

# AI半導体技術の動向と展望 ～GPUから次世代チップへのシフト～

株式会社日本総合研究所 先端技術ラボ

2025年8月27日

本資料は作成日時点で弊社が一般に信頼できると思われる資料に基づいて作成されたものですが、情報の正確性・完全性を保証するものではありません。情報の内容は、経済情勢等の変化により変更されることがあります。本資料の情報に起因してご閲覧者様及び第三者に損害が発生したとしても、執筆者、執筆にあたっての取材先及び弊社は一切責任を負わないものとします。なお、本資料の著作権は株式会社日本総合研究所に帰属します。

## AI半導体技術の動向と展望 ～GPUから次世代チップへのシフト～

項目	見出し・取り上げテーマ	時間目安
1. はじめに	先端技術ラボ、本日スピーカーの紹介	—
2. AI半導体技術の動向と展望 ～GPUから次世代チップへのシフト～	<ul style="list-style-type: none"> <li>□ ロジック半導体とは？</li> <li>□ AI利用に向けた次世代チップの概観</li> <li>□ ASIC/ニューロモルフィックチップの仕組み</li> <li>□ 多様化する演算技術</li> <li>□ AIチップが抱える現在の課題</li> <li>□ 展望：ロードマップ</li> </ul>	40分
3. 質疑応答	挙手にてご質問ください	20分

# 先端技術ラボ紹介

## 先端技術の調査・研究組織「先端技術ラボ」

SMBCグループの技術の目利き役。先端技術の本質を追求し、デジタル社会における先進ビジネスを技術起点から牽引します。

技術の進展スピードがますます速くなる中、ビジネス・サービス創出におけるテクノロジー活用の重要性はこれまでになく高まっており、先端技術の本質を正しく見極める力が企業の競争力を大きく左右するようになりました。

日本総研では、IT分野における先端技術の調査・研究を行う専門集団として先端技術ラボを組織し、技術の目利き役として、中期的に大きなビジネスインパクトをもたらすことが予測される有望な先端技術について、技術動向調査と技術検証・評価を継続的に取り組んでいます。

これらの活動から得られた知見を基に、SMBCグループの事業の高度化に貢献するとともに、先端技術に関する中立的かつ実践的な情報の発信にも努めています。



先端技術ラボ 執務スペース

# 先端技術ラボのミッション

## ▶ 先端技術トレンドの調査・提言

先端技術の動向を幅広く収集・分析し、独自の分析・考察を加えたレポートを随時発表しているほか、学会・研究会、専門委員会、セミナーなどでの発表・講演にも積極的に取り組んでいます。

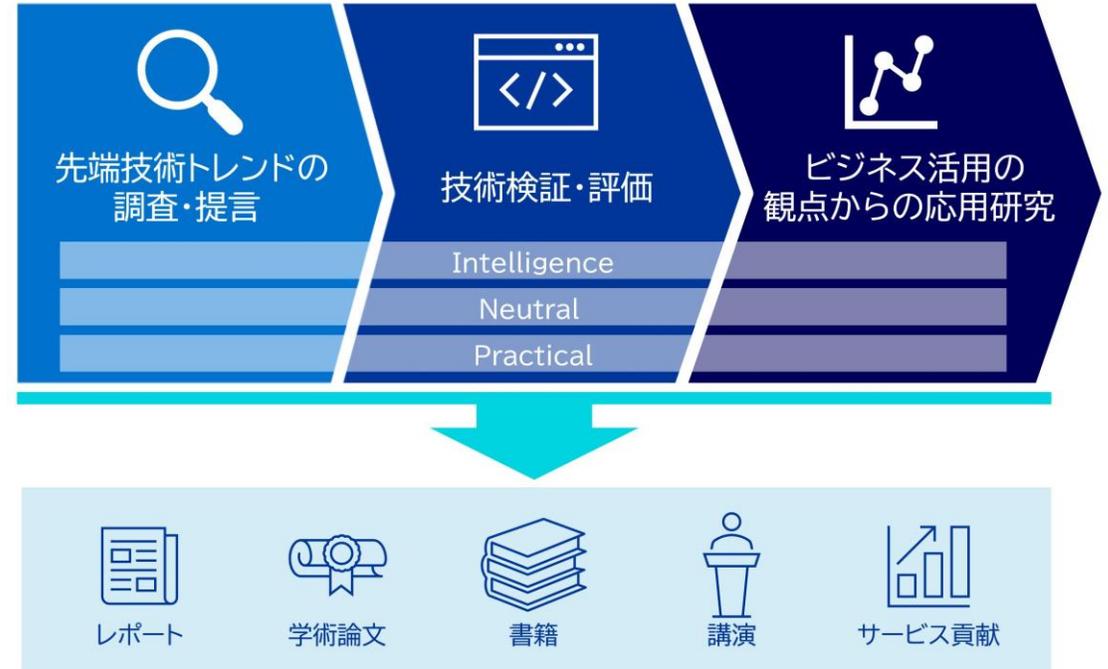
## ▶ 技術検証・評価

今後の実用化が見込まれる先端技術に関する先行研究や先端手法について、技術実装に基づいた実践的な検証・評価を行っています。

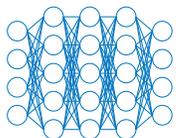
## ▶ ビジネス活用の観点からの応用研究

研究機関や大学といったアカデミアが公表する基礎研究を踏まえながら、各専門分野の有識者との協働による応用研究も進めています。

また、応用研究と技術検証を進めた上で、SMBCグループ各社のITサービスへの適用評価等を実施し、その成果について、学会・研究会への発表を通じた還元を行っています。



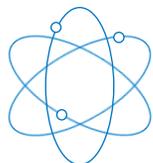
## 主な研究・取り組み領域



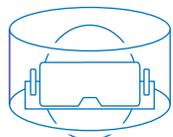
**人工知能(AI)** 金融機関は多種多様な大量のデータ・情報等を保有しており、人工知能(AI)技術との親和性も高いこと等を背景に、金融ビジネスへの活用が進んでいます。先端技術ラボでは、AI領域の様々な要素技術を調査・研究し、SMBCグループの金融ビジネス創出を技術面からリード・サポートし、日本総研が提供するITシステムへのAI実装に繋がっています。



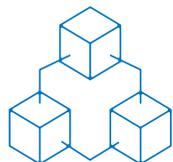
**ブレインテック** 脳科学(Brain Science)の知見とテクノロジー(Technology)を組み合わせたブレインテックは、医療や健康管理での活用の他、マーケティングや教育など、多方面での活用が期待されています。ブレインテック領域において、研究・開発されている黎明期のデバイスの調査や、その技術検証・評価を通じて、中長期的な視野に立った取り組みを進めています。



**量子技術** 量子力学の原理を利用して、現在のコンピュータでは不可能な演算処理を実現する量子コンピュータが注目を集めています。本格的な量子コンピュータのハードウェアは研究・開発段階にありますが、将来の製品化・普及期に向けて、量子コンピュータ上での業務アプリケーションの開発に求められる量子ソフトウェアの応用研究や高度な専門能力の向上を継続的に取り組んでいます。



