

「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」

研究開発項目① 革新的AIエッジコンピューティング技術の開発

研究開発項目② 次世代コンピューティング技術の開発

研究開発項目③ 高度なIoT社会を実現する横断的技術開発

事業原簿

担当部	国立研究開発法人 新エネルギー・産業技術総合開発機構 IoT 推進部
-----	--

【第1部】

研究開発項目① 革新的AIエッジコンピューティング技術の開発
(事業期間:2018~2022 年度)

—目次—

概要	(1-10)
プロジェクト用語	(1-13)
1. 事業の位置付け・必要性について	1-13
1.1 事業実施の背景	1-1-1
1.2 政策的位置づけ	1-1-2
1.3 アウトカム効果	1-1-3
1.4 国際的なポジション	1-1-3
1.5 海外状況のまとめ	1-1-4
1.6 NEDO が関与する意義	1-1-5
1.7 今回の事業の位置づけ	1-1-5
2. 研究開発マネジメントについて	1-2-1
2.1 事業の目的	1-2-1
2.2 研究開発目標と根拠	1-2-1
2.3 研究開発スケジュール	1-2-4
2.4 プロジェクト費用	1-2-5
2.5 マネジメント体制	1-2-5
2.6 実施体制	1-2-7
2.7 動向・情勢の把握と対応	1-2-712
3. 研究開発成果について	1-2-12
3.1 研究開発テーマ「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」	1-3-1
3.1.1 開発全体概要	1-3-1
3.1.2 研究開発項目の開発内容、最終目標、根拠、達成度	1-3-1
3.1.3 目標の達成度	1-3-2
3.1.4 成果と意義	1-3-3
3.1.5 成果の普及	1-3-4
3.2 研究開発テーマ「動的再構成技術を活用した組み込み AI システムの研究開発」	1-3-5
3.2.1 研究開発サブテーマ「DRP アーキテクチャ、コンパイラの研究開発」	1-3-5
3.2.1.1 概要	1-3-5
3.2.1.2 最終目標と根拠	1-3-5
3.2.1.3 目標の達成度	1-3-6
3.2.1.4 成果と意義	1-3-6
3.2.1.5 成果の普及	1-3-6
3.2.2 研究開発サブテーマ「DNN 単位演算回路を加速処理する先進的なアーキテクチャ・回路技術の研究開発」	1-3-7
3.2.2.1 概要	1-3-7
3.2.2.2 最終目標と根拠	1-3-7
3.2.2.3 目標の達成度	1-3-7
3.2.2.4 成果と意義	1-3-8
3.2.2.5 成果の普及	1-3-8
3.2.3 研究開発サブテーマ「軽量 DNN 変換機能の研究開発」	1-3-8
3.2.3.1 概要	1-3-8
3.2.3.2 最終目標と根拠	1-3-9
3.2.3.3 目標の達成度	1-3-10
3.2.3.4 成果と意義	1-3-10
3.2.4 研究開発サブテーマ「DNN によるエンドポイント学習用ソフトウェアの研究開発」	1-3-11
3.2.4.1 概要	1-3-11
3.2.4.2 最終目標と根拠	1-3-12
3.2.4.3 目標の達成度	1-3-12
3.2.4.4 成果と意義	1-3-13
3.2.4.5 成果の普及	1-3-13
3.2.5 研究開発サブテーマ「DNN によるエンドポイント学習用ハードウェアの研究開発」	1-3-14
3.2.5.1 概要	1-3-14
3.2.5.2 最終目標と根拠	1-3-14
3.2.5.3 目標の達成度	1-3-14

3.2.5.4	成果と意義	1-3-15
3.2.5.5	成果の普及	1-3-16
3.2.6	研究開発サブテーマ「競合学習機構による汎用・超軽量エンドポイント学習技術の開発」	1-3-17
3.2.6.1	概要	1-3-17
3.2.6.2	最終目標と根拠	1-3-17
3.2.6.3	目標の達成度	1-3-17
3.2.6.4	成果と意義	1-3-18
3.2.6.5	成果の普及	1-3-18
3.2.7	研究開発サブテーマ「実製品への AI 組込みを容易化するツールの研究開発」	1-3-19
3.2.7.1	概要	1-3-19
3.2.7.2	最終目標と根拠	1-3-19
3.2.7.3	目標の達成度	1-3-19
3.2.7.4	成果と意義	1-3-21
3.2.7.5	成果の普及	1-3-22
3.3	研究開発テーマ「FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発」	1-3-23
3.3.1	概要	1-3-23
3.3.2	最終目標と根拠	1-3-24
3.3.3	目標の達成度	1-3-25
3.3.4	研究開発サブテーマ「再構成可能低遅延低消費電力 AI コンピューティングアーキテクチャの研究開発」	1-3-26
3.3.4.1	概要	1-3-26
3.3.5	研究開発サブテーマ「再構成可能低遅延低消費電力 AI プロセッサチップの研究開発」	1-3-29
3.3.5.1	概要	1-3-29
3.3.5.2	最終目標と根拠	1-3-30
3.3.5.3	目標の達成度	1-3-30
3.3.5.4	成果と意義	1-3-30
3.3.5.5	成果の普及	1-3-30
3.3.6	研究開発サブテーマ「ソフトウェアフレームワークの開発」	1-3-31
3.3.6.1	概要	1-3-31
3.3.6.2	最終目標と根拠	1-3-31
3.3.6.3	目標の達成度	1-3-31
3.3.6.4	成果と意義	1-3-31
3.3.6.5	成果の普及	1-3-31
3.3.7	研究開発サブテーマ「ロボティクス応用のためのソフトウェア開発」	1-3-32
3.3.7.1	概要	1-3-32
3.3.7.2	最終目標と根拠	1-3-32
3.3.7.3	目標の達成度	1-3-32
3.3.7.4	成果と意義	1-3-32
3.3.7.5	成果の普及	1-3-33
3.4	研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」	1-3-34
3.4.1	研究開発サブテーマ「画像・信号処理 AI 基盤技術（仮想エンジンアーキテクチャ）開発」	1-3-34
3.4.1.1	概要	1-3-34
3.4.1.2	最終目標と根拠	1-3-35
3.4.1.3	目標の達成度	1-3-35
3.4.1.4	成果と意義	1-3-35
3.4.1.5	成果の普及	1-3-35
3.4.2	研究開発サブテーマ「アルゴリズム記述ツール開発」	1-3-35
3.4.2.1	概要	1-3-35
3.4.2.2	最終目標と根拠	1-3-35
3.4.2.3	目標の達成度	1-3-36
3.4.2.4	成果と意義	1-3-36
3.4.2.5	成果の普及	1-3-36
3.4.3	研究開発サブテーマ「Computer Vision/AI 基本ミドルウェア開発」	1-3-36
3.4.3.1	概要	1-3-36
3.4.3.2	最終目標と根拠	1-3-36
3.4.3.3	目標の達成度	1-3-36
3.4.3.4	成果と意義	1-3-36
3.4.3.5	成果の普及	1-3-37
3.4.4	研究開発サブテーマ「リアルタイム SLAM 技術開発（SLAM ライブラリ）」	1-3-37

3.4.4.1	概要	1-3-37
3.4.4.2	最終目標と根拠	1-3-37
3.4.4.3	目標の達成度	1-3-37
3.4.4.4	成果と意義	1-3-37
3.4.4.5	成果の普及	1-3-37
3.4.5	研究開発サブテーマ「AI エッジ LSI 搭載車載 ECU 試作・評価」	1-3-37
3.4.5.1	概要	1-3-37
3.4.5.2	最終目標と根拠	1-3-37
3.4.5.3	目標の達成度	1-3-37
3.4.5.4	成果と意義	1-3-38
3.4.5.5	成果の普及	1-3-38
3.4.6	研究開発サブテーマ「量子化 DNN 技術開発」	1-3-38
3.4.6.1	概要	1-3-38
3.4.6.2	最終目標と根拠	1-3-38
3.4.6.3	目標の達成度	1-3-38
3.4.6.4	成果と意義	1-3-38
3.4.6.5	成果の普及	1-3-38
3.4.7	研究開発サブテーマ「DNN 変換ツール開発」	1-3-39
3.4.7.1	概要	1-3-39
3.4.7.2	最終目標と根拠	1-3-39
3.4.7.3	目標の達成度	1-3-39
3.4.7.4	成果と意義	1-3-39
3.4.7.5	成果の普及	1-3-39
3.4.8	研究開発サブテーマ「エッジ環境最適化技術開発」	1-3-39
3.4.8.1	概要	1-3-39
3.4.8.2	最終目標と根拠	1-3-39
3.4.8.3	目標の達成度	1-3-39
3.4.8.4	成果と意義	1-3-40
3.4.8.5	成果の普及	1-3-40
3.4.9	研究開発サブテーマ「AI エッジ LSI 試作開発」	1-3-40
3.4.9.1	概要	1-3-40
3.4.9.2	最終目標と根拠	1-3-40
3.4.9.3	目標の達成度	1-3-40
3.4.9.4	成果と意義	1-3-40
3.4.9.5	成果の普及	1-3-40
3.5	研究開発テーマ「ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発」	1-3-41
3.5.1	概要	1-3-41
3.5.2	最終目標と根拠	1-3-43
3.5.3	成果と意義	1-3-43
3.5.4	成果と意義	1-3-43
3.5.5	成果の普及	1-3-43
3.6	研究開発テーマ「5G 時代を見据えた高度自律的学習機能搭載のための AI エッジコンピューティング技術の研究開発」	1-3-45
3.6.1	研究開発サブテーマ「ネットワーク圧縮と BaseNetwork 共通化による演算量削減技術の開発」	1-3-46
3.6.1.1	概要	1-3-46
3.6.1.2	最終目標と根拠	1-3-47
3.6.1.3	目標の達成度	1-3-47
3.6.1.4	成果と意義	1-3-48
3.6.1.5	成果の普及	1-3-48
3.6.2	研究開発サブテーマ「人工意識による高度自律的学習機能の開発の研究開発」	1-3-49
3.6.2.1	概要	1-3-49
3.6.2.2	最終目標と根拠	1-3-49
3.6.2.3	目標の達成度	1-3-49
3.6.2.4	成果と意義	1-3-52
3.6.2.5	成果の普及	1-3-53
3.6.3	研究開発サブテーマ「AI エッジ統合制御システムの開発」	1-3-54
3.6.3.1	概要	1-3-54
3.6.3.2	最終目標と根拠	1-3-54
3.6.3.3	目標の達成度	1-3-54
3.6.3.4	成果と意義	1-3-55
3.6.3.5	成果の普及	1-3-56

3.7	研究開発テーマ「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」	1-3-57
3.7.1	概要	1-3-57
3.7.2	最終目標と根拠	1-3-59
3.7.3	目標の達成度	1-3-63
3.7.4	成果と意義	1-3-66
3.7.5	成果の普及	1-3-67
3.8	研究開発テーマ「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」	1-3-68
3.8.1	研究開発サブテーマ①「分散マイクロカーネル OS の研究開発（分散 MK OS）」、及び、サブテーマ②「分散 MK OS 内蔵 SOC の研究開発」	1-3-69
3.8.1.1	概要	1-3-69
3.8.1.2	最終目標と根拠	1-3-69
3.8.1.3	目標の達成度	1-3-69
3.8.1.4	成果と意義	1-3-70
3.8.1.5	成果の普及	1-3-70
3.8.2	研究開発サブテーマ「Hybrid-scheduling/Load-balancing アルゴリズムの研究開発」	1-3-70
3.8.2.1	概要	1-3-70
3.8.2.2	最終目標と根拠	1-3-70
3.8.2.3	目標の達成度	1-3-70
3.8.2.4	成果と意義	1-3-71
3.8.2.5	成果の普及	1-3-72
3.8.3	研究開発サブテーマ「ソフトウェアマッピング支援ツールの研究開発」	1-3-72
3.8.3.1	概要	1-3-72
3.8.3.2	最終目標と根拠	1-3-72
3.8.3.3	目標の達成度	1-3-72
3.8.3.4	成果と意義	1-3-73
3.8.3.5	成果の普及	1-3-73
3.8.4	研究開発サブテーマ「ヘテロジニアス・マルチ・メニーコア向けハードウェア記述標準モデル SHIM の研究開発と国際規格化」	1-3-74
3.8.4.1	概要	1-3-74
3.8.4.2	最終目標と根拠	1-3-74
3.8.4.3	目標の達成度	1-3-74
3.8.4.4	成果と意義	1-3-74
3.9	研究開発テーマ「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」	1-3-75
3.9.1	研究開発サブテーマ「多分岐結合型推論プロセッサの研究開発」	1-3-75
3.9.1.1	概要	1-3-75
3.9.1.2	最終目標と根拠	1-3-75
3.9.1.3	目標の達成度	1-3-75
3.9.1.4	成果と意義	1-3-75
3.9.1.5	成果の普及	1-3-75
3.9.2	研究開発サブテーマ「多様な AI プロセッシングアーキテクチャの評価と性能比較」	1-3-76
3.9.2.1	概要	1-3-76
3.9.2.2	成果と意義	1-3-76
3.9.3	研究開発サブテーマ「ヘテロジニアス向けコンパイラの研究開発」	1-3-76
3.9.3.1	概要	1-3-76
3.9.3.2	最終目標と根拠	1-3-76
3.9.3.3	目標の達成度	1-3-76
3.9.3.4	成果の普及	1-3-77
3.9.4	研究開発サブテーマ「OS の抽象化とミドルウェア最適化技術ならびに統合開発環境の研究開発」	1-3-77
3.9.4.1	概要	1-3-77
3.9.4.2	最終目標と根拠	1-3-77
3.9.4.3	目標の達成度	1-3-77
3.9.4.4	成果と意義	1-3-77
3.9.4.5	成果の普及	1-3-77
3.9.5	研究開発サブテーマ「リアルタイム空間理解、判断アルゴリズムの研究開発」	1-3-78
3.9.5.1	概要	1-3-78
3.9.5.2	最終目標と根拠	1-3-78
3.9.5.3	目標の達成度	1-3-78
3.9.5.4	成果と意義	1-3-78
3.9.5.5	成果の普及	1-3-78
3.9.6	研究開発サブテーマ「多分岐結合型推論プロセッサ性能評価用ベンチマークの研究開発」	1-3-79
3.9.6.1	概要	1-3-79

3.9.6.2	最終目標と根拠	1-3-79
3.9.6.3	目標の達成度	1-3-79
3.9.6.4	成果と意義	1-3-79
3.9.6.5	成果の普及	1-3-79
3.10	研究開発テーマ「完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発」	1-3-80
3.10.1	研究開発サブテーマ「完全自動運転に向けた SoC の研究開発」	1-3-80
3.10.1.1	概要	1-3-80
3.10.1.2	最終目標と根拠	1-3-80
3.10.1.3	目標の達成度	1-3-81
3.10.1.4	成果と意義	1-3-81
3.10.1.5	成果の普及	1-3-81
3.10.2	研究開発サブテーマ「全自動運転に向けたコンパイラ・OS の研究開発」	1-3-82
3.10.2.1	概要	1-3-82
3.10.2.2	最終目標と根拠	1-3-82
3.10.2.3	目標の達成度	1-3-82
3.10.2.4	成果と意義	1-3-82
3.10.2.5	成果の普及	1-3-82
3.10.3	研究開発サブテーマ「完全自動運転に向けたミドルウェアの研究開発」	1-3-83
3.10.3.1	概要	1-3-83
3.10.3.2	最終目標と根拠	1-3-83
3.10.3.3	目標の達成度	1-3-83
3.10.3.4	成果と意義	1-3-83
3.10.3.5	成果の普及	1-3-84
3.10.3.6	その他	1-3-84
3.10.4	研究開発サブテーマ「完全自動運転に向けたアプリケーションの研究開発」	1-3-84
3.10.4.1	概要	1-3-84
3.10.4.2	最終目標と根拠	1-3-84
3.10.4.3	目標の達成度	1-3-84
3.10.4.4	成果と意義	1-3-85
3.10.4.5	成果の普及	1-3-85
3.10.4.6	その他	1-3-85
3.11	研究開発テーマ「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」の研究開発成果	1-3-86
3.11.1	概要	1-3-86
3.11.2	最終目標と根拠	1-3-86
3.11.3	目標の達成度	1-3-88
3.11.4	成果と意義	1-3-89
3.11.5	成果の普及	1-3-91
3.12	研究開発テーマ「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」	1-3-92
3.12.1	研究開発サブテーマ「実施項目 0 : AI エッジで必要となるセキュリティ評価分析基盤の研究開発」	1-3-92
3.12.1.1	概要	1-3-92
3.12.1.2	最終目標と根拠	1-3-92
3.12.1.3	目標の達成度	1-3-92
3.12.1.4	成果と意義	1-3-93
3.12.1.5	成果の普及	1-3-94
3.12.2	研究開発サブテーマ「実施項目 1 : AI エッジ入出力セキュリティ評価シミュレータの開発」	1-3-95
3.12.2.1	概要	1-3-95
3.12.2.2	最終目標と根拠	1-3-96
3.12.2.3	目標の達成度	1-3-96
3.12.2.4	成果と意義	1-3-98
3.12.2.5	成果の普及	1-3-98
3.12.3	研究開発サブテーマ「実施項目 2 : AI エッジ内部実装保護技術の研究開発」	1-3-99
3.12.3.1	概要	1-3-99
3.12.3.2	最終目標と根拠	1-3-100
3.12.3.3	目標の達成度	1-3-101
3.12.3.4	成果と意義	1-3-102
3.12.3.5	成果の普及	1-3-104
3.12.3.6	事業化までのマイルストーン	1-3-105

3.12.4	研究開発サブテーマ「実施項目3: AI エッジの個体管理を支えるための人工物メトリクスの研究開発」	1-3-105
3.12.4.1	概要	1-3-105
3.12.4.2	最終目標と根拠	1-3-106
3.12.4.3	目標の達成度	1-3-106
3.12.4.4	成果と意義	1-3-106
3.12.4.5	成果の普及	1-3-106

4. 実用化・事業化に向けての見通し及び取り組みについて 1-4-1

4.1	「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」の研究開発における実用化・事業化の見通し（日本電気株式会社）	1-4-1
4.1.1	概要	1-4-1
4.1.2	実用化・事業化への課題と対応策	1-4-1
4.1.3	実用化・事業化の体制	1-4-1
4.1.4	市場規模と経済効果	1-4-2
4.1.5	ベンチマーク	1-4-2
4.1.6	事業化までのマイルストーン	1-4-3
4.2	動的再構成技術を活用した組込み AI システムの研究開発における実用化・事業化の見通し	1-4-4
4.2.1	動的再構成技術を活用した組込み AI システムの研究開発における実用化・事業化の見通し（ルネサスエレクトロニクス株式会社）	1-4-4
4.2.1.1	概要	1-4-4
4.2.1.2	実用化・事業化への課題と対応策	1-4-4
4.2.1.3	実用化・事業化の体制	1-4-4
4.2.1.4	市場規模と経済効果	1-4-4
4.2.1.5	ベンチマーク	1-4-5
4.2.1.6	事業化までのマイルストーン	1-4-5
4.2.2	競合学習機構による汎用・超軽量エンドポイント学習技術の開発における実用化・事業化の見通し（SOINN 社）	1-4-6
4.2.2.1	概要	1-4-6
4.2.2.2	実用化・事業化への課題と対応策	1-4-6
4.2.2.3	実用化・事業化の体制	1-4-7
4.2.2.4	市場規模と経済効果	1-4-7
4.2.2.5	ベンチマーク	1-4-7
4.2.3	事業化までのマイルストーン	1-4-7
4.3	FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発における実用化・事業化の見通し	1-4-8
4.3.1	概要	1-4-8
4.3.2	実用化・事業化への課題と対応策	1-4-8
4.3.3	実用化・事業化の体制	1-4-8
4.3.4	市場規模と経済効果	1-4-8
4.3.5	ベンチマーク	1-4-8
4.3.6	事業化までのマイルストーン	1-4-8
4.4	研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」の研究開発における実用化・事業化の見通し	1-4-9
4.4.1	低進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し（ArchiTek 株式会社、株式会社ソシオネクスト）	1-4-9
4.4.1.1	概要	1-4-9
4.4.1.2	実用化・事業化への課題と対応策	1-4-9
4.4.1.3	市場規模と経済効果	1-4-9
4.4.1.4	ベンチマーク	1-4-9
4.4.1.5	事業化までのマイルストーン	1-4-9
4.4.2	進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し（株式会社ソシオネクスト）	1-4-10
4.4.2.1	概要	1-4-10
4.4.2.2	実用化・事業化への課題と対応策	1-4-10
4.4.2.3	実用化・事業化の体制	1-4-10
4.4.2.4	市場規模と経済効果	1-4-10
4.4.2.5	ベンチマーク	1-4-10
4.4.2.6	事業化までのマイルストーン	1-4-10
4.4.3	進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し（株式会社豊田自動織機）	

