

低消費電力・小型化を特長とする OKI ASICを活用したAIシステム

宮村 紀秀

近年のAI技術の発展に伴い、労働生産性の向上など社会課題を解決するAIソリューション製品が各社から提供されている。それら製品の主流は、エッジ（現場）からクラウドサーバーにデータを転送し、クラウドサーバー側で学習・推論を行うクラウド型AIである。クラウド型AIでは、クラウドサーバーとエッジ間の通信が必要であるため、「リアルタイム性の欠如」という課題がある。

本稿では、低消費電力でかつ小型なASIC（Application Specific Integrated-Circuit: 特定用途向けIC）を活用したAIシステムの技術の概略と今後の展開について紹介する。ここでのAIシステムは、AI推論を実行するハードウェア基板と、その基板上で動作する実行コードを生成するソフトウェア開発環境を指す。本AIシステムを提供することで、電力の安定供給が困難なケースや設置スペースが限られるケースなど従来では適用が困難なエッジ環境にもAI技術を展開できる。

OKI ASICの紹介¹⁾

ASICとは、特定の用途に特化して作り込むタイプの半導体デバイスである。一般にAI推論を実行するハードウェアとしてGPU（Graphics Processing Unit）やFPGA（Field-Programmable Gate Array）などがある。ASICはGPUやFPGAと比較して汎用性に欠けるというデメリットがあるが、低消費電力でかつ小型であるというメリットを備えている。

OKIではAIシステム用に独自ASICを開発した。OKIのASICは、フル稼働時でも冷却用のヒートシンクやファンが不要となるレベルまで消費電力を抑えている。冷却装置を取り付ける必要がないため、ASICを搭載する基板も小型化できる。低消費電力でかつ小型、ファンレスという特長から、以下のようなエッジ環境においても利用が可能となる。

- ①電力の安定供給が困難なバッテリー駆動環境
（低消費電力であるため）
- ②機器の内部など限られた省スペース環境
（小型、低消費電力であるため）
- ③水中などの密閉空間環境
（ファンによる外気冷却を必要としないため）

- ④病院などの静音が求められる環境
（ファンによる騒音がないため）

本ASICのスペックは表1に示すとおりである。

表1 ASICのスペック

項目	詳細
CPU	メイン： SIMD エンジン搭載×4 コア
AI 推論	独自 AI アクセラレーター搭載
画像処理	独自画像処理アクセラレーター搭載

AI推論の演算は年々大規模化している。これらの演算にも対応するため、メインCPU（Central Processing Unit）としてSIMD（Single Instruction, Multiple Data）エンジン搭載のコアを4つ備えている。また、AI推論用に独自のAIアクセラレーターを備えている。本AIアクセラレーターについては次の章で詳細に述べる。さらに、画像処理用に、独自の画像処理アクセラレーターを備えており、データ圧縮・伸張、色空間変換、ボックスフィルター、画像回転などの計算を効率化できる。ASICの動作時でもヒートシンクは不要であり、また、待機時には通信機能を維持したままに大幅に電力を削減する機能を備えている。写真1は独自に開発したASICの外観である。



写真1 ASIC 外観

AI アクセラレーター

前述したように、OKIで開発したASICには、OKI独自開発のAIアクセラレーターを搭載している。本章ではAIアクセラレーターについて紹介する。前章において、ASICは汎

用性に欠けるというデメリットがあると述べたが、本AIアクセラレーターは回路規模を増やさずに基本的な演算を網羅することで高い汎用性を実現している。図1はAIアクセラレーターの機能概略である。

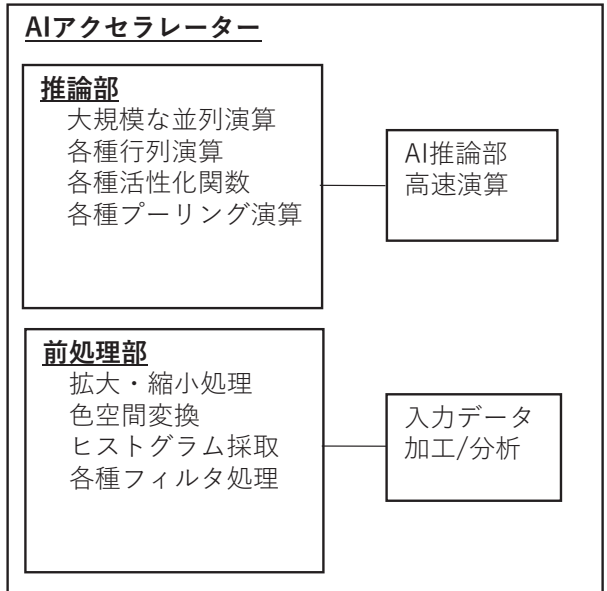


図1 AIアクセラレーター機能概略

AIアクセラレーターは推論部と前処理部から構成される。

●推論部:

汎用性を高めるために公開されている画像認識や超解像処理などのNN (Neural Network) モデルを参考にして、さまざまなNNモデルに対応できるように設計している。具体的には、NNモデルの主たる構造である大規模な並列積和演算はもちろんのこと、各種行列演算、各種活性化関数、各種プーリング関数も計算できる。

●前処理部:

それぞれのNNモデルには決められた入力フォーマット(サイズ、色空間、スケーリングなど)が存在する。例えば、カメラから取得した入力画像データが必ずしも利用するNNモデルの定めた入力フォーマットに合うとは限らない。この前処理部は、取得したデータをNNモデルの入力フォーマットに変換するために加工/分析する部分である。以下の処理が存在する。

・拡大・縮小処理:

画像データをNNモデルの入力データサイズに変換

・色空間変換:

画像データ色空間をNNモデルの入力色空間に変換
(例) 色ごとの輝度表現RGBから輝度/色差表現YCbCr

・正規化 (Normalization):

最小値“0”、最大値“1”とする特徴量スケーリング

・標準化 (Standardization):

平均“0”、分散“1”とする特徴量スケーリング

・ヒストグラム採取:

値ごとの出現頻度分布を求め、正規化に必要な最小値、最大値、標準化に必要な平均値、標準偏差を算出

推論部、前処理部で処理できない計算に関しては、主にメインCPUで計算できるが、データ圧縮・伸張、ボックスフィルタ、画像回転などの計算には、前章の独自画像処理アクセラレーターも利用できる。

外部接続インターフェース

前章まではAI推論に関するハードウェアを中心に説明した。エッジ環境におけるAIソリューションにおいては、単にAI推論が行えるだけでなく、AI推論の入力となる画像などを取得する必要がある。本章ではOKIで開発したASICの周辺との外部接続インターフェースについて紹介する。表2は本ASICが搭載しているインターフェースである。

表2 外部接続インターフェース

種別	詳細
ホスト	USB、有線/無線 LAN
カメラ	USB
その他	USB(マイク)、I ² C(温湿度センサー)、SPI(水圧水位計)、AD 変換(電流計)、PCIe(拡張デバイス) など

外部のホストコンピューターと通信するために、USB[®] *1)、有線/無線LAN (Local Area Network) を備えている。カメラとのインターフェースとしては、USBを備えている。USB以外にも、入力センサーのインターフェースとして広く使われているI²C[®] *2)、SPI (Serial Peripheral Interface)、AD (Analog-to-Digital) 変換を備えており、画像だけでなく、音、温湿度、水圧水位、電流などさまざまな情報を取得するセンサーにも接続できる。また、拡張デバイス用にPCIe[®] *3)も備えており、拡張ストレージデバイスやほかのAIアクセラレーターも接続できる。さらに、スピーカー、デバッガー、LCD (Liquid Crystal Display) パネルなども接続できるインターフェースも備えている。さまざまなデバイスと接続できるため、幅広い分野のAIソリューションに適用できる。

ハードウェア基板構成

本稿執筆時点でASICを搭載する基板の開発を進めている。

*1) USB(Universal Serial Bus)は、USB Implementers Forumの登録商標です。 *2) I²C(Inter Integrated Circuit)は、NXP Semiconductorsの登録商標です。
*3) PCIe(Peripheral Component Interconnect Express)は、PCI-SIG(PCI Special Interest Group)の登録商標です。

図2はASICを搭載した評価基板のブロック図である。この評価基板ではさまざまな機能を含んでいるが、必要な機能を絞り込むことで実運用の基板のサイズをさらに小さくできる。今後、カメラを利用する画像処理AIやマイクを利用した音声処理AIなど、適用するAIソリューションごとに必要な機能を絞り込み、基板サイズを小さくしていく予定である。一例として、カメラを利用したAIソリューションの場合、基板に搭載するカメラサイズを基板サイズの目標としており、そのサイズは60mm×60mm程度である。

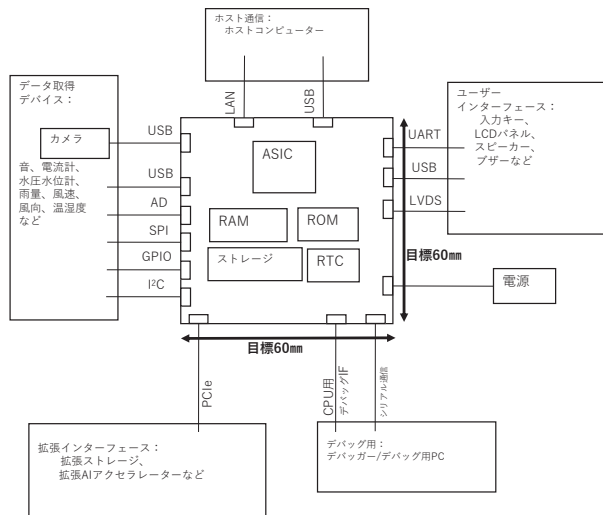


図2 評価基板ブロック図

実行コード生成ツール

本章ではAIアクセラレーターを利用するために開発中の実行コード生成ツールについて紹介する。従来、AIアクセラレーターのようなハードウェア回路を利用するには、その機能や制御方法を理解する必要があった。そのうえで、制御レジスターの設定や終了割り込み制御、必要であればCPUキャッシュ上のデータをメモリーに書き戻すなどの処理を実行するコードに埋め込まなければならない。一方、AI推論の開発においては、既存の学習済みNNモデルから目的に近いものを選択し、学習データを収集・学習させる形が一般的となっており、実際のAI推論の演算内容をコード化しない。そのため、AI推論の演算時にハードウェア回路を制御するコードを挿入した実行コードを生成するのに労力がかかる。この労力を抑えるため、実行コード生成ツールは、学習済みNNモデルからOKI ASIC上で動作する実行コードを自動的に生成する。

