う、いわゆる決定論モデルとは異なる。複数のデータ群を確率分布として扱い、不確実性も含めることで、元データ群に対して、より妥当な結果を得ることができる方法として、ディープラーニング等で取り入れられつつある。

表2 量子コンピュータの特徴比較

		アニー	-リング型
	量子ゲート型	量子アニーリング	シミュレーテッドアニーリング
目的	現行コンピュータの 上位互換となる高速処理	組合せ最適化問題、サンプリング (イジングモデルから最適値を与える組合せ解を導出	
デバイス	超伝導量子ビット回路 スピン量子ビット(Si)	超伝導量子ビット回路	Si半導体(CMOS)
動作温度	約 -273℃	約 -273℃	室温
開発機関	IBM、Google、東大	D-WAVE(商用)、 Google、AIST	日立、富士通

アニーリングとは、焼きなまし と言われる処理 で、ある温度に加熱後にゆっくりと冷却することで、内部組織の均質化等を行う処理のこと。

組合せ最適化問題とは、離散的変数群のイジングモデルから目的関数の最小(大)値とその変数群を求める問題のこと。



2章 人工知能を支えるハードウェア 分野の置かれた状況

2 -1 市場規模

AIを支えるハードウェアの市場規模予測は難しいが、ここでは複数の予測データを基にしたAIハードウェアの世界市場規模予測を示す(図4)。AIハードウェアとしては、現時点では、ディープラーニング用を含むCPU、GPU、ストレージ等の関連チップが大部分を占めている。市場全体は2020年以降から大幅に増加しており、2025年の世界市場は約6.7兆円となり、その後も、さらに伸長すると見られる。

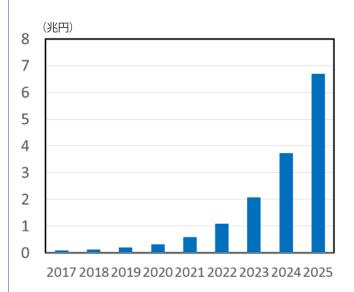


図4 AIハードウェアの世界市場予測

出所:複数の予測データを基にNEDO技術戦略研究センター作成(2018)

2 -2 技術開発の動向

AIを支えるハードウェア技術の研究開発において、今後 進展すると考えられる、以下の3分野の次世代コンピュータ 技術について、その動向を示す。

- 1) 脳型 (ニューロモーフィック) コンピュータ
- 2) 量子コンピュータ (ゲート型、イジングマシン型)
- 3) 確率モデルコンピューティング

(1) 特許出願の動向

図5に、各分野の特許出願件数の推移を示す。全体的には増加傾向にあり、特に、脳型 (ニューロモーフィック) コンピュータは2010年ごろから増加傾向となっている。



図5 各分野の特許出願件数の推移

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017)を基に NEDO 技術戦略研究センター作成(2018)

図6~図9に、各分野における出願件数のシェア上位国と上位機関をついて示す。いずれの分野も米国の出願が多く、Google、IBM、Microsoftなどの世界的に著名な企業

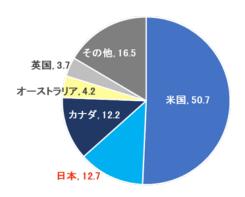
が上位にある。また、量子コンピュータ (特にイジングマシン型) ではカナダの出願が多く、D-Wave の特許出願が牽引している。



順位	出願人	出願件数
1	IBM(米)	167
2	Google (米)	100
3	Qualcomm(米)	90
4	国家電網公司(中)	62
5	三星電子(韓)	56
6	Microsoft (米)	55
7	浙江大学(中)	47
8	Siemens(独)	35
9	NEC(日)	34
10	Braincorp (米)	32

図6 脳型コンピュータの特許出願国別シェア及び上位出願機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基に NEDO 技術戦略研究センター作成 (2018)



順位	出願人	出願件数
1	D-Wave Systems (カナダ)	46
2	IBM(米)	42
3	Microsoft(米)	32
3	Northrop Grumman Systems (米)	32
5	東芝(日)	20
6	NTT(日)	19
7	Google (米)	9
8	Hewlett-Packard (米)	7
8	NewSouth Innovations (豪)	7
8	NEC(日)	7

図7 量子コンピュータ (ゲート型) の特許出願国別シェア及び上位出願機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基にNEDO技術戦略研究センター作成(2018)



順位	出願人	出願件数
1	D-Wave Systems (カナダ)	57
2	日立(日)	11
3	Microsoft (米)	9
4	Google (米)	7
4	情報・システム研究機構(日)	7
6	Rigetti Computing (米)	5
6	NTT(日)	5
8	ハーバード大学(米)	4
8	Northrop Grumman Systems (米)	4