.....	1-4-11
4.4.3.1 概要	1-4-11
4.4.3.2 実用化・事業化への課題と対応策	1-4-11
4.4.3.3 市場規模と経済効果	1-4-11
4.4.3.4 ベンチマーク	1-4-11
4.4.3.5 事業化までのマイルストーン	1-4-11
4.5 ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発の研究開発における実用化・事業化の見通し（沖電気工業株式会社、総合警備保障株式会社、ジャパンマリンユナイテッド株式会社）	1-4-12
4.5.1 概要	1-4-12
4.5.2 実用化・事業化への課題と対応策	1-4-12
4.5.3 実用化・事業化の体制	1-4-12
4.6 「ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発」における実用化・事業化の見通し（株式会社アラヤ、KDDI 株式会社）	1-4-13
4.6.1 「ネットワーク圧縮と BaseNetwork 共通化による演算量削減技術の開発」における実用化・事業化の見通し（株式会社アラヤ）	1-4-13
4.6.1.1 概要	1-4-13
4.6.1.2 実用化・事業化への課題と対応策	1-4-13
4.6.1.3 実用化・事業化の体制	1-4-13
4.6.1.4 市場規模と経済効果	1-4-14
4.6.1.5 ベンチマーク	1-4-14
4.6.1.6 事業化までのマイルストーン	1-4-15
4.6.1 「人工意識による高度自律的学習機能の開発の研究開発」における実用化・事業化の見通し（株式会社アラヤ/国立大学法人 東京大学）	1-4-15
4.6.1.1 概要	1-4-15
4.6.1.2 実用化・事業化への課題と対応策	1-4-16
4.6.1.3 実用化・事業化の体制	1-4-16
4.6.1.4 市場規模と経済効果	1-4-16
4.6.1.5 ベンチマーク	1-4-17
4.6.1.6 事業化までのマイルストーン	1-4-18
4.6.2 「AI エッジ統合制御システムの開発」における実用化・事業化の見通し（KDDI 株式会社）	1-4-19
4.6.2.1 概要	1-4-19
4.6.2.2 実用化・事業化への課題と対応策	1-4-19
4.6.2.3 実用化・事業化の体制	1-4-20
4.6.2.4 市場規模と経済効果	1-4-22
4.6.2.5 ベンチマーク	1-4-22
4.6.2.6 事業化までのマイルストーン	1-4-23
4.7 「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」における実用化・事業化の見通し	1-4-24
4.7.1 概要	1-4-24
4.7.2 実用化・事業化への課題と対応策	1-4-26
4.7.3 実用化・事業化の体制	1-4-28
4.7.4 市場規模と経済効果	1-4-28
4.7.5 ベンチマーク	1-4-28
4.7.6 事業化までのマイルストーン	1-4-28
4.8 「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」における実用化・事業化の見通し（イーソル株式会社）	1-4-29
4.8.1 概要	1-4-29
4.8.2 実用化・事業化への課題と対応策	1-4-29
4.8.3 実用化・事業化の体制	1-4-29
4.8.4 市場規模と経済効果	1-4-29
4.8.5 ベンチマーク	1-4-29
4.8.6 事業化までのマイルストーン	1-4-29
4.9 「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」における実用化・事業化の見通し	1-4-30
4.9.1 多分岐結合型推論プロセッサの研究開発における実用化・事業化の見通し（株式会社エヌエスアイテクス）	1-4-30
.....	1-4-30
4.9.1.1 概要	1-4-30
4.9.1.2 実用化・事業化への課題と対応策	1-4-30
4.9.1.3 実用化・事業化の体制	1-4-30
4.9.1.4 市場規模と経済効果	1-4-30

4.9.1.5	ベンチマーク	1-4-30
4.9.1.6	事業化までのマイルストーン	1-4-30
4.9.2	OS の抽象化とミドルウェア最適化技術ならびに統合開発環境の研究開発における実用化・事業化の見通し (株式会社ユーリカ)	1-4-31
4.9.2.1	概要	1-4-31
4.9.2.2	実用化・事業化への課題と対応策	1-4-31
4.9.2.3	実用化・事業化の体制	1-4-31
4.9.2.4	市場規模と経済効果	1-4-31
4.9.2.5	ベンチマーク	1-4-31
4.9.2.6	事業化までのマイルストーン	1-4-31
4.9.3	リアルタイム空間理解、判断アルゴリズムの研究開発における実用化・事業化の見通し (株式会社日立製作所)	1-4-32
4.9.3.1	概要	1-4-32
4.9.3.2	実用化・事業化への課題と対応策	1-4-32
4.9.3.3	実用化・事業化の体制	1-4-32
4.9.3.4	市場規模と経済効果	1-4-32
4.9.3.5	ベンチマーク	1-4-32
4.9.3.6	事業化までのマイルストーン	1-4-32
4.10	「完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発」における実用化・事業化の見通し (株式会社アクセル、株式会社ティアフォー社)	1-4-33
4.10.1	概要	1-4-33
4.10.2	実用化・事業化への課題と対応策	1-4-33
4.10.3	実用化・事業化の体制	1-4-33
4.10.4	市場規模と経済効果	1-4-34
4.10.5	ベンチマーク	1-4-34
4.10.6	事業化までのマイルストーン(株式会社アクセル)	1-4-36
4.10.6	事業化までのマイルストーン(株式会社ティアフォー)	1-4-37
4.11	「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し	1-4-38
4.11.1	「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し(NSITEXE)	1-4-38
4.11.1.1	概要	1-4-38
4.11.1.2	実用化・事業化への課題と対応策	1-4-38
4.11.1.3	実用化・事業化の体制	1-4-38
4.11.1.4	市場規模と経済効果	1-4-38
4.11.1.5	ベンチマーク	1-4-38
4.11.1.6	事業化までのマイルストーン	1-4-38
4.11.2	「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し(セコム)	1-4-39
4.11.2.1	概要	1-4-39
4.11.2.2	実用化・事業化への課題と対応策	1-4-39
4.11.2.3	実用化・事業化の体制	1-4-39
4.11.2.4	市場規模と経済効果	1-4-39
4.11.2.5	ベンチマーク	1-4-39
4.11.2.6	事業化までのマイルストーン	1-4-39
4.11.3	「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し(日立製作所)	1-4-40
4.11.3.1	概要	1-4-40
4.11.3.2	実用化・事業化への課題と対応策	1-4-40
4.11.3.3	実用化・事業化の体制	1-4-40
4.11.3.4	市場規模と経済効果	1-4-40
4.11.3.5	ベンチマーク	1-4-40
4.11.3.6	事業化までのマイルストーン	1-4-40
4.12	「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」における実用化・事業化の見通し	1-4-41
4.12.1	「実施項目 0 : AI エッジで必要となるセキュリティ評価分析基盤の研究開発」における実用化の見通し(産業技術総合研究所)	1-4-41
4.12.1.1	概要	1-4-41
4.12.1.1	実用化・事業化への課題と対応策	1-4-41
4.12.1.2	実用化・事業化の体制	1-4-41

4.12.1.3	市場規模と経済効果	1-4-42
4.12.1.4	ベンチマーク	1-4-43
4.12.1.5	事業化までのマイルストーン	1-4-44
4.12.2	「実施項目1: AI エッジ出力セキュリティ評価シミュレータの開発」における実用化・事業化の見通し (電子商取引安全技術研究組合)	1-4-45
4.12.2.1	概要	1-4-45
4.12.2.2	実用化・事業化への課題と対応策	1-4-45
4.12.2.3	実用化・事業化の体制	1-4-45
4.12.2.4	市場規模と経済効果	1-4-45
4.12.2.5	ベンチマーク	1-4-45
4.12.2.6	事業化までのマイルストーン	1-4-46
4.12.3	「実施項目2: AI エッジ内部実装保護技術の研究開発」における実用化・事業化の見通し (株式会社コネクテックジャパン)	1-4-46
4.12.3.1	概要	1-4-46
4.12.3.2	実用化・事業化への課題と対応策	1-4-46
4.12.3.3	実用化・事業化の体制	1-4-47
4.12.3.4	市場規模と経済効果	1-4-47
4.12.3.5	ベンチマーク	1-4-47
4.12.3.6	事業化までのマイルストーン	1-4-47
4.12.4	「実施項目2: AI エッジ内部実装保護技術の研究開発」における実用化・事業化の見通し (産業技術総合研究所)	1-4-47
4.12.4.1	概要	1-4-47
4.12.4.2	実用化・事業化への課題と対応策	1-4-48
4.12.4.3	実用化・事業化の体制	1-4-49
4.12.4.4	市場規模と経済効果	1-4-49
4.12.4.5	ベンチマーク	1-4-51
4.12.4.6	事業化までのマイルストーン	1-4-51
4.12.5	「実施項目2: AI エッジ内部実装保護技術の研究開発」における実用化・事業化の見通し (IIJ イノベーションインスティテュート)	1-4-52
4.12.5.1	概要	1-4-52
4.12.5.2	実用化・事業化への課題と対応策	1-4-52
4.12.5.3	実用化・事業化の体制	1-4-52
4.12.5.4	市場規模と経済効果	1-4-53
4.12.5.5	ベンチマーク	1-4-53
4.12.5.6	事業化までのマイルストーン	1-4-53
4.12.6	「実施項目3: AI エッジの個体管理を支えるための人工物メトリクスの研究開発」における実用化・事業化 の見通し(産業技術総合研究所)	1-4-54
4.12.6.1	概要	1-4-54
4.12.6.2	実用化・事業化への課題と対応策	1-4-54
4.12.6.3	実用化・事業化の体制	1-4-54
4.12.6.4	市場規模と経済効果	1-4-54
4.12.6.5	ベンチマーク	1-4-55
4.12.6.6	事業化までのマイルストーン	1-4-55

●特許論文等リスト (添付資料) 1-添-1

◎研究開発テーマ「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」	1-添-1
◎研究開発テーマ「動的再構成技術を活用した組み込み AI システムの研究開発」	1-添-3
◎研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」	1-添-5
◎研究開発テーマ「ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発」	1-添-7
◎研究開発テーマ「5G 時代を見据えた高度自律的学習機能搭載のための AI エッジコンピューティング技術の研究開発」	1-添-12
◎研究開発テーマ「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」	1-添-13
◎研究開発テーマ「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」	1-添-18
◎研究開発テーマ「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」	1-添-19
◎研究開発テーマ「完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発」	1-添-23
◎研究開発テーマ「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」	1-添-26
◎研究開発テーマ「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」	1-添-32

概要

最終更新日

2020年11月19日

プロジェクト名	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		プロジェクト番号	P16007		
担当推進部/ PMまたは担当者	IoT推進部 PM:伊藤隆夫 (2020年7月現在) IoT推進部 PM:大杉伸也 (2018年7月~2020年4月) IoT推進部 PM:千田和也 (2016年4月~2018年6月)					
0. 事業の概要	来るべきポストムーア時代のIoT社会を築くため、大量データの効率的かつ高度な利活用を可能とする、基盤技術開発の開発が必要となる一方、IoT社会の到来が近づくにつれ、データ量の爆発的な増加とその処理に伴う消費電力の増加という、新たな社会課題にも直面している。これらの社会課題解決と日本の情報産業の再興を目的として、本プロジェクトでは、データ量削減のため、ネットワークの末端で中心的なAI処理を行う「AIエッジコンピューティング*技術」、消費電力を劇的に低減するため、これまでの延長線上にない新原理の技術開発を推進する「次世代コンピューティング技術」、それらを共通的に支えるための「共通基盤技術」の開発を実施している。					
1. 事業の位置 付け・必要性 について	本プロジェクトでは、社会課題の解決と我が国の情報産業の再興を目的とし、ポストムーア時代におけるコンピューティング技術開発を行う。本プロジェクトで取り組むポストムーア時代を見据えたコンピューティング技術開発は、Society 5.0の実現につながる Connected Industries を実現するために必要不可欠なものであるため、NEDOが主導して取り組む意義が極めて大きい。					
2. 研究開発マネジメントについて						
事業の目標	<p>【研究開発項目①】革新的AIエッジコンピューティング技術の開発 (期間:2018-2022年度) エッジにおけるAI処理を実現するための小型かつ省エネながら高度な処理の能力を持った専用チップ及びコンピューティング技術等を開発する。</p> <p>【研究開発項目②】次世代コンピューティング技術の開発 (期間:2018-2027年度) ・既存の技術の延長にない、新原理等による高速かつ低消費電力化を実現する次世代コンピューティング技術を開発する。</p> <p>【研究開発項目③】高度なIoT社会を実現する横断技術開発 (期間:2016-2020年度) ・IoT社会を支えるIoT情報基盤を築くため、大量データの効率的かつ高度な利活用を可能とする収集、蓄積、解析、セキュリティ等に関する横断的技術開発を行う。</p>					
事業の計画内容	主な実施事項	2016fy	2017fy	2018fy	2019fy	2020fy
	革新的AIエッジコンピューティング技術の開発			←		
	次世代コンピューティング技術の開発			←		
	高度なIoT社会を実現する横断技術開発	←				
事業費推移	会計・勘定	2016fy	2017fy	2018fy	2019fy	2020fy
	一般会計					
	特別会計(需給)	2,762	3,776	8,390	8,879	10,012
	開発成果促進財源					
	総NEDO負担額	2,762	3,776	8,390	8,879	10,012
	(委託)				8,174	9,332
(助成)				705	680	

開発体制	経産省担当原課	商務情報政策局 情報産業課
	プロジェクトリーダー	研究開発項目① 東京工業大学 教授 本村真人 研究開発項目②-(1) 産業技術総合研究所 デバイス技術研究ユニット長 川畑史郎 研究開発項目②-(2) 産業技術総合研究所 特別顧問 金山敏彦 研究開発項目③ 東京大学 教授 森川博之
	プロジェクトマネージャー	IoT 推進部 PM: 伊藤隆夫
	委託先 (助成事業の場合「助成先」とするなど適宜変更) (組合が委託先に含まれる場合は、その参加企業数及び参加企業名も記載)	代表実施者: ルネサスエレクトロニクス(株)、東京大学、東京大学生産技術研究所、KDDI(株)、(株)ソシオネクスト、沖電気工業(株)、日本電気(株)、イーソル(株)、熊本大学、東京理科大学、(株)デバイス&システム・プラットフォーム開発センター、(株)フィックスターズ、(株)エヌエスアイテクス、産業技術総合研究所、(株)Preferred Networks、(株)日立製作所、奈良先端科学技術大学院大学、日本電気(株)、技術研究組合光電子融合基盤技術研究所、早稲田大学、大阪大学、理化学研究所、中央大学、日本 IBM(株)、(株)東芝、技術研究組合 NMEMS 技術研究機構、東京工業大学、横浜国立大学、立命館大学、アラクスネットワークス(株)、等 112 社
情勢変化への対応	<p>本プロジェクトは、IoT 推進のための横断技術開発事業として、IoT 社会を高度化するための、情報の収集、蓄積、解析、セキュリティに分類される各種技術の開発を推進するプロジェクトとして 2016 年度に開始されたもの。一方で、ムーアの法則の限界や、情報社会が拡大したことに伴う爆発的な情報量の増加等、情報社会がもつ問題が一層表面化したことを受け、クラウドコンピューティングからエッジコンピューティングへの転換、並びに既存の技術の延長にない新原理の技術を先立って開発していくため、2018 年に「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」事業として、名称を変更すると共に、研究開発項目①、②を追加し (IoT 推進のための横断技術開発事業は、研究開発項目③として整理) 事業内容を拡充して事業を実施している。</p> <p>なお、研究開発項目①②については、2020 年度に実施するステージゲート審査の結果により、2021 年度以降の研究開発テーマの継続・一部見直し・早期完了・中止に加え、研究開発体制の見直しやテーマの統合、事業規模の拡大、縮小等の包括的な事業の見直しを行う。また 2022 年度から 2023 年度にかけては、本事業全体の見直しを実施する。</p> <p>最長 10 年間の研究開発期間を実施するに当たり、特に有効であると考えられる技術開発に支援を集約するために、これらの見直しに際しては、国内外における研究開発の動向や政策動向を踏まえ、外部有識者の評価も実施することとする。</p>	
評価に関する事項	事前評価	2015 年度実施 担当部 電子・材料・ナノテクノロジー部 2017 年度実施 担当部 IoT 推進部
	中間評価	-
	事後評価	-
3. 研究開発成果について	<p>研究開発項目① 2018 年度に事業を開始し、主に研究開発体制の確立、研究拠点の構築を進めた。各研究開発を進める中で、将来的な目標として設定する既存技術に対して電力効率で 10 倍以上を実現するための見込みを得た。</p> <p>研究開発項目② 2018 年度に事業を開始し、主に研究開発体制の確立、研究拠点の構築を進めた。各要素技術の開発を進める中で、将来的な目標として設定する既存技術に対して電力効率で 100 倍以上を実現するための見込みを得た。</p> <p>研究開発項目③</p>	

	2016 年度に事業を開始。サンプル提供や、実際の産業現場における実証実験等、構築した IoT システムを用いた実用化・事業化に向けた取組を行い、本研究開発項目の目標である、電力効率で 10 倍以上を達成すると共に、システムとしての実用性の検証を継続している。	
	投稿論文	777 件
	特 許	出願済み：237 件
	その他の外部発表	233 件（フォーラム、展示会等）
4. 成果の実用化・事業化に向けた取組及び見直しについて	<p>研究開発項目① キックオフミーティングやサイトビジットを実施し、本研究開発項目の関係者間のコミュニケーションを強化した。また、技術推進委員会を複数回実施し、その結果を基に研究開発計画の見直しや予算の増減、追加公募等を行った。その他、サンプル提供等によるユーザー評価を実施し、研究開発にフィードバックしている。</p> <p>研究開発項目② 関連する技術に関するシンポジウムの開催や、開発成果を活用するコンソーシアムの構築等を行い、研究開発と並行して早期のユーザーニーズの把握に努めた。また、事業化に当たり必要となる有効性を示す指標の整理を進めている。</p> <p>研究開発項目③ 事業化を見据えた研究開発体制の構築に加え、サンプル提供や、実際の産業現場における実証実験等、構築した IoT システムを用いた取組を行い、事業終了後、速やかに実用化・事業化に移れるよう研究開発を実施している。</p>	
5. 基本計画に関する事項	作成時期	2016 年 3 月 作成
	変更履歴	<p>2018 年 3 月、事業名称の変更、研究開発項目①及び②の追加による大幅改訂。なお、従来からの「IoT 推進のための横断技術開発プロジェクト」は研究開発項目③として実施する。</p> <p>2019 年 3 月、研究開発項目③の助成事業の追加等による改訂。</p> <p>2020 年 1 月、研究開発項目②の分割、ステージゲートに関する整理の記載。並びに西暦、和暦の修正等による改訂。</p>

プロジェクト用語集

用語	説明
AI エッジコンピューティング	AI を用いたエッジコンピューティング。エッジ領域において、AI 処理を行う。
AI エッジ統合制御システム	本事業において研究開発する、AI モジュールを組み込むシステム(通信・センサー・CPU 等)の呼称。
AMALTHEA	特に車載をターゲットとした、マルチコア向けオープンプラットフォームに関する一連の EU プロジェクト。2011 年に AMALTHEA(8M ユーロ)が開始され、現在は 2018 年より PANORAMA(19M ユーロ)が実施中。
APU	Accelerated Processing Unit: SoC 上の各アクセラレータ。
ARM	組み込み向けプロセッサの種類。
ASIC	Application Specific Integrated Circuit: 特定用途向け集積回路。
AS-SIMT アーキテクチャ	Application Specific SIMT のアーキテクチャ。GPGPU に代表される、同一命令を、多スレッドで同時実行する計算機アーキテクチャを、あるアルゴリズムに対して専用化する設計方法。並列計算による高速化が可能なアルゴリズムは多く存在するが、設計を共通化できることにメリットがある。
Autoware	ROS(Robot Operating System)をベースとした自動運転プラットフォームの名称。
Autoware.AI	ROS(Robot Operating System)をベースとした自動運転プラットフォームの名称。(Autoware.Auto と区別する場合に利用)
Autoware.Auto	ROS2 をベースとした自動運転プラットフォームの名称。
C/C++	高級汎用プログラミング言語のひとつ。
CNN アクセラレータ	畳み込み演算を駆使する DNN(CNN: Convolutional Neural Network) の高性能化をねらったアクセラレータ。畳み込みは GPGPU との親和性が高いが、専用化によってさらに低レイテンシ、低消費電力を実現できる可能性がある。
CUDA	NVIDIA 社から提供されている、NVIDIA 社の GPGPU 向けの計算プラットフォーム及びプログラミングモデル。
DDS	Data Distribution Service: リアルタイム pub/sub programming を実現する API を、プログラマへ提供。
DNN	Deep Neural Network: ニューラルネットワークのうち、隠れ層を 2 層以上積み重ねたネットワークのこと。DNN を利用した機械学習を総称して、Deep Learning と呼称することが多い。
DSP	Digital Signal Processor: 積和演算の高速な処理に特化したマイクロプロセッサ。
Eclipse Cyclone DDS	Eclipse Foundation で開発している DDS 名称。
ECU	各種電子制御をするユニット。
EKF	Extended Kalman Filter (拡張カルマンフィルタ): センサーに対する Filter のひとつであり、ノイズの低減等の効果がある。
FastRTPS	eProsima 社が開発しているオープンソースの DDS の名称。
FPGA	Field Programable Gate Array: アプリケーションを論理回路上にマッピング・プログラミングでき、信号が論理回路上を伝達することにより演算処理が実行可能な集積回路。プログラマブルで電力効率が高いなどがメリット。
GENESIS DevEnv	本事業で研究開発した技術を実用化した、エッジビジョン AI システムのプロトタイプ開発サービスの名称。

Github	GitHub 社によって管理されているソフトウェア開発のプラットフォーム。ここにソースコードをホスティングすることで、複数人のソフトウェア開発者と協働してコードをレビューしたり、プロジェクトを管理しつつ開発を行ったりすることができる。
GPU	Graphics Processing Unit: 数値計算(乗算・加算)を非常に高い並列度で処理する集積回路。グラフィック処理が主だが、近年、AI 学習に用いられている。特徴の一つとして、電力消費は大きい。
Halide	画像処理やテンソル計算の高性能計算を対象とした、ドメイン固有言語のひとつ。
Inet socket 通信	TCP ベースのネットワーク通信。
IntelHLS	Intel 社が提供している、Intel 製 FPGA 向けの高位合成ツールの名称。
IP	Intellectual Property: 知的財産。特に、SoC 関連の場合は、LSI や IC などの回路コンポーネントを纏めた設計情報を表す。
libc++	標準 C++ライブラリ。
LiDAR	Light Detection and Ranging: レーザー光を発射することによって距離情報を測定するセンシング技術や、それをういたセンサーの総称。
LSI	大規模集積回路。
Mcube	東京大学により独自開発された、マイクロカーネル OS。Mcube OS ともいう。
Mcube kernel	Mcube OS の中核となるプログラム。
MEC	Mobile Edge Computing: 計算リソースやストレージを移動体通信網のエッジに相当する基地局などに配備することで、サービスやアプリケーションにおいて応答の低遅延化などを実現する技術。
MPPA256	256 コアのメニーコアプロセッサの名称であり、本事業では試作で利用。
NoC	Network on chip: メニーコアもしくはクラスタ間の通信の名前。
ODD	Operational Design Domain: 運行設計領域。安全な自動運転が可能な条件。天候や走行、歩行者の存在を認めるか否かなどで、様々な種類の ODD が存在する。Level-N とは直交概念。
ONNX	Open Neural Network Exchange: DNN モデルを表現するための中間表現フォーマットのひとつ。
OpenCL	Open Computing Language: マルチコア CPU や GPU、Cell プロセッサ、DSP などによる異種混在の計算資源(ヘテロジニアス環境など)を利用した並列コンピューティングのためのクロスプラットフォームな API。
OpenCV	OSS で公開されている、コンピュータビジョン処理のためのライブラリのひとつ。
OS	オペレーティングシステム。
OSS	Open Source Software: ソースコードが一般に公開され、かつソースコードの利用・修正・再配布などが可能なソフトウェアの総称。
PredictiveCoding	脳が環境のモデルを用いて感覚入力を予測し、予測と実際の感覚入力との誤差を最小化するように情報処理を行なっていると、脳の情報処理様式に関する仮説。
Publisher/Subscriber プログラム	メッセージを使用する、プログラミングモデル。送信者、購読者を特に指定しない特徴がある。ROS が提供されているプログラミングモデル。
RCLC	C 言語のアプリ実装に共通機能を提供する ROS2 の API。
RCLCPP	C++のアプリ実装に共通機能を提供する ROS2 の API。
RISC-V	UC Berkeley を中心に開発されたオープンな RISC プロセッサ仕様(命令セットアーキテクチャ)。Rocket コアを中心とするフリーの実装も存在。