**xR・メタバース** 専用デバイスを通して仮想空間を作り出す技術xR(AR・MR・VRの総称)や、インターネット上の仮想空間で自身のデジタル・キャラクター(アバター)としてコミュニケーションを実現するメタバース(仮想空間を提供するサービスの総称)は、個人のエンターテインメントに留まらず、企業におけるお客さまとの新しいコミュニケーション・ビジネスの場となる等、今後も現時点では想像もしていない利活用へと進化を遂げる可能性があります。メタバース特有の価値の提供や業務効率化が期待されており、ビジネスシーンでの活用に向けた調査・提言を実施しています。



**ブロックチェーン・Web3.0** 暗号資産の基盤技術として登場したブロックチェーンは、次世代インターネットと称されるWeb3.0(ブロックチェーン技術を基盤とした分散型ウェブ技術)のトレンドへと進展しました。企業や社会の構造変容へ影響を与え、技術戦略にインパクトをもたらす可能性を持つ技術領域についても、先端技術の調査を進め社会実装の可能性を探求しています。

## 先端技術ラボの特徴



**中立的・実践的** 高い専門性を持つメンバーの考察や提言を重視し、ITベンダーとは異なり、特定の製品やサービスにとらわれない中立的な立場から、対象とする技術・テーマを取り決めています。また調査や技術検証・評価においては、SMBCグループでの新ビジネスや想定ユースケースを基に実践的なアプローチで取り組んでいます。

**金融領域に限定しない** SMBCグループにおける組織としては金融ビジネスへの応用が見込まれる技術を重視していますが、SMBCグループが非金融ビジネスへ事業拡大する中で、従来の発想にとらわれない技術シーズへの探索が求められています。量子コンピュータやブレインテックなど、中長期的にビジネスインパクトが大きいと予想される技術へのチャレンジも進めています。

**テクノロジー面からの情報発信と提言** 先端技術の調査および検証・評価の成果は、SMBCグループへの技術戦略や活用に向けた知見に留めることなく、一部はレポート公表や専門誌への記事執筆、セミナー登壇を通じて社会還元しています。また技術検証・評価の結果、新規性の高い成果については学会・研究会で発表を行い、アカデミアへの貢献にも努めています。

**柔軟で自律した働き方とカルチャー** 自律した働き方を支える仕組みとして、先端技術ラボでは「20%ルール」という制度があります。業務時間の20%を将来的に芽が出そうな技術領域の調査や技術検証に充てられるというもので、社員の自由な発想に基づく活動をサポートしています。

# 先端技術動向の調査・レポート

<https://www.jri.co.jp/company/business/system/advtechlab/research/>



先端技術ラボ

🔍

検索

- 先端技術ラボ
- 先端技術ラボ
- 先端技術リサーチ
- 学術論文
- メディア掲載・書籍
- 技術コミュニティ活動
- 講演・イベント登壇
- 受賞歴
- 活動概要
- 採用情報
- サービスに関するお問い合わせ

## 先端技術リサーチ

IT分野における最新技術・トレンド等をリサーチした先端技術レポートを公開しています。

- 2025年08月04日 **ローカルLLMの可能性 - オープンなモデルが拓くAI活用の展望 -**  
伊藤 蓮、近藤 浩史
- 2025年07月01日 **AI半導体の最新動向 ～ポストGPU時代の注目技術～**  
會田 拓海、間瀬 英之
- 2025年06月25日 **NeuroTech×Art ①① 視覚芸術編**  
大城 武史
- 2025年06月10日 **金融×生成AI -事例から導く変革の最前線と実践戦略- / Finance × Generative AI -Case Study and Strategy-**  
落合 達堂 / Yodo Ochiai
- 2025年04月10日 **生成AIの台頭で大きく注目されるAIガバナンス ～欧米のAI・デジタル動向の考察と示唆～【後編】**  
先端技術ラボ 田谷洋一、株式会社三井住友フィナンシャルグループ シリコンバレー・デジタルイノベーションラボ 緒方雄二

2.2 AIチップとは | AIチップの技術概要

AIの演算環境は、汎用チップの利用からAIに特化したASICや既存の電子回路とは異なる次世代の半導体/演算装置に移りつつある

- AIの演算処理に特化したチップはAIチップやAIアクセラレータと呼ばれ、消費電力の削減や処理速度の向上を目指して開発が進められている。
- 2025年現在では、AI演算に特化したASICの実用が進んでおり、ニューロモルフィックチップは実証実験段階。

名称	汎用向け		特定用途向け		次世代AIチップ	
	CPU	GPU	FPGA*1	ASIC	データセンター向け NPU	ニューロモルフィックチップ
特徴	汎用かつ複雑な演算に強み コンピューターの中核としてプログラムを実行	並列演算に強み AIの行列演算に利用	特定用途向けに汎用部品を組み合わせて設計 チップ製造後に回路を再構成できる	特定用途向けのカスタム設計 回路の再構成は不可	ニューラルネットワーク演算に特化 データセンターやモバイル端末向けに独自設計	人間の脳構造を模倣し、アナログ信号を用いた回路設計
用途例	コンピューター	大規模演算 (GPGPU*2) グラフィック処理	基地局 医療画像機器	家電製品・自動車制御	AI学習・推論処理	さらに高度なAI処理

\*1: Field Programmable Gate Array.

【凡例】 ■ 実用化済 ■ 開発・検証中

# 説明者紹介

會田 拓海 (Takumi Aita) aita.takumi.m2@jri.co.jp

本編説明を担当

株式会社日本総合研究所 先端技術ラボ アナリスト

- 2021年日本総合研究所入社。
- 金融×IT、ブロックチェーン領域などを中心に、先端技術の動向調査、業務適用に向けた応用研究を担当。

間瀬 英之 (Hideyuki Mase) hideyuki.mase@jri-america.com

SMBCグループ 兼 JRI America シリコンバレー・デジタルイノベーションラボ Director/Senior Researcher

- 2025年よりSMBCグループ シリコンバレー・デジタルイノベーションラボに駐在。
- 国際金融規制に対するシステムの企画・開発、プロジェクト管理などを経て、2018年より先端技術ラボにて量子コンピュータ、AIなどの先端技術に関する動向調査業務に従事。
- 共著書に「量子コンピュータまるわかり」（日経BP 日本経済新聞出版，2023年）、「金融デジタルイノベーションのすべて」（金融財政事情研究会，2021年）。