図8 量子コンピュータ (イジングマシン型) の特許出願国別シェア及び上位出願機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基に NEDO 技術戦略研究センター作成 (2018)



順位	出願人	出願件数
1	IBM(米)	122
2	Microsoft(米)	69
3	Mathworks (米)	65
4	Siemens(独)	30
5	Hewlett-Packard (米)	26
6	NTT(日)	22
6	NEC(日)	22
8	Google (米)	20
9	Xerox(米)	19
10	Yahoo(米)	17

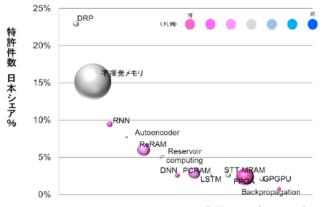
図9 確率モデルコンピューティングの特許出願国別シェア及び上位出願機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ, 2017) を基に NEDO 技術戦略研究センター作成 (2018)

AIに関連するハードウェア・ソフトウェアの要素技術についても特許出願は行われている。

例えば、DRP (Dynamic Reconfigurable Processor:動的再構成可能なプロセッサ)は、ディープニューラルネットワークで必要とされる並列処理等において、動的な演算器の再構成を行うことで演算器使用の効率と汎用性を高めることが期待されている。また、STT-RAMやReRAMは次世代のメモリ技術であるが、メモリ素子の構造から脳型コンピュータにおけるニューロンーシナプスモデルに近い機能を実現するものとして研究対象になっている。

図10に、これらの要素技術に関する特許件数の日本シェアを示す。DRPや不揮発性メモリでは、日本のシェアは20%前後ある。



期間:2007年~ 2016年

図10 AIコンピュータに関する各種要素技術別特許出願 件数の日本シェア

出所: Derwent Innovation ™での検索結果を基に NEDO 技術戦略研究センター作成 (2018)

(2) 論文発表の動向

図11に、各分野の論文発表件数の推移を示す。全体的には増加傾向にあり、特に2011年前後を境に、量子コンピュータ(ゲート型、イジングマシン型)の論文件数が増加傾向にある。



図11 各分野の論文発表件数の推移

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017)を基にNEDO技術戦略研究センター作成(2018)

図12~図15に、各分野における論文件数のシェア上 位国と上位機関を示す。いずれの分野も米国、中国が上 位を占めているが、量子コンピュータ (イジングマシン型) に おいてはカナダの論文件数が多く、D-Waveが研究機関別で2位となっている。



順位	研究機関	論文件数
1	中国教育部(中)	131
2	中国科技院(中)	128
3	西南大学(中)	108
4	フランス国立科学研究センター(仏)	105
5	南洋理工大学(シンガポール)	104
6	浙江大学(中)	96
7	ピッツバーグ大学(米)	79
8	華中科技大学(中)	76
9	香港城市大学(中)	66
10	カルフォルニア大学サンディエゴ(米)	64

図12 脳型コンピュータの論文発表国別シェア及び上位発表機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基に NEDO 技術戦略研究センター作成 (2018)



順位	研究機関	論文件数
1	中国科技院(中)	176
2	ウォータールー大学(カナダ)	155
3	シンガポール国立大学(シンガポール)	153
4	中国科学技術大学(中)	141
5	シンガポール量子技術センター(シンガポール)	131
6	オックスフォード大学(英)	116
7	フランス国立科学研究センター(仏)	112
8	清華大学(中)	103
9	アメリカ国立標準技術研究所(米)	99
10	メリーランド大学(米)	98

図13 量子コンピュータ(ゲート型)の論文発表国別シェア及び上位発表機関

出所:平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基に NEDO 技術戦略研究センター作成 (2018)



順位	研究機関	論文件数
1	南カルフォルニア大学(米)	23
2	D-Wave Systems (カナダ)	16
3	情報科学研究所/南カルフォルニア大学(米)	15
4	ハーバード大学(米)	14
5	国立情報科学研究所(日)	12
5	アメリカ国立標準技術研究所(米)	12
5	メリーランド大学(米)	12
8	テキサスA&M大学(米)	11
8	オックスフォード大学(英)	11
8	中国科学院(中)	11
8	インスブルック大学(オーストリア)	11
8	シドニー大学(豪)	11

図14 量子コンピュータ (イジングマシン型) の論文発表国別シェア及び上位発表機関

出所: 平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基に NEDO 技術戦略研究センター作成 (2018)



順位	研究機関	論文件数
1	フランス国立科学研究センター(仏)	243
2	マサチューセッツ工科大学(米)	161
3	中国科学院(中)	154
4	インペリアル・カレッジ・ロンドン(英)	152
5	清華大学(中)	149
6	チューリッヒ工科大学(スイス)	139
7	カーネギーメロン大学(米)	126
8	オックスフォード大学(英)	122
9	テキサス大学(米)	110
10	イリノイ大学(米)	105
10	ジョージア工科大学(米)	105