RMW	ROS Middleware: DDS の通信機能をアプリに抽象化する API。
ROS	Robot Operating System: ロボット用のソフトウェアプラットフォーム。
ROS 2	機能安全対応など製品化の対応を目指した次世代 ROS。
ROSBAG	ROS の通信で交わされるメッセージを記録するツール。
RTL	Register Transfer Level: レジスタからレジスタへの転送の振る舞いを記述するような、ハードウェアを記述する際に利用される抽象度のひとつ。
RTOS	Real Time Operating System 資源管理において、時間資源の優先度に基づく配分と実行時間の予測可能性を提供することに特化しているような OS。
SIMD 演算器	Single Instruction, Multiple Data 演算器: 1命令で複数のデータを処理可能な、プロセッサ内の高速演算器の総称。
SLAM	Simultaneous Localization and Mapping: 環境地図作成と自己位置推定を同時に行うアルゴリズムの名称。自動運転やロボットなどといった、マシンビジョンのアプリケーションに多く利用されている。
SoC	System-on-Chip: CPU および特定のアプリケーションが実行できる専用回路(FPGA を含む)、メモリ、インターフェース回路を備えた集積回路。システムチップともいう。
Society 5.0	日本が目指すべき未来社会の姿として提唱される、サイバー空間(仮想空間)とフィジカル空間(現実空間)を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会(Society)。
Soft-NMS 方式	ディープラーニングを利用した物体検出では、一つの物体に対して複数の検出結果が得られる。NMS とは、それらの複数の検出結果に対して、信頼度をもとに検出結果の出力を抑制し、一つの物体に対し一つの検出結果が得られるようにする処理。 NMS の中でも、単純な閾値処理ではなく連続的な関数を利用することにより精度向上を実現する方式。
SRAM	Static Random Access Memory: 6 個以上のトランジスタから構成された揮発性メモリ。情報の保持には電圧印加が必要。
SSS-Net	本事業により研究開発された、軽量化 DNN モデルの名称。
Subnetwork 方式	ディープラーニングのモデルのひとつであり、より詳細な特徴を抽出するためのモデルを組み込む方式。詳細な特徴を抽出することで、Subnetwork を組み込んだ部分に対して、精度向上が可能。
TensorFlow、PyTorch	DNN モデルを実装、学習、実行するための開発フレームワークの 1 つ。
ToF	Time of Flight: 赤外線を照射し、被写体からの反射光が戻ってくるまでの時間を測定することによって、被写体までの距離情報を測定する技術の総称。
UGV	Unmanned Ground Vehicle: 運転手が乗車することなく、無人状態で地上を走行できる車両。
Unix Domain Socket	単一マシン上の高効率なプロセス間通信に用いられる、機能・インターフェースの一種。
Upsampling 方式	ディープラーニングで畳み込み演算を繰り返すと、通常、特徴マップの大きさが小さくなる(Downsampling)が、畳み込み演算とは異なる演算を行うことにより、小さくなった特徴マップを大きくする処理。

	Upsampling を行うことにより、小さな物体を検出するための大きな特徴マップにおいて詳細な特徴が得られ、精度向上が可能。
Verilog HDL	ハードウェア記述言語のひとつ。
Veriloggen	東京大学が中心となって開発し、OSS で公開されている、マルチパラダイム型ハードウェア開発フレームワークの名称。
VivadoHLS	Xilinx 社が提供している、Xilinx 製 FPGA 向けの高位合成ツールの名称。
WorldModel	エージェントが自分の経験を元に学習した外界の環境のダイナミクスのモデルであり、現在の環境の状態とエージェント自身のアクションを元に、アクションの結果として環境がどのように変化するかを予測する。予測を元にアクションを選択することでエージェントが効率的にタスクを遂行できるようになる。 また、一般に WorldModel は、環境の情報を低次元のコンパクトな表現に変換した上で時間変化をモデル化しており、コンパクトな表現にすることで計算量・計算時間を削減している。
アクセラレータ	コンピュータの処理能力を加速するもの。
アテンション	主に画像認識や自然言語認識などで、認識に重要な情報に着目するための手法として開発されてきたニューラルネットワーク技術。
枝刈り	DNN における、学習パラメータの一部を除去する技法。
エッジコンピューティング	ネットワークの末端(エッジ領域)におけるデバイス(センサーやコントローラなどが搭載される端末)上で行う計算の総称。
エッジビジョン AI	エッジデバイス上で実行される、イメージセンサーなどに代表される可視光データを入力とした変換や認識、機械学習処理の総称。
演算量削減技術	深層ニューラルネットワークにより構成されるモデルを、よりコンパクトなものに置き換えることで、モデルの学習・推論の演算量を削減する技術。
学習	機械学習において、訓練データなどを用いて、機械学習モデル内のパラメータを決定するフェーズ。
環境シミュレータ	現実の物理環境を模擬したコンピュータ内の仮想環境で、深層学習・深層強化学習・人工意識の学習や評価のために用いるシミュレータ。 現実の物理環境では、データ取得やエージェントの行動に大きな時間・労力・費用などのコストが発生し、また予期しない要因による周囲の環境への悪影響のリスクも無視できない。環境シミュレータを用いることにより、これらのコストやリスクを大幅に削減することができる。 特に、エージェントの環境とのインタラクションにかかる時間を削減して、学習を高速化できることが大きな利点。
機械学習	主に、人間が計算機に対し明示的にルールを与えるのではなく、計算機に多くの例をデータとして与えて、そのデータの中に潜む有用な規則、ルール、知識表現、判断基準などを自動的に抽出する手法などの総称。
キャッシュメモリ	プロセッサ上に搭載されている高速小容量なメモリ。直近に使用されたデータや頻繁に使用されるデータをキャッシュメモリ上に自動的に格納することによって、メモリアクセスの遅延を隠蔽することが可能。
クラスタ	複数のコアをまとめた単位。
原子スイッチ	金属イオンが伝導可能な固体電解質を、銅およびルテニウム電極で挟んだ2端子構造を備えたスイッチ。印加電圧の極性によりその抵抗が変化し、信号のルーティングを行うクロスバススイッチや情報を不揮発に保持するメモリに利用。

高位合成	C/C++などのような高級プログラミング言語を入力として、ハードウェア記述言語に変換し出力する技術。
コンパイラ	高級言語で記述されたソースコードから、より低級なコードに変換するためのソフトウェア。変換と同時に、プログラムの等価性を保ったまま高速なプログラムに変換する場合も多く存在する。
自律	「作業の目的」を元に状況に応じて、適切に判断・行動してタスクを遂行する行動スタイル。一方、「作業の手順」を元に適切な行動を行う行動スタイルは「自動」と呼ぶことができる。「自律」は「自動」をより高度化させた概念。
人工意識	既存の知識を内部モデルとして持つことで、新しい問題を少ない入力情報で解く機能が、意識の重要な機能のひとつと考えられる。この機能をロボット・ドローン・仮想環境内のエージェントなどの人工的なシステムに搭載したもの。
深層強化学習	強化学習とは、逐次的な意思決定ルールを学習する機械学習の一分野であり、毎時刻ごとにエージェントは環境の状態を元に適切なアクションを選択し、得られる報酬の合計を最大にすることを目的としている。そのためには、環境の状態やアクションを評価することが必要となり、この評価に深層ニューラルネットワークを用いるもの。
推論	機械学習において、学習済みの機械学習モデルを利用して、未知のデータを入力して出力を得るフェーズ。
スーパースカラ	プロセッサにおいて、複数の命令を、複数の実行ユニット上で並列に動作させるアーキテクチャの総称。
スクラッチパッドメモリ	プロセッサ上に搭載されている高速小容量なメモリ。キャッシュメモリと比較して、スクラッチパッドではデータの格納やメインメモリとの転送を、プログラム上で明示的に記述する必要がある場合に呼称されることが多い。
畳み込み層	CNNの構成要素。一つの畳み込み層には複数のチャンネルが含まれる。入力データ上に分布する空間的な特徴を保持した演算が可能。特に画像認識のタスクを行うDNNで頻用される。
チャンネル	フィルターとの畳み込み演算の結果を保持するニューロンの集まり。
電力対性能	一般的に、消費電力を上げれば、よりニューラルネットワークを速くさせることができる。性能評価のためには、同じ電力でどれだけ速くニューラルネットワークを動作させることができるかを考える必要があるため、評価指標として用いることが可能。
ドメイン固有言語	Domain Specific Language, DSL: 特定の用途向けに特化して設計されたプログラミング言語の総称。
ニューラルネットワーク	生物の脳の認知機能と構造に基礎を置いている機械学習モデル。神経細胞を数理的にモデル化した人工ニューロン素子のネットワークによって構成。
パイプラインスケジューリング	プロセッサ内の演算器パイプラインにおいて、プログラム中のどの演算を、どの演算器上で、どのような順番で実行するかを決定すること。

複数深層学習	<p>複数のタスクを共通の入力データを用いて深層ニューラルネットワークで学習させること。</p> <p>例えば、物体検知と意味的領域分割は異なるタスクであるが、共通の入力画像データで学習させることが可能だが、もし入力データだけでなく深層ニューラルネットワークの途中の層の出力についても複数のタスクで共通に用いることができれば、ニューラルネットワークの一部を複数のタスクで共通のものにすることが可能になり、演算量を削減できる。</p>
ペイロード	通信における通信したい内容本体。
ヘテロジニアスコンピューティング	非同一の複数プロセッサでコンピュータを構築すること。CPU と GPU の組み合わせもこの発想のひとつ。
ミドルウェア	アプリケーションプログラムを特定のハードウェアで動作させる際に、その仲立ちとなって動作するソフトウェアならびにその集合。
モデルベース並列化	MBP: システムモデルや MATLAB/Simulink モデルのようなモデルのレベルにおいて、モデル内のブロック単位でプロセッサコアに割り当て、並列化するツール。C 言語プログラムなどのコードレベルと比較し、システム全体の視点から並列化できるので、特にヘテロジニアスアーキテクチャに向いている。
有限状態機械	Finite State Machine, FSM: 有限個の状態と、各状態における動作の組み合わせによって構築される、数学的な計算モデル。
量子化	DNN におけるアクティベーションや学習パラメータのビット精度を低減し、等価あるいは近似的な演算に変換する技法。
ループ変形	プログラム中におけるループの形や実行順序を変形するような、プログラム変換手法のひとつ。

1. 事業の位置付け・必要性について

1.1 事業実施の背景

IoT端末(デバイス)により、コンピュータが処理可能なデジタルデータに変換されることにより、大量のデータを活用することができ、新たなサービスモデルの創出や、産業や社会システムの構造を大きく変える可能性がある。

特に、少子高齢化・労働力不足、地球環境・エネルギー問題、社会インフラの維持・メンテナンス、地域経済の維持等の課題解決に向け、IoT、AIの利活用に期待が大きい。

他方で、IoT 技術が社会のあらゆる分野に実装されることで、インターネットに繋がる機器は大幅に増大し、これまでデジタル化されていなかったデータがネットワークに加速度的に流入することで、情報の収集・蓄積、流通、解析、制御等のあらゆるプロセスにおいて機器が消費する電力が大幅に増大することが見込まれている。

民間の試算では、全世界で IoT 機器の台数が 2025 年には 416 億台(2020 年との比較で約 1.5 倍)に達し、流通するデータ量も年間 79.4 兆 G バイト(2020 年との比較で約 4 倍)になると予測され、これに伴い機器の消費電力は増大される。また、その市場規模は産業用 IoT の世界市場においてだけでも 2025 年段階で 1106 億ドル(2020 年との比較で約 1.5 倍)に達するとの民間予測があり、経済効果においては IoT 技術の適用先と効果が非常に多岐の分野にわたるため、全体で 2025 年に 3.9~11.1 兆ドルに及ぶと試算されている。

このようななか、ムーア則の終焉が叫ばれ、半導体微細化による情報処理性能の向上は限界を迎えつつある。エッジやクラウド等において、既存技術の延長にない新原理の技術により、著しく増加するデータ処理に要する電力の劇的な低減が必要とされる。

また、民間の試算では、IoT 機器の台数が 2025 年には 416 億台(2020 年との比較で約 1.5 倍)に達し、流通するデータ量も年間 79.4 兆 G バイト(2020 年との比較で約 4 倍)になり、2025 年には世界で生成されるデータの約 30%がリアルタイムデータになるという予測があり、AIエッジでの自律制御、機会学習を可能とする、新たなコンピューティング技術の実現も期待されている。

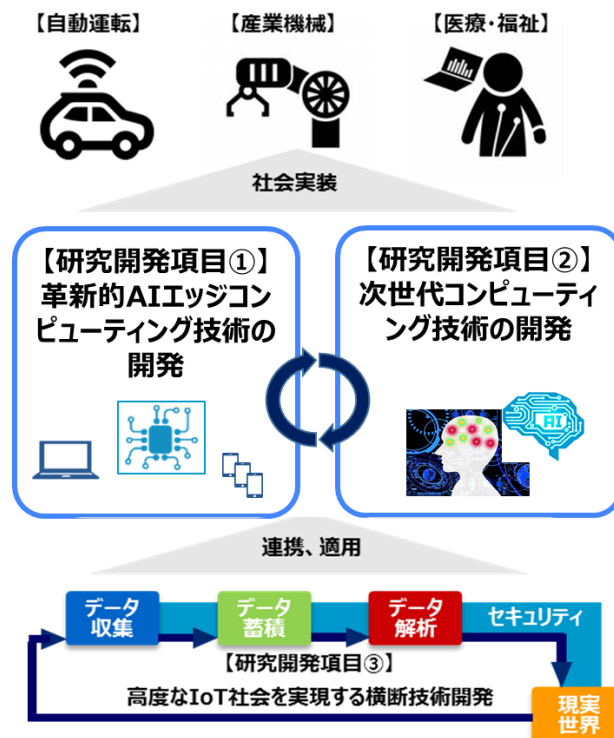


図 3.1.1 IoT が実現する CPS(Cyber Physical System)

1.2 政策的位置づけ

政府においては、「日本再興戦略」改定 2015(2015 年 6 月 30 日閣議決定)において、IoT・ビッグデータ・人工知能時代の到来により、ビジネスや社会の在り方そのものを根底から揺るがす「第四次産業革命」とも呼ぶべき大変革が進みつつあり、未来の幅広い分野における産業創造や社会変革に対応するため、新たな時代を支える共通基盤技術(IoT、ビッグデータ解析、人工知能、センサー等)に関して研究開発等を実施することが期待されるとしている。また、「日本再興戦略」改定 2015 を受けて、2015 年 10 月に IoT 推進コンソーシアムが設置され、官民共同で IoT を活用した未来への投資を促すべく、新たなビジネスモデルの創出、IoT 推進のための技術開発・実証に係る、規制改革等の提言等の取組が推進されている。

さらに、2015 年 6 月に閣議決定された「科学技術イノベーション総合戦略 2015」においても、「現在発展しつつある個別のシステムが更に高度化し分野や地域を超えて結び付き、あらゆるものがネットワーク化されることにより、必要なもの・こと(サービス)を、必要な人に、必要な時に、必要なだけ提供でき、社会の様々なニーズに対し、きめ細やかに、かつ、効率良く対応できる「超スマート社会」ともいうべき社会が向かう方向性と考えられる」とした上で、「超スマート社会(Society5.0)」の実現に向け、国際競争力強化や生産性の向上を図り、持続的な社会基盤づくりにつなげていくことの重要性が指摘されている。

総合科学技術・イノベーション会議が策定に向けて検討を行ってきた 2016～20 年度の第 5 期科学技術基本計画について 2016 年 1 月に閣議決定されたが、ここでも超スマート社会サービスプラットフォームの構築に必要な技術のうち、特に国として速やかな強化を図る技術として、デバイス技術、エッジコンピューティング、ビッグデータ解析技術、AI技術、ネットワーク技術、サイバーセキュリティ技術、IoT システム構築技術等の IoT 関連の基盤技術が挙げられている。

本研究開発項目の開始後においても、内閣府が 2019 年 6 月に掲げた我が国の産業力強化に向けた統合的な政策パッケージである「AI 戦略 2019」では、2022 年までに、情報処理に係る消費電力性能を従来比 10 倍以上に向上させる革新的 AI チップ技術の研究開発方針が示され、2020 年 5 月に経済産業省により掲げられた「産業技術ビジョン 2020」では、ポストムーア時代の次世代コンピューティング技術(エッジ)、Intelligence of Things を支えるキーテクノロジー群等の先端技術の研究開発に必要な措置が講ずることとされている。

さらに、内閣府が 2020 年 7 月に閣議決定した「統合イノベーション戦略 2020」では、コロナ渦、国内外の変化を踏まえた我が国の課題としてデジタル化等を挙げ、AI 等の基盤技術において、世界最先端の研究を推進することとされている。

表 1.2 IoT デバイス、AI エッジコンピューティングの研究開発に係る関連政策

	政策	概略	'15~'20	'25~'50
内閣府	統合イノベーション戦略2020 (2020.7.17)	・コロナ禍、国内外の変化を踏まえた我が国の課題(危機感とスピード化を持ち、デジタル化を加速、イノベーションを創出) に対して、重点的に取り組むべき施策(Society5.0)の具体化を提示。 ・戦略的に進めていくべき主要分野=基盤技術(AI、バイオ、量子技術、マテリアルなど世界最先端の研究開発、人材育成他)、応用分野(安全安心、環境エネルギー、健康・医療ほか)	2019	2020
経済産業省	産業技術ビジョン2020 (2020.5.29)	・2050年に向けた5つのグローバルメカトレンドと世界動向を踏まえ、日本が包摂する本質的課題を仮説として特定し、2050年の産業技術の方向性、2050年までに実現すべきことを取りまとめ ・R&D投資の重点化(デジタルにおいては、全ての基盤となるポストムーア時代の次世代コンピューティング技術(エッジ)とIntelligence of Thingsを支えるキーテクノロジー群のR&Dを強化)		ターゲット時期 (中長期: 2025、次の30年)
内閣府	AI戦略 (2019.6.11)	AI社会原則を踏まえ、Society 5.0の実現を通じて世界規模の課題の解決に貢献するとともに、我が国自身の社会課題も克服するため、さらには、その先の我が国の産業競争力の向上に向け、「人、産業、地域、政府全てにAIを」普及させるために策定。		
経済産業省	Connected Industries (2017.10.2)	・「Society5.0」実現に向け、様々な業種、企業、人、機械、データなどがつながり、AI等によって、新たな付加価値や製品・サービスを創出、生産性を向上させることにより、高齢化、人手不足、環境・エネルギー制約などの社会課題を解決。これらを通じて、産業競争力の強化を目指す。		
	Society5.0 (2016.1.22)	・サイバー空間とフィジカル(現実)空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する人間中心の社会 ・IoTで全ての人とモノがつながり、様々な知識や情報が共有され、新たな価値が生まれる社会など4つの社会像の実現を目指す		
内閣府	第5期科学技術基本計画 (2016.1.22)	10年先を見通した5年間の科学技術の振興に関する総合的な計画。基本方針・4本柱のひとつとして「超スマート社会の実現(Society5.0)」を掲げる		ターゲット時期 (10年先)
	日本再興戦略改定2015 (2015.6.30)	未来投資による生産性革命の実現に向け、ビジネスや社会の在り方そのものを根底から揺るがす「第四次産業革命」とも呼ぶべき大変革に対し、生産性向上のための施策として、鍵となる施策として「IoT、ビッグデータ解析、AI、センサー等による産業構造・就業構造の変革の検討」を掲げた。		
	科学技術イノベーション総合戦略2015 (2015.6.19)	「未来の産業創造・社会変革」に先行し、あるべき経済・社会システムを構想し、SIPを含め研究開発を組み合わせ(システム化)、産業競争力を生み出す価値の連鎖(バリューチェーン)を形成。社会実装に向け2020年までの成果目標を設定。我が国の強みを活かしてIoT、ビッグデータ等を駆使した新産業の育成を施策の一つとした。		ターゲット時期 (5年先)

1.3 アウトカム効果

各研究開発項目の下で実施する個別の研究開発テーマ毎に、エネルギー消費効率あるいは電力効率(単位電力あたり性能)について、事業開始時点における同等の技術と比較した目標を設定する。

想定する社会実装先(自動運転、産業機械、医療・福祉等)でのAI関連ハードウェア世界市場において、研究開発成果の一部が市場に出る2032年に約7,500億円、さらに成果の普及が加速する2037年に約1.6兆円の市場獲得し、それに付随するソフトウェア及びサービス産業等により、更なる波及効果の創出を目指す。さらに、エッジやクラウド等の省電力化を実現し、2037年において約2,729万t/年のCO2削減を目指す。

1.4 国際的なポジション

現状では、米国系ソリューションプロバイダ(Google社、Apple社、Facebook社、Amazon社等)が世界のコンピューティング分野を牽引している。ハードウェアからソフトウェアまでの一体開発に強みがあり、クラウド領域を中心に市場を含め世界を席巻している。各陣営による買収合戦も激化しており、新たな製品やサービスが目まぐるしいスピードで社会に投入されている。そのような中で、処理をするデータ量の増大から、エネルギー効率やリアルタイム性の観点等から、クラウドからエッジへデータ処理の主軸が移行しつつあり、エッジにおけるデータ処理の高機能化が求められている。そのひとつが人工知能(AI)を取り入れたものである。

