ROHM SEMICONDUCTOR

Electronics for the Future

オンデバイス学習AIチップ 紹介資料

2022年9月27日
ローム株式会社
マーケティング・コミュニケーション部

＊「tinyMicon MatisseCORE™」「RapidScope ™」は、ローム株式会社の商標または登録商標です。
＊本資料は発行日付時点の情報です。予告なく変更することがあります。

AIと機械学習の概要

人工知能(Artificial Intelligence)

人間の機能の一部を実現する
画像認識など

機械学習(Machine Learning)

AIが機械的に学習する手法

ディープラーニング：学習を深層化したもの

ニューラルネットワーク(Neural Network)

機械学習の一つ

AIの学習と推論

猫と犬の画像認識AIなら、

学習 ・・・ AIがたくさんの画像を見て、犬と猫の特徴を覚えること

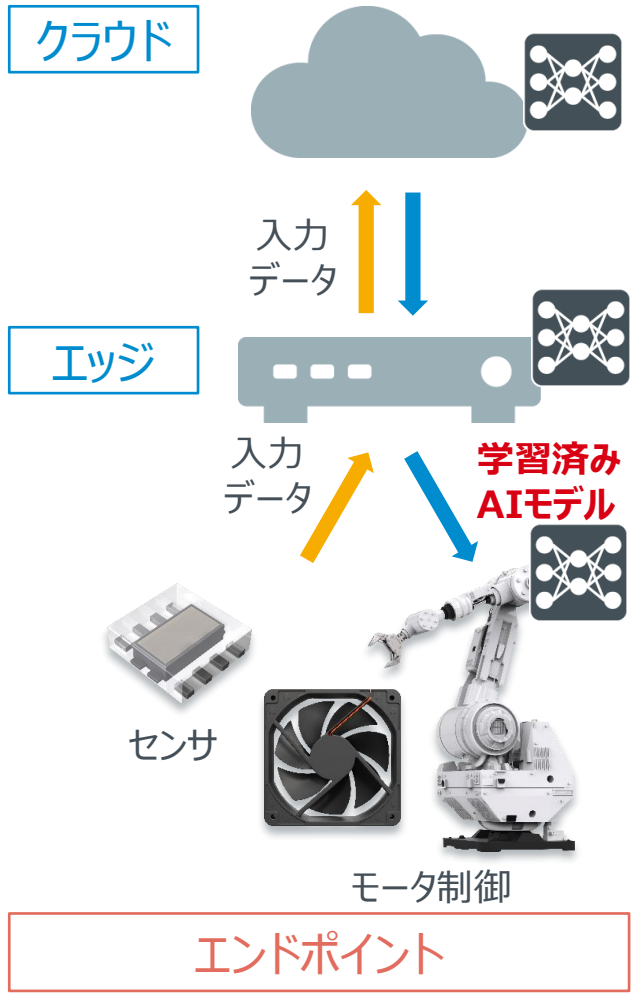
推論 ・・・ AIが画像を見て、猫か犬か判定すること

学習には演算パワーが必要!

AIはクラウドからエッジ、エンドポイントへ広がっている

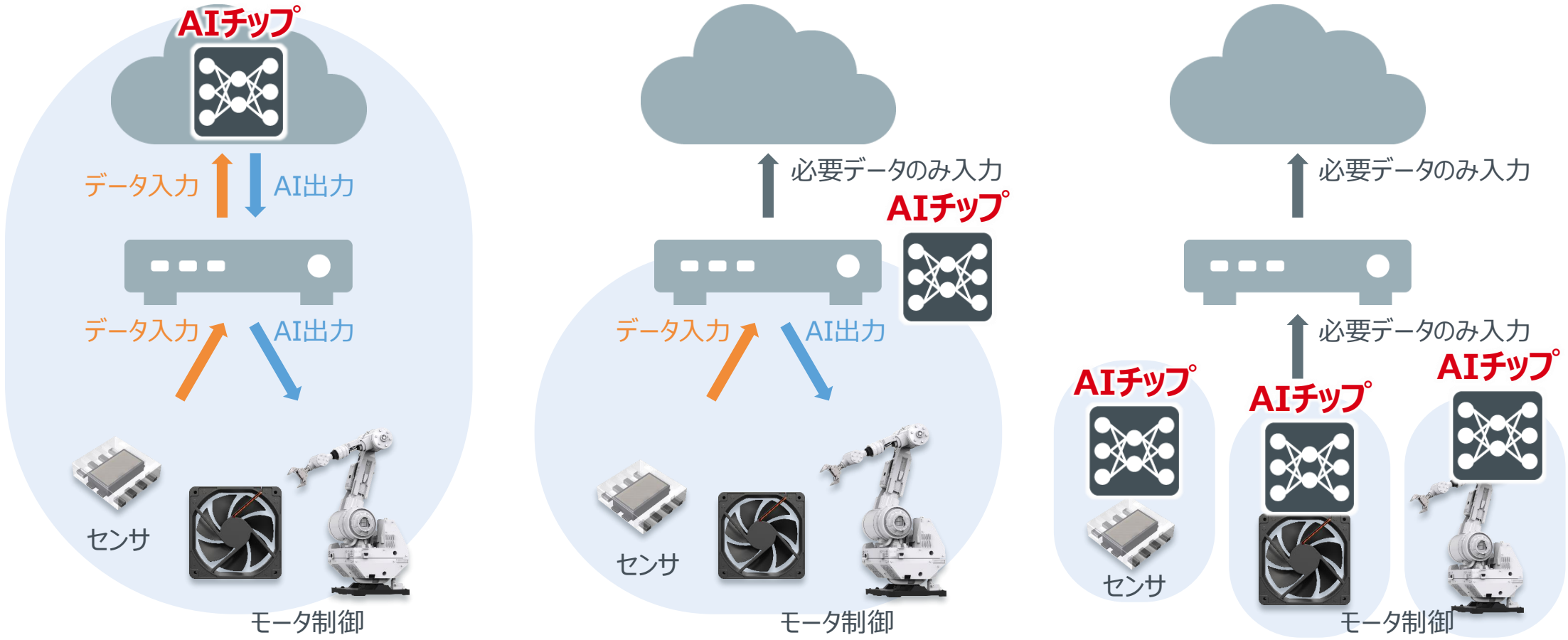
	従来型クラウドAI	エッジAI	エンドポイントAI
AI機能	学習も推論もクラウド	学習はクラウド 推論はエッジ	学習はクラウド 推論はエンドポイント
要求性能	• 優れた学習能力 • 高度なセキュリティ	• ネットワーク負荷軽減 • 短い応答時間 • 低消費電力	• ネットワーク負荷ゼロ • 極めて短い応答時間 • 超低消費電力
課題	• 通信コスト、電力が増大 • 応答時間の変動あり • セキュリティコストがかかる	• エッジに高性能なFPGAやGPUが必要 • 少ないながら応答時間の変動あり	• 組み込みMCUの性能に応じたAIモデルに限定される