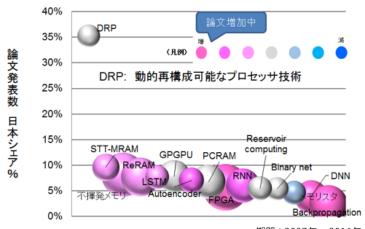
図15 確率モデルコンピューティングの論文発表国別シェア及び上位発表機関

出所: 平成29年度「非ノイマンコンピューティング及び量子技術に関する技術動向調査」(三菱ケミカルリサーチ,2017) を基にNEDO技術戦略研究センター作成(2018)

図16に、AIに関連するハードウェア・ソフトウェアの要素 技術における論文数の日本シェアを示す。日本シェアが最も 高いのがDRPであり、約35%のシェアとなっている。メモリ 関係など他の要素技術は5~10%程度が多い。

(3) 標準化の動向

AIを支えるハードウェアに関する前述の技術領域におけ る標準化は現時点で行われていない。AIの用語と概念に ついては、ISO/IEC JTC 1/SC 42 Artificial Intelligence で、2018年4月より標準化検討が開始されている。



期間:2007年~2016年

図16 AIコンピュータに関する各種要素技術別論文発表件数の日本シェア

出所: Web of Science [™]での検索結果を基に NEDO 技術戦略研究センター作成 (2018)

2

-3 国内外の研究開発及び政策の動向

(1) 米国の研究開発状況

米国では他の地域と同様に、大学や研究機関のプレイヤーが多いが、特に企業のプレイヤーが多くの成果を発表しているのが特徴である。

①脳型コンピュータ

表3に米国における脳型コンピュータの主な取組例を示す。 IBMは、2014年に脳型コンピュータとなるデジタル型の ニューロモーフィックチップ TrueNorthを世界で初めて発 表した。このチップは、100万個のニューロン、2.56億個のシナプス結合に相当し、54億個のトランジスタ、4096のコアで形成され、非同期式動作により消費電力は70mWと低い。これは、米国政府のDARPA (Defense Advanced Research Projects Agency:米国国防高等研究計画局)によるプログラムSyNAPSE (Systems of Neuromorphic Adaptive Plastic Scalable Electronics)の中で開発された。

Intelは2017年に新たなデジタル型のニューロモーフィックチップLoihiを発表した。こちらは14nmプロセスで製造され、1,024個のニューロンからなるクラスタが128個搭載され、脳の基本的な仕組みを模したデジタル回路となっている。

表3 米国における脳型コンピュータの開発状況

タイプ	機関	名 称	アーキテクチャ、特徴	概 要	時 期
デジタル型	IBM	True North	データフロー型ニューロン100万 個、シナプス結合2.56億個、54億個 のトランジスタ、SRAM	順方向認識処理のみ。 チップ当たり70mW、16チッ プ接続可能	2014年 発表
	Wave Computing	DPU	16nm FinFET, 16,000コア、32GB、 512Gバイト DDR4	独自のデータフロー型	2017年 4Q量産
	Intel	Loihi	14nmプロセス、1,024個のニューロ ンからなるクラスタ128個	デジタル回路のニューロモー フィックチップ	2017年 発表
アナログ型	IBM		PCM(GeSbTe)	シナプスをPCMで構成、蓄積-発火型ニューロンの挙動を再現。1チップ上に400万セルを集積化、FPGAと組み合わせて実装	2016年 5月
	UCSB デンソー		メモリスタ(Al2O3/TiO2)	CMOS型ニューロンの上に 12x12クロスバー構造シナ プスを形成、2030年車載用 実用化目標	2015年 12月
	HP,Univ.Utah Michigan Univ.		メモリスタ	32nmプロセスでチップ面積 85.4mm ² メモリスタを用いた画像認識 への取組	2016年 12月

②量子コンピュータ

表4に米国(含むカナダ)における量子コンピュータの主 な取組例を示す。

D-Wave (カナダ) が2011年に世界で初めて量子アニーリング理論に基づく量子コンピュータ (スピンを使った量子アニーリングマシン) を製品化しており、2017年には2,048量子ビットを持つ「2000Q」を発表している。極低温環境が必要であるが、既存のコンピュータを大きく超える性能を実現している。従来、演算時間が膨大になっていた巡回セールスマン問題やナップサック問題 (ナップサックに食料、用具等の品物を詰める際、ナップサックの容量の範囲内でできるだけ有用な品物を選ぶこと) などのいわゆる「組合せ最適

化問題」を高速に解くことが期待されている。いくつかの機関で導入され、航空制御システムのバグ特定、大都市の交通量最適化制御、タンパク質構造解析、宇宙探査機軌道最適化等の分野で適用検討を進められている。特定の用途分野に限定されるものではなく、広く適用可能性が模索されている。