# AI半導体技術の動向と展望

## ～GPUから次世代チップへのシフト～

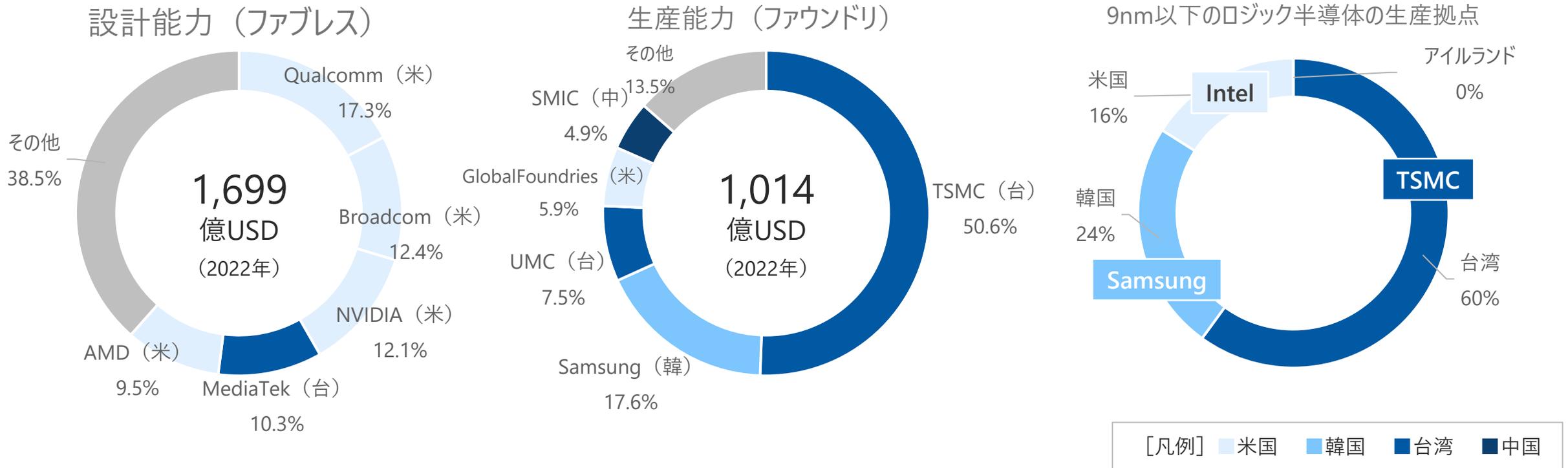
## 半導体はデータの記憶や電圧変換などさまざまな用途で使われるが、AI演算の省電力化・高速化という観点では、演算処理を担うCPU/GPUをはじめとしたロジック半導体の開発競争が続いている

- ロジック半導体は、米国・台湾企業、メモリ半導体は韓国企業が世界市場において大きなシェアを占める。
- 日本企業は、パワー半導体、マイコン、イメージセンサーの分野で一定のシェアを獲得している。

種類	概要	日本製の世界シェア (2022年)	製造している日本企業
<b>ロジック半導体</b>	本レポートのスコープ CPU/GPUなど、コンピューター内で演算や制御を行う処理装置	-	-
DRAMメモリ	コンピューターに使用される揮発性メモリ	-	-
NANDメモリ	SSDやSDカードに使用される不揮発性メモリ	18.9%	キオクシア
パワー半導体	家電製品や自動車などに使用される高い電圧・大きな電流を扱う半導体	25.5%	三菱電機、富士電機、東芝、ルネサスエレクトロニクス、ローム
マイコン (MCU)	演算/制御/記憶/入出力機能を集積した小さなコンピューター	16.5%	ルネサスエレクトロニクス
イメージセンサー	産業/車載/モバイル用カメラなどに使用されるセンサー	44.0%	ソニーセミコンダクタソリューションズ

# 先端ロジック半導体の設計能力は米国、生産能力は台湾の存在感が強く、とりわけ最先端の技術を用いた9nmプロセス\*1以下のロジック半導体の生産拠点は、世界4か国のみに限られる

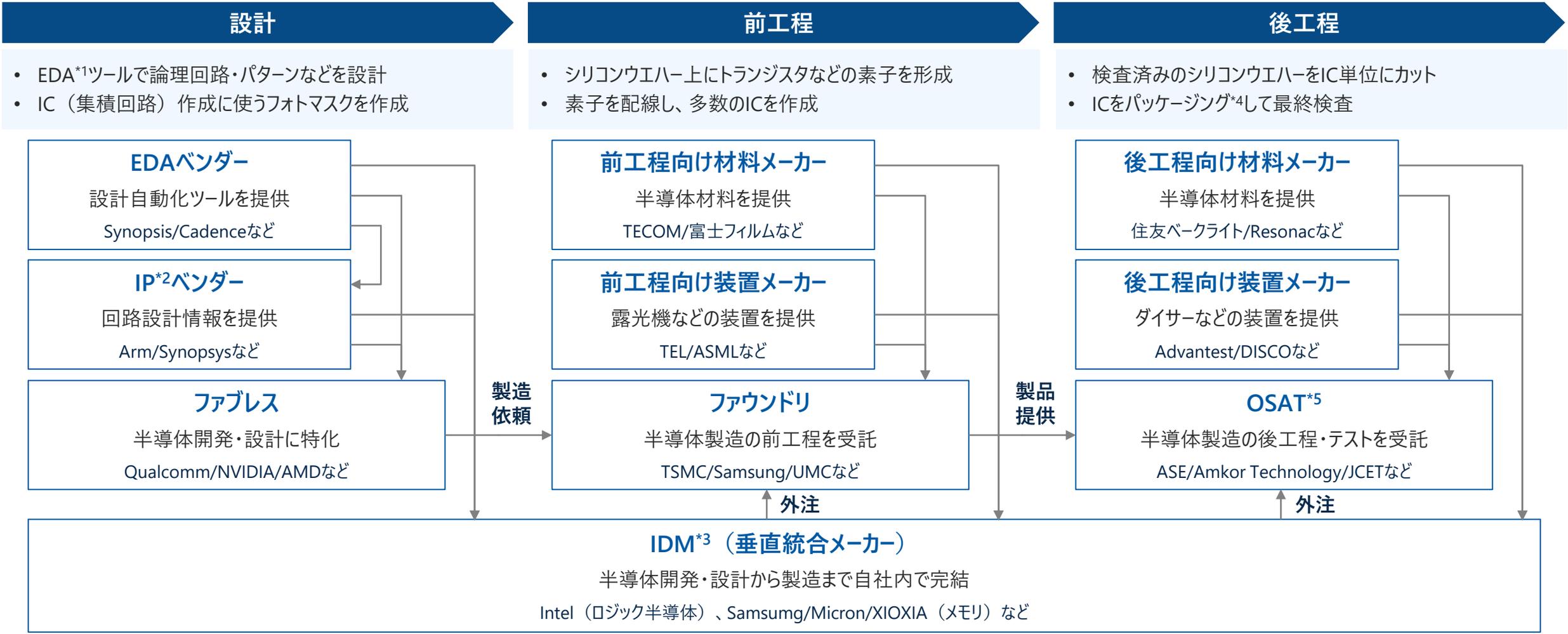
- 米国企業が半導体設計を担い、台湾や韓国などアジアの企業が製造を受託する構造が定着。
- アイルランドには、1980年代からIntel社が拠点を構え、2023年には当時の最先端プロセス「Intel4」を量産する工場を開設。Intel4は同社の「Core Ultra」や「Loihi2」\*2に使用されている。



\*1: プロセスは、一般に配線幅と同義として使われてきた用語だが、近年は半導体チップの面積に基づく値を示す。\*2: 別冊参照。

出所：[1] 半導体・デジタル産業戦略（経済産業省 商務情報政策局），[2] <https://eetimes.itmedia.co.jp/ee/articles/2310/02/news046.html>, (参照 2025-03-06)

# 近年は半導体設計を担うファブレスや前工程を担うファウンドリに注目が集まるが、半導体はステークホルダーが水平分業しながら設計・製造されており、日本は材料・装置メーカーが強い



\*1: Equipment Data Acquisition / Electric Design Automation. \*2: Intellectual Property. \*3: Integrated Device Manufacturer. \*4: ICの電極とパッケージをワイヤーで接続し、樹脂を注入して封止 \*5: Outsourced Semiconductor Assembly and Test.

