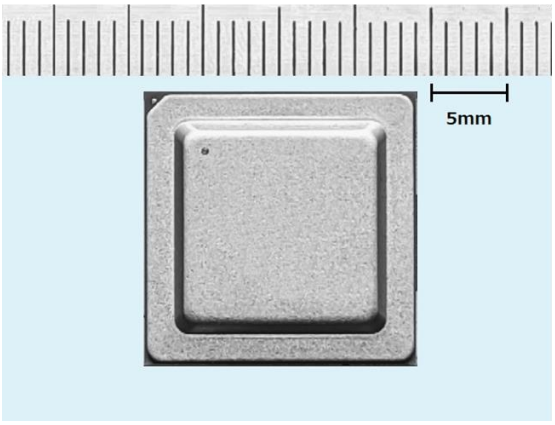
News Release
2020.6.18

国立研究開発法人新エネルギー・産業技術総合開発機構
株式会社ソシオネクスト
ArchiTek 株式会社
株式会社豊田自動織機

# AI エッジ LSI で AI 認識・画像処理効率 10 倍、SLAM 時間 1／20 を達成
## ―ハイブリッド量子化 DNN 技術、進化型仮想エンジンアーキテクチャ技術により実現―

**NEDOは、進化型・低消費電力AIエッジLSIの研究開発事業に取り組んでおり、今般、(株)ソシオネクスト、ArchiTek(株)、(株)豊田自動織機は、同事業でAI認識処理を行うハイブリッド量子化DNN技術、画像処理を行う進化型仮想エンジンアーキテクチャ技術(aIPE)およびリアルタイムSLAM処理技術を開発しました。これらの技術を導入した進化型・低消費電力AIエッジLSIを試作評価したところ、AI認識処理と画像処理が汎用GPUと比較してそれぞれ10倍以上の電力効率化、リアルタイムSLAMの自己位置推定処理時間がCPUと比較して1／20となる短縮を達成しました。**

**今後、このAIエッジLSIが物流やマシンビジョン、セキュリティ・見守り、車載センシングシステムなどに適用されることにより、低消費電力、低遅延、低コストの要件を満足するエッジコンピューティングシステムの構築が可能となります。その結果、急増するデータの高度な利活用促進が加速され、ネットワークの末端(エッジ)側で処理の分散化を実現するAIエッジ技術として、超低消費電力社会の実現が期待できます。**




**・テストチップの主な機能**

| Function | Test Chip |
|---|---|
| CPU | Arm Cortex-A53 Quad Core 1.25GHz |
| AI Processor | aIPE (processor/hardware accelerator) |
| | QNN engine |
| ISP | Image Signal Processor 1080 60fps x 2 |
| Package | 18mm□、0.8 pitch |

**・社会での応用適用例**

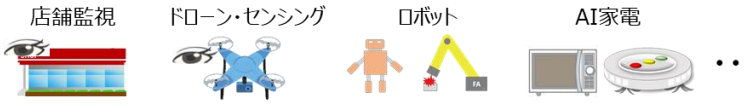



図1　試作した進化型・低消費電力AIエッジLSI

## 1. 概要

IoT社会の到来によりデータ量が爆発的に増加する中、それらのデータの高度な利活用を促進するためには、従来のクラウドによるデータ処理だけではなく、ネットワークの末端(エッジ)側において低消費電力で高度な情報処理を行う「エッジコンピューティング技術」の確立が求められています。その実現のために、エッジ側で処理の分散化を実現するハードウェアからアプリケーションまでの総合技術(トータルソリューション)として「超低消費電力エッジコンピューティング技術開発」が求められています。

こうした背景の下、国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)と株式会社ソシオネクスト、ArchiTek株式会社(アーキテック)、株式会社豊田自動織機は、NEDO事業※1において、人工知能(AI)認識処理、および各種画像処理、リアルタイムSLAM※2処理の技術を開発することを目的としたAIエッジ技術の研究開発テーマを推進してきました。そして今般、AI認識処理を行うハイブリッド量子化ディープ

ニューラルネットワーク（DNN）技術※3、各種画像処理を高速に並列実行可能な進化型仮想エンジンアーキテクチャ技術（aIPE）※4、およびリアルタイムSLAM処理技術を開発しました。これらの技術を取り入れた進化型・低消費電力AIエッジLSI※5を試作して評価した結果、AI認識処理と画像処理において汎用GPU※6と比較してそれぞれ10倍以上の電力効率化に成功し、リアルタイムSLAMの自己位置推定処理においては汎用CPUと比較して1／20の処理時間の短縮を達成しました。