IBMはゲート型の量子コンピュータIBM Qを開発し2016年5月からクラウド上でのオンラインサービスを提供している。現時点で20 量子ビットの装置を提供しており、2017年には50量子ビットの装置を開発している。

Googleも2018年に、同じくゲート型で72量子ビットの量子コンピュータの開発を発表している。

表4 米国(含むカナダ)における量子コンピュータの開発状況

タイプ	機 関	概 要	時 期
イジング	Google	100量子ビットの量子コンピュータ開発	2017年以降
マシン型	D-Wave Systems (カナダ)	世界初の量子コンピュータを発表2,048量子ビット搭載D-Wave2000Qを発売	· 2011年 · 2017.1.24
ゲート型	Google	72量子ビットの量子コンピュータを発表	2018年
	Intel	49量子ビットの量子コンピュータチップを発表	2018年
	IBM	20量子ビットの量子コンピュータのオンラインサービスを提供、 50量子ビットの量子コンピュータを発表	2017年

③ノイマン型AIコンピュータ

表5に、米国におけるノイマン型AIコンピュータのハードウェア開発、表6にソフトウェア開発の主な取組例を示す。

米国では、多様な企業を中心にディープラーニングに関

連した先進的な取組が進んでおり、論文や特許に留まらず、ハードウェア、ソフトウェア、システムのレベルで技術開発をリードしている。

表5 米国におけるノイマン型 AI コンピュータ (ハードウェア) の開発状況

機関	名 称	アーキテクチャ	実績•特徴	出荷時期
NVIDIA	Pascal	153億個トランジスタ、 16nm FinFET 16GBのDRAMを720GB/s接続	AI研究のスタンダード、産総 研ABCIにも採用予定	出荷済み
Google	TPU (Tensor Processing Unit)	Google製NNシミュレータ TensorFlowのハードウェア化	低精度の浮動小数点演算で 高速化 Google翻訳、AlphaGo, スト リートビューやクラウドサービ スで採用中	外販しない
MICROSOFT	Catapult	17万のALMと1,600のDSPを集積 したALTERA製FPGA	Bingサーチに使用	
Intel (Nervana)	Lake Crest	HBM2 独自のインターコネクト、 1Tバイト/秒	深層学習フレームワーク "Neon"用意	2017年 特定顧客向出荷

出所:公開資料を基に NEDO 技術戦略研究センター作成 (2018)

表6 米国におけるノイマン型 AI コンピュータ (ソフトウェア) の開発状況

名 称	開発者、頒布	言語•プラットフォーム	普及状況
TensorFlow	Google	Linux、MacOS、GPU可	ベンチマークも含め、標準
	オープンソース	Python、C++	Google翻訳などにも使われる
Torch	Deep Mind (Google) オープンソース	Lua (言語) 、 Google Cloud Platform	AlphaGOが使用、並列化が強力
Caffe	UC Berkeley	Windows, MacOS, Linux、	最も古く、コミュニティが大きく、サポートOSが多い。
	オープンソース	GPU可、C++, (Python)	画像認識専用で高速
Neural Network	Mathworks	Matlabの一部、GPU可	制御システムや組込みソフトウェアの開発ツールであ
Toolbox	商用		るMatlabのモジュールと組み合わせることができる

(2) 米国の公的研究開発支援プログラム

表7に、米国における主な公的研究開発支援プログラムを示す。DARPAから継続的に大きな資金が投入されている。また、IBMなど企業の取組が中心である。

表7 米国における主な公的研究開発支援プログラム

プログラム名と概要	機関	期間
SyNAPSEプロジェクト 概要 人間並みの規模の脳型コンピュータ・システムを消費電力1kW、体積2ℓ以下で実現すること を目標としている。DARPAから5,350万ドルの資金提供を受け開発。IBMから2014年にAIチッ プ「TrueNorth」が発表された。	DARPA	2008-2016
BRAIN initiative (Brain Research through Advancing Innovative Neurotechnologies) 概要 「アポロ計画」「ヒトゲノム計画」に匹敵する巨大科学プロジェクトとしてオバマ大統領が発表。スタンフォード大学は、Neurogridプロジェクトを実行。デジタル回路による実装ではなく、デジタル/アナログ混載システムにより、脳型コンピュータの実装を目指した。	NIH DARPA NSF	2016-2025
QEOプログラム (Quantum Enhanced Optimization) 概要 高度な性能を持つ量子アニーリングマシンの開発計画。ハイリスク・ハイリターンであるが、 焦点を明確に絞って膨大な研究資金を投入。米国を中心に世界中からトップクラスの研究者 を結集、量子アニーリングによる高度量子コンピューティングで米国の主導権を狙う。	IARPA	2017-2026