## AI利用の需要拡大を受け、従来型・汎用型とは異なる、より省電力かつ高速に処理できる演算用チップの開発が進んでいる

- AI分野では、大規模かつ複雑なデータセットの利用、AI演算の需要の高まりなどから、処理速度の向上や生産コストの削減が期待される。
- 電力不足も懸念されており、世界全体のデータセンターにおける電力消費量は4600億kWh（2022年）から1兆kWh（2026年）に急増するとの予測<sup>\*1</sup>もある。
- AI演算の省電力化・高速化に向け、半導体の技術開発、非電子回路による演算環境の研究が行われている。

	技術	概要	ページ
電子処理	AI特化のASIC <sup>*1</sup>	ニューラルネットワークの演算に特化したチップ（NPU <sup>*2</sup> を含む）	p.16
	ニューロモルフィックチップ	アナログ信号処理を用いて演算を行うチップ	p.21
非電子処理	光チップ	電子信号の代わりに光信号を用いて演算を行うチップ	p.24
	量子チップ	量子力学の原理を情報処理に応用して演算を行うチップ	
	バイオプロセッサ	人工多能性幹細胞（iPS細胞）を培養して創出した組織体を用いて演算を行うチップ	

<sup>\*1</sup>: IEA（International Energy Agency）発表。<sup>\*2</sup>: Application Specific Integrated Circuit. 特定用途向けのカスタム集積回路。<sup>\*3</sup>: Neural Processing Unit.

# AIの演算環境は、汎用チップの利用からAIに特化したASICや既存の電子回路とは異なる次世代の半導体/演算装置に移りつつある

- AIの演算処理に特化したチップはAIチップやAIアクセラレータと呼ばれ、消費電力の削減や処理速度の向上を目指して開発が進められている。
- 2025年現在では、AI演算に特化したASICの実用が進んでおり、ニューロモルフィックチップは実証実験段階。

	汎用向け	特定用途向け				
		既存チップ			AIチップ	次世代AIチップ
名称	CPU	GPU	FPGA <sup>*1</sup>	ASIC	データセンター向け / NPU	ニューロモルフィックチップ
特徴	<ul style="list-style-type: none"> <li>汎用かつ複雑な演算に強み</li> <li>コンピューターの中核としてプログラムを実行</li> </ul>	<ul style="list-style-type: none"> <li>並列演算に強み</li> <li>AIの行列演算に利用</li> </ul>	<ul style="list-style-type: none"> <li>特定用途向けに汎用部品を組み合わせて設計</li> <li>チップ製造後に回路を再構成できる</li> </ul>	<ul style="list-style-type: none"> <li>特定用途向けのカスタム設計</li> <li>回路の再構成は不可</li> </ul>	<ul style="list-style-type: none"> <li>ニューラルネットワーク演算に特化</li> <li>データセンターやモバイル端末向けに独自設計</li> </ul>	<ul style="list-style-type: none"> <li>人間の脳構造を模倣し、アナログ信号を用いた回路設計</li> </ul>
用途例	<ul style="list-style-type: none"> <li>コンピューター</li> </ul>	<ul style="list-style-type: none"> <li>大規模演算 (GPGPU<sup>*2</sup>)</li> <li>グラフィック処理</li> </ul>	<ul style="list-style-type: none"> <li>基地局</li> <li>医療画像機器</li> </ul>	<ul style="list-style-type: none"> <li>家電製品・自動車制御</li> </ul>	<ul style="list-style-type: none"> <li>AI学習・推論処理</li> </ul>	<ul style="list-style-type: none"> <li>さらに高度なAI処理</li> </ul>

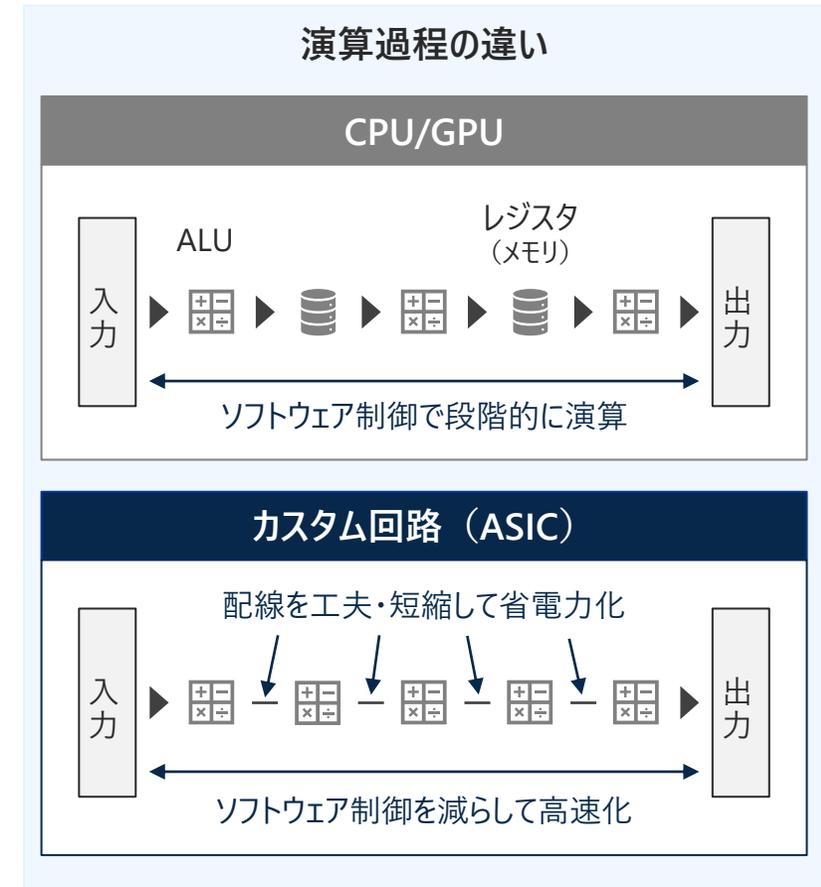
[凡例] ■ 実用化済 ■ 開発・検証中

\*1: Field Programmable Gate Array.