AIの広がりイメージ



クラウド型AIシステムとエンドポイント型AIシステムの比較

	クラウド型AIシステム	エッジ型AIシステム	エンドポイント型AIシステム
クラウドコンピュータ	AIチップ	必要データのみ入力	必要データのみ入力
エッジコンピュータ	データ入力 / AI出力	AIチップ データ入力 / AI出力	必要データのみ入力 AIチップ
エンドポイント	センサ モータ制御	センサ モータ制御	AIチップ センサ AIチップ モータ制御
	クラウドコンピュータのAIに負荷が集約される	エッジコンピュータのAIに学習・推論の負荷を分散できる	エンドポイントのAIに負荷を分散できる



エンドポイント型AIシステムの応用例

今回の
主ターゲット

画像認識

複雑なAIの実現に、
高性能なGPU/FPGAが必要

各種機器の故障予知

比較的小さなAIで十分
高い精度よりはサイズ、コスト重視

機械の故障が最初に現れやすいモータの故障予知

課題

- 設置機器ごとに学習が必要
- 環境の変化があると再学習が必要
- IC製品一つ一つ設計が必要

→ **効率化したい**

これらの課題を**独自技術(＝オンデバイス学習)**により解決する

**省電力、高速応答、超小型で、
リアルタイム学習ができる新しいAIソリューションを開発**

オンデバイス学習(デバイス上での機械学習)

オンデバイス学習アルゴリズム

慶應義塾 Keio University

オンデバイス学習 デバイス上で高速にAI学習する技術

→ チップ内で学習可能で、学習用データの準備が不要
→ クラウド等での事前学習が不要
→ **現場で学習できる**ので、ばらつきや環境の変化に強い

＋

実デバイスでの回路技術

AIアクセラレータ(AxlCORE-ODL)

→ **低コストなハードウェア回路**でAIを実現

小型CPU MatisseCORE

→ AIの構成を**ソフトで柔軟に変更**可能

特徴

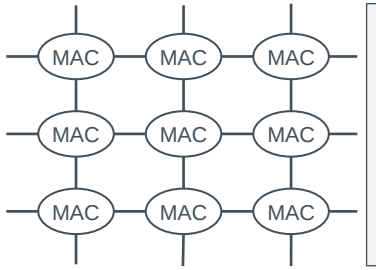
① **低コスト**
= エッジデバイス向けを想定(解析的重み計算)

② **追従能力**
= 変化するパターンに対応(軽量忘却機構)

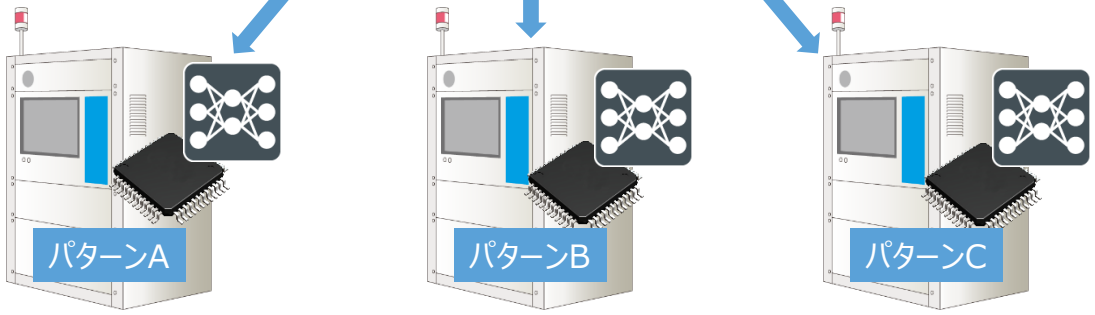
③ **高精度**
= 正常パターンが複数あっても精度を維持
(アンサンブル手法)

④ **安定動作**
= 組み込まれて常時稼働
(過学習の抑制と出力の安定化)

オリジナルのAI回路



設置場所ごとに異なった環境をその場で学習できる
(現場学習)



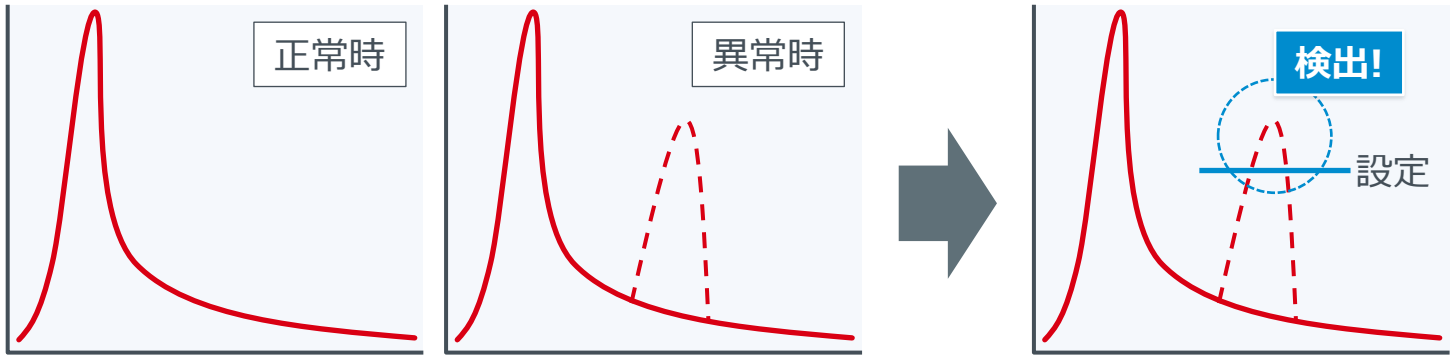
設置場所A　設置場所B　設置場所C

場所ごとの事前データ収集不要！
各デバイスへのダウンロード不要！

補足　AIによる故障予知(故障予兆検知)、異常検出とは？

AIは、正常動作時からの変化を数値化することで、未知の異常であっても検出できる

従来手法による故障予知



異常時の変化(特定のピークが現れる等)が分かっていれば検出できる

どんなデータが入力されるか、
異常時の変化がどうなのかを
人間が設定しなければ検出できない

AIによる故障予知

想定外の異常、あるいは異常時の変化が未知であっても、異常を検出できる

どんなデータが入力されても、
AIが正常を学習することで、
異常を検出できる(推論できる)