図3は実行コード生成と実行環境の概略である。

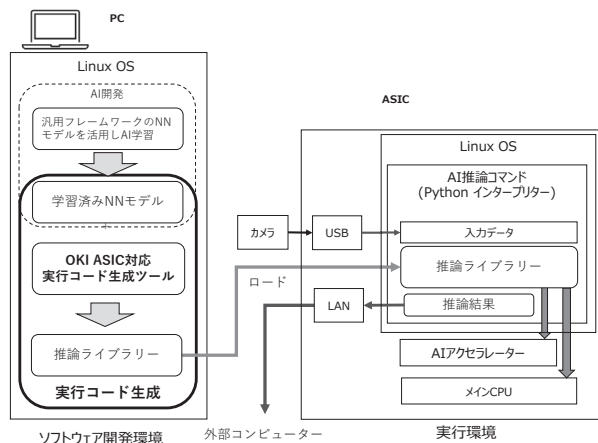


図3 実行コード生成と実行環境の概略

実行コード生成ツールは、Linux[®] *4) OS上で動作し、TensorFlow[®] *5)などのNNモデルの汎用的フレームワークをほぼすべてサポートする。このツールを利用すれば、学習済みNNモデルから推論ライブラリーを生成できる。この推論ライブラリーはOKI ASIC上でAIアクセラレーターでの演算、CPUでの演算を利用してAI推論を実行するコードである。

次に、画像を利用したAI推論を例に、この推論ライブラリーを実行する方法について説明する。

STEP1:推論ライブラリーをASIC上にロードする。

STEP2:AI推論コマンドを実行する。

Linux OS上のPython[®] *6) インタープリターでAI推論コマンドを実行できる。このAI推論コマンドでは、以下のSTEP2-1、STEP2-2を実行することにより推論結果を得る。
STEP2-1:カメラから入力画像データを取得する。
STEP2-2:推論ライブラリーを呼び出し、AI推論を実行する。
STEP3:推論結果を送信する。

LANなどを通じてその推論結果を外部コンピューターに送信する。

STEP4:推論結果に基づいた処理を実行する。

適用するソリューションの例

適用するAIソリューションとしては以下のものを考えている。

・人物検知:

ソーラー式のバッテリー駆動のカメラにAIシステム（OKI ASICを搭載した基板）を設置する。カメラの画像から、不審者や侵入者がいないかを検知する。不審者や侵入者がいた場合、管理者に通知する。バッテリー駆動と

*4) Linuxは、米国及びその他の国におけるLinus Torvaldsの登録商標です。 *5) TensorFlowは、Google Incの登録商標です。 *6) Pythonは、AI・機械学習などで幅広く利用されている汎用的なプログラミング言語であり、Python Software Foundationの登録商標です。

いった電力供給が限られた環境で利用することで、本AIシステムの低消費電力であるという特長を活かせる。

・故障検知

機器の内部にAIシステム（マイクとASICを搭載した基板）を設置する。機器内部の動作異音を検知することで、故障を検知する。機器の内部という限られたスペースに設置することで、本AIシステムの小型であるという特長を活かせる。

まとめ・今後の予定

本稿においては、AIシステムで利用するOKI開発のASIC及びAIアクセラレーター、外部接続インターフェース、実行コード生成ツール、ハードウェア基板構成、適用するソリューションの例について紹介した。

ハードウェア基板と実行コード生成ツールについては現在開発中である。ハードウェア基板は小型化を進めていく。また、実行コード生成ツールに関しては、AI推論部分だけでなく、周辺のソフトウェアに関しても実行コードを生成できるように開発を進めていく。そのうえで、適用可能なAIソリューションを増やしていきたいと考えている。 ◆◆

参考文献

1) 齋木祥廣、渡辺啓介、若林円:エンタープライズソリューション事業製品の次世代プラットフォーム技術、OKIテクニカルレビュー 第242号、Vol90 No.2、pp40-43、2024年2月

●筆者紹介

宮村紀秀:Norihide Miyamura.技術本部 研究開発センター
ハードウェア研究開発部

TiPo 【基本用語解説】

AIアクセラレーター

NN処理を高速化するために設計されたハードウェア。

NN (Neural Network)

脳の神経回路網を模した数理モデル。

推論ライブラリー

ハードウェア資源を効率的に利用して推論できるように学習済モデルをコード化したもの。