今後、今回の技術を用いたAIエッジLSIを物流やマシンビジョン、セキュリティ・見守り、車載センシングシステムなどに適用することにより、低消費電力、低遅延、低コストの要件を満足するエッジコンピューティングシステムの構築が可能となり、急増するデータの高度な利活用促進が加速され、ネットワークの末端（エッジ）側で処理の分散化を実現するAIエッジ技術として、超低消費電力社会の実現が期待できます。

## 2. 今回の成果

今回、NEDOと（株）ソシオネクスト、ArchiTek（株）および（株）豊田自動織機が新たに開発したハイブリッド量子化DNN技術、進化型aIPEおよびリアルタイムSLAM処理技術は以下の特長を持ちます。

（1）ハイブリッド量子化DNN技術

深層学習（ディープラーニング）を実行するのに必要なパラメーター※7やアクティベーション※8を低ビット化する技術です。開発したハイブリッド量子化DNN技術は、従来浮動小数点を32ビット／16ビット、整数を8ビットで表していたネットワーク構造を、バックボーンネットワーク※9は3進数を2ビット、2進数を1ビット、ヘッドネットワーク※10は整数8ビットと、複数の量子化精度を混在させる技術です。これにより、認識精度の低下を抑えながら、低消費電力を実現しました。さらに、学習環境でTensorFlow※11に適合した量子化ライブラリ、推論環境でハイブリッド量子化エンジン、および学習環境から推論環境への変換処理技術を開発しました。これら技術を用いると、高速で低消費な電力でAI認識処理を実行できます。試作したAIエッジLSIを測定した結果、ハイブリッド量子化DNN技術により、AI認識処理において汎用GPUと比較して10倍以上の電力効率化に成功しました。