各国のAI関連の政策動向を見てみると、米国においては2018年5月のAIサミットにおいて、国防総省を中心としてAI Next Campaign(~2023年)が張られ、AI、量子、5G通信領域での戦略策定の動きが活発化した。2018年7月のDARPAにおけるAI探索プログラムの発表、さらには、2019年2月のAIイニシアチブに関する大統領令も発せされた。欧米ではEUにおいて2018年4月にAI活用政策指針・AIに関する協力宣言がなされ、これを受けて2018年12月にAI協調計画が発表された。EUの動きに連動し、英独仏でそれぞれAI Sector Deal(英・2018年4月)、ハイテク戦略2025(独)、AI国家戦略(独・2018年11月)、AI国家戦略(仏・2018年3月)が立案された。中国においては、次世代人工知能発展計画(2017年7月)から次世代AIガバナンス原

則(2019年6月)を制定するなど次世代 AI 産業の発展促進に関する動きが活発である。韓国でも AI 国家戦略 2019、台湾においては人工知能に関する科学研究戦略において五大戦略を定めている。このように各国とも積極的にこの分野において政策的な後押しをしている。

一方で、世界の取り組みにおいても、未踏な部分があると考えられる。今後は、第 4 次産業革命技術の更なる進展に伴い、従来のクラウドコンピューティングからネットワークの末端(エッジ)で中心的な情報処理を行うエッジコンピューティングへの分散が不可欠になると考えられるが、クラウドで確立された技術とは異なり、エッジで情報処理を実現するための小型デバイスとコンピューティング技術については各国各社手探りの部分も多く、サイズやエネルギー性能に加え、ソフトウェアとの組み合わせであるコンピューティング技術として最適化できていないと思われる。

また、次世代コンピューティングの領域では、先に述べたような様々な開発の動きがあるものの、例えば販売事例のある量子アニーリングコンピュータであっても、実用的な組み合わせ最適化問題を解くためには量子ビット数を大幅に増加させる必要があるだけでなく、結合や正確性の問題、ミドルウェアを含むソフトウェア等の開発等の課題を抱えており、社会実装まではまだ遠い状況である。

1.5 海外状況のまとめ

第 4 次産業革命の第 1 幕にあたる、これまでのネット上のデータ(バーチャルデータ)獲得競争においては、米国系企業(Google、Amazon、Apple 等)が世界を牽引しており、そのプラットフォームを海外企業に握られているため、我が国はその恩恵を享受できていない。

また、第 4 次産業革命の第 2 幕として、健康・医療・介護、製造現場、自動走行等の産業応用に向けたフィジカル領域における「リアルデータ」を巡る競争へ突入しており、欧米で大規模で、長期的な政府系プロジェクトが推進されている。

例えば、欧州では「AUTONOMICS for Industry 4.0」(ドイツ、2013～2017)、「Smart Service World I」(ドイツ、2016～2019)、「Smart Service World II」(ドイツ、2018～2021)、「Internet of Things and Platforms for Connected Smart Objects」(欧州委員会、2016～2019)、「Large Scale Pilots」(欧州委員会、2017～2020)、「IoTUK」(英国、2015～2018)等のプロジェクトが実施され、

製造プロセスにおけるエネルギーの削減や原料の消費における効率化および環境負荷の軽減、IoT による革新的な支援サービスを促進するための技術開発、業界横断型のソリューション開発、スマート製品向けプラットフォームのエコシステム構築、実社会での IoT の導入に向けた大規模なパイロット試験、産業分野および公共分野における IoT 技術やサービスの利活用推進等が行われている。

また、米国では「Cyber-Physical Systems」(NSF、2009～)、「Big Data R&D Initiative」(NSF/NIH/DOD/DOE/DARPA/USGS、2012～2016)、「SMART AMERICA Challenge」(White House、2013～2014)、「Global City Teams Challenge」(NIST、2014～)、「Digital Manufacturing and Design Innovation Institute/MxD」(DoD、2014～)等のプロジェクトが実施され、科学、基盤技術の研究開発支援から BigData から知見を引出す技術開発、Smart Manufacturing、Smart City 等の実証、デジタル製造設計によるコスト削減のための開発等が行われている。また、中国では「中国製造 2025」(國務院、2015～2025)が実施されており、情報化と産業化の融合をその理念とし、スマート製造・グリーン製造を目標に推進されている。加えて、各国においては要素技術の研究開発に係るプロジェクトも実施されている。

これらの動きと相まって、ドイツの Industrie 4.0 においては 関係企業・団体等を構成員としたコンソーシアムが立ち上げられるとともに大手企業が旗振り役となり中小企業を含む産学官の垂直連携・水平連携 体制の構築が推進され、米国においては製造・IT 分野における企業が多数

参画し、民間主導で業界をまたいで垂直・水平連携した IIC (Industrial Internet Consortium) が 2014 年に設立される等、垂直・水平連携により製造業全体の生産性向上、製造と IT 技術の融合による新たなバリューチェーン改革、標準化に向けた取組が行われている。

上記に対し、我が国特有の強みである多様で活用性の高いリアルデータの蓄積等を活かし、日本にイノベーションを生み出す「リアルデータプラットフォーム」を創出・発展させるため、IoT 推進の横断的技術開発、AI エッジコンピューティング技術開発、次世代コンピューティング技術開発の各成果を次々に具現化していく必要がある。

1.6 NEDO が関与する意義

データトラフィックの爆発的な増加を支えてきたムーア則が終焉を迎えつつあるなか、従来の微細化の進展に依拠した CPU 中心のコンピューティングアーキテクチャでは、リアルタイムデータ処理、自律制御への対応が困難となっており、集中処理と、エッジ処理との制御を伴う新たなコンピュータアーキテクチャの実装が必要とされる。

5G から 6G へ向かうモバイル技術の進展に伴い、クラウドに加えてエッジコンピューティングの重要性が増している。

コンピューティングとネットワークが融合したシステム全体で最適化する方向性が強まっており、従来のコンピューティングやデータ形式等の大規模な変革と産業のゲームチェンジを想定し、このゲームチェンジを日本として契機とするべきである。

このようなゲームチェンジに対して、ハイパースケール企業が不在の日本において、民間単独で対応することは困難なため、産官学の事業者が互いのノウハウを持ち寄り、協調して研究開発を行うことが必要であり、さらに実証推進においても国の関与が必要と考えられることから NEDO が実施すべきものである。

1.7 今回の事業の位置づけ

本研究開発項目では、情報提供依頼 (RFI) の結果や政策的観点から、以下、(A) から (F) の重点課題を設定し、公募を実施した。実施期間中においては、最終目標である、開発成果を組み込んだシステムレベルでのエネルギー消費効率あるいは電力効率 10 倍達成に向けて、適宜外部委員等による助言等を行った。また、先導調査研究として開始した 4 テーマについては、研究開発期間終了に伴うテーマ事後評価を実施している。

IoT のさらなる高度化を実現するために、従来に比べて格段に省エネルギーで高度なデータ活用を可能とする次世代技術を産学官の連携体制で開発する。

(A) 不揮発性素子等のスイッチング機構を用いたコンピューティング技術

フラッシュメモリとは異なる原理で ON/OFF 状態を保持する新構造の不揮発性素子を用いたスイッチング機構を、メモリ、通信回路、デジタル信号処理回路等に導入することで、低消費電力の AI 処理を可能とするコンピューティング技術を開発。

(B) リコンフィギュラブルデバイスによるコンピューティング技術

与えられたタスクや設置環境に合わせて、チップ内の回路構成を動的に変更することにより、常に高速かつ低消費電力での動作を実現する技術等を開発。

(C) 演算処理量の軽量化を実現する AI 組込みコンピューティング技術

チップがデータ処理する際に、データを軽量化してから処理するよう、チップに入力される計算プログラムを自動的に変更することにより、チップでの情報処理量を可能な限り軽減する技術等を開発。

- (D) エッジコンピューティング向けリアルタイムソフトウェア制御技術
 多数のプロセッサを高い実行効率で制御できるのみならず、エッジコンピューティングで要求されるリアルタイム性を満足し、かつ省エネ性能に優れた OS 技術等を開発。
- (E) 多数の分岐ノードを有する AI アルゴリズム処理を高性能化するコンピューティング技術
 与えられたタスクを小さな処理単位に分解・整列し、プロセッサの利用状況と処理単位の実行優先度を動的に判断して実行制御できる並列プロセッサ技術等を開発。
- (F) エッジデバイスのセキュリティ技術及びその評価技術
 オープンアーキテクチャの CPU 命令セットである RISC-V を用いて、産業用途で多用されているデータ格納方式に対応したプロセッサや、当該プロセッサ上でアプリケーションをセキュアに動作させる実行環境、当該プロセッサ上で AI 処理・並列処理などで多用されるライブラリを活用可能にする技術等の開発。
 エッジデバイスへの攻撃に対するセキュリティ技術等を開発するとともに、既存の技術も含めて、その性能を横断的に評価するための評価技術等を開発。
 また、RISC-V コアやそれをベースとしたセキュリティ基盤技術等に関するソフトウェアスタック群の開発。

取組区分	実現区分	既存ハード	新アーキテクチャ	新デバイス	
1. 専用チップ (AIアクセラレータ、SoCの開発)		CPU・FPGA等 ※開発対象外	(B) リンフィギュラブルデバイスによるコンピューティング技術 ルネサス PFN	(E) 多数の分岐ノードを有する AI アルゴリズム処理を高性能化するコンピューティング技術 エヌエスアイテクス 東京大学	(A) 不揮発性素子等のスイッチング機構を用いたコンピューティング技術 NEC
2. コンピューティング技術 (OS、コンパイラ、ツール等、開発環境の開発)			(C) 演算処理量の軽量化を実現する AI 組み込みコンピューティング技術 ソシオネクスト 沖 KDDI フィックスターズ		
			(D) エッジコンピューティング向けリアルタイムソフトウェア制御技術 イーソル		
			(F) エッジデバイスのセキュリティ技術及びその評価技術 イーソル		
3. セキュリティ基盤 (エッジ向けセキュリティ)			TRASIO 産総研		

図 1.7 「革新的 AI エッジコンピューティング技術の開発」の事業体制

2. 研究開発マネジメントについて

2.1 事業の目的

IoT 社会の到来によるデータ量の爆発的な増加に対応し、急増したデータの高度な利活用を促進するために、ネットワークの末端(エッジ)で高度かつ低消費で情報処理を行う「エッジコンピューティング」の確立が必要である。

また、エッジ領域における、省電力化や高性能化、ハード・ソフトを組み合わせたコンピューティング技術の最適化は、いまだ手探り状態であり、日本が強みを有する「現場」で生成されるデータを処理する技術に関する研究開発を実施し、成果を社会実装することにより、我が国の産業競争力の向上が見込まれる。

そのため、エッジにおける AI 処理を実現するための小型かつ省エネながら高度な処理の能力を持った専用チップ及びコンピューティング技術等を開発することを目的とした。

2.2 研究開発目標と根拠

以下の通り、エッジコンピューティングにおける AI 処理を実現するための小型かつ省エネルギーながら高度な処理能力をもった専用チップと、それを用いたコンピューティング技術に関する研究開発を、社会課題を見据えた上で実施する。またエッジ側におけるセキュリティ基盤技術の研究開発も実施する。

達成目標

各研究開発テーマにおいては、以下を達成することを目標とする。なお、研究開発期間によっては、必要に応じて中間、最終目標を、以下に示す目標を基準としつつ変更して設定する。

<中間目標(2020 年度)>

開発成果を組み込んだ要素技術に係る検証あるいはシミュレーションにより、エネルギー消費効率あるいは電力効率(単位電力あたり性能)が、事業開始時点における同等の技術と比較し、10 倍以上となる見込みを示す。

<最終目標(2022 年度)>

開発成果を組み込んだシステムレベルでの検証を行い、エネルギー消費効率あるいは電力効率(単位電力あたり性能)が、事業開始時点における同等の技術と比較し、10 倍以上となることを示す。

上記目標は、事業終了時点で社会実装先に求められる諸性能を満たすことを前提に、事業開始時に広く普及している技術と比較し評価する。研究開発対象が社会実装するために必要不可欠なセキュリティ技術等の場合については、対象技術自体のエネルギー効率等を加味して評価する。

また、先導調査研究枠は、以下のうち、いずれかを達成することを目標とする。

- ・ 事業終了時点において、研究開発事業等への移行に向け、根拠データの取得等により、技術の確立の見通しを付けることを目標とする。
- ・ 周辺技術や関連課題に係る開発及び研究開発に直結する調査については、イノ

バージョンの創出や本事業における更なる成果最大化に繋げることを目標とする。

・ 表 2.2 研究開発テーマ一覧

開発課題	事業形態	研究開発テーマ	研究代表	中間目標(2020 年度末)	達成状況
(A)	委託	不揮発省電力 FPGA コアを用いた低遅延 AI 処理 コンピューティング技術の研究開発	日本電気株式会社	ASSoC に向け、原子スイッチの製造技術や回路設計等を実施する。	達成見込み。
(B)	委託	動的再構成技術を活用した組み込み AI システムの研究開発	ルネサスエレクトロニクス株式会社	1 次 TEG の検討、仕様策定、試作、及び評価等を行い、加えて、ツール等の研究開発等を実施する。	達成見込み。
(B)	委託	進化型・低消費電力 AI エッジ LSI の研究開発	株式会社ソシオネクスト	開発した仮想エンジンアーキテクチャ技術や量子化 DNN エンジン技術を搭載した LSI を試作し、各種評価等を実施する。	一部実施項目を前倒しで達成見込み。
(B)	委託	FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発	株式会社 Preferred Networks	可変精度演算コアと AISC 用 FPGA IP を組み合わせたプロセッサアーキテクチャを設計し、10TOPS/W を実現できるプロセッサの開発等を実施する。	新型コロナウイルス感染症の影響により、一部実施項目において、多少の遅延が発生。
(C)	委託	エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発	株式会社フィックスターズ	軽量化 DNN を開発し、加えて、ベンダ非依存のオープンソース高位合成基板のプロトタイプに開発等を実施する。	達成見込み。
(C)	委託	ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発	沖電気工業株式会社	ソフトテンソルプロセッサ技術や高効率モデリング技術、広範囲大規模認識技術を開発し、20 倍以上の電力効率(現状の GPU の電力効率を 30GOP/J と想定)の改善等を実施する。	達成見込み。
(C)	委託	5G 時代を見据えた高度自律的学習機能搭載のための AI エッジコンピューティング技術	KDDI 株式会社	深層強化学習の Base Network 層共通化及び FPGA 実装 (Base Network 層共通化及び FPGA 化効果による、消費電力 (対 GPU) 従来比 1/10 目標) 等を	達成見込み。

開発課題	事業形態	研究開発テーマ	研究代表	中間目標(2020 年度末)	達成状況
		の研究開発		実施する。	
(D)	委託	スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発	イーソル株式会社	チップ内/チップ間/ノード間の統合的動的通信最適化機構を開発し、従来のミドルウェア型メッセージパッシングに対して 10 倍のメッセージパッシング効率の実現等を実施する。	達成見込み。
(E)	委託	動的多分岐・結合トレース型 AI プロセッサのエコシステム開発	株式会社エヌエスアイテクス	AI をエッジシステムで利用する際に必要となる、ニューラルネットワークおよびその前後処理、ルールベース AI の補完処理、時分割処理にも対応可能な、高速かつ高効率・低消費電力なエコシステムを、ハードウェア、ソフトウェア、開発環境も含めて実現するべく、基本アーキテクチャを決定し、協調設計の推進等を実施する。	達成見込み。
(E)	委託	完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発	国立大学法人東京大学	FPGA を使った SoC プロトタイプング実装および評価等を実施する。	達成見込み。
(F)	委託	セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発	セキュアオープンアーキテクチャ・エッジ基盤技術組合 (TRASIO)	RISC-V アプリケーションコア部との回路接続とその回路を制御するためのソフトウェアを開発し、並行して開発したセキュア MUC への実装等を実施する。	達成見込み。
(F)	委託	AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発	国立研究開発法人産業技術総合研究所	AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術について、各種セキュリティ要求仕様と評価分析手法の策定等を実施する。	達成見込み。

2.3 研究開発スケジュール

研究開発スケジュールを図 2.3 に示す。

1. 事業の位置付け・必要性 (1) 事業の目的の妥当性

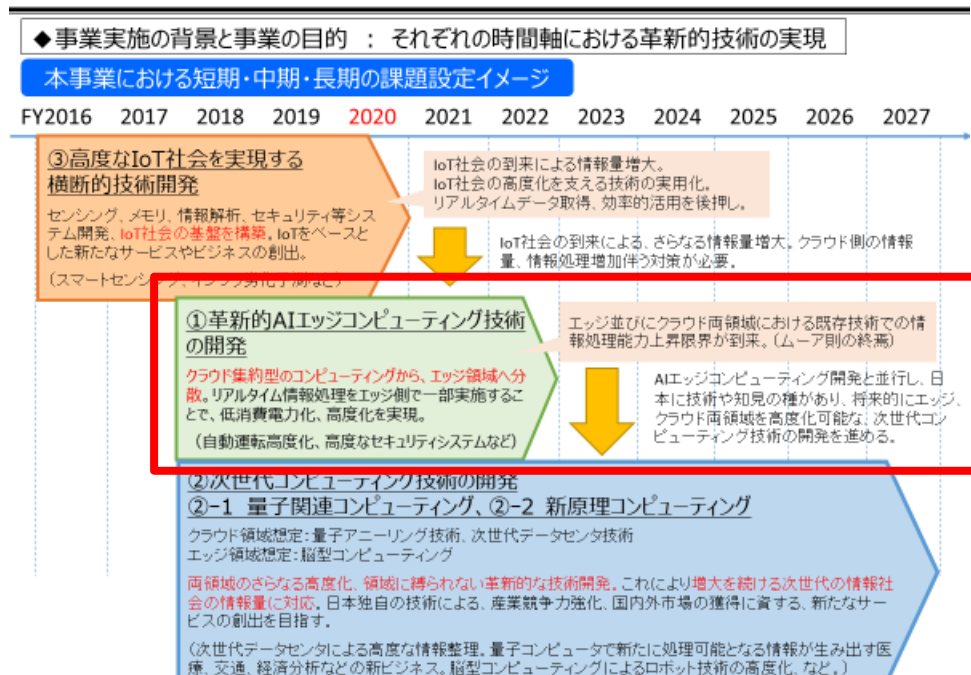


図 2.3 研究開発スケジュール

【2018 年度】

本研究開発項目の実施体制を構築すべく、情報提供依頼(RFI)の結果や政策的観点から重点課題を設定し、2 回の公募を実施した。結果、研究開発枠 12 テーマ等を選択し、研究開発に着手した。

【2019 年度】

各事業者のサイトビジットおよび技術推進委員会を通じて進捗状況を確認し、最終目標である、開発成果を組み込んだシステムレベルでのエネルギー消費効率あるいは電力効率 10 倍以上の達成に向けて外部委員等による助言等を行った。また、先導調査研究として開始した 4 テーマについては、研究開発期間終了に伴うテーマ事後評価を実施した。

【2020 年度】

既存の研究開発テーマのスコープには含まれていない、誰もが使いやすい組み込みシステム向け RISC-V のソフトウェアスタック群の研究開発に関する追加公募を実施し、結果 1 テーマを選択した。

既存の研究開発テーマの進捗は、最終目標に向けて概ね順調に進捗。ステージゲート審査等を実施し、2021 年度以降への研究開発の継続可否に加え、研究開発体制の見直しや、事業規模の拡大、縮小等の包括的な事業の見直しを行い、最終目標の達成に向けてより一層邁進していく。

2.4 プロジェクト費用

2018年度から3年間のプロジェクト費用は、表の通りである。2018年度及び2019年度は執行ベースの金額であり、2020年度は契約ベースの金額としている。

表 2.4 プロジェクト費用 [単位:百万円(四捨五入)]

研究開発項目	年度			
	2018	2019	2020	合計
①革新的AIエッジコンピューティング技術の開発	2,282	4,393	4,512	11,187
②次世代コンピューティング技術の開発	1,866	2,021	3,324	7,211
③高度なIoT社会を実現する横断的技術開発	4,093	2,344	2,029	8,466
その他調査事業等	149	121	147	417
合計	8,390	8,879	10,012	30,337

2.5 マネジメント体制

本研究開発項目では、技術的成果及び政策的効果の最大化、産業社会への還元等を目指し、プロジェクト全体の企画・推進・管理を行うためにプロジェクトマネージャー(以下 PM という)およびサブ・プロジェクトマネージャー(以下 SPM という)を設置した。

加えて、効率的な研究開発の推進を図る観点から各実施者の研究開発能力を最大限に引き出し、性格の異なる各研究機関のベクトルを束ねるため、当事業には NEDO が指名する研究開発責任者(プロジェクトリーダー、以下 PL という)を設置し、各実施者はプロジェクトリーダーの下で研究開発を実施する体制をとっている。

PL は、プロジェクトをより効率的かつ効果的に遂行するために、プロジェクトの技術目標等の達成に向けた取り組み、研究開発の進捗状況の把握、プロジェクトの実施体制の構築・改変及び事業者間等の予算配分に係る助言、プロジェクトの成果の評価等に係る業務の全部又は一部について、PM/SPM と緊密な連携を図っている。