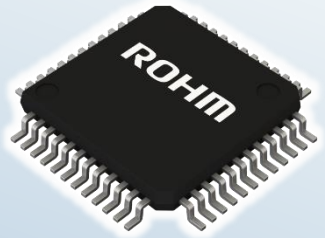
オンデバイス学習アルゴリズムは、リアルタイムに現場で実現する(クラウドサーバー不要)

ロームのエンドポイント向け、試作オンデバイスAIチップ「BD15035」の概要

オンデバイス学習に必要なAIアクセラレータ、CPU、入力I/Fを1チップ化

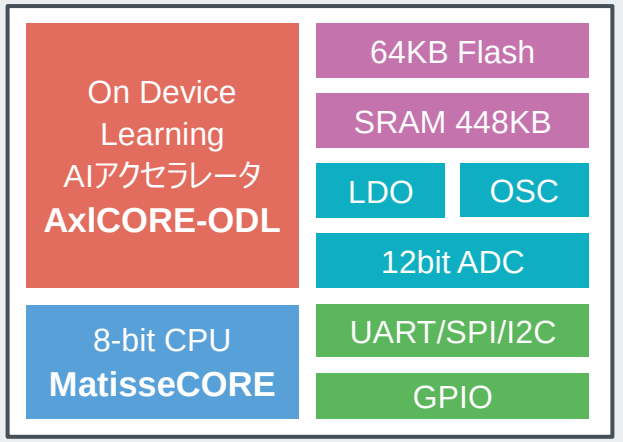
主な回路・機能

- AIアクセラレータ「AxlCORE-ODL」搭載
  - AIのベースにオンデバイス学習アルゴリズム採用
    (3層ニューラルネットワーク)
  - FFT、フィルタ処理が可能
- 8-bit CPU「tinyMicon MatisseCORE™」搭載
- 入力I/FにUART/SPI/I2C、12bit ADCを搭載



BD15035
TQFP48V Package
(9.0 mm × 9.0 mm × 1.2 mm)

BD15035回路



特長

AI機能
- オンデバイス学習可能：
  事前学習やクラウドサーバーでの解析不要
  (3層ニューラルネットワーク)

超低消費電力
- 数10mWの消費電力：電池駆動やエンドポイントでの動作が可能

小型チップ
- AI機能を超小型AIアクセラレータとして再構築
- 小型、高効率8-bit CPU

高速処理
- AIアクセラレータによる高速処理でCPUへの負荷が少ない

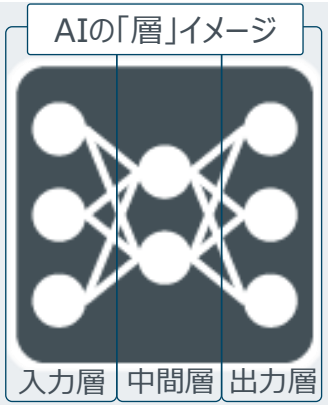
機器が設置された現場でリアルタイムの故障予知(故障予兆検知)を実現

各AIチップとロームのエンドポイント向けAIチップ 性能比較

	クラウドコンピュータ向け AIチップ	エッジコンピュータ向け AIチップ	従来 エンドポイント向け AIチップ	ローム エンドポイント向け AIチップ
要求性能	• 優れた学習能力 • 高度なセキュリティ	• ネットワーク負荷軽減 • 短い応答時間 • 低消費電力	• ネットワーク負荷ゼロ • 極めて短い応答時間 • 超低消費電力	• ネットワーク負荷ゼロ • 極めて短い応答時間 • 超低消費電力
ハードウェア構成	高性能GPU/ 機械学習専用プロセッサ	組み込みGPU/FPGA	MCU	AIアクセラレータ ＋Matisse搭載MCU
消費電力	**20W ～ 200W**	**2W ～ 10W**	**20mW ～ 1000mW**	**約30mW** ※特定アプリケーション動作時の実測値
応答時間	**数秒 ～ 数十秒**	**数秒**	**ミリ 秒**	**ミリ 秒**
学習	**可**	**不可** **※学習済みのAIモデルを使用**	**不可** **※学習済みのAIモデルを使用**	**可**
推論	可	可	可	可

消費電力わずか数10mWのAIチップで、学習・推論が可能
エンドポイントでリアルタイムの故障予知を実現

補足　AIのマッピングイメージ



ディープラーニング(数十層の中間層を持つ)

※応用例

- 人の代わりに囲碁や将棋を指す
- 気象情報を予測する
- 監視カメラと画像の人を識別する　等

3層ニューラルネットワーク

※応用例

- 人の動きを識別できる

例)イメージセンサで人が倒れているか、起きているか程度

3層ニューラルネットワークのAIチップ「BD15035」

（メモリ等のスペックによる制約による）

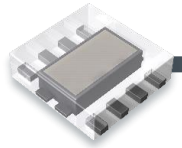
- 加速度や電流、音声等の識別による故障予知

AIチップ「BD15035」の処理・役割分担

入力データに対するAIチップ出力までのフロー

数多の入力データに対して学習を行い、
「いつもと違う！」を数値化する

入力データ



各種センサ信号など

入力I/F
- 12bit ADC
- UART/SPI/I2C

前処理回路
- FFT
- フィルタ処理

AI学習・推論(AI回路)
- 3層ニューラルネットワーク

異常度

$$L = \sum_{k=0}^{n} |x - y|^2$$

後処理回路
- 出力判定

エッジ
コンピュータ側

AI回路が演算しやすいように
前処理を行う

Matisseに設定された異常の
閾値で出力をフィルタリングできる

BD15035

On Device Learning AIアクセラレータ **AxlCORE-ODL**

64KB Flash

SRAM 448KB

LDO　OSC

12bit ADC

UART/SPI/I2C

GPIO

8-bit CPU **MatisseCORE**

AIアクセラレータ「AxlCORE-ODL」

演算パワーの必要な機能を担当
- 入力データの前処理
- AI演算

8-bit CPU「MatisseCORE」

アプリケーションごとに変更が必要な機能を担当
- AIモデルの設定
  (ニューラルネットワークのノード数等)
- 前処理回路へのデータ入力(補助)
- 学習・推論の切替や制御
- 出力判定の閾値の設定