出所:公開資料を基に NEDO 技術戦略研究センター作成 (2018)

(3) 欧州の研究開発状況

①脳型コンピュータ

欧州では、HBP (Human Brain Project) と呼ばれる プロジェクトが進められており、脳型コンピュータシステムで 実験を行うためのプラットフォームを提供している。簡単な スパイクニューロンモデルによるエネルギー効率のよい大 規模なニューロンネットワークシミュレーションが可能となっ ている。

HBPは、2013年からの10年間のプロジェクトであり、本部はスイス連邦工科大学に置かれている。イギリスのマンチェスター大学の"SpiNNaker"システムと、ドイツのハイデルベルク大学の"BrainScaleS"システムの2つの補完的な大規模システムが成果として発表されて

いる。BrainScaleSはアナログ型で高速な演算を実現、 SpiNNakerはデジタル型で、リアルタイム性の高い演算を 実現している。

②量子コンピュータ

量子コンピュータでは、CEA-Leti (フランス) やデルフト 工科大学 (オランダ) 等で研究開発が進んでいる。

③ノイマン型AIコンピュータ

英国と韓国のGraphcoreが、DNNモデルをプロセッサ に内蔵したハードウェアを開発しており、GPUを凌ぐ性能を 発表している。

(4) 欧州の公的研究開発支援プログラム

表8に、欧州における主な公的研究開発支援プログラムを示す。長期的に大規模な資金が投入されている。米国とは異なり、大学が取組の中心である。

表8 欧州における主な公的研究開発支援プログラム

プログラム名と概要	機関	期間
Human Brainプロジェクト 概要 全欧州規模で実施される、最大規模の研究及び革新的開発を促進するためのフレームワークプログラム。脳型コンピュータとしてSpiNNaker、BrainScaleSプロジェクト、また量子コンピューティング、量子シミュレーション等に出資。	EU	2013-2022
BrainScaleSプロジェクト (Brain-inspired multiscale computation in neuromorphic hybrid systems) 概要 アナログ回路でニューロンやシナプスを再現する研究開発に取り組む。脳の1,000倍~10万倍程度の速度で動作する加速系のデジタル/アナログ混載システムを目指し、ニューロン512本とシナプス12万8,000個を再現したIC「HICANN (High Input Count Analog Neural Network)」を開発。	ハイデルベルグ 大学、他	2011-2014
SpiNNakerプロジェクト 概要 EUのHuman Brain Projectの一環として実施されたプロジェクト。人間の脳をコンピュータ で再現することを目標とし、SyNAPSEの専用チップではなく、ARMプロセッサコア等を用い てシステムを構築している。	マンチェスター 大学、他	2005-2014
The UK National Quantum Technologies Programme 概要 2013年にイギリス工学・物理科学研究会議 (EPSRC) に5年間で2億7,000万ポンド (約378億円) を量子コンピュータ関連研究費として出資。	オックスフォード大学、他	2013-2017
Quantum Technologies flagship概要欧州委員会の研究プロジェクト「Quantum Technologies FET-Flagships」は10年間で10億 ユーロ(約1,200億円)を量子コンピュータ関連に拠出。	EU	2018-2027

出所:公開資料を基に NEDO 技術戦略研究センター作成 (2018)

(5) 中国・韓国の研究開発状況

中国では、国策としてコンピュータや半導体の研究開発に注力している。特にスーパーコンピュータでは近年世界ランク1位であり、従来のインテルCPUから中国製CPUを採用しても1位を維持している。この流れの中で、精華大学等において脳型コンピュータの研究開発が進められており、IBMのTrueNorthチップの発表に続いて、翌年には同様の構造のTianji (天機)チップを学会発表するなど、精力的に開発が行われている。

韓国では、韓国科学技術院(KAIST)でCMOSのマルチコアプロセッサと組み合わせた脳型インテリジェントSoC (System on Chip)といった研究開発を進めている。サムスン電子やSKハイニックスにおいても米国企業や研究機関と連携を図りながら、研究開発を行っている。

(6)日本の研究開発状況

日本では、脳型コンピュータは要素技術の開発が中心となっている。また、光や半導体などを用いた大規模な量子コンピュータ (イジングマシン型) の開発で先行している。

①脳型コンピュータ

表9に、日本における脳型コンピュータの主な取組例を示す。

②量子コンピュータ

表10に、日本における量子コンピュータの主な取組例を示す。量子コンピュータにおいては、量子アニーリング理論

(1998年東工大)、超伝導量子ビット(1999年NEC)といった先駆的な成果が出ているが、近年の製品レベルの研究成果は米国、カナダ等の北米企業が先行している。また、量子コンピュータには分類されないが、イジングマシンの中にはSi半導体回路によるアニーリングも発表されている。CMOSで構成するスピン回路によりイジングモデルを実現した日立のCMOSアニーリングや、デジタル回路によるアニーリングを用いた富士通のデジタルアニーラなど、室温で最適化問題に適用できる実用的な技術が開発されている。