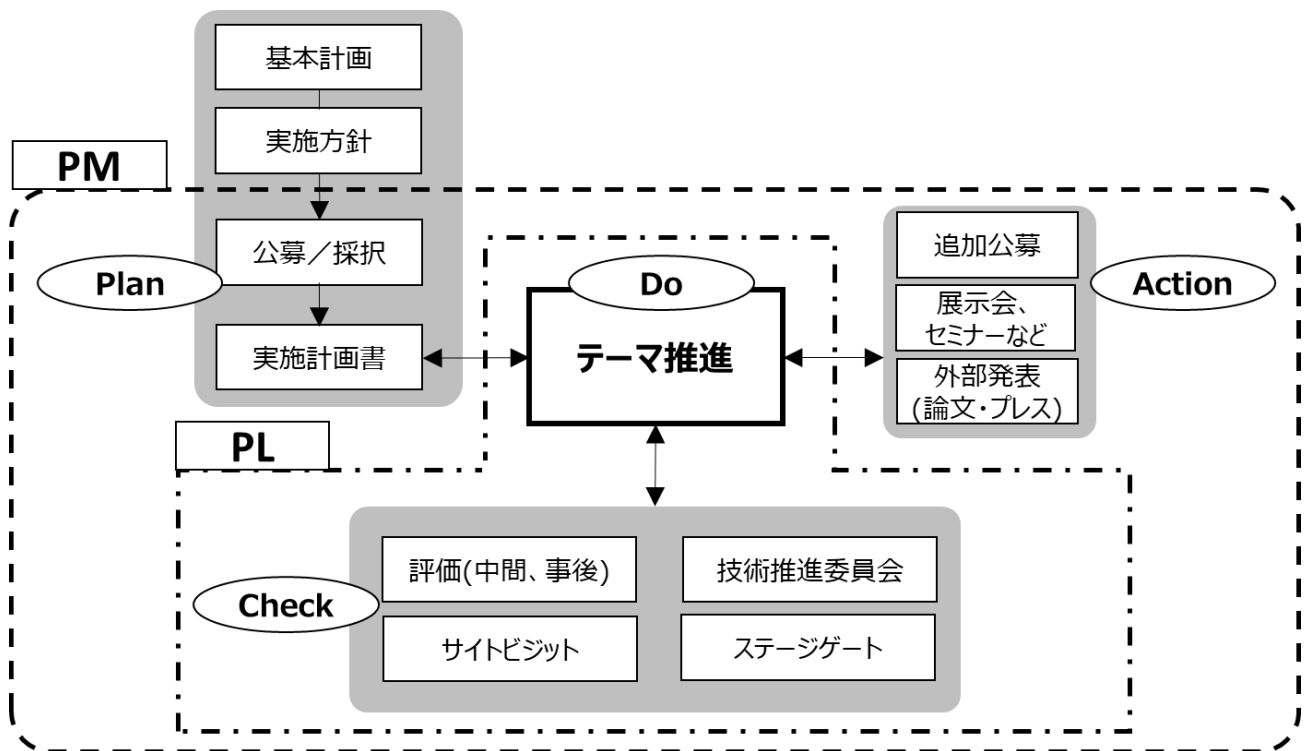


図 2.5.1 マネジメント体制

2018年度～2020年度は、プロジェクトを実施するための体制構築や実用化・事業化のもととなる研究開発を実施した。

2021年度～2022年度は、実用化・事業化をより意識した研究開発を実施しつつ、社会実装に向けた計画のブラッシュアップを実施する予定。その一環として、実用化・事業化に重きを置いたステージゲート審査を2020年度に実施し、2021年度以降の研究開発の実施の可否について判断を行う。

	2018年度	2019年度	2020年度
公募	▼ 4月 ▼ 9月		
サイトビジット	→	▼ 7-9月	
技術推進委員会	4月	▼ 11月 ▼ 2月	▼ 1月
追加公募			▼ 4月
評価		11月 ▼ 事後(先導研究)	中間 ▼ 12月
ステージゲート			▼ 12月

図 2.5.2 研究開発マネジメントスケジュール

2.6 実施体制

本事業は図 2.6.1 に示す実施体制にて推進している。

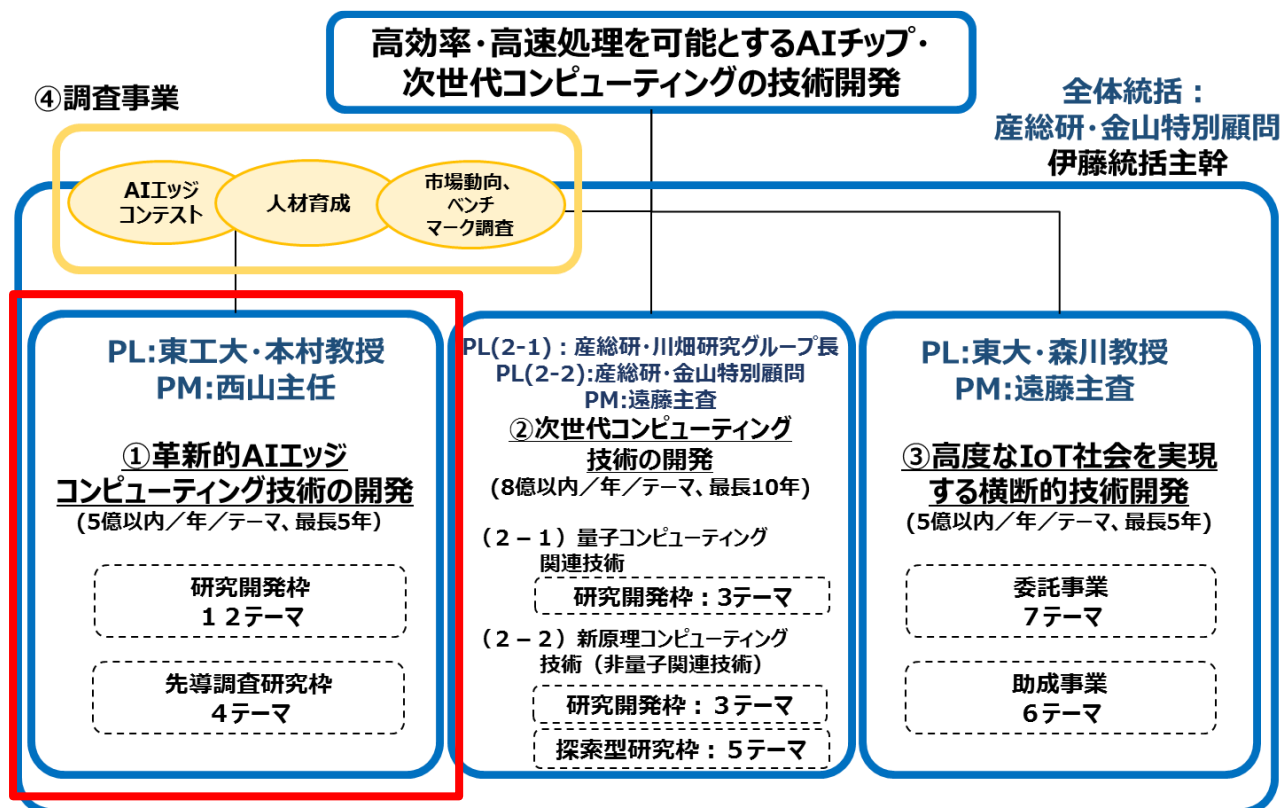
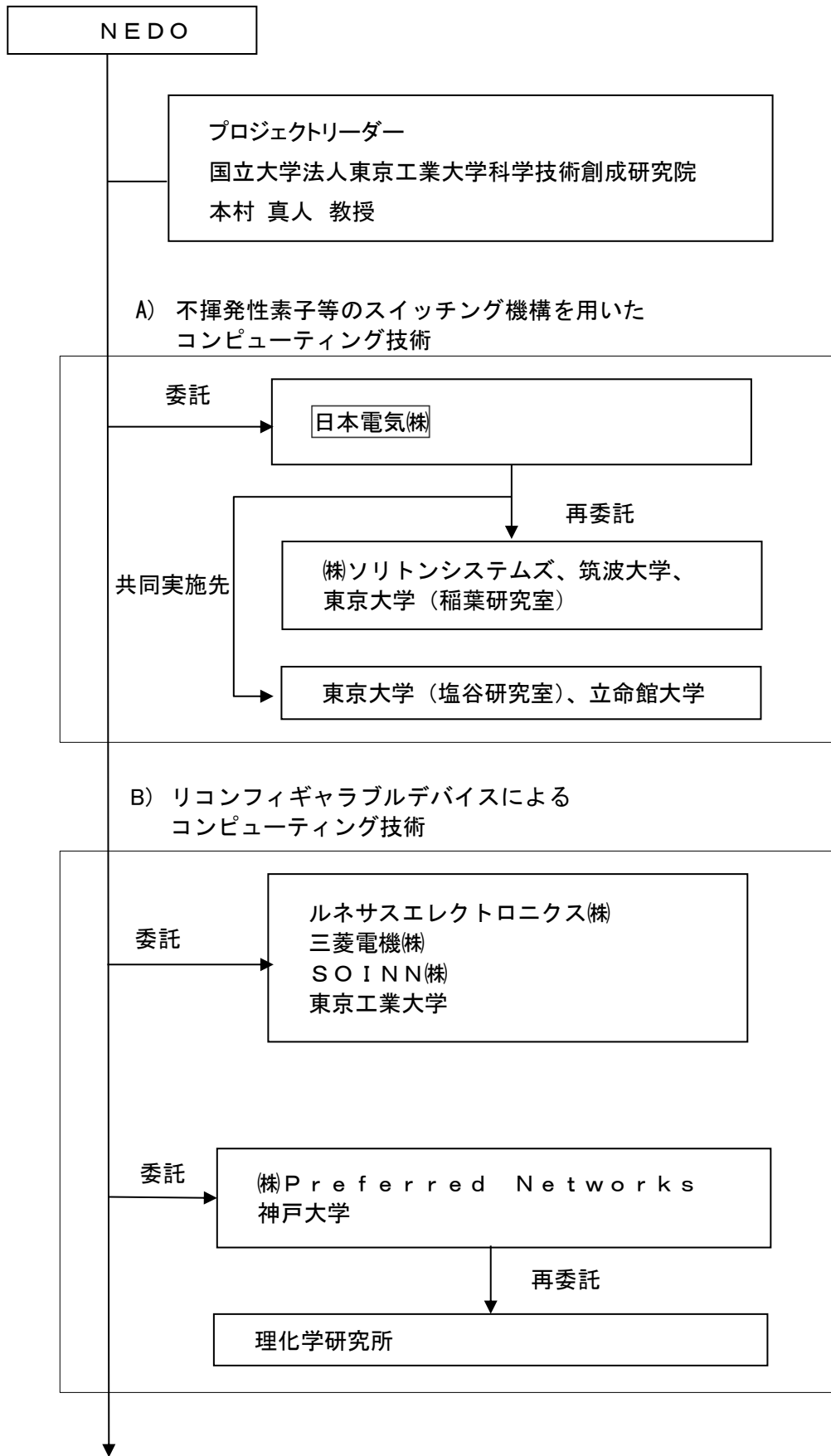


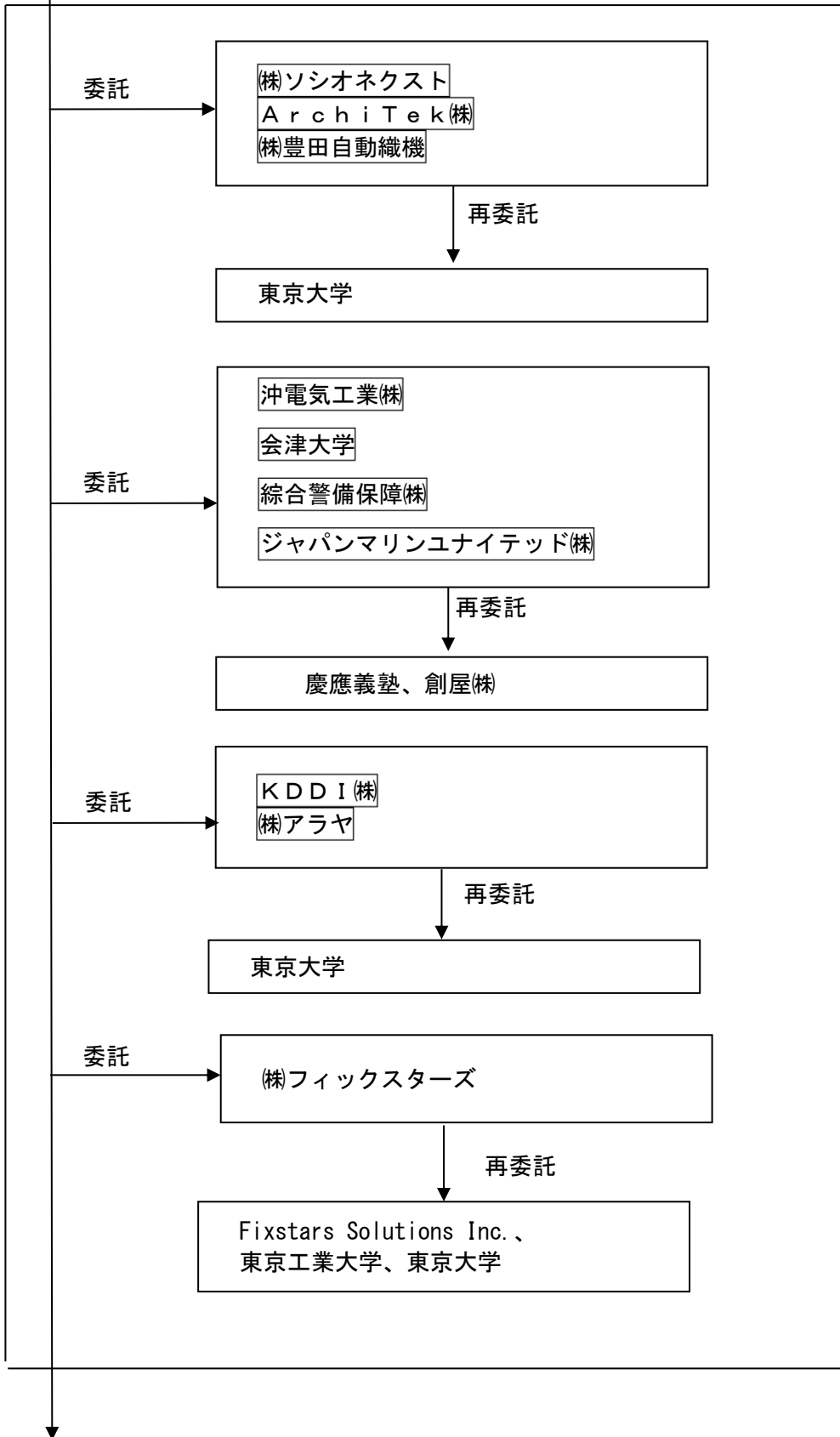
図 2.6.1 実施体制(2020年11月時点)

図 2.6.2 に各テーマにおける実施体制の詳細を示す。

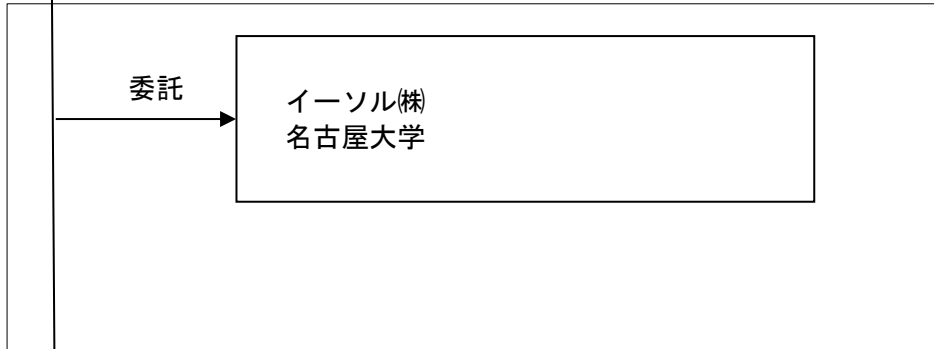
●研究開発項目① 革新的 AI エッジコンピューティング技術の開発



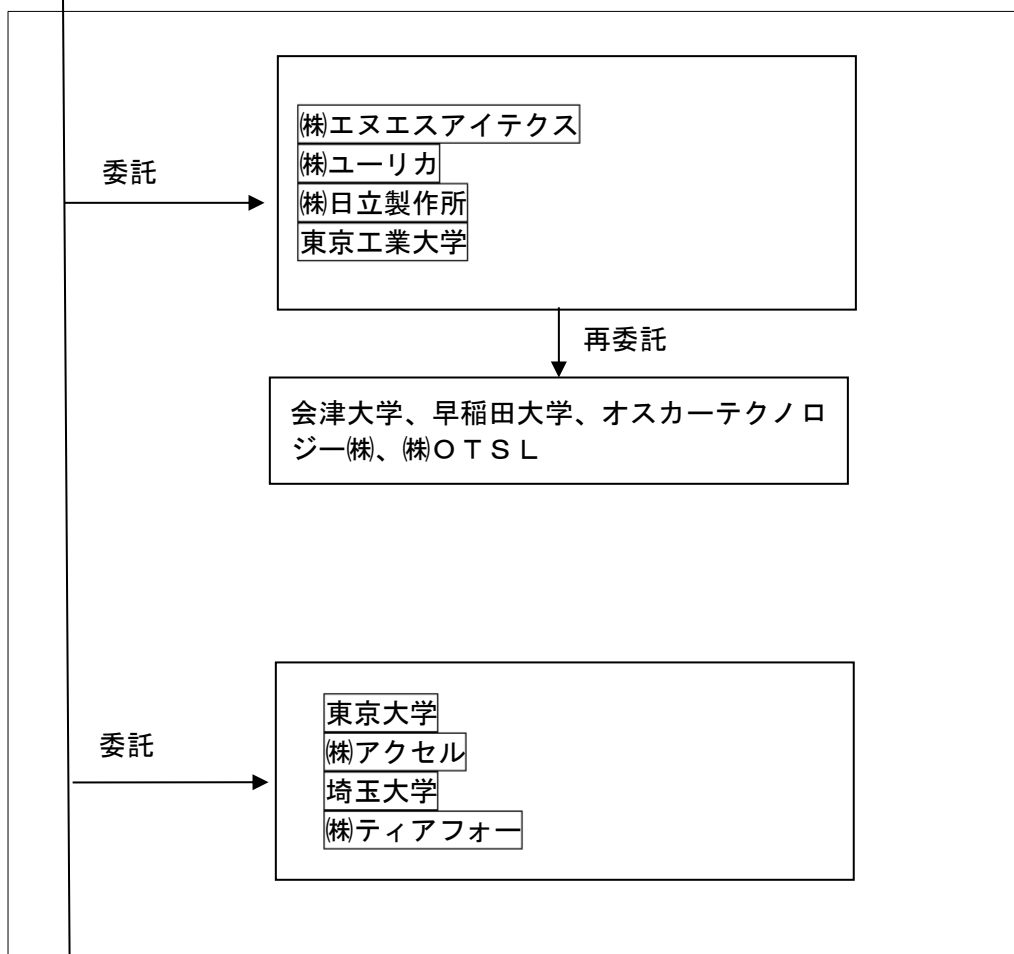
C) 演算処理量の軽量化を実現する AI 組込みコンピューティング技術



D) エッジコンピューティング向けリアルタイムソフトウェア制御技術



E) 多数の分岐ノードを有する AI アルゴリズム処理を高性能化するコンピューティング技術



F) エッジデバイスのセキュリティ技術及びその評価技術

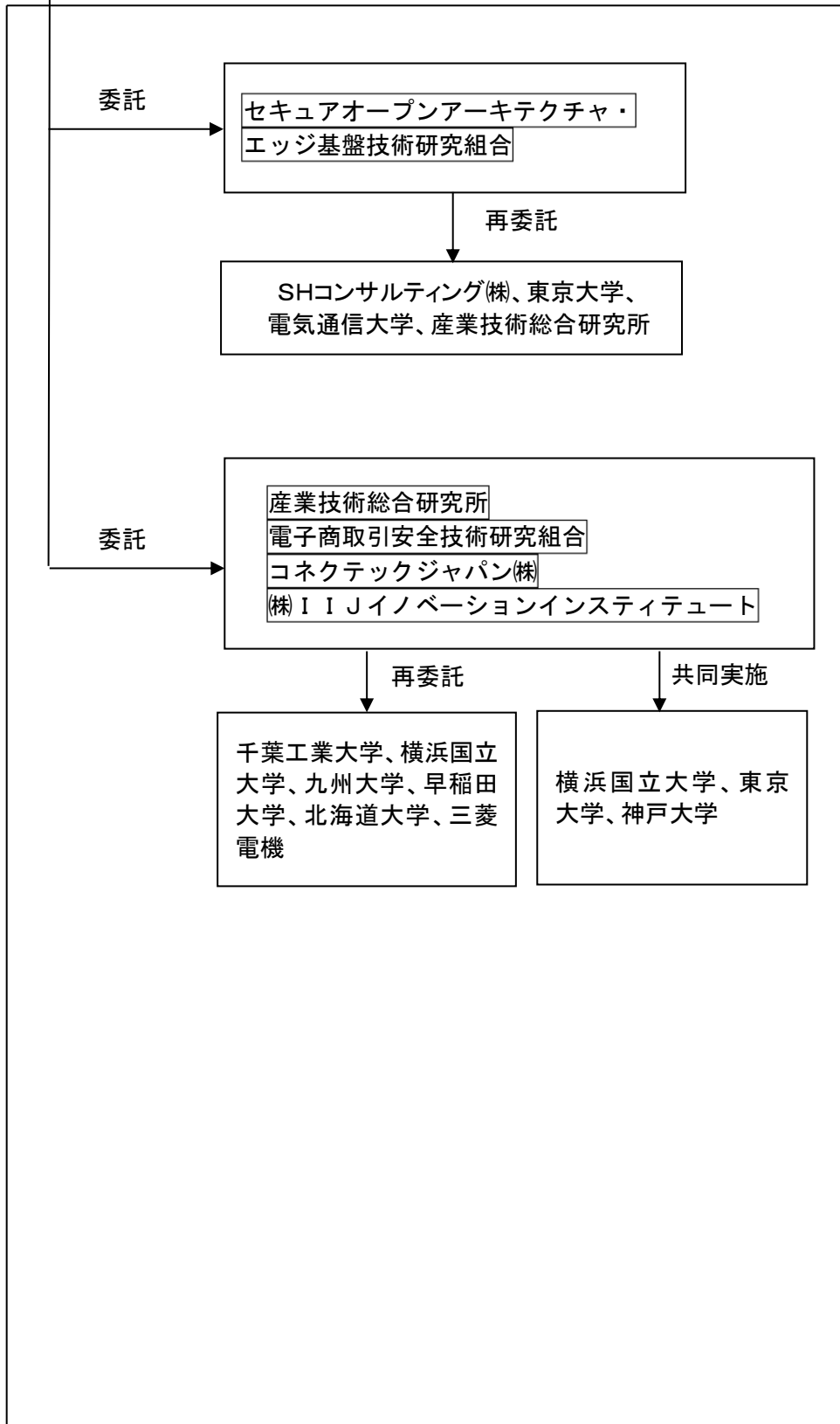


図 2.6.2 実施体制(2020 年時点)