## ニューラルネットワークの構造を電子回路で再現し、AI演算に特化したASICの開発・導入が進む

- ASICとは、特定用途向けに設計された回路で、汎用向けのCPUと比べて省電力・高速な処理が期待される。
- AI演算向けには、データセンターで大規模なデータの学習・推論に利用されるASICや、クライアント/モバイル端末上で音声・画像処理などに利用されるASICがある。

<p>概要</p>	<p>ASICとは、特定用途に特化してカスタム設計された回路。 汎用型チップはさまざまな場面の利用に備えて冗長性を確保しているのに対し、ASICには不使用回路が存在しないため、消費電力が抑えられる。</p>	
<p>用途</p>	<p><b>データセンター向け</b> 学習・推論</p>	<p><b>モバイル端末向け (NPU)</b> (主に) 推論</p>
<p>特徴</p>	<p>大規模データを用いて学習するため、ALU*1の演算結果を一時保存する大容量のメモリが必要。</p>	<p>推論の工程は、学習に比べて演算に必要なメモリが小さく、小型化できる。 モバイル端末にSoC*2として搭載し、学習済みのモデルを利用することで、モバイル端末のAI処理を実行する。</p>
<p>製品例</p>	<ul style="list-style-type: none"> <li>AWS Trainium (Amazon)</li> <li>Cloud TPU (Google)</li> </ul>	<ul style="list-style-type: none"> <li>Neural Engine (Apple)</li> <li>Core Ultra (Intel)</li> <li>Qualcomm Hexagon NPU</li> </ul>

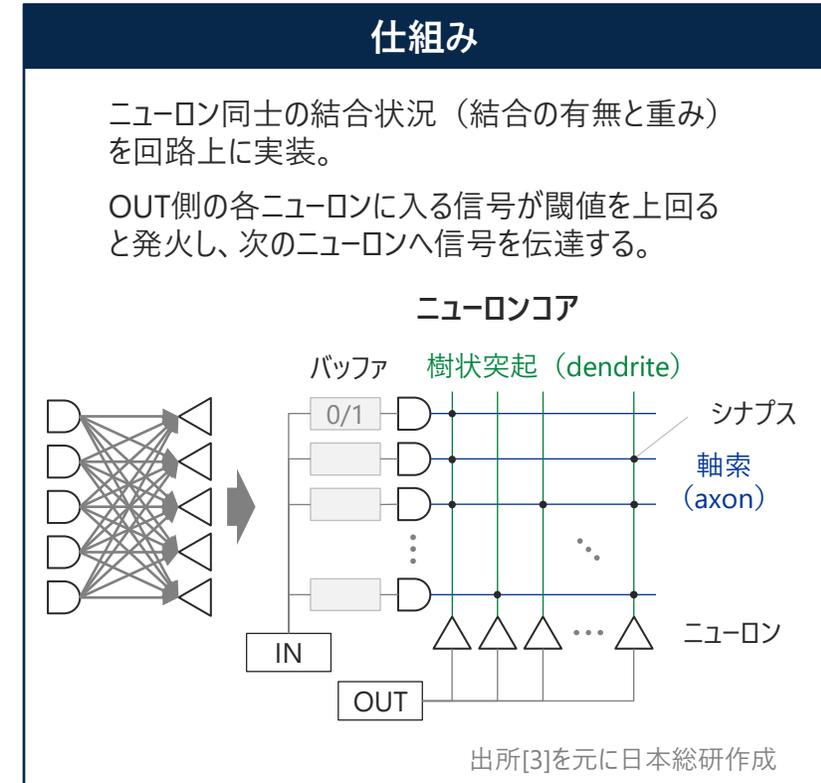


\*1: Arithmetic Logic Unit. 算術演算、論理演算などを行う。\*2: System on a Chip. システムに必要な部品を1つのチップ上に実装した集積回路。

# アナログ信号処理を用いることで推論処理の省電力化・高速化を図るニューロモルフィックチップは、演算の仕組みは実現に近づきあるものの、信号変換時の電力消費や発熱など課題があり、実用化には時間を要するとみられる

- ニューロモルフィックチップは、脳構造を模倣した回路<sup>\*1</sup>で構成され、アナログ信号の強弱を用いて演算を行う。
- 外部での学習結果を基に実装され、主に推論に利用される。
- 光/量子/バイオ技術を用いたチップと比べ、実証の段階は進んでいるとみられるが、実用化に向けた課題は多い。

概要	<p>ニューロモルフィックチップとは、脳の神経細胞（ニューロン）と神経細胞同士の結合（シナプス）を回路で再現し、アナログ信号によって処理を高速化するチップ。</p> <p>アナログ信号の強弱を用いて演算を行い、必要な回路にのみ信号を伝達する仕組みをもつため、省電力かつ高速な処理が期待される。</p>
用途	リアルタイム処理、省電力が求められるエッジデバイス（センサー、ロボットなど）
特徴	<p>既存の演算処理はデジタル信号で行われるため、ニューロモルフィックチップに接続するにはデジタル/アナログ信号の変換が必要。</p> <p>変換処理時の電力消費や発熱、ソフトウェアの整備などが課題とされる。</p>
取り組み例	<ul style="list-style-type: none"> <li>NorthPole (IBM)</li> <li>Loihi2 (Intel)</li> </ul>



\*1: SNN (Spiking Neural Network) の仕組みを用いることが多い。

## ニューロモルフィックチップは、CPU/GPU/ASICなどのデジタル信号を用いる既存チップに比べて消費電力や発熱が少ないという強みがある一方、アナログ信号由来の誤差が生じる恐れがある

- 論理回路を用いて演算する既存チップは、データや演算結果が演算部と記憶部を往復するノイマン型アーキテクチャに分類され、汎用性や演算精度が高い。
- ニューロモルフィックチップは演算部と記憶部が一体化し、非ノイマン型アーキテクチャに分類される。記憶部とデータをやり取りせず、必要な回路にのみ電流が流れるため、省電力・高速化が期待される。

	既存チップ	ニューロモルフィックチップ
構造	<p>演算部と記憶部が分割され、命令やデータを読み書きしながら逐次的に演算を行う。</p> <p>汎用性が高い一方、演算部と記憶部を結ぶバス（配線）が電力を消費する。</p>	<p>演算部と記憶部が同一チップ内に実装され、演算中に外部メモリとデータ伝達が生じない。この仕組みは、IMC（In Memory Computing）などとも呼ばれる。</p> <p>アナログ信号処理に誤差が生じるため、既存チップに対して演算精度は劣る。</p>
分類	ノイマン型アーキテクチャ	非ノイマン型アーキテクチャ ※演算結果を外部記憶装置に逐一出力しない
回路の方式	常に回路全体に電流が通っている同期型回路	回路が必要な時に局所的に利用されるイベント駆動型回路
想定用途	高精度が求められるシミュレーションや数値計算	AIを用いた音声・画像認識