表9 日本における脳型コンピュータの研究開発状況

タイプ	機関	名称	アーキテクチャ、特徴	概 要	時 期
デジタル型	東芝	TDNN	神経細胞の興奮、結合係数を1 ビットに圧縮し、32kシナプス を1.9mm角に集積	興奮レベルをゲート遅延で表現し、加 算する信号処理で極端な省電力化	2016年 11月 学会発表
	トプスシステムズ NEDO次世代 人工知能	SMYLE deep	データフロー型 8 コア75MHz	低い動作周波数で消費電力を抑えつつ、最大480fpsで超高速画像認識処理が可能	2017年2月 プレス リリース
アナログ型	デンソー UCSB		メモリスタ(Al2O3/TiO2)	CMOS型ニューロンの上に12x12クロス バー構造シナプスを形成、2030年車載 用実用化目標	2015年 12月
	パナソニック、他		FeMEM(強誘電体メモリ)	CMOS型ニューロンの配線層にシナプス	2013年6月
	東北大		スピントロニクス	スピントロニクス素子36個とFPGAとの 組合せ	2016年 12月
	産総研、他 (NEDO IoT横断PJ)		アナログ型 抵抗変化素子	対TrueNorth 電力効率100倍、チップ 面積 1/20、28nmプロセス、100万個以 上のシナプスを集積化する技術を開発	2016年6月 開発開始
	NEC / 東大			ブレインモルフィックAI技術 東大合原教授が中核	2016年9月 共同開発発表

出所:公開資料を基にNEDO技術戦略研究センター作成(2018)

表 10 日本における量子コンピュータの研究開発状況

タイプ	機関	概 要	時 期	備 考
イジング	産総研、理研	3次元実装による高集積技術、1Mビット		NEDO(IoT横断PJ)で開発中
マシン型	NTT, NII	光パラメトリック発振器によるコヒーレントイジングマシン、2,048量子ビット室温動作	2016年	ImPACT
	富士通	デジタルアニーラ発表(Si)	2016年	1QB(カナダ)と協業
	日立	CMOSアニーリングマシン発表(Si)	2015年	
	東工大	量子アニーリング理論を発表	1998年	
ゲート型	慶大、東大、他	デバイス技術開発(Si系等)		
	NEC	超伝導量子ビットを開発	1999年	

③ノイマン型AIコンピュータ

表11に、日本におけるノイマン型AIコンピュータの主な取 組例を示す。

(7)日本の公的研究開発支援プログラム

表12に、日本における主な公的研究開発支援プログラムを示す。基礎的な研究はJSTを中心に、実用化に向けた研究はNEDOを中心に取り組まれている。

表11 日本におけるノイマン型 AIコンピュータの研究開発状況

機関	名 称	アーキテクチャ、特徴	概要	出荷時期
DMP、産総研、 NEDO (loT横断PJ)	EP1	GPUのシェーダプロセッサにDNN 高効率処理HWを接続	1W電力により、広く組み込みシステムでの適用を目指す	2025年
富士通	DLU	京の成果のTofuインターコネクト で10万チップの接続		2018年度
Deep Insights	DI-1	磁界結合と光通信100Tバイト/秒 7nmプロセス、100万コア		2018年度
理研、 NEDO(次世代人工知能PJ)		低精度計算と高並列化、NVIDIA T100の70倍が目標、40nm TSMC で製造	NVIDIAは、倍精度→半精度で4倍に しかならないが、16倍を達成する	2017年度 チップ試作
PFN (オープンソース)	Chainer	Linux、GPU可、Python	コミュニティが、ほぼ国内のみ。画像 処理から自然言語処理、ロボット制御 (NEDO次世代人工知能PJで拡張中)	

出所:公開資料を基に NEDO 技術戦略研究センター作成 (2018)

表 12 日本における主な公的研究開発支援プログラム

プログラム名と概要	機関	期間
高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発 /次世代コンピューティング技術の開発 概要 アニーリングマシンコンピューティング技術やニューロモーフィックコンピューティン グ技術に関する開発を実施予定である。	NEDO	2018-2027
量子情報処理プロジェクト 概要 量子もつれを用いた計測、標準、通信、情報処理技術の4つの分野についての日本の 独創的なアプローチに基づいた研究開発。	JST	2009-2013
高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発 / 高度なIoT社会を実現する横断的技術開発 概要 IoT推進のため横断技術開発プロジェクトのサブプログラムとして、組合せ最適化処理 に向けた革新的アニーリングマシンの研究開発を行う。CMOSアニーリングチップや 超伝導量子ビット回路を開発する。	NEDO	2016-2020
量子人工脳を量子ネットワークでつなぐ高度知識社会基盤の構築 概要 光パラメトリック発振器ネットワークを用いてイジングマシンとしての大規模量子人工 脳を開発する。様々な組合せ最適化問題、機械学習の分野での適用を目指す。	JST	2014-2018