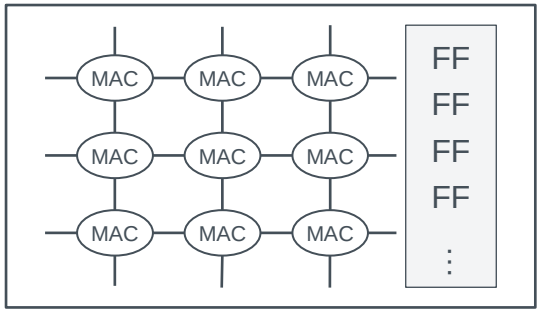
詳細

ロームの技術:
AIアクセラレータ「AxlCORE-ODL」による、小型化と高速化

**慶應義塾大学から提供されたオンデバイス学習の回路(AI回路)を、
AIアクセラレータで再設計してゲート数を削減**

オリジナルのAI回路

500万ゲート

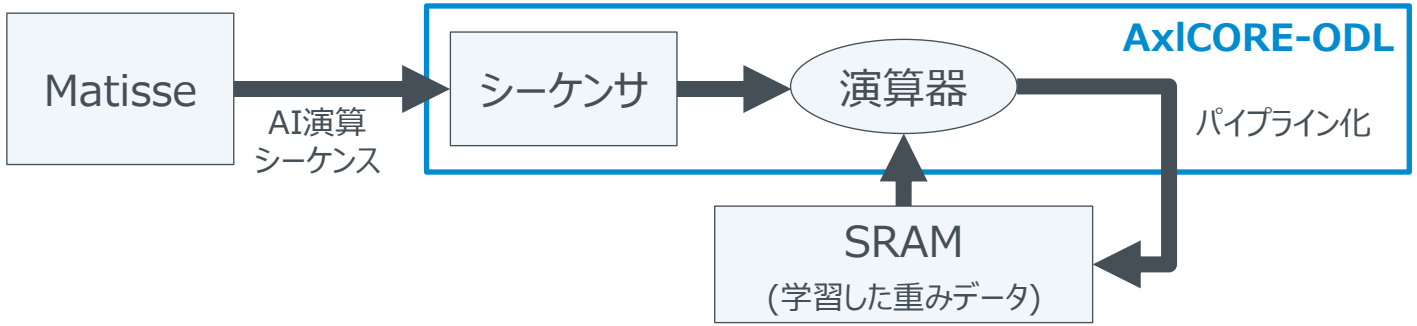


ゲート数250分の1
0.4%まで小型化

ロームのAIチップ(AIアクセラレータ+Matisse)

2万ゲート

特許2件申請中



- 固定小数点32bit
- 多数の積和演算器(MAC)とFFで構成され回路が非常に大きい
- AIの構造が固定

✓ **bfloat 16bit浮動小数点演算**を採用、バイナリ演算に比べ精度がよい
(多くのAIチップは、高速化・省メモリのために、1-2bitのバイナリ演算で精度が落ちる)

✓ **MatisseでAI演算シーケンスを設定**することで、演算器を１つに集約

✓ **AIの構造(入力データ数、アルゴリズム)を可変にした**
- 処理時間やメモリ使用量のバランス、アルゴリズムの改良が可能

✓ SRAMからのデータ取得、演算、保存をパイプライン化して**処理速度を3倍**に高めた

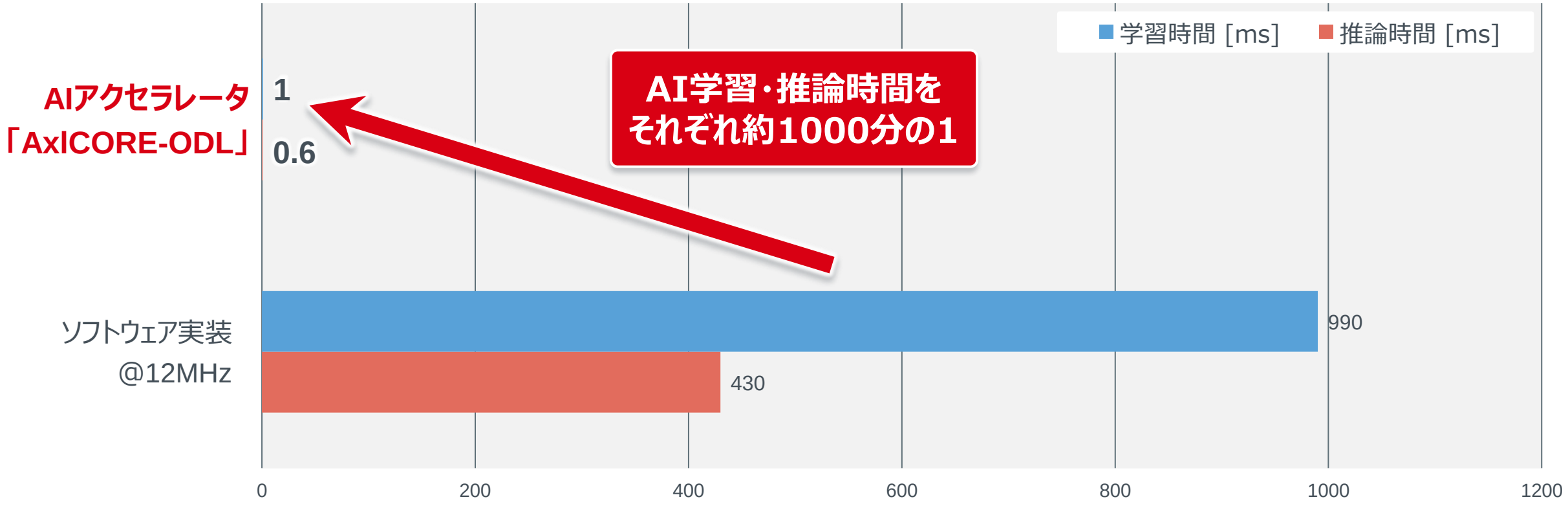
✓ オンデバイス学習アルゴリズムにより、チップ上で**3層ニューラルネットワークの学習**が可能

✓ 教師なし学習が可能なオートエンコーダで、**事前学習なし**で異常検知可能

詳細

ロームの技術:
AIアクセラレータ「AxICORE-ODL」による、小型化と高速化

学習・推論の実行時間比較(ニューラルネットワーク: 入力層96ノード、中間層12ノードを設定)



- ✓ **CPU負荷が少ない**。低コストな8-bit CPUでも、十分なアプリケーション処理能力を確保
- ✓ **高速サンプリング対応**。10kHz程度の高周波域に現れる異常の検出に対応
- ✓ 時系列データの前処理として必要な、**FFTもAIアクセラレータ(回路に内蔵)で実現可能**