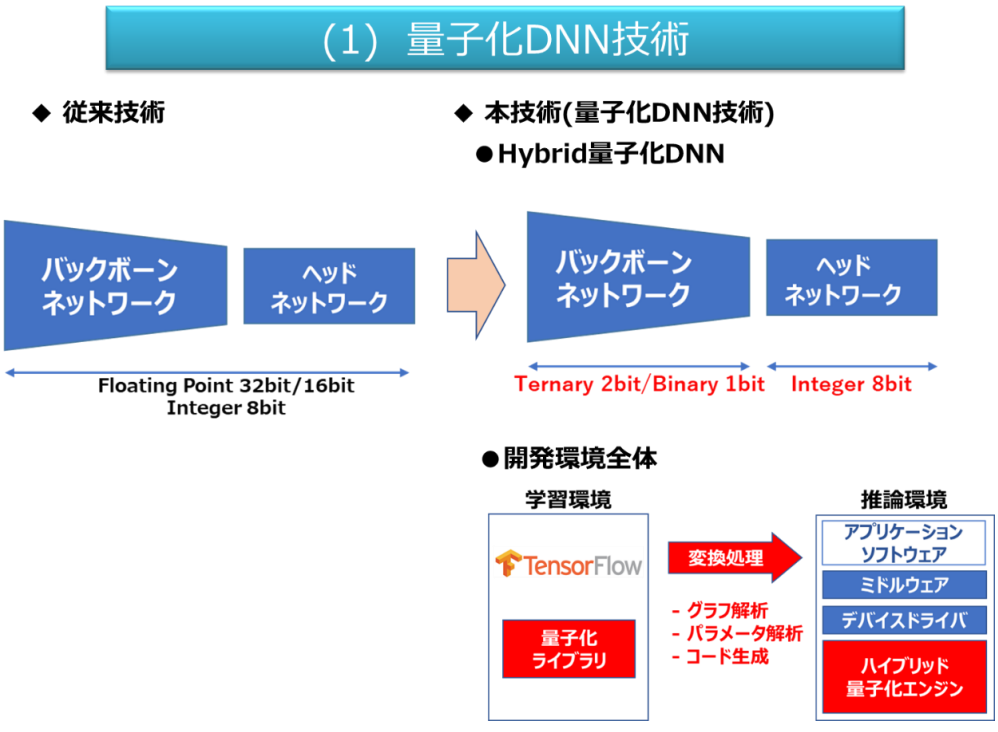



図2　開発したAI認識処理を行う量子化ディープニューラルネットワーク(DNN）技術の概要

(2)進化型aIPEおよびリアルタイムSLAM処理技術

従来の仮想エンジンアーキテクチャ技術(aIPE)では、最小限のハードウェア部品を組み合わせてアルゴリズムを構築し、それらを実行する時間スロットに密に配置することで、高効率化と柔軟性と多様性を実現してきました。今般、開発した進化型aIPEは、最適化向上およびAI拡張機能の実装のためのアーキテクチャ改良により、高速・低消費電力で画像処理を実行します。開発した主な技術は、画像処理およびリアルタイムSLAM処理の高速化に関する技術です。SLAMアルゴリズムを最大限活用できるアーキテクチャによりLiDAR※12センサーを使用するSLAM処理を効率よく実行します。これらに加えて、①ハードウェア調停機構の改良およびメモリコントロール機能を最適化する技術、②ディープラーニングのネット定義を柔軟に可能にして、畳み込み回路などの要素部品を自由にプラグインする技術、③LiDARからVisual SLAMへの機能向上、④量子化演算器を開発しました。これら技術を従来のaIPEに取り入れ、論理設計から実装に耐えうるIPへ変換することにより、アルゴリズムの進化や幅広いAIエッジ応用に対応できるプラットフォームとなりました。

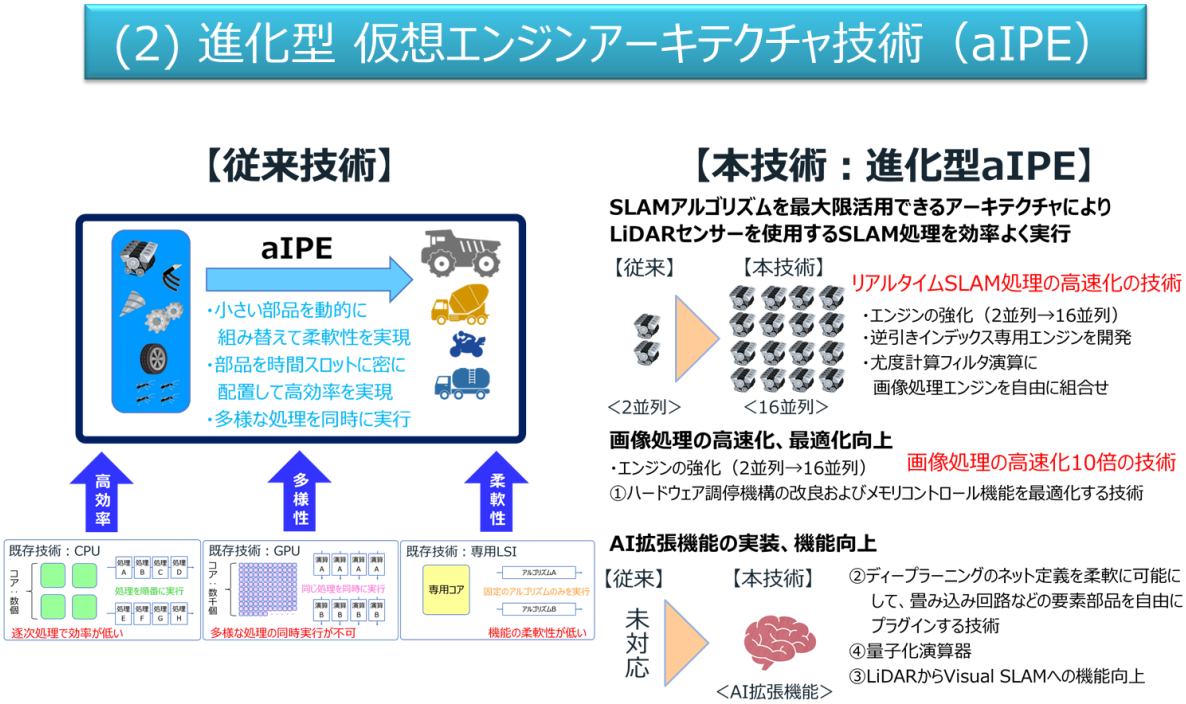



図 3　各種画像処理を並列実行可能な仮想エンジンアーキテクチャ技術の概要

試作したAIエッジLSIで測定した結果、進化型aIPEにより、画像処理において汎用GPUと比較して10倍以上の電力効率化に成功しました。

また、試作したLSIの進化型aIPE上で並列して動作可能なリアルタイムSLAM処理ライブラリを開発し、エッジコンピュータによるSLAM処理で課題となっていた高速移動ロボットの高精度な自己位置の推定処理において、CPUと比較して1／20の処理時間に短縮できることを確認しました。