3章

人工知能を支えるハードウェア 分野の技術課題

(1) AIを支えるハードウェアの現状と課題

①脳型コンピュータ

脳型コンピュータの技術課題には、1) ニューロンやシナプスの数に対応した回路規模の拡大、2) 実際の神経細胞の機能をどこまで忠実に実装するか、3) 学習の手段と実用的な負荷の実現、4) 使用、普及におけるプラットフォームや環境構築、などがある。特に、1) の規模拡大の課題は、微細化、省電力性能、計算精度、実行速度などに分割される。

ノイマン型 AIコンピュータでは、NVIDIA などの GPU が現在は最も実用性が高いが、集積化に限界もあり、100 億の脳サイズまでの到達は難しいと考えられる。

デジタル型脳型コンピュータは1つ1つのニューロンセルの構造は単純にして、多くのニューロンセルを並べる構造を取る。規則性の高い構造のため、設計も高密度実装も容易である。

アナログ型脳型コンピュータは、各種の不揮発メモリの新 しい応用領域と期待されている。アナログなので完全なコ ピーはできないため、経時安定性、微細化したときのノイズ などに課題があると考えられる。デジタル型、アナログ型の 脳型コンピュータでは、今後はデバイス開発とともに学習手 法の開発が重要になると推察する。

技術課題に取り組むにあたって、実現する機能に対して どのようなアーキテクチャでコンピュータを構成し、どのような ソフトウェアで学習、演算を行うかによって、ハードウェアに 求められる性能や規模は異なってくる。ハードウェア開発の 実用を想定したターゲット設定では、これらの研究開発との 連携が重要である(図17)。

脳型(ニューロモーフィック)コンピュータの実現には、脳科学の進展を 基にしたソフトウェアとハードウェアの最適化が求められる。

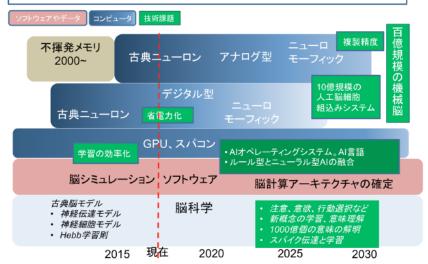


図17 脳型コンピュータ研究のフレームワーク

②量子コンピュータ

a. ゲート型

量子ビットは不安定であり、コヒーレンス時間と呼ばれる量子状態の保持時間が短いため、量子ビットを集積した上でコヒーレンス時間を長く保つ技術や量子ビットの誤りを訂正する技術が大きな課題である。現状では、およそ100μ秒のコヒーレンス時間が実現されている。表13に示すように、現在は誤り訂正技術で一定のコヒーレンス状態が実現できており、さらなる改善によりメモリ等の演算や量子ビット集積の段階に進んでいる。

b. イジングマシン型

イジングマシン型は量子ビットの配列と量子ビット間の相 関性設定から解くべき問題をイジングモデルとして構成し、 量子効果により最適解を求めるもので、アニーリング処理を 行う量子アニーリングが主である。性能向上には量子ビット 規模の拡大とその結合数の拡大が課題となるが、実問題 のイジングモデル化やハードウェア実装等が複雑であり、そ の簡略化手法も課題である。

③確率モデルコンピューティング

確率モデルとベイズ統計を使ったコンピューティングは画像処理等の応用が研究されてきているが、近年のディープラーニングにおいて確率モデル技術を取り入れることで、大量の学習データへの対応やデータの不確実性の扱いを含んだリスク把握、欠損データの生成といったディープラーニングの課題への対応が期待される。課題としては、確率モデルの選択を含む導入技術の最適化やプログラミングの複雑さやフレームワークの確立などがある。

難度
 技術レベル
 1 単一量子ビットの操作
 フルゴリズムによる複数の量子ビット操作
 3 誤り訂正制御による量子ビット非破壊操作
 4 量子ビット保持時間を超えるロジカル量子ビット(メモリ)
 5 単一ロジカル量子ビットの操作
 アルゴリズムによる複数のロジカル量子ビット演算
 2030年~
 誤り訂正可能な量子コンピュータ