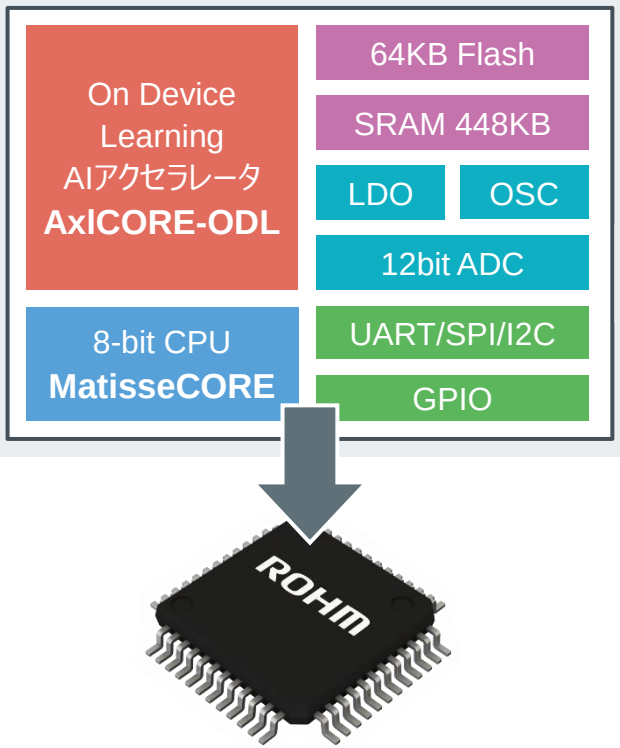
AIチップ「BD15035」の評価ボード

試作オンデバイスAIチップ「BD15035」

- AIアクセラレータ「AxlCORE-ODL」搭載
- 高効率8-bit CPU「tinyMicon MatisseCORE™」搭載

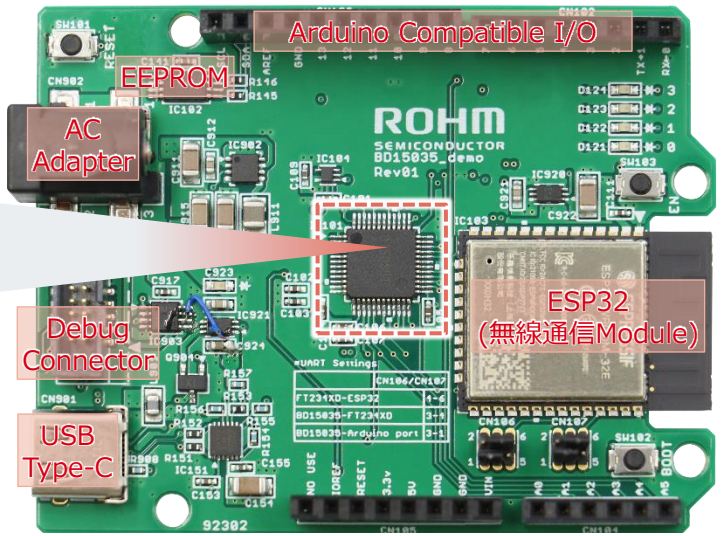
デモ動画は コチラ

BD15035



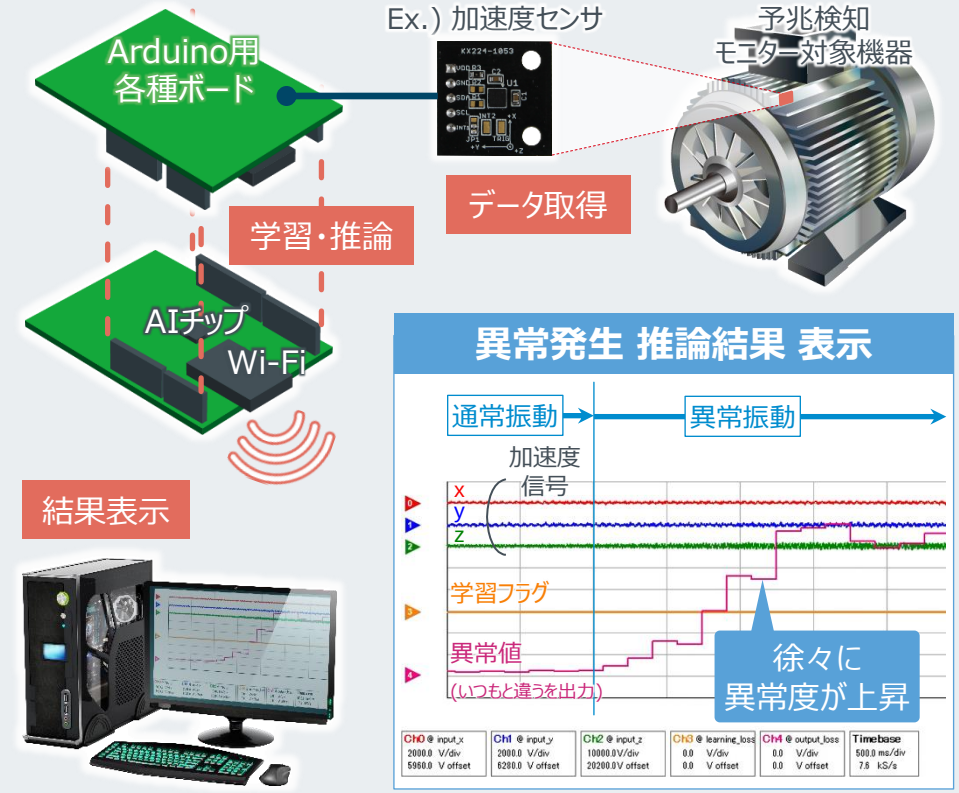
TQFP48V Package
(9.0 mm × 9.0 mm × 1.2 mm)

評価ボード



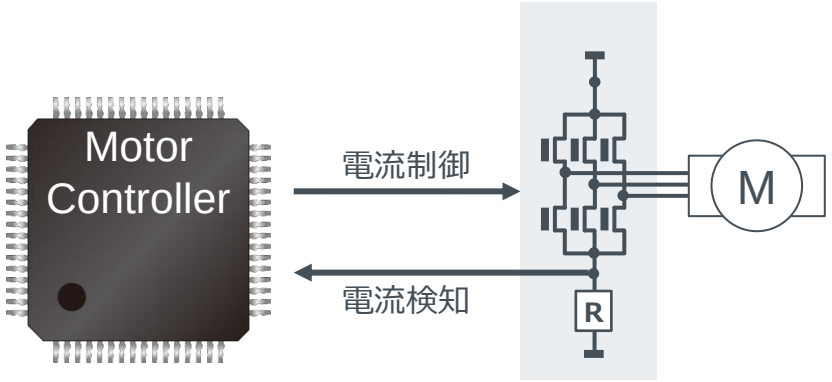
- Arduino用ボード 接続可能
- Wi-Fi/Bluetooth®モジュール搭載
- 64kbit EEPROM搭載

評価ボードの使用イメージ
(加速度センサを使った場合)