2.7 動向・情勢の把握と対応

各研究テーマの個別動向・進捗把握に努め、ステージゲート審査委員会や技術推進委員会等の外部有識者で構成される有識者を活用し、適宜、研究開発テーマの加速・減額や拡大・縮小を実施した。

具体的な例として、2019 年度に実施した技術推進委員会において、RISC-V の課題について議論し、現行研究開発テーマの範囲に含まれていない、エコシステム形成のためのソフトウェアスタックについて、2020 年度に追加公募を実施した。(表 2.7)

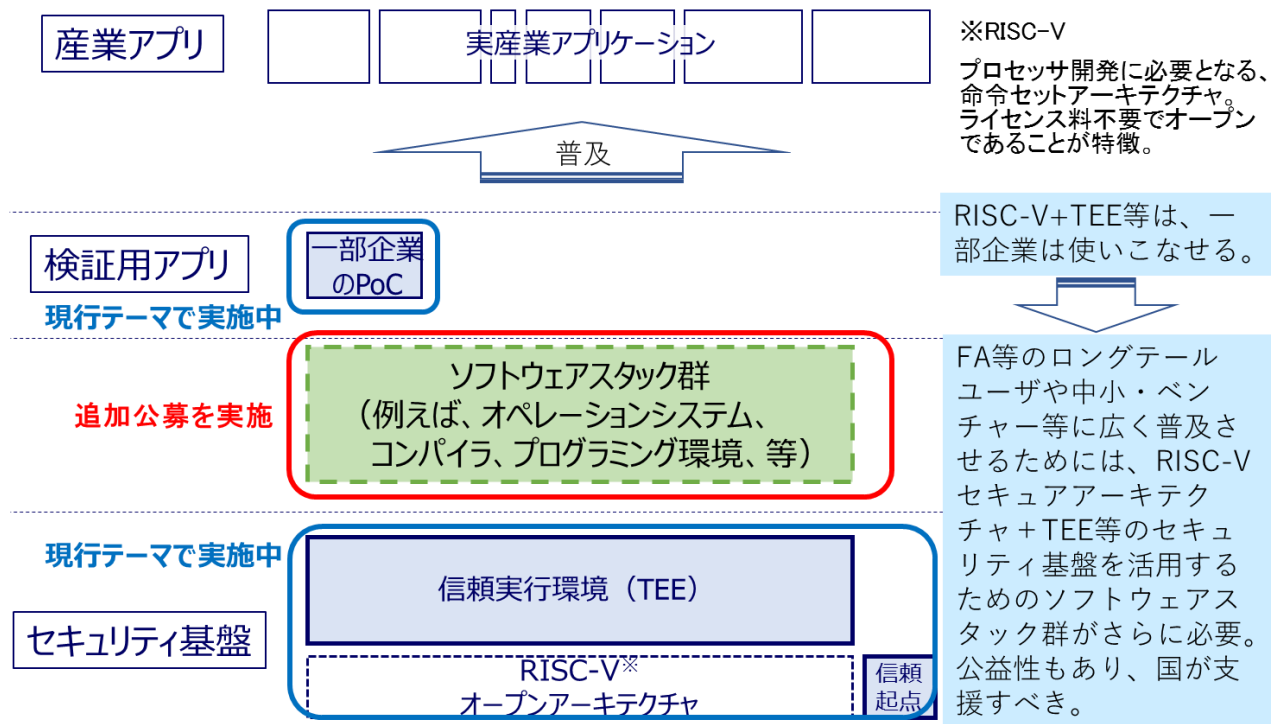


表 2.7 2020 年度追加公募の概要

3. 研究開発成果について

本研究開発項目では、従来チップに比べて電力効率10倍を目標としている。個別の研究開発テーマにおける詳細は以下に示す。

3.1 研究開発テーマ「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」

3.1.1 開発全体概要

本研究開発では、社会の隅々まで AI 技術を行き渡らせるため、高い処理能力と高い電力効率を備えたシステムチップ(原子スイッチ SoC または ASSoC)を開発する。電力消費が大きい AI 学習はクラウド上で行い、学習の結果得られたアルゴリズムを用いた AI 推論処理をシステムチップにおいて実行する。開発するシステムチップは、AI 推論処理を実行するための不揮発かつ高効率な原子スイッチ FPGA コアに加えて、CPU、メモリ、さらにはそれらを接続する原子スイッチプログラマブルバスから構成される。また、アプリケーションをマッピングするための設計ツールの開発も実施する。

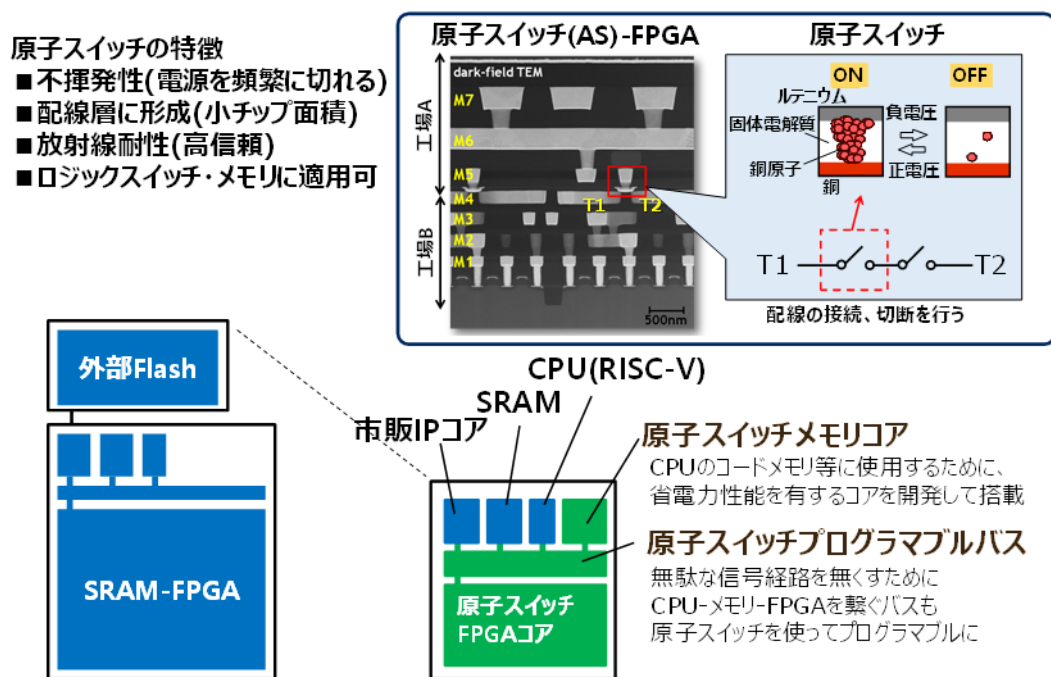


図 3.1.1 原子スイッチの適用イメージ

3.1.2 研究開発項目の開発内容、最終目標、根拠、達成度

AI 処理を行うロジック LSI の最も基本的な形態の一つは CPU である。CPU は書換え可能なソフトウェアによって目的に応じた様々な処理が実現できる。一方で、ソフトウェア上の命令を逐次的実行するため、リアルタイム性能、処理能力・電力効率が低い。一方、ASIC は処理手順を電子回路として全てハードウェア化したものであり、処理速度や電力効率が低いものの、固定の処理しかできない。そのため、日々発展している AI 処理への全面的な採用は難しい。

また、近年、AI 処理用のコンピュータに大量のデータの並列処理が可能な GPU が採用されている。GPU の場合、データ処理のスループットは上げられるが、逐次実行という基本的なコンピューティング技術は CPU と同じであるため、低遅延化(ハードリアルタイム化)への効果が限定的である。低遅延化の要求が強いアプリケーション領域への対応として、電子回路自体をプログラマブル可能な FPGA に、AI 処理を行わせる試みが進んでいる。FPGA は、比較的小規模な AI

処理の推論部のみであれば、逐次実行は導入せずに、細粒度並列処理を導入することで、低遅延で低消費電力な回路実装が可能だからである。しかし、現状の最先端 FPGA は、回路情報を記憶するために揮発性の SRAM を使った回路構成となっており、逐次実行のノイマン型プロセッサである CPU や GPU よりも電力効率は高いものの、数 10W 以上の電力を消費し、極めて消費電力制限の厳しい IoT 端末やエッジサーバには適用が難しかった。

本研究開発では、AI 処理を行う電子回路として不揮発かつ低消費電力である ASFPGA (Atom Switch FPGA) を用い、ASFPGA をシステム LSI に混載した ASSoC (Atom-Switch SoC) の開発を行う。ASSoC ではメモリア、プログラマブルバスにも原子スイッチを利用し、従来のシステムチップに比べて 10 倍の電力効率を目指す。原子スイッチを用いた低消費電力な FPGA とシステムの低消費エネルギー化(高効率化)を実現する以下の要素技術群の研究開発を実施する。

- (1) メモリア、FPGA コアの両者に原子スイッチを用いるための原子スイッチの製造技術
- (2) CPU 等のコア間を接続するプログラマブルバス回路技術
- (3) 頻りに電源を切るための混載 FPGA コア制御技術
- (4) 用途に合わせた多様な FPGA コアの仕様を実現する設計技術。
- (5) AI アルゴリズムを FPGA コアにマッピングする設計ツール

3.1.3 目標の達成度

学習型スマートコンセントレータとスマートコンセントレータから構成される学習型スマートセンシ開発期間は 2019 年より 2021 年であり、2020 年度末までに前節で述べた要素技術群の開発を行い、ASSoC に向けた原子スイッチの製造技術および回路設計を完了する。2021 年度においては ASSoC の製造を行い、様々なユースケースの応じた AI 推論アルゴリズムを含むアプリケーションを実装して、ASSoC の低電力性能を実証する。

これまでに得られた成果は下記の通りであり、進捗は計画通りである。

実施項目① AI 向け省エネ FPGA コア及び省エネメモリア開発(担当: 日本電気株式会社)

①-(a) 省エネルギー書換え原子スイッチ製造技術

技術ノード 28nm 向けの原子スイッチにおいては、スケールングによる回路動作電圧の低下に対応するため書換え電圧を低減する必要があり、また、書換え電圧の低減は省エネルギー化に繋がる。電圧ばらつきの小さいスプリット電極構造(サドンデス構造)を開発した。

①-(b) 28nm メモリア回路技術

実証テーマのアプリケーション動作に必要なメモリア容量およびアクセス時間の仕様を検討・決定した。

①-(c) ON 抵抗混在安定化原子スイッチ製造技術

技術ノード 28nm 向けの原子スイッチにおいて、固体電解質の膜厚と書き込み電流の相関についての評価を実施・定式化した。

実施項目② ASSoC の開発

実施項目②-1 端末-エッジ連携 AI 処理用 AI チップ設計(担当: 日本電気株式会社)

②-1-(a) 原子スイッチ AISoC チップの混載要素技術開発

ASSoC に用いる CPU として RISC-V の検討を行い、実装面積・電力評価を行った。

②-1-(b) ファブリックコンパイラ技術開発および統合設計
ファブリックコンパイラの全体仕様と、FPGA マクロの構成要素であるロジックセルの設計方法について検討を行った。

②-1-(c) 原子スイッチ AISoC チップのツール開発
AI アルゴリズム実装に向けた乗算器・メモリマクロ・組み合わせ論理回路を最適に分割・配置する自動パーティショニングおよびフロアプランニングツールの実装を完了した。

②-1-(d) 適用領域拡大のための共通化仕様の策定
自動車部品メーカーとの仕様を協議した。

②-2 端末-エッジ連携 AI 処理用 AI チップを用いたリアルタイム映像・音声処理の検討(担当:株式会社ソリトンシステムズ)

リアルタイム物体検出の AI アルゴリズムを検討し、市販の SRAM 型の SoC-FPGA(CPU 混載 FPGA)をターゲットにハードウェア設計を実施した。評価基板を製作し、でリアルタイム物体検出の AI アルゴリズムが動作することを確認した。

②-3 スマート農業を対象とした AI チップによる実時間映像解析の検討(担当:筑波大学)
スマート農業を対象とした AI チップによる実時間映像解析の検討として、FPGA および ASSoC への実装に適したアルゴリズム・精度に関する検討を行った。

②-4 介護介助ロボット向け AI チップの仕様策定と電源制御の検討(担当:東京大学)

(a)産業用ロボット向けトルク制御器オンラインパラメータ推定機能の実装評価

トルク制御関節制御のために、一軸試験機を作成し、関節トルク、角度、角速度などの学習用データを取得した。

(b)ロボット体内分散通信システム状態学習機能の AI 処理機能による実装評価

腱駆動ロボットと実施項目(d)において開発を行うロボット義足のコントローラの開発を行った。

(c)多用途柔軟センサ分布ユニットの組込センサ処理・通信機能の実装評価

FPGA のみで推論処理を行うことに関して開発を行い、ネットワークの量子化、実装を行った。

(d)介護介助ロボット向け AI チップ適用評価

パラメータ調整とユーザーの行動推論を行うための教師データを収集の準備を行った。

(e)介護介助ロボット向け AI チップ適用評価(省電力性評価)

実施項目(b)における低電力化制御の検討において高頻度の電源制御では低消費電力化の効果が低いため、義足のように着席時や直立時など待機状態の長いアプリケーションにおいては、待機状態時にモータの制御ブロックへの電源供給を停止することなどを検討した。

3.1.4 成果と意義

本研究開発よって、電力供給が限られた IoT 端末・エッジサーバにおいても高度な AI 処理を実行することが可能となり、社会システムの自動化や効率化が実現できる。認識精度の高い小型監視システムにより安心・安全な社会が実現できたり、農業の自動化が進むことで人手不足が解消されたり、あるいは介護介助ロボットによって介護人の負担が軽減されるようになる。

3.1.5 成果の普及

開発する ASSoC を、再委託先において映像音声配信端末、ドローン、ロボット義足に適用検討を行う。

AI 技術の進展により様々なタスクを自動的もしくは遠隔的に行う需要が高まっており、ここにおいて、ライブ中継の需要は、既存の放送分野・公共分野のみならず、多くの IoT 機器への実装されていくことが見込まれる。また、今後移行される 5G のネットワークにおいては基地局近傍にエッジサーバが設置され、無線通信機器からのアクセスを受け、AI を含むリアルタイムの信号処理を行うことが可能になる。その際、エッジサーバに対してデータを送信する側となる IoT 機器は、物体認識などの高負荷となる AI 信号処理そのものはエッジサーバに任せ、自らは複数画像の統合・切り出しや音声分離等、AI のための「前処理」をリアルタイムかつ低消費電力で行うことが求められる。IoT 機器側において ASSoC により前処理を行うことにより、IoT 機器－サーバー間の通信に必要となる帯域が削減され、低消費電力化が達成されるのみならず、通信のリアルタイム性、信頼性が向上することが見込まれる。

ドローンの社会実装を考えたとき、安定飛行や特定問題への最適化に加え、消費電力効率を大きく改善する必要がある。ASSoC の電力効率がシステムレベルでどれだけ有意に作用するかを検証を本研究開発で行う。軽量ドローンにおいては、ASSoC の導入により演算部の電力効率が 10 倍以上となることで、システムレベルでも電力効率の向上が期待できる。ロボット義足は、システムを低消費電力化することで長時間の使用が可能になり、また、過熱が防げることで断熱材の削減が可能になる。人間に装着可能なロボットは、将来的に介護分野にも広く適用できるため、社会課題の解決だけでなく、産業面、省エネ効果面でも大きな効果が見込める。

3.2 研究開発テーマ「動的再構成技術を活用した組み込み AI システムの研究開発」

本研究開発項目では、「研究開発開始当時入手可能な AI チップの 10 倍の電力性能比の動作見込を、要素技術のシミュレーションで実証」という 2020 年度目標に対し、DRP-AI テストチップ(TEG)試作による電力効率向上を達成し、期待通りの成果を上げつつある。主な成果を以下に示す。

- ・DRP-AI チップの演算性能引き上げ・電力性能向上を志向した 1 次および 1.5 次 TEG の設計を完了。1 次 TEG の評価では、電力効率の向上や実行効率の向上を実証した。(ルネサス)
- ・認識精度の劣化を柔軟に抑え、かつハードウェア特性も考慮し高速な推論を可能にする DNN 軽量化技術を開発。画像認識のサンプル DNN で約 90%の枝を削減。(三菱電機)
- ・シフト演算・ノイズ演算などの先進的単位演算方式の研究を進め、新アーキテクチャに基づくチップ設計試作を進めている。またスパース化等により ImageNet 向け DNN の 75%パラメータ削減も実現した。(東工大)
- ・超軽量の「競合学習機構」では、DRP 上での稼働を想定し大幅に軽量・高速化したアルゴリズム、及びその制御・予測系タスクへの拡張適用のための、長さの異なる時系列データの一括学習、再生機構を開発した。(SOINN)

個別の研究開発テーマにおける詳細は以下に示す。

3.2.1 研究開発サブテーマ「DRP アーキテクチャ、コンパイラの研究開発」

3.2.1.1概要

生産現場などにおいて、AI を実行するためには、特に産業の現場では、機器の認識や制御のリアルタイム性に加え、機器内の限られた場所に搭載するため、低電力化・小型化も同時に必要となる。本プロジェクトでは、(a) 低電力化に必要な DNN 圧縮手法に対する認識精度や実効的な電力効率を分析する環境を構築して最適な圧縮・DRP ハードウェア構成を探索して、(b) 実際に DRP アーキテクチャやコンパイラを開発して、(c) テストチップの設計・試作までを行い実証することを、1 次 TEG、1.5 次 TEG、2 次 TEG という 3 つのステップに分けて進める。

3.2.1.2最終目標と根拠

たとえばロボットなどのバッテリー駆動かつ発熱による電力制約の厳しい組み込み機器(数 W 程度)において、リアルタイムかつ高度な処理(高解像画像認識・判断・制御)を DNN (Deep Neural Network)をはじめとした AI 処理で行おうとした場合、必要性能が 10 TOPS (1 秒あたり 10 兆オペレーション)程度とすると、10 TOPS/W の電力効率が求められる。そこで本プロジェクトでは、動的再構成機能を持つ DRP (Dynamically Reconfigurable Processor)をベースとし、多様な AI 処理にフレキシブルに対応できる高い柔軟性を持ちつつ、従来技術比 10 倍以上(10 TOPS/W)の電力性能を有する組み込み向け AI チップの研究開発を行う。

2020 年度中間目標:1.5 次 TEG 評価により、推論の 10 TOPS/W 動作の見込みを実証する。

2022 年度最終目標:2 次 TEG 評価により、推論の 10 TOPS/W 動作を実証する。

3.2.1.3 目標の達成度

現在の主な内容の達成度を以下に示す。2020 年度末に向けて概ね順調に進めている。

項目	研究目標	成果	達成状況
1	1 次 TEG の検討、仕様策定、試作、および評価	1 次 TEG のコンセプト検討、仕様策定から試作を経て、評価ボード上での動作を確認。 ・コア数を 4 に増加させて絶対性能を向上 ・AI アクセラレータの量子化対応による処理の軽量化により低電力化 ・AI アクセラレータの性能を活かす肝となるオンチップバス幅および外部メモリ帯域を拡幅して積和演算器の実行効率を向上	○
2	DRP 向けのコンパイラの研究開発	1 次 TEG 向けにツールキットとしてまとめた。 ・浮動小数点演算に対応したコンパイラを開発 ・ニューラルネットワークのソフト処理レイヤーの実装、ニューラルネットワーク入力前後処理のための画像処理ライブラリを開発	○
3	1.5 次 TEG の検討、仕様策定、試作、および評価	1 次 TEG の設計や AI のシミュレーション結果をもとに電力削減方法を検討して、1.5 次 TEG 向けのアーキテクチャを策定。設計と検証を行い、TEG をテープアウトした。 ・基本演算精度を浮動小数点(FP16)から整数化 ・命令キャッシュ機構の導入による命令メモリの削減、およびメモリアクセス電力の削減	△ (2020 年度完了予定)

3.2.1.4 成果と意義

成果の概要は「目標の達成度」に記載。

3.2.1.5 成果の普及

本成果を普及させるためには、まずは本成果を使った AI アクセラレータをマイコンなどの半導体製品へ搭載する必要がある。実際の製品への搭載とその製品を顧客から選んでいただくために、更なる低電力化を進めて電力性能を高めると共に、ソリューション開発などアプリケーションとしての使い勝手の改善を進めていく予定である。

3.2.2 研究開発サブテーマ「DNN 単位演算回路を加速処理する先進的なアーキテクチャ・回路技術の研究開発」

3.2.2.1 概要

DNN を効率的に処理するためには、積和演算を筆頭とする DNN の単位演算(他には非線形演算、プーリング処理など)を高エネルギー効率で加速処理することが必要である。多数の積和演算器にデータをスムーズに流して加速する次元を超えてこの部分を将来的にも強化していくために、以下のような総合的な研究開発により、先進的な単位演算回路を実現していくことを目指している。

- ① 効率的な低ビット精度対応技術: 必要な認識精度に応じて必要な演算ビット精度は変わる。低ビット精度時に回路面積や電力消費の点で無駄が生じず、高ビット精度時に性能が必要以上に低下しないように、DNN 処理データフローの発展方向性を洞察し単位演算群の構成を検討する。
- ② メモリバンド幅削減技術: 係数メモリの中で(あるいは密着させて)単位演算を行うことで、メモリコアの外部読み出しバンド幅や消費電力を削減する手法を検討する。
- ③ 確率的動作技術: DNN 演算の中にノイズを積極的に取り込むことで、予測精度が上がったり、演算データ量を減らせたりすることが知られている。このような技術を積極的に単位演算に取り込み、演算効率向上を検討する。
- ④ 非行列乗算志向の単位演算器: 対数量子化により乗算を加算に置き換える、テーブル引きで乗算を排除する、あるいは行列を分解して単純な演算に置き換える等、DNN の処理内容に踏み込むことで単位演算器群を大幅に単純化する技術を検討する。
- ⑤ 数値表現の工夫による回路削減技術: 例えば、1 ビットの情報を 2 ビットで表現することで、加算の桁上げを排除して高速化する冗長二進という技術がある。このような数値表現の工夫と、乗算と加算をビットレベルに分解して再構築する工夫とを組み合わせることで、DNN の演算回路を軽量化できる可能性がある。このような可能性を検討する。

3.2.2.2 最終目標と根拠

最終目標としては、研究テーマ「DRP アーキテクチャ、コンパイラの研究開発」の最終目標である 10TOPS/W 実現に貢献するアーキテクチャ技術に関する助言、検討を行う。さらに、DRP アーキテクチャ・コンパイラの電力性能比の向上に貢献可能な、その時点での世界最高レベルの DNN 単位演算回路の高エネルギー効率化・加速技術の提案と効果実証を行う。

わかりやすい指標として、世界最高の LSI 技術の国際会議である ISSCC での発表を目標に置いて進めている。

3.2.2.3 目標の達成度

以下に説明する試作チップを集積回路系のトップレベル国際会議に投稿する準備を進めており、現時点の目標達成度は高い。

3.2.2.4 成果と意義

本テーマは、東京工業大学の本村のグループ(すずかけ台キャンパス)と、同じく東京工業大学の中原のグループ(大岡山キャンパス)で共同して進めている。

本村グループでは、本プロジェクト内で種々の検討を進めてきた単位演算方式のアイデアの中から、中原グループのアドバイスも得て、シフト演算とノイズ演算など、空間フィルタの畳み込みに頼らない新しい CNN(Convolutional Neural Network)演算方式とニアメモリ型・直積(デカルト積)型の演算器アレイ方式とを特徴とする DNN 推論加速アーキテクチャを 2019 年度後期に発案した。今年度前半には試作チップの評価を行い、有用性を実証できる見込みである。

中原グループでは、モデル軽量化手法であるシャッフル、シフト、スパース化の検討を行い、ResNet18 ベースのモデルを ImageNet で学習させ、85%の認識精度を達成した。また、Ultra96 FPGA ボードに推論回路の実装を行い、100MHz 動作で 20FPS 達成することを確認した。シフト演算とシャッフル演算を組み合わせるとスパース化を適用した S3-Net を開発し、ImageNet データセットを用いて学習を行い 75%パラメータを削減しつつ Top-1 認識率で 68.8%を達成した。タセットを用いて学習を行い 75%パラメータを削減しつつ Top-1 認識率で 68.8%を達成した。

3.2.2.5 成果の普及

本村グループの研究成果は、ルネサスエレクトロニクス社の 2 次 TEG やそれ以降の DRP 技術の AI 処理高効率化に貢献することを狙っている。中原グループの成果は、上記のように本村グループのアーキテクチャ検討における方向性を決定する基本的な知見を提供するとともに、ルネサスエレクトロニクス社の DRP コア上で実装する AI 処理自体の実効効率向上にも貢献している。このように、東工大における本プロジェクトの研究開発は、ルネサスエレクトロニクス社の技術発展と社会投入を通じた成果の普及を狙っている。

3.2.3 研究開発サブテーマ「軽量 DNN 変換機能の研究開発」

3.2.3.1 概要

AI チップで高い電力性能目標を実現するためには、アーキテクチャに閉じた最適化手法だけでは難しくなっている。そこで、本研究開発では、AI チップ向けアーキテクチャの開発に留まらず、AI チップの回路、アーキテクチャ、ニューラルネットワーク変換機能、ツールまでの垂直連携を重視した開発を行うことで、アーキテクチャ進化だけでは実現困難な大幅な電力削減・メモリ容量削減を実現する。本研究開発テーマはニューラルネットワーク変換機能に関するものである。

本研究開発テーマでは、認識精度の劣化を柔軟に抑え、AI チップ上でも高性能に動作可能な、DNN 軽量化技術を適用した DNN 変換機能を研究開発する。

i) 認識精度の劣化を柔軟に抑えることのできる DNN 軽量化技術

DNN 軽量化による認識精度の劣化を抑えるためには、DNN 軽量化による認識精度劣化を単純に小さくするだけでなく、ハードウェア制約の下でも認識精度劣化を抑えることが必要である。DNN 軽量化方法は、実装するハードウェア制約に応じて、大小様々に変更を加える必要が生じるものと考えられるが、そのような変更に対応して認識精度の劣化を抑えられる DNN 軽量化技術を適用した DNN 変換機能を開発する。

例えば従来の枝刈りでは、認識精度をいかに劣化させずにより多くの枝を刈ることに注目しており、ハードウェア制約が考慮されていないことが多い。例えば、演算器の数や演算器と記憶デバイスとの間のインタフェースに応じて、ニューラルネットワークの枝の密度にはハー

ドウェアに適した値があると考えられる。そのような制約を定式化し、学習アルゴリズムにおける損失関数に付与する等によって、学習処理の途上でハードウェア制約の考慮された枝刈りが決定されるようになる。このように枝刈りされたニューラルネットワークは、ハードウェア制約を考慮された上での学習処理が行われた結果であるため、実際にハードウェアに実装した際の識別精度の劣化を抑えることができる。

ii) ハードウェア上でも高性能に動作可能な DNN 軽量化技術

ハードウェア実装時の処理オーバーヘッドを削減するため、ハードウェア構成を踏まえた DNN 軽量化技術を適用した DNN 変換機能を開発する。軽量化を実装するための追加機能の処理量を出来る限り少なく抑え、ハードウェアで高効率に演算できる DNN 軽量化技術を DNN 変換機能に適用する。また、非効率な記憶デバイスへのアクセスを極力抑えるため、非効率アクセスの原因となる不規則性をニューラルネットワークの構造から除き、効率的に記憶デバイスにアクセスできる DNN 軽量化技術を DNN 変換機能に適用する。

例えば従来の量子化技術では、ビット幅削減演算やそのための記憶デバイスへのアクセスが考慮されていない。これらの処理を高効率に行うために、ハードウェアに搭載される演算器の種類や数、およびそれらを用いたときにどのような演算であれば高効率に処理できるかを見極め、オーバーヘッド処理を構成する必要がある。ハードウェアの特徴に適した DNN 軽量化技術を DNN 変換機能に適用し、高性能な動作を可能とする。

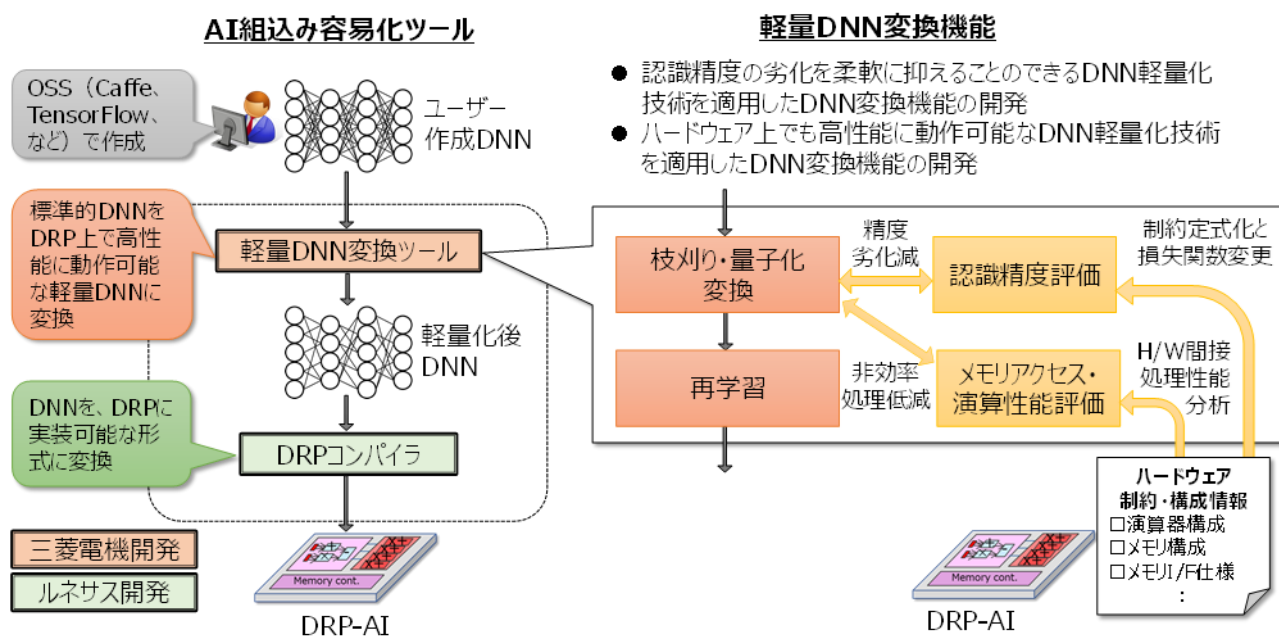


図 3.2.3.1 軽量 DNN 変換機能と位置付

DNN 変換機能の開発と試作ツール化により、AI チップ上での有効性検証と、機能改良を行う。AI チップ向けの AI 組み込み容易化ツールの一つとして構成し、推論時の精度維持と枝刈りによる演算量削減を行い、AI チップのハードウェアとの連携により電力効率向上を実現する。

3.2.3.2 最終目標と根拠

DNN 軽量化技術は、推論時の精度維持と枝刈りによる演算量削減はトレードオフの関係にあり、比較条件や制約条件によっても結果は異なってくる。そのため、以下の条件で最終目標を設定する。

- (a) 評価基準
認識率の精度劣化を 1%未満に維持しながら、行える枝刈り率を評価基準とする。ここで枝刈り率とは、DNN 全体の枝数に対する削減した枝数の割合とする。
- (b) 評価条件
画像処理で使用されることの多い複数の DNN 構造、およびデータセットで比較・評価を行う。
- (c) 比較対象
定量的な評価を行えるように、公開されているアルゴリズムでかつ H/W 実装を考慮した DNN の軽量化技術との比較を行う。
表 3.2.3.2 に研究開発の目標値を以下に示す。

表 3.2.3.2 研究開発目標

枝刈り率	
DNN 構造とデータセット	目標値
VGG16, CIFAR-10	88.5% ^[2] 以上
ResNet-56 CIFAR-10	13.7% ^[1] 以上
ResNet-110 CIFAR-10	32.4% ^[1] 以上
ResNet-164 CIFAR-10	35.6% ^[1] 以上
ResNet-164 CIFAR-100	29.7% ^[2] 以上

[1] H. Li, et al., “Pruning filters for efficient convnets,” ICLR 2017.

[2] Z. Liu, et al., “Learning efficient convolutional networks through network slimming,” ICCV 2017.

3.2.3.3 目標の達成度

研究開発目標の達成状況を以下に示す。

表中の達成度の「○」は目標達成を、「△」は目標達成見込みを、「×」は目標未達を示す。

表 3.2.3.3 研究開発の達成状況

DNN 構造とデータセット	枝刈り率		電力性能	達成度	今後の課題
	目標値	成果			
VGG16 CIFAR-10	88.5% ^[2] 以上	90.9%	10 倍見込	○	
ResNet-56 CIFAR-10	13.7% ^[1] 以上	83.3%	(評価中)	(評価中)	
ResNet-110 CIFAR-10	32.4% ^[1] 以上	89.9%			
ResNet-164 CIFAR-10	35.6% ^[1] 以上	82.7%			
ResNet-164 CIFAR-100	29.7% ^[2] 以上	83.6%			

3.2.3.4 成果と意義

本成果については、研究開発テーマ「実製品への AI 組込みを容易化するツールの研究開発」で開発中の統合ツール中に組み込み、普及を図ることを検討している。

3.2.4 研究開発サブテーマ「DNN によるエンドポイント学習用ソフトウェアの研究開発」

3.2.4.1 概要

本研究開発テーマでは、DNN の一部を学習しなおし、低演算量・小規模メモリで実現可能なエンドポイント学習機能向けソフトウェアを研究開発する。

エンドポイントデバイスに搭載可能な小容量の記憶デバイスで学習を行うためには、学習アルゴリズムの低演算量化や省メモリ化が必要となるが、通常ではクラウドやサーバで行われる処理であることを踏まえると、単純なアルゴリズムの改善だけでエンドポイント学習に耐えうる水準まで演算量や使用記憶領域量を削減することは困難である。

そのため、効率的に学習するための方式が必要であり、効率的に学習する従来学習方式として、例えば転移学習や強化学習などがある。これらは必ずしもエンドポイント学習を目的とした方式ではないが、上記課題を解決するための技術としての候補となる。

転移学習は、既存の学習データで学習したニューラルネットワークをまず用意しておいて、新たな学習データが得られたら、その新たな学習データに対してのみ学習処理を適用する手法である。学習済みのニューラルネットワークを用いることで、新たに学習するデータが少ない場合であっても、未学習のニューラルネットワークを最初から学習するのに比べて識別精度を高くできる。また、新たなデータを学習する際に、学習済みニューラルネットワークの一部の層を凍結し学習処理を施さなくても識別精度を高くできることが知られている。さらには一部の層を凍結した方が、識別精度が高くなることもある。しかしながら、学習アルゴリズムそのものは通常の学習アルゴリズムと同一であり、演算量や使用記憶領域量は新たな学習データの量や凍結する層の数に応じて削減できるのみとなり、推論と比べた場合には長い計算時間や大きなメモリ量を必要とするには変わりがない。

強化学習は、AI の出力に対するロボットなどの行動結果をフィードバックデータとして得て、それに基づき学習する方式である。逐次的に学習する点や AI 自身の出力に基づいて得られたデータを利用できることから、比較的少ない量のデータで学習でき、演算量や使用記憶領域量を削減することができる。しかしながら、高い精度の AI 出力を得るためには、過去に使用したデータも保持しておいて再度学習することが望ましく、使用データ量と精度はトレードオフの関係にある。また、ニューラルネットワークの学習アルゴリズムそのものは通常の学習アルゴリズムに付随機能を追加したものとなる。そのため、推論と比べた場合には長い計算時間や大きなメモリ量を必要とする。さらに、AI 自身の出力に基づいたフィードバックデータを得るための機能がソフトウェアとハードウェアの両方で必要となる。

本プロジェクトにおいては、従来の転移学習や強化学習などをそのまま適用したのでは難しいエンドポイント学習を実現するため、転移学習や強化学習などの候補技術をベースとした更なる効率化を図った学習方式を検討する。また、これらの技術をエンドポイントデバイスで実現するためには、デバイスが学習データを取得する仕組みや保存する仕組み、ニューラルネットワークの更新方法、新たな学習で得られたニューラルネットワークの妥当性を検証する環境とそれをデバイス上で実行するための方法などを含めたシステム開発が必要となる。

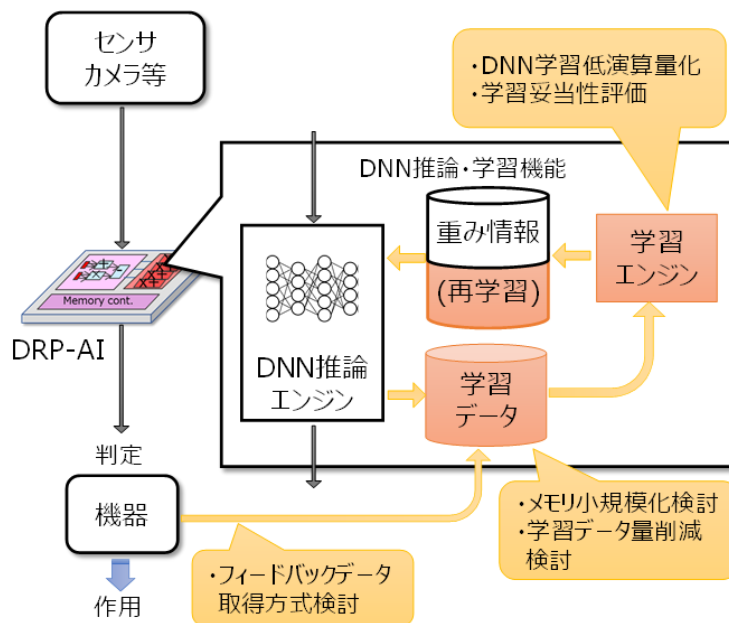


図 3.2.4.1 エンドポイント学習用ソフトウェア構成

3.2.4.2 最終目標と根拠

2018 年度～2020 年度はエンドポイント学習用ソフトウェアのコンセプト検討、必要なソフトウェアの要素技術検討・評価環境構築を行い、2020 年度末にベンチマークを行い、エンドポイント学習の価値判断と開発継続判断を行う。エンドポイント学習の実装価値があると判断された場合、2021 年度以降に試作ツールを開発し、エンドポイント学習が可能であることを実証する。

3.2.4.3 目標の達成度

研究開発目標の達成状況を以下に示す。

表中の達成度の「○」は目標達成を、「△」は目標達成見込みを、「×」は目標未達、「—」は評価対象外、を示す。

表 3.2.4.3 研究開発の達成状況

比較項目	本開発成果	比較対象	達成度	今後の課題
ターゲットアプリケーション	予兆検知 異常検知	異常検知	—	—
エンドポイント学習基本方式	深層学習 (+転移学習)	NN (+逐次学習)	—	—
学習時演算量 (削減率)	1/2		○	
学習データセット量 (削減率)	1/10		○	
学習データ取得方法	強化学習応用手法の 有効性を確認		○	
ニューラルネットワークの更新方法	転移学習手法をベー スに検討・評価中		△ (2020 年 度可否 判断)	枝刈りされた NN からの更新方法 を検討する

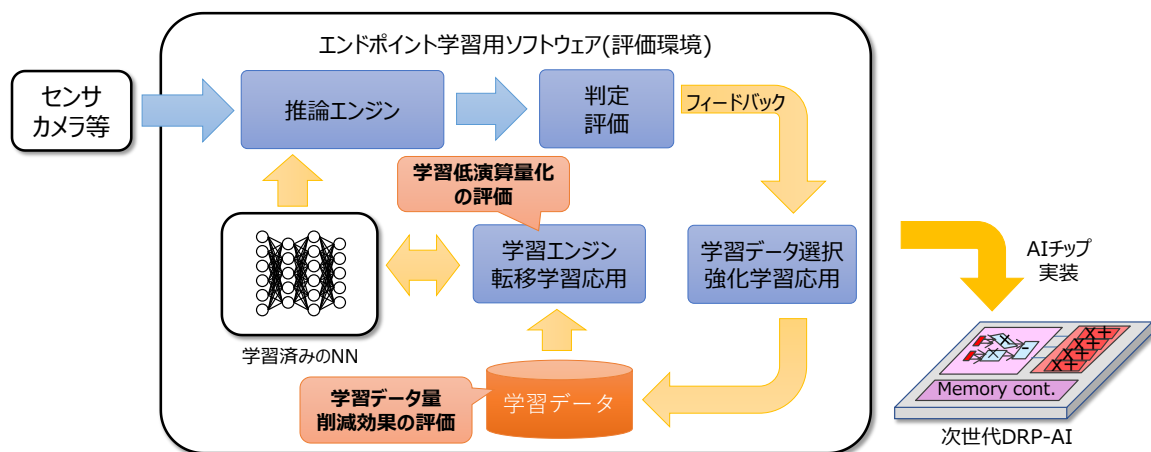


図 3.2.4.3 エンドポイント学習用ソフトウェア評価環境

3. 2. 4. 4成果と意義

AI チップ事業適用時期に大きな市場規模が見込まれる分野の一つである予防保全・故障予兆検知をターゲットシステムとしとして、エンドポイントで波形データからの異常検知アプリケーションを実現する基本システムの評価環境を構築した。

評価環境には、DNN 推論機能、フィードバックデータ取得機能、およびエンドポイント学習機能を搭載した。またエンドポイント学習機能の要素技術として、学習処理の低演算量化を目的とした転移学習応用技術と、学習データ量削減を目的とした、強化学習のデータ選択手法の応用技術を実装し評価した。

転移学習応用では、事前学習した DNN モデルの一部レイヤーを凍結することで、転移学習を適用しなかった場合に比べて、学習に要する時間を 1/2 に削減できることを確認した。また強化学習のデータ選択手法を応用し、エンドポイントで得られるデータの推論結果に基づき、誤差が大きいものを優先的にサンプリングして学習データとすることで、エンドポイントで使用する学習データ 1/10 で、異常検知の精度劣化がほぼないことを確認した。これにより今回選定した波形データからの異常検知アプリケーションにおいて、検討した要素技術の適用によりエンドポイント学習での演算量と学習データ量の削減効果を確認した。

3. 2. 4. 5成果の普及

学習に関する継続検討の可否判定を行った後、成果の普及を検討する。基本的にはルネサスのハードウェアへの搭載による普及を図る。

3.2.5 研究開発サブテーマ「DNNによるエンドポイント学習用ハードウェアの研究開発」

3.2.5.1概要

生産現場にAIが普及するためには、タスクの変化や動作環境の変化などにも、機器が自律的に対応できることが求められる。しかしながら、現状ほぼ全てのエンドポイント向け AIハードは推論専用であり、自律的な学習には適していない。そこで、本プロジェクトでは、現場の環境やタスクの変化にも自律的に対応可能にするエンドポイント学習システムの開発を目指す。

3.2.5.2最終目標と根拠

エンドポイント学習システムに関しては、現状ほぼ未開拓の領域であるため、2018年度～2020年度における開発としては、エンドポイント学習の価値判断から開始する。具体的には、④DNNによるエンドポイント学習用ソフトウェアの研究開発を担当する三菱電機と連携しながら、必要な機能の見極め、理論やシミュレーションによる検証を行い、エンドポイント学習の価値の有無を判断することを中間目標とする。

2022年度最終目標:2020年度の結果で価値があると判断された場合には、TEG(2021年度以降に予定している2次TEG、あるいは現在開発中の1次/1.5次TEG)でのオンチップ学習の効果を確認する。

3.2.5.3目標の達成度

現在の主な内容の達成度を以下に示す。2020年度末に向けて概ね順調に進めている。

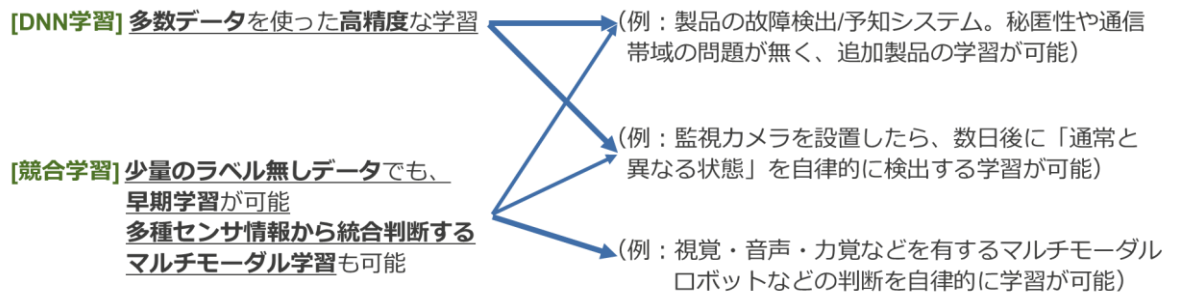
表 3.2.5.3 研究開発の達成状況(2)

項目	研究目標	成果	達成状況
1	エンドポイント学習のコンセプト、ターゲットシステムの検討	初期検討では、センサデータに基づく機器の異常検知をターゲットにした。環境変化にロバストなカメラなど、より高次元データにも拡張を検討中。	○
2	エンドポイントDNN学習ハードの基本構成検討	再学習の実行はCPUあるいはDRPでのバックグラウンド実行が適しているという仮説構築	○
3	エンドポイントDNN学習のハードウェア性能見積(理論・シミュレーション)	Back-propagation 演算の実機および簡易シミュレーションによる見積を実施。Full-connection層では、データ収集時間以下で可能な見込み。DRPの場合は1桁程度の高速化の可能性が得られた。	△ (2020年度中に実施予定)
4	エンドポイント学習の価値の有無を判断	「DNNによるエンドポイント学習用ソフトウェアの研究開発」でのシステム検討結果と3.の性能評価を合わせて判断予定。	△ (2020年度中に実施予定)

3.2.5.4 成果と意義

(a) オンチップ学習のコンセプト、ターゲットシステムの検討

主な用途と適用可能性については、関連する研究開発テーマ「④DNN によるエンドポイント学習用ソフトウェアの研究開発(三菱電機)」 「⑥競合学習機構による汎用・超軽量エンドポイント学習技術の開発(SOINN)」での内容やAI技術動向、ユーザ需要などの情報から用途の抽出を行った。



【主なエンドポイント学習の価値】

- 通信機能の有無によらず、機器ごとに異なる設置場所の環境やセンサの特性変動などにロバストな機能を提供可能。
- 特に画像認識のクラウド学習の場合、膨大なデータ通信コストやクラウド処理コストに加え、通信前に機器にデータを貯蔵する仕組みによるコスト増が問題となり、その解決に貢献可能。

(b) 2. ハードの基本構成の検討

各用途のシステムを想定した際のハードウェアリソースの制約ならびに学習速度の制約の観点から、適切な学習ハードの構成および性能要求を検討した。結果として再学習の実行はAI-MACよりもCPUあるいはDRPでのバックグラウンド実行が適していると仮説を立てた。

次に、仮説検証の一つとして、ハードウェアでの実現性を見るため、性能見積もりを実施した。

(c) 3. エンドポイントDNN学習のハードウェア性能見積(理論・シミュレーション)

エンドポイント学習の価値や実現可能性判断に必要な要素となる、エンドポイント学習のボトルネック抽出およびハードウェア実装時の実行時間の見積等を行った。特にすべてのDNNベースのオンチップ学習システムで共通して必要となる、再学習時の誤差逆伝搬(back-propagation)演算がボトルネックになると考えられるため、その性能評価から優先して実施した。

対象とするネットワークとして、研究開発テーマ「④DNNによるエンドポイント学習用ソフトウェアの研究開発」において、最初のターゲットを「1次元データを使った故障検出」を「一部のレイヤのみ再学習する追加学習」を想定している。したがって、まずはfull-connection構成かつ最終層1層の再学習に必要な演算時間の検討から行った。

具体的には、DRPに適したデータフローの検討、誤差逆伝搬ベンチマークプログラムの作成、実機評価(CPU)および DRP 設計環境を用いた簡易シミュレーション(DRP)によるベンチマークプログラムの試行実験を実施した。

(d) エンドポイント学習価値の判断に向けた、今後の取り組み

(c)の結果から、DRPの方が高速化・低電力化を期待できるが、価値判断の検証段階であれば CPU/DRP いずれも用いても可能な範囲であるため、まずは CPU を用いてオンチップ学習の実現性・有効性を評価する予定である。また、オンチップ学習の需要・必要なシステムについては、研究開発テーマ「④DNN によるエンドポイント学習用ソフトウェアの研究開発」と連携しながら検討しており、有用な結果・用途が得られた場合、2021 年度以降で DRP へのオフロードによる高速・低電力化の可能性を検討する方針である。

3.2.5.5成果の普及

特にオンチップ学習技術は、ユーザ環境で安定して使えるのかの検証が十分なされていない技術領域であり、かつ動作精度の定量的評価も難しいことなども、普及に向けた課題となる。オンチップ学習の効果ありと判断された段階で、実証実験および普及に向けた技術課題(評価方法の構築など)の解決を進めていく予定である。

3.2.6 研究開発サブテーマ「競合学習機構による汎用・超軽量エンドポイント学習技術の開発」

3.2.6.1 概要

現在主流でDNNにも用いられる誤差逆伝搬学習方式は、多大な演算量やメモリ量を要するだけでなく、教師ラベルつき学習用データが大量に必要ななど、実用面で課題が多い。これに対し、競合学習方式をベースとし SOINN 社が特許を有する機械学習アルゴリズムとしての SOINN は、DNN に対し多数の優位性を持ち、現在も機能・性能の両面で進歩を続けている。

例えば、SOINN は基本的に教師なし学習手法であるため、教師データがゼロの状態からでも学習を開始できる。また SOINN はオンライン学習機能を有し、演算が非常に軽量で、必要となるメモリ量も DNN に比較して圧倒的に少ない。さらに学習済みの SOINN を類似の他タスクに転用する転移学習が可能である。これらの機能は、生産ラインの現場などでの AI の有効活用のためには必須といえる。実際、これまでに SOINN 社が受注したシステム開発案件は、ほぼすべて、そうした機能の積極活用が指定されている。本研究開発は、SOINN 本来の機能や性能の改善を図るとともに、SOINN を DRP 上で稼働させることを目的とする。

3.2.6.2 最終目標と根拠

SOINN を DRP 上で稼働できれば、安価で省電力、省演算、省スペースという超小型 AI の大量生産が可能となる。本研究開発の最終目標は、そうした極めて使い勝手の良い“SOINN on DRP”を生産現場や社会の隅々にまで届け、SOINN on DRP にあらゆる機器や装置の制御、多様な現象の予測や異常検知、さらには分類・識別などの機能を担わせることにある。SOINN と DRP は夫々他にない際立った特長を有しており、双方のシナジーにより高い競争力を有する SOINN on DRP の実現は、十分根拠があると考えられる。

3.2.6.3 目標の達成度

以下に、現在の達成度を示す。2020 年度末に向けて順調に進行している。

項目	目標	成果	達成状況
⑥	SOINN 社独自の競合学習をベースとした新たなオンチップ逐次学習の実装アルゴリズムの検討、シミュレーションによる有効性評価	1. DRP 上での稼働を想定し大幅に軽量・高速化したアルゴリズムを開発。 2. 上記を制御・予測系タスクに拡張適用するための、長さの異なる時系列データの一括学習、再生機構を開発。	◎ (ここまでの重要成果物として知財3件出願。更なる成果を2021年2月追加申請予定)

これまでに、改良 SOINN について3件の特許申請を進めている(12月申請予定)ほか、改良 SOINN を DRP で稼働させた場合の優位性の定量評価を進めている。

さらに、研究成果の極めて有効な適用先の一つであるロボットを対象に、ロボットアームの3次元の挙動を改良 SOINN 搭載の DRP ボードで学習させ、実際にロボットを稼働させるためのアルゴリズム開発と検証を進めている。ロボットの制御には、理想的には2msec以下の高速な制御サイクルが要求されるため、そのクリアも課題である。現状の DRP ボードでは、通信速度の制約上そのサイクルでの通信は困難が予想されるが、SOINN on DRP の自体の推論速度は2msec以下が期待できる。

3.2.6.4 成果と意義

近年、IoT デバイスやセンサが普及し、日々莫大なデータが収集されているが、殆どのデータは記録として保存されるにとどまっている。すべてのデータをそのまま記録・保存するのは極めて冗長で高コストである。本来データは、分類整理され、必要な時に、タイムリーに活用できて初めて意味をなす。そこで、機器や装置に SOINN on DRP を埋め込み、ネットワークやセンサの情報からデータを溜めずに直接自己学習ができれば、個々の機器や装置が現場で自ら学習して機能することとなり、その恩恵は極めて大きい。

少子高齢化が進む我が国では、生産労働人口の不足が深刻である。将来 SOINN on DRP は、産業用ロボットなどの生産現場向けのみならず、現場ごとに臨機応変な対応が必要で、これまで省力化が困難であった流通業、農業、水産業、医療介護などの分野でも有効に活用できると考える。すなわち、本研究成果は幅広い分野・業種の生産性向上に大きく寄与し、新産業の創出にもつながると考える。

3.2.6.5 成果の普及

本研究開発とは別に、数年前より SOINN 社と川崎重工・精密機械・ロボットカンパニー社との間で旧 SOINN アルゴリズムを活用した共同研究開発が進んでいる。2019 年国際ロボット展では、川崎重工社の世界初の職人技を現場で学習してロボットに再現させる“Successor”に旧 SOINN アルゴリズムが試験採用され、デモ展示された。

しかしこの展示は、まだ可能性が示されたレベルであり、実用レベルに至るには克服すべき課題が多数あることも指摘された。本研究の成果物である「改良」SOINN on DRP では、それらの課題が解決できると考える。このほかにも、本研究の成果は、これまでに SOINN 社に頂いている、ドローンや建設用重機の自動運転、大規模プラントの多地点異常検知、家電製品への適用といった分野にも幅広く極めて有効に活用できると考える。

上記に加え、SOINN on DRP を画像処理と組み合わせた、画像検査モジュール群の研究開発も進めている。まず 2020 年初頭より、CPU 稼働で旧 SOINN アルゴリズムを活用したモジュール群を“SWITCH VISION”の名称(商標登録申請済)で販売したところ、建設業、食品業、技術商社などから多くの引き合いを得ている。画像検査系の業務は多大なニーズがあり、かつ DNN にはない当社の強みを活かせる分野であり、こちらも本研究成果の大きな普及を見込んでいる。

下図はその一例である。左は不良品を含まない乾燥食品の実サンプル、右はベルトコンベアを流れる不良品を含むサンプルの判定結果例である。右で濃い黄色の箇所ほど、不良品の可能性が高い。こうした検査システムを SOINN on DRP 化できれば、生産ラインの要所に多数配置し、一層の省人化や不良品の早期除去、不良品が混入するリスクの大幅低減が図れる。

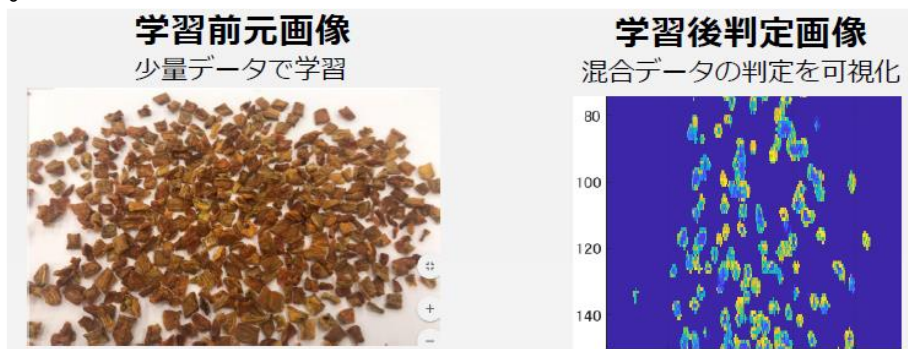


図 3.2.6.5 不良品を含まない乾燥食品サンプル(左)と判定結果例(右)

3.2.7 研究開発サブテーマ「実製品への AI 組み込みを容易化するツールの研究開発」

3.2.7.1 概要

エンドポイント AI の省フットプリント化および学習技術とそれらを組み込むハードウェアである動的再構成可能な AI チップは密接な連携関係にある。そこで、本プロジェクトで構築した基盤技術が相互に連携し、ソフトウェアからハードウェアまで各レイヤ間を跨る最適化および拡張性を持たせた統合ツールを開発する。特に、実用フェーズを想定すると、主なユーザは、AI 技術者から組み込み SW 開発者になると想定。AI 技術者のような複雑な設定が無くとも自律的に AI をハードに実装できることを重視した。

開発中のツールは、主に以下の2つである。

- (1) AI フレームワークで記載されたモデルやパラメータ情報を、ハードウェア(AI アクセラレータ)情報に変換するための「AIトランスレータ」
- (2) 圧縮プログラムの活用を、組み込みユーザ向けに容易化する「AI 統合ツール」

3.2.7.2 最終目標と根拠

年度ごとの目標は以下の通りである。

- ・2018 年度： ツールコンセプト、ツール間インタフェース仕様の策定
 - ・2019 年度： 機能仕様策定、基本ツールの開発
 - ・2020 年度： 基本ツールの評価、1.5 次 TEG 向けツール改良
- また、3年目/5年目の目標は、それぞれ以下である。

2020 年度中間目標： 評価結果から要改良機能のフィードバック

2022 年度最終目標： 各研究開発項目で開発した技術を統合したツールの完成、
実施項目 1 で開発した 2 次 TEG による動作実証

ユーザが AI ハードを使いこなすためには、AI モデルを適切に変換・圧縮・ハード実装可能なツールの充実が不可欠である。したがって、最終目標としてはツールの完成に加え、AI チップでの動作実証まで接続することを目指すこととした。

3.2.7.3 目標の達成度

現在の主な内容の達成度を以下に示す。2020 年度末に向けて概ね順調に進めている。

[トランスレータ]

項目	研究目標	成果	達成状況
1	トランスレータの仕様、インターフェースの検討	各種の AI フレームワークと AI アクセラレータや DRP を繋ぐための形式として ONNX を選定。	○
2	トランスレータのプロトタイプの開発と基本動作の実証	上記方針を基に、①の 1 次 TEG 向けのプロトタイプを一旦まとめて下記の機能を含めたツールとして統合。 <ul style="list-style-type: none"> ・FP32 から FP16 への変換 ・サイクル数推定機能 下記のニューラルネットワークモデルで変換フローが動作することを確認。 <ul style="list-style-type: none"> ・一般物体認識の TinyYOLOv2 ・クラス分類の Resnet50 	○
3	トランスレータの改良	①の 1.5 次 TEG の仕様に合わせたツールの改良方針を検討中。	△ (2020 年度完了予定)

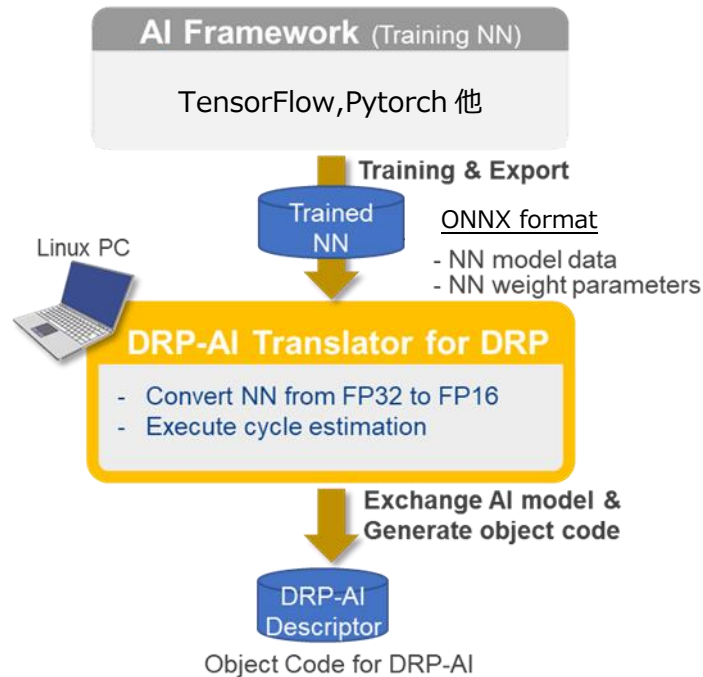
[統合ツール]

項目	研究目標	成果	達成状況
1	統合ツールのコンセプト、インターフェース仕様の検討	組み込み SW 開発者を想定し、複雑な設定パラメータ設定を不要にする方式や、性能予測の見える化などの技術導入方針を選定。	○
2	統合ツールの機能仕様の策定・設計	上記方針を基に、ツールの機能仕様と開発を完了。 <ul style="list-style-type: none"> ・ユーザモデルを統合ツールで使用するための、モデル変換インターフェース設計 ・AI 専門家でなくても圧縮技術の手順や効果の認識が容易で、かつパラメータ最適化などをサポートするツールなどを導入。 ・要求性能などに応じて、複数の圧縮ツールを接続/選択できる機能 	○
3	統合ツールのツールチェーンの基本動作の実証	<ul style="list-style-type: none"> ・1.5 次 TEG 対応として三菱電機製圧縮モジュールも含めて統合ツールの基本動作を確認。 ・動作確認結果からツール GUI や内部ツールが使用するモデル情報 (ONNX) の対応課題を発見。GUI への対策を実施するとともに、内部ツール間での対応状況の整合を検討中。 	△ 今後、内部ツール改版や ONNX 対応部分のアップデートなどを開発

3.2.7.4成果と意義

[トランスレータ]

2018年度は、各種のAIフレームワークとAIアクセラレータやDRPをONNX形式で繋ぐための変換ツールであるAIトランスレータの仕様を作成し、一部の試行実装を開始した。



2019年度はこの仕様に従って①の1次TEG向けのプロトタイプを一旦まとめてツールとして統合した。1コア向けの一般物体認識のTinyYOLOv2と、クラス分類のResnet50に対応した変換フローが動作することを確認した。

2020年度は①の1.5次TEGに向けてツールを改良する。

[統合ツール]

2018年度はコンセプトおよびツール間インターフェース仕様を策定した。

2019年度はこのコンセプトに沿って機能仕様の策定を行い、基本ツールGUIとして以下のような機能を開発した。1)「High/Mid/Low」といったシンプルな圧縮率の設定 2)圧縮率/精度/サイズ/電力効率等を比較可能なGUI 3)複数の圧縮設定を同時に実行する機能

GUIとしてはユーザの使用手順を考慮し、手順に沿ったGUIレイアウトとなるよう開発を行った。

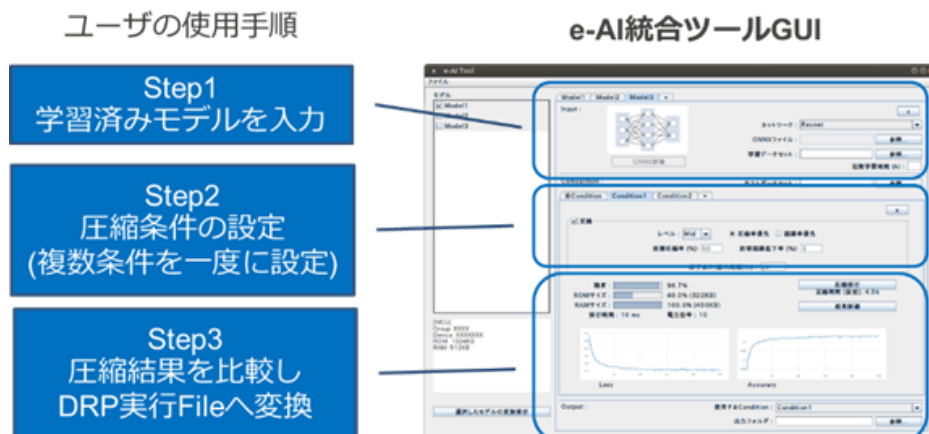


図 3.2.7.4 統合ツールの使用手順

2020 年度は 1.5 次 TEG 用ツール改良として、研究開発テーマ「軽量 DNN 変換機能の研究開発」で開発中の圧縮モジュールとの連結部分を開発。加えて基本的なツール動作の評価を実施した。動作確認結果からツール GUI や内部ツールが使用するモデル情報 (ONNX) の対応に課題が見えており、GUI への対策を実施するとともに、内部ツール間での対応状況の整合を検討中。

また引き続き、圧縮モジュールや DRP-AI トランスレータの改定に対応し、ツールとしての完成度を高める。

3.2.7.5 成果の普及

開発したツール (AI translator, 統合ツール) に関しては、本研究開発の AI チップだけでなく、ルネサスの DRP-AI を搭載した現行製品群にも展開可能な技術である。成果の普及として、ツール部分を切り出し、早期展開も検討している。

3.3 研究開発テーマ「FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発」

3.3.1 概要

本研究開発、「FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発」の目的は、我々(株式会社 Preferred Networks(以下 PFN)・神戸大学・理化学研究所(以下理研))が開発を進めてきた可変精度演算コアと、最近急速に発展したカスタム LSI (ASIC) 用 FPGA IP を組合せることで、専用演算回路の高い電力性能と FPGA の柔軟性をあわせもつ、エッジ AI 向けのチップを開発し、それを使うことでまだ高コストであり消費電力も大きいロボティクス応用の飛躍的普及を実現することである。

可変精度演算コアでは、INT8相当から FP64 までの広い範囲で、消費電力、チップ面積を抑えつつ、演算能力が語長の 2 乗に反比例する回路を実現する。これにより、INT8 相当では FP64 の 64 倍の電力性能を実現し、7nm を使った場合で 10 Tops/W、5nm では 20Tops/W 程度を達成する。さらに、これらの演算コアを固定されたチップ内ネットワークではなく、最