表13 量子コンピュータ(ゲート型)の研究開発段階

(2) 各技術における主な課題

表14に、AIを支えるハードウェアの各技術における主な課題をまとめた。

表14 主な技術課題

	技術課題
ディープラーニング (CMOS)	回路規模拡大、高速化、計算精度の最適化技術低消費電力化技術認知、判断、言語応答等より高次の学習、リアルタイム化技術
脳型コンピュータ	上記に加えて □ニューロンとシナプスの最適モデル探索(可塑性など)□ニューロンとシナプスの数に対応する回路規模の拡大●学習方法とアルゴリズム、アーキテクチャ技術●データ、学習、推論のプラットフォームや環境構築
量子コンピュータ (ゲート型)	量子ビット集積化技術(コヒーレンス時間拡大含む)2次元量子ビット制御や誤り訂正、誤り耐性技術演算アルゴリズム、アーキテクチャ技術
量子コンピュータ (イジングマシン型)	量子ビット集積化技術多体問題演算の適用技術実問題のイジングモデル化/実装簡略化、ミドルウェア技術
確率モデル	確率モデルを含むデータ構造、アルゴリズム、アーキテクチャ技術確率モデル開発プラットフォーム

4章

おわりに

AI技術は、科学技術基本計画 (2016年) で示された超スマート社会 (Society5.0) の中で、IoT やビッグデータ解析、高度なコミュニケーションを支える基盤技術と並ぶ中核技術と位置づけられ、高度な社会基盤構築や、様々な社会課題解決への貢献、産業分野への波及効果が大きいと期待されている。そして、現在のディープラーニングを基盤とする AI は今後も進化していくことが期待され、これを支えるハードウェア技術の進展は AI の進化において重要な要素である (図18)。

これまでのCMOS半導体によるノイマン型コンピュータの性能を凌駕する脳型コンピュータや量子コンピュータなど、次世代コンピュータの取組は始まっているが、社会実装に

至るには技術と応用それぞれに課題も多く、研究開発に時間を要する段階にある。AIコンピュータとして示した、ノイマン型 AIコンピュータ、脳型コンピュータ、量子コンピュータの実用化では、いずれも現行のCPUによるコンピュータと演算機能を分担するハイブリッド型または併用型としてコンピュータ性能の加速、拡張のアクセラレータ技術となり普及していくものと見られている。

現在のAIコンピュータ開発は、米国やこれに続く中国の研究開発投資・人材育成が旺盛であり、技術進展も速いため、日本がこれらと同等の取組をしていくことは難しい。しかしながら、AI開発では学習に使う各適用分野(エッジ領域)のデータ量とコンピュータとしての経済性が社会実装の強い推進力となるため、日本としてはモノづくりやサービス等の適用分野におけるデータを活用し、競争力強化に繋がるようなハードウェア技術開発を進めることが重要である。

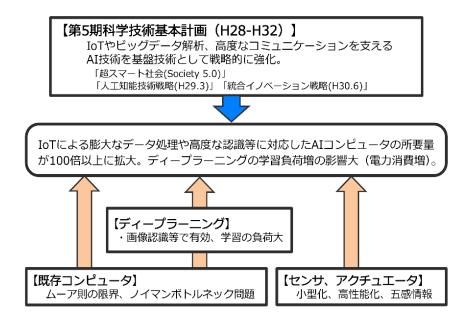


図18 AIコンピュータの必要性

出所: NEDO 技術戦略研究センター作成 (2018)

技術戦略研究センターレポート

TSC Foresight vol. 30

人工知能を支えるハードウェア分野の技術戦略策定に向けて

2018年10月31日発行

TSC Foresight Vol.30 人工知能を支えるハードウェア分野 作成メンバー

国立研究開発法人 新エネルギー・産業技術総合開発機構 技術戦略研究センター(TSC)

- センター長 川合 知二
- センター次長 竹上 嗣郎
- 電子・情報・機械システムユニット
 - ・ユニット長 伊藤 智
 - •主任研究員 大窪 宏明
 - •研究員 村嶋 清孝

砂口 洋毅

有馬 宏和

・フェロー 松井 俊浩 学校法人情報セキュリティ大学院大学教授(平成29年3月まで)

林 秀樹 元住友電気工業株式会社理事、IEEE Life Fellow、応用物理学会フェロー

中屋 雅夫 元株式会社半導体理工学研究センター代表取締役社長

山口 佳樹 国立大学法人筑波大学准教授

遠藤 直樹 東芝デジタルソリューションズ株式会社

- ◆ 本書に関する問い合わせ先電話 044-520-5150 (技術戦略研究センター)
- ◆ 本書は以下URL よりダウンロードできます。 http://www.nedo.go.jp/library/foresight.html

本資料は技術戦略研究センターの解釈によるものです。 掲載されているコンテンツの無断複製、転送、改変、修正、追加などの行為を禁止します。 引用を行う際は、必ず出典を明記願います。