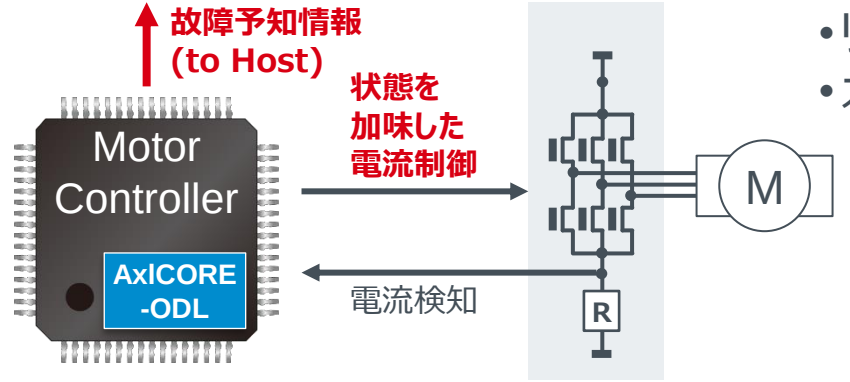


想定するユースケース1:
オンデバイス学習AI機能付きモータコントローラ

既存のモータ制御環境



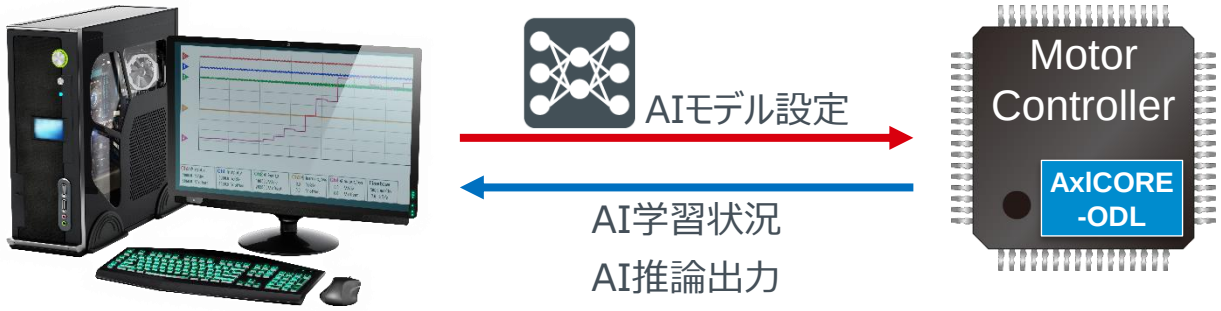
AI機能付きモータ制御環境



- リアルタイム高速推論
- オンデバイス学習

**従来のモータコントローラICに対して、
低コストかつ簡単(追加部品無し)にAI機能を導入可能**

AIモデルの構築から評価まで、簡単に実現するツールも開発中



複雑なモデルの設計や多数のパラメータの調整が不要

- AIの出力をモニタしながら、AIモデルを簡単に構築
- 最小限のパラメータでAIモデルを調整
  (入力データ数、異常判定のしきい値)
- ボタン一つで、デバイス上で再学習

想定するユースケース2:
オンデバイス学習AI機能付きエッジ向け汎用マイコン

汎用マイコン ← **エンドポイントAIをアドオン**

- オンデバイス学習AIアクセラレータを
  ペリフェラルとして持つ汎用マイコン
- 高速なAI演算を小さなハードウェアで実行、
  アプリケーション機能はソフトウェアで自由に実装可能

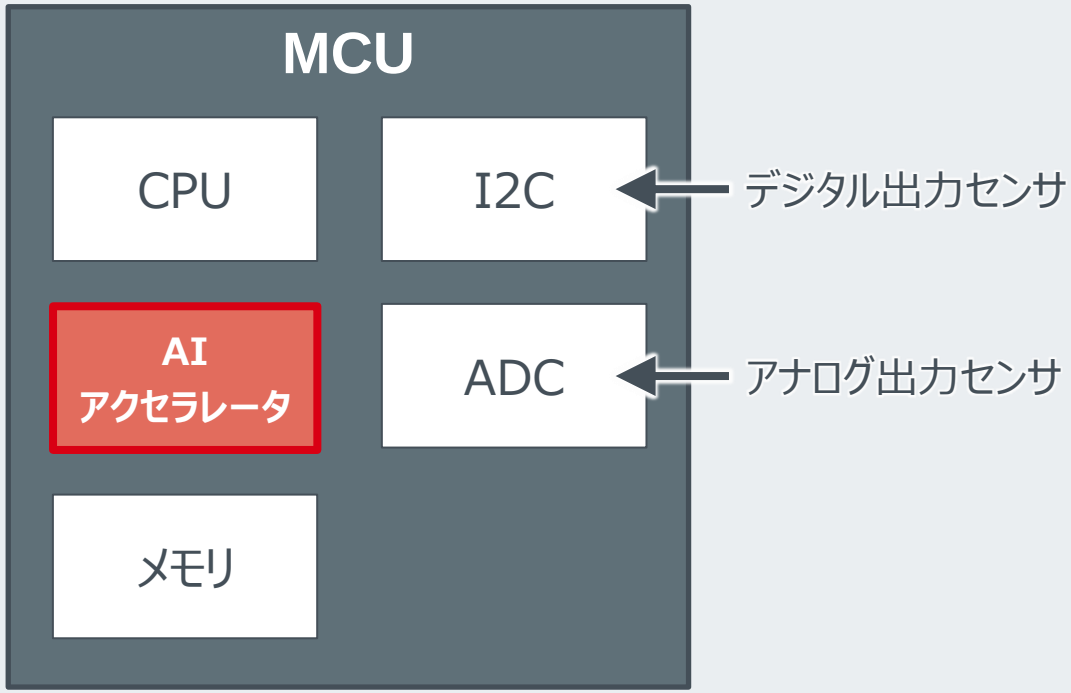
**産業機器、車載機器、家電などの
エッジ/エンドポイントMCUに
簡単にAI機能(故障予兆検知など)を追加**