## 電子処理と比べて高効率な演算環境を見据えて、光や量子、バイオ技術を用いた研究開発もみられる

- 従来の電子集積回路（電子IC）以外の手法を用いて、省電力・高速化を図る研究開発も進められている。
- 光チップは既存の電子ICと組み合わせる技術が一部実用化されているものの、光チップ単体の利用は研究途上。量子チップやバイオプロセッサの実現にはさらに時間を要する。

	光チップ / 光集積回路（光IC）	量子チップ〈ゲート式〉	バイオプロセッサ（脳オルガノイド）
概念	電子信号の代わりに光信号を用いて演算処理を行う。	量子力学の原理を情報処理に応用して演算処理を行う。 <small>※詳細は当部発行レポート「<a href="#">量子コンピュータの動向と展望</a>」を参照。</small>	人工多能性幹細胞（iPS細胞）を培養して創出した組織体を用いて演算を行う。 <small>※詳細は当部発行レポート「<a href="#">ブレインテック最新動向2025</a>」を参照。</small>
データの扱い	光信号	量子ビット	電気信号
長所	光導波路で信号を伝達して、発熱や損失を抑えるため、省電力・高速化が期待される。	特定のタスクを高速に演算できる。	電力効率が高い。
短所	光の制御や回路の微細化が難しい。	量子ビット数の増加と誤り抑制が課題。	生体組織の寿命や倫理面で、既存の演算装置を置き換えるには大きな障壁がある。
開発段階	電子ICと組み合わせた開発は進む一方で、光IC単体での実用化時期は未定。	超伝導、半導体、イオントラップなど、いくつかの実現方式が研究されているが、実用化には少なくとも数年を要する。	基礎研究の段階にあり、商用化の取り組みはみられるが、本格運用には課題が多い。
取り組み例	<ul style="list-style-type: none"> <li>• 沖電気工業</li> <li>• Lightstandard社（中国）</li> </ul>	<ul style="list-style-type: none"> <li>• Willow（Google）</li> </ul>	<ul style="list-style-type: none"> <li>• Brainoware（米インディアナ大学ブルーミントン校）</li> <li>• Neuroplatform（スイススタートアップのFinalSpark社）</li> </ul>

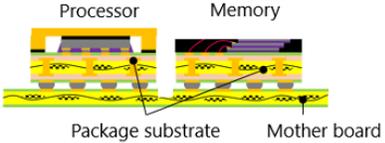
# 半導体の性能向上が求められる一方で、プロセス微細化というアプローチは物理的な限界に達しつつあり、複数のチップを集積するチップレットをはじめとした先端パッケージング技術が高度化している

- 半導体はIC\*1を微細化し、単位面積あたりの性能を高めてきた。近年は、優れた製造技術をもつファウンドリが微細なICの製造からパッケージングまで一貫する形で先端ロジック半導体を生産している。
- 先端パッケージング技術の1つであるチップレットとは、複数チップをまとめて1つのパッケージに搭載する技術で、2.x/3次元パッケージという実装方法がみられる。製造中に低品質のチップを取り除けるため、歩留まり改善につながる。
- 従来の電子ICと光ICを組み合わせるパッケージ化し、消費電力の低減を図る研究も進む。

**従来型 (FC-BGA\*2)**

各機能をもつチップを個々に搭載し、マザーボードを介して情報を伝送する。

現在のプロセッサやSoCなどに利用される主流の実装方法。



Processor Memory  
Package substrate Mother board

画像出所: Resonac HP [10]

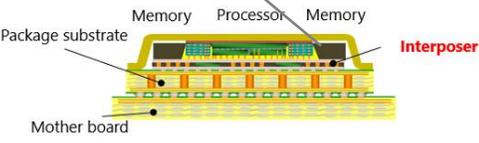
チップレット化

**2.xDパッケージ (2.x次元実装)**

複数のチップを1つのパッケージに搭載し、インターポージャーを介して情報を伝送する。

チップ間の距離が短いため、処理が高速化し、伝送損失も低減する。

異種/他社チップの混載も可能



Memory Processor Memory  
Package substrate Interposer  
Mother board

画像出所: Resonac HP [10]

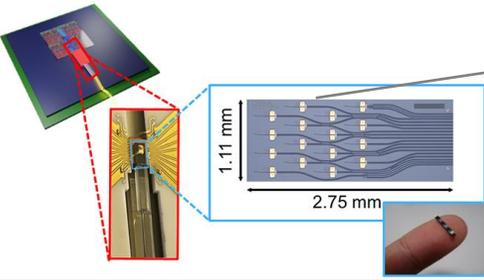
**研究例 (NTT)**

電気信号の代わりに光信号を用いて、高速・低消費電力・大容量の情報伝送を実現することを目的とする。

電気信号と光信号の双方を利用する情報処理技術は、光電融合技術と呼ばれる。その進展に必要な光チップレット開発に向け、光半導体チップの研究が進められている。

メンブレン化合物半導体技術

発光した光を閉じ込める技術を用いて、小型かつ低消費電力の光送信を実現。



1.11 mm  
2.75 mm

画像出所: NTT HP [11]

\*1: Integrated Circuit. 集積回路。\*2: Flip Chip-Ball Grid Array. ワイヤの代わりにチップ裏面の突起（導電性バンプ）を介して基板と接続する。

出所: [10] <https://www.resonac.com/jp/solution/tech/next-gen-semiconductor-packages.html>, (参照 2025-03-27), [11] [https://www.rd.ntt/iown\\_tech/post\\_6.html](https://www.rd.ntt/iown_tech/post_6.html), (参照 2025-03-27)

## AI演算を支えるソフトウェアはNVIDIA製GPUのみに対応するCUDAが広く利用されている一方で、AMDやIntelはNVIDIA以外のプロセッサに対応するソフトウェア開発を主導し、一極集中を防ぐ動きがみられる

- AI演算で多用される行列演算は、一般に並列処理を強みとするGPUで処理され、ソフトウェアが整備されている。
- 他のAIアクセラレータ（ASICやNPUなど）向けのソフトウェアの開発は、GPUと比較して進んでいない。

	CUDA   NVIDIA	ROCm   AMD	oneAPI   Intel	AWS Neuron SDK
概要	画像処理、機械学習、数値計算など幅広い並列演算を支援するソフトウェア。GPGPU*1向けソフトウェアの中で普及率は最も高く、寡占状態。	CUDA非対応のAMD製GPUでも並列演算を行うために開発された、GPGPU向けソフトウェア。	単一のコードで異種のプロセッサに対応した演算処理を記述するソフトウェア。ただし、プロセッサにより発揮する性能は異なる。	クラウド（Amazon EC2）で深層学習/生成AIを利用するための開発環境。TensorFlowやPyTorchなどのフレームワークと統合されている。
対応するプロセッサ	NVIDIA製GPU	AMD製GPU	CPU/GPU/FPGA	AWS Trainium
強み	利用者が多く、資産・ノウハウの蓄積があり、最適化が進んでいる。C/C++を用いて直接GPUを制御できる。	オープンソースのため、利用者は柔軟にカスタマイズできる。HIP*2を用いると、CUDA向けのコードをROCmに移植可（一部）。	特定のハードウェアベンダーに依存しない開発環境を提供する。異種ハードウェアを使う際に、コードの再利用が図れる。	他のAWS製品と連携しやすい。上記プロセッサに特化することで、高いパフォーマンスが実現できる。
課題	NVIDIA製品に依存するため、他のプラットフォームへの移植が困難。	CUDAに比べ、ROCmに対応する機械学習用パッケージが少ない。	ライブラリやツールはまだ充実していない。	AWSに依存した環境でのみ利用可。

\*1: General-purpose computing on GPU. グラフィック処理以外の演算処理にGPUを利用すること。\*2: Heterogeneous-compute Interface for Portability.

# 世界各国は自国内での半導体の製造・安定供給を目指しており、製造能力や研究開発能力・体制の強化、サプライチェーンの強靱化に向けて資金を投入している

	各国・地域*1の戦略目標	取り組み・方針（一部）
<b>米国</b>	<ul style="list-style-type: none"> <li>「CHIPSプラス法（The CHIPS and Science Act of 2022）」に基づき、2030年までに<b>米国の先端ロジック半導体製造における世界シェアを約20%に引き上げ</b></li> </ul>	<ul style="list-style-type: none"> <li>半導体企業の製造拠点建設や設備投資に対する経済支援</li> <li>米国内における大量生産可能な先端パッケージング施設の設置</li> </ul>
<b>欧州</b>	<ul style="list-style-type: none"> <li>EUは同域内における<b>次世代半導体の世界シェアを20%以上に引き上げ</b></li> </ul>	<ul style="list-style-type: none"> <li>半導体の研究開発補助金プログラム「欧州半導体イニシアチブ」による設計・製造能力の強化、企業・人材育成の支援</li> </ul>
<b>中国</b>	<ul style="list-style-type: none"> <li>「中国製造2025」（2015年）を掲げ、<b>2030年に半導体自給率を75%に引き上げ</b></li> </ul>	<ul style="list-style-type: none"> <li>半導体の研究開発プロジェクトに対する法人税優遇</li> <li>生産能力を有する企業（SMIC*2）に対する経済支援</li> </ul>
<b>日本</b>	<ul style="list-style-type: none"> <li>日本国内における次世代半導体技術の確立、サプライチェーンの強靱化</li> <li>2030年、<b>日本国内の半導体製造企業の合計売上高15兆円超に引き上げ</b></li> </ul>	<ul style="list-style-type: none"> <li>米国と連携した半導体技術の習得</li> <li>次世代半導体のR&amp;D拠点（LSTC）・製造拠点（Rapidus）の設置</li> </ul>
<b>英国</b>	<ul style="list-style-type: none"> <li>国内企業の成長</li> <li>サプライチェーンの混乱リスクの緩和</li> <li>国家安全保障の確保</li> </ul>	<ul style="list-style-type: none"> <li>半導体設計関連のツールやプロトタイプ製造に対する経済支援</li> <li>官民連携でサプライチェーンリスクを明確にし、対策を検討</li> </ul>
<b>韓国</b>	<ul style="list-style-type: none"> <li>「半導体超強大国達成戦略」（2022年）を掲げ、システム半導体（パワー半導体、車載半導体、AI半導体など）の<b>世界シェアを10%に引き上げ</b></li> <li>素材、部品、製造装置について、<b>2030年までに国内調達比率を50%に引き上げ</b></li> </ul>	<ul style="list-style-type: none"> <li>半導体企業への大型投資</li> <li>半導体研究に強い大学/大学院を支援し、専門人材の育成を促進</li> </ul>
<b>台湾</b>	<ul style="list-style-type: none"> <li>半導体産業の<b>追加生産額を2.6兆台湾元（日本円で11.8兆円相当）引き上げ</b></li> <li>2025年までに<b>2nm半導体の量産化、2030年までに1nm半導体の実現</b></li> </ul>	<ul style="list-style-type: none"> <li>産学共同の半導体研究開発センターを設立</li> <li>高度な半導体材料・製造技術を開発する「オングストローム半導体計画」を進行</li> </ul>

\*1: GDP順。\*2: 中芯国際集成电路製造（Semiconductor Manufacturing International Corporation）

出所：[1] 半導体・デジタル産業戦略（経済産業省 商務情報政策局），[12-19, 21] 地域・分析レポート（JETRO），[20] 英国政府HP

# AIチップを含めた半導体は、その高性能化とともに消費電力の増大や微細化による性能向上の限界、設計・製造ベンダーの寡占が課題とされ、新技術開発や設備投資などによる解決が期待されている

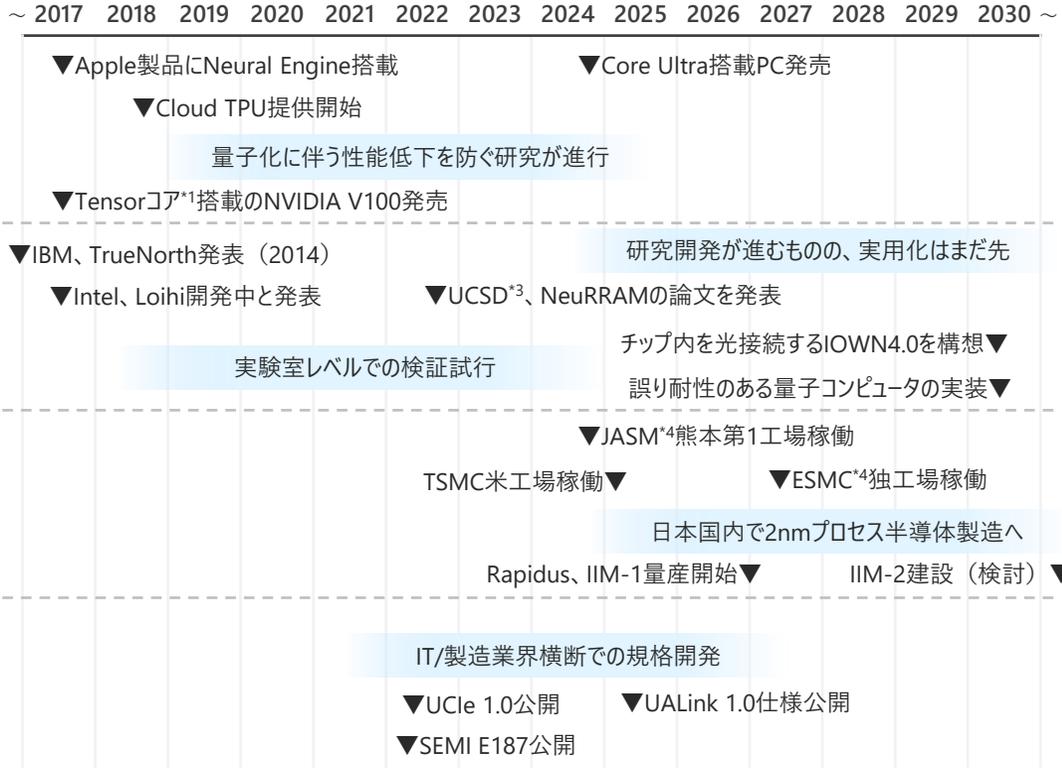
	現状	課題	取り組み
処理速度	高いAI需要などを受け、従来より高速に処理できる演算環境が求められている。	微細化による性能向上率が低下し、集積回路にみられるムーアの法則*1が崩れつつある。	汎用チップでなく特定用途向けチップの利用。 ニューラルネットワークにおける量子化技術*2を用いた低精度ビットでの演算。
熱/電力	半導体の性能向上のため、微細化だけでなく、チップレット化や2.x/3次元積層が進んでいる。	半導体の高性能化に伴い消費電力が増加。 高密度の集積により、発熱も増大している。	ニューロモルフィックチップや光チップなど、 発熱・電力消費を抑える技術の開発。
供給網	9nmプロセス以下の先端ロジック半導体を設計・製造できる企業は一部の国・地域に偏在する。	少数ベンダーによる寡占は地政学リスクを伴い、 半導体の供給が不安定になる恐れがある。	製造技術をもつ半導体企業の工場を誘致 (米、独、日)。 国家プロジェクトの推進 (Rapidus)。
標準化	半導体製造企業ごとに独自手法で半導体を製造している。	互換性が限られ、チップレット技術による製造 や演算環境構築の際に、特定企業の製品への依存度が高まる。	国際標準化団体、業界団体による標準化の推進。

\*1: 集積回路上のトランジスタの数が約2年ごとに2倍になるという予測。 \*2: データ圧縮技術の一つ。高精度の浮動小数点を低精度の整数に変換するため、誤差が大きくなる。

# 処理性能の向上とそれに伴う発熱・電力消費の抑制に対し、従来の電子ICの代替技術の研究は進むものの、2030年頃まではASICやGPUの発展が続くとみられる

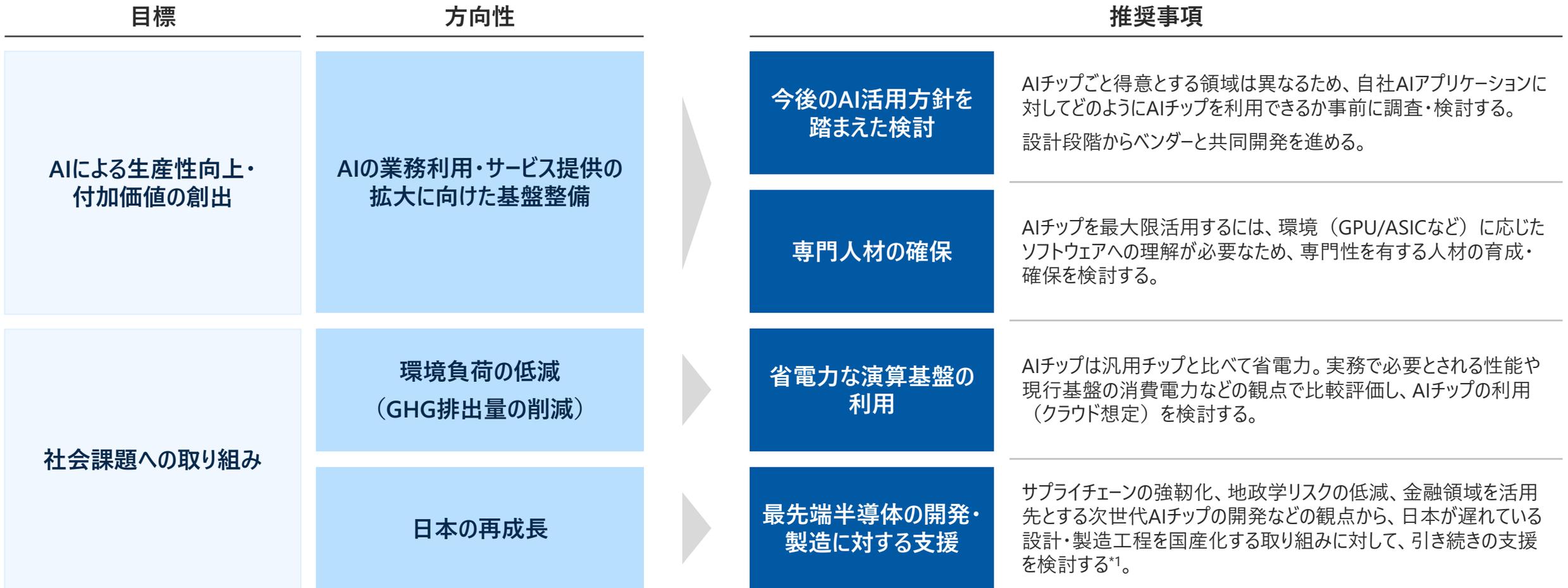
- ニューロモルフィックチップや光/量子などの新型チップの実用化時期は未定であり、既存チップの発展がしばらく続く。それに伴い、最先端ロジック半導体の製造技術が引き続き重要になるため、各国は供給網構築を進めている。
- 日本ではRapidusが最先端ロジック半導体の国産化を目指しており、2027年頃に量産開始を見込む。

取り組み		展望	~ 2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030 ~
処理速度	特定用途向けチップの利用	非ノイマン型に近いASICの開発・利用が続く。							▼Apple製品にNeural Engine搭載							
	低精度ビット演算	少ないビット数で高速・高性能な推論を実現。							▼Cloud TPU提供開始							
熱/電力	ニューロモルフィックチップ	D/A変換*2など実用化には多くの課題を残す。														
	光/量子/バイオ技術の利用	難易度は高く、実用化は早くても2030年以降。														
供給網	工場誘致	TSMCをはじめ、有力企業の誘致が進む。														
	研究開発拠点の国内設置	最先端半導体の製造に向けた拠点整備が進む。														
標準化	標準化の推進	標準規格やガイドラインが普及し、規格準拠品の利用、認証取得などが進む。														

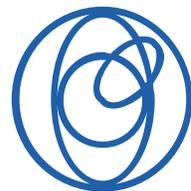


\*1: 深層学習に特化し、半精度/低精度演算で高い性能をもつ回路。 \*2: デジタル/アナログ変換。 \*3: University of California, San Diego. \*4: TSMCとの合併会社。

# AIの本格運用に向け、基盤となるサービス/ハードウェアや専門人材の確保、AIチップの利用による環境負荷の低減、最先端半導体の国産化に向けた支援などの検討を提案する



\*1: 既に複数の半導体関連企業（Rapidus, Western Digital, XIOXIAなど）に対し、SMBCからの出資・融資の検討が進んでいる。



**日本総研**

The Japan Research Institute, Limited