## 3. 今後の予定

NEDO と(株)ソシオネクスト、ArchiTek(株)、(株)豊田自動織機は、今後、進化型 aIPE とハイブリッド量子化 DNN 技術を統合し、併せてリアルタイム SLAM 処理技術の高度化、コンピュータービジョンと AI 基本ミドルウェアライブラリの開発、およびクラウド環境とエッジ環境の最適化技術のさらなる開発を進め、より一層低消費電力で動作する進化型・低消費電力 AI エッジ LSI を構築していきます。これにより、産業検査、運転支援、ドローンなどへの適用拡大に向けて、高度な AI を低消費電力で実行できる技術の確立を目指

します。

なお、今回開発した進化型 aIPE を取り込んだプラットフォームは、ArchiTek（株）が、2020 年 10 月から IP（回路情報）を提供する予定です。

【注釈】

※1 NEDO事業

事 業 名：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／革新的AIエッジコンピューティング技術の開発／進化型・低消費電力AIエッジLSIの研究開発

事業期間：2018年度～2020年度

※2 リアルタイムSLAM

Simultaneous Localization and Mapping（自己位置推定と環境地図作成の同時実行）。移動体の自己位置推定と環境地図作成を同時に行う技術の総称です。

※3 量子化ディープニューラルネットワーク（DNN）技術

Deep Neural Network（DNN）とは、ニューラルネットワークをディープラーニングに対応させて4層以上に層を深くしたもので、量子化DNNではDNNの演算アルゴリズムの処理を低ビット化することにより計算量を減らし低消費電力演算を行うものです。

※4 仮想エンジンアーキテクチャ技術（aIPE）

ArchiTek Intelligence® Pixel Engineの略称です。画像処理やAI処理に必要な機能を直行的に抽出してハードウェアで実現し、それらハードウェア部品を組み合わせることで高度で応用範囲の広いアプリケーションが実行可能な柔軟なエンジンで、ArchiTek（株）が独自に開発した技術です。ハードウェア部品は少ない資源で幅広く利用できるよう、合理化・強化したことがポイントで、小型、低コスト、低消費電力を実現します。

※5 エッジLSI

利用者に近いネットワークの末端に位置するIoT機器で使用される半導体チップです。クラウドやサーバーなどと比較し、利用できる電力や発熱、コストなどの制限が厳しいチップです。

※6 汎用GPU

Graphics Processing Unit（画像処理用演算プロセッサ）。リアルタイム画像処理に特化した演算装置です。

※7 パラメーター

演算に用いる多数の固定数値です。

※8 アクティベーション

ニューラルネットワークにおける、中間レイヤーの入力値です。

※9 バックボーンネットワーク

画像入力からフィーチャーマップを抽出するネットワークです。

※10 ヘッドネットワーク

バックボーンネットワークからフィーチャーマップを受け取り、後段処理を実行するネットワークです。

※11 TensorFlow

さまざまな機械学習の分野で使用するためのオープンソフトウェアライブラリで、多次元のデータ構造（テンソル）を、流れるように処理することができる深層学習（ディープラーニング）のためのライブラリです。

※12 LiDAR

「light detection and ranging（光による検知と測距）」の頭文字をとった言葉で、レーザー光を照射し、物体に当たって跳ね返ってくるまでの時間を計測し、物体までの距離や方向を測定する技術です。

## 4. 問い合わせ先

（本ニュースリリースの内容についての問い合わせ先）

NEDO IoT 推進部 担当：広瀬、西山 TEL：044-520-5211

ソシオネクスト　広報担当：泉　TEL：045-568-1006
問い合わせフォーム　https://www.socionext.com/jp/contact/

ArchiTek　担当：CFO　藤中　E-mail：pico@architek.co.jp

豊田自動織機　広報部　担当：宮崎　TEL：0566-27-5157

（その他NEDO事業についての一般的な問い合わせ先）
NEDO　広報部　担当：坂本、佐藤　TEL：044-520-5151　E-mail：nedo_press@ml.nedo.go.jp