メリット

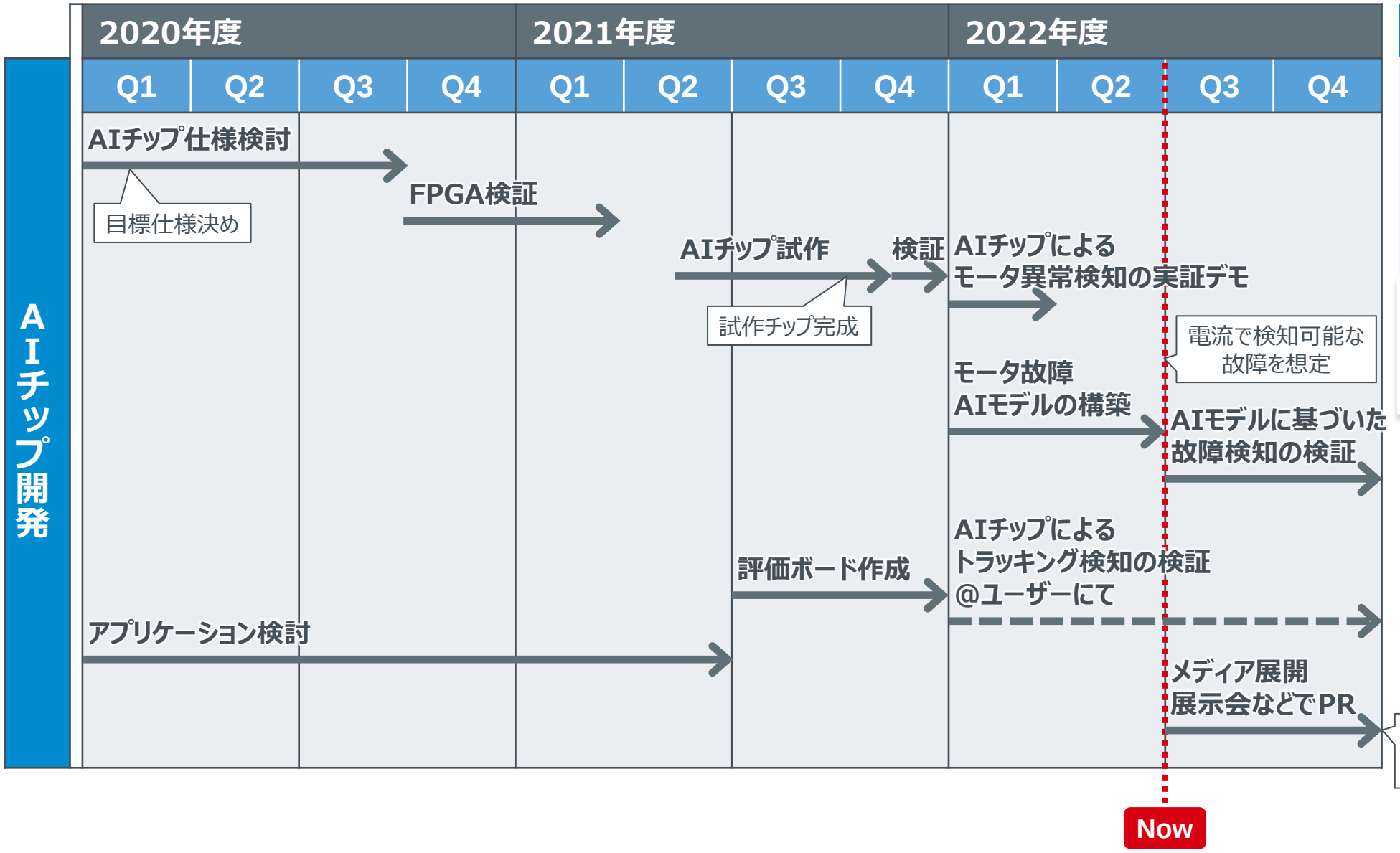
- 必要なAI演算はハードウェアで演算できるため
  ソフトウェアの負荷が少なくアプリケーション機能に制限がない
- 今あるアプリケーションマイコンの置き換えで
  AI機能を追加できる
- 推論だけでなく学習もデバイス側でできるため、
  設置場所ごとの最適化が簡単にできる

様々なタイプのセンサ入力を同時に処理可能

加速度、電流、温度、照度、マイク



試作チップ開発から、今後製品化までのスケジュール



今後の予定

2023年度
ローム量産チップ開発着手

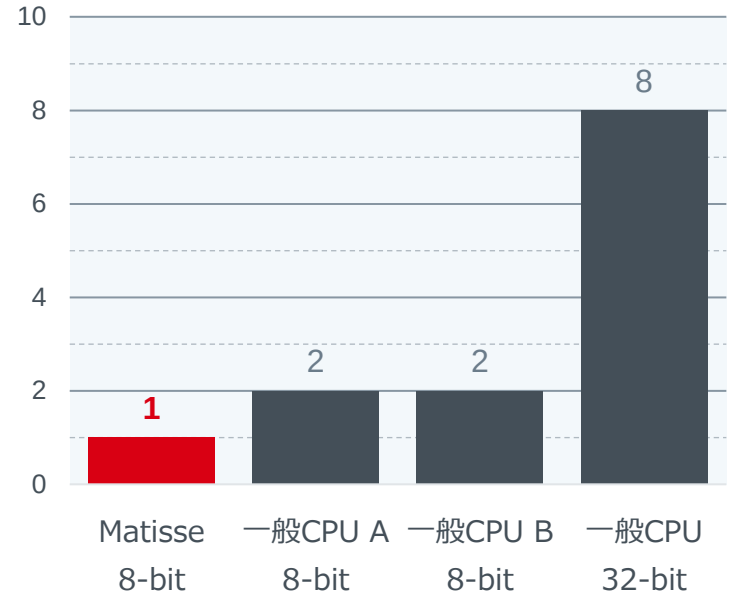
2024年度
量産開始

**モータ・センサ向け
IC製品への搭載により、
故障予知/予知保全に貢献**

ご参考：tinyMicon MatisseCORE™
～ 小型で高速の8-bit CPU ～

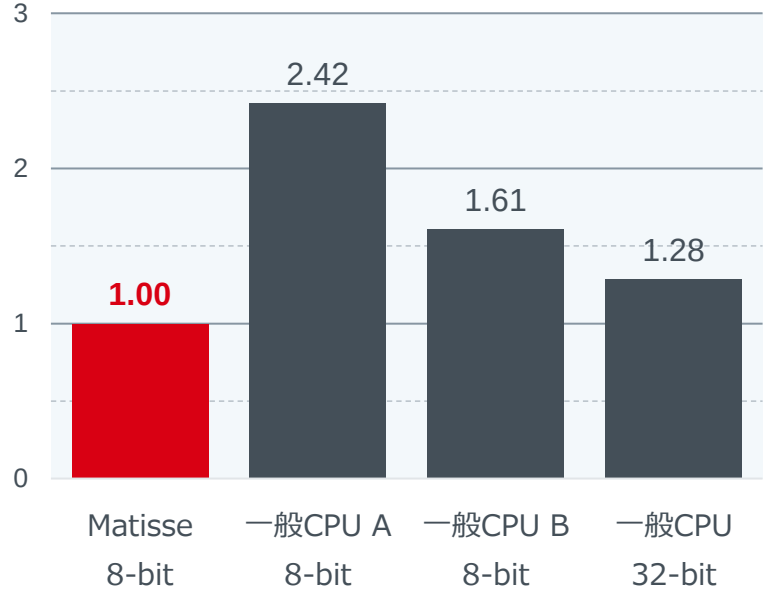
一般的な小型CPUとの性能比較(Matisseを1とした場合)

ゲートサイズ比較



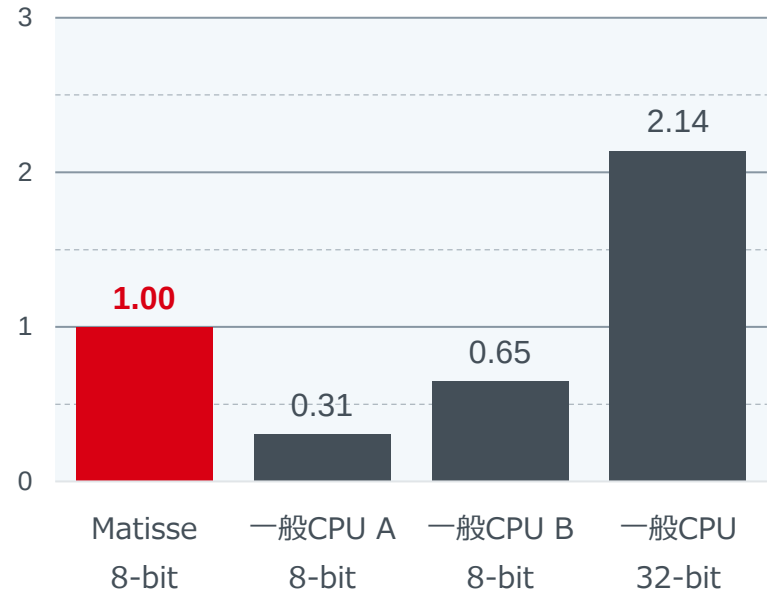
優れた省面積性

ROMサイズ比較 (8-bit演算プログラム時)



コンパクトな プログラムコードサイズ

処理性能比較 (Dhrystone)



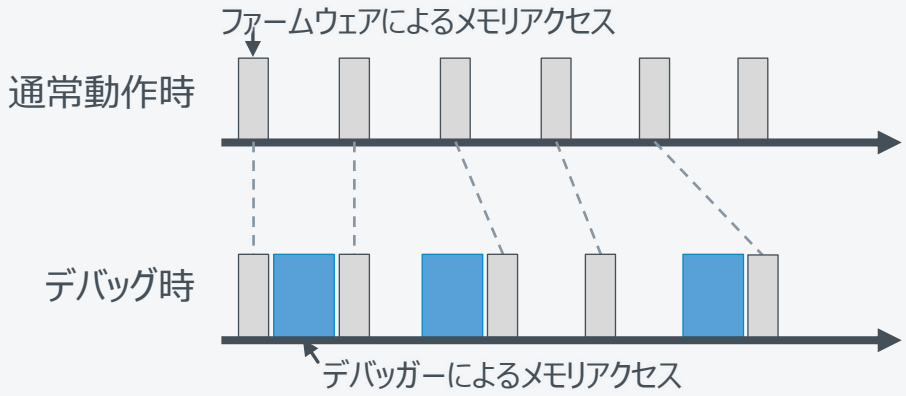
高速な演算処理

Matisseは、コンパクトなチップ面積とプログラムコードサイズ、高速な演算処理を高い水準で実現（さらに車載ASIL-Dまで対応も可能）

ご参考：tinyMicon MatisseCORE™

～ 組み込みユースに最適なリアルタイムデバッグ機能 ～

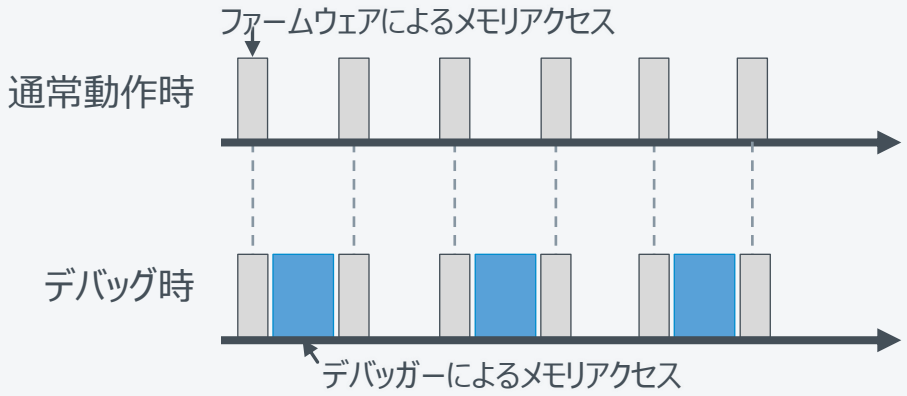
従来手法



デバッガーによるメモリアクセスがファームウェアの動作を妨げる

デバッグによりプログラムの
動作が大きく変わる場合がある

Matisseのリアルタイムデバッグ機能



ファームウェアの動作を妨げないようにデバッガーのメモリアクセスを調整

プログラムの動作を一切変えずに
デバッグを実行可能

Matisseは、完全ロスレスでCPU内部情報の取り出し/変更が可能

- リアルタイムデバッグにより、**通常動作時とデバッグ時の動作が変わらない**
- モータ制御など、**動きを止めてデバッグすることができないアプリケーションでも簡単にデバッグが可能**
- 通常は見ることができない**IC内部の変数をリアルタイムに取り出して波形表示できる**

波形表示用ソフトウェア
RapidScope™

Electronics for the Future

ROHM Co.,Ltd.

ご注意事項

- 本資料に記載されている内容はロームの製品（以下「ローム製品」といいます）のご紹介を目的としています。
- ローム製品のご使用にあたりましては、別途最新の仕様書およびデータシートを必ずご確認ください。
- 本資料に記載されております情報は、何ら保証なく提供されるものです。万が一、当該情報の誤りまたは使用に起因する損害がお客様または第三者に生じた場合においても、ロームは一切の責任を負うものではありません。
- 本資料に記載されておりますローム製品に関する代表的動作および応用回路例は、一例を示したものであり、これらに関する第三者の知的財産権およびその他の権利について権利侵害がないことを保証するものではありません。
- 上記技術情報の使用に起因して紛争が発生した場合、ロームはその責任を負うものではありません。
- ロームは、ロームまたは他社の知的財産権その他のあらゆる権利について明示的にも黙示的にも、その実施または利用を許諾するものではありません。
- 本資料に記載されております製品および技術のうち、「外国為替及び外国貿易法」その他の輸出規制に該当する製品または技術を輸出する場合、または国外に提供する場合には、同法に基づく許可が必要です。
- 本資料の記載内容は 2022年9月 現在のものであり、予告なく変更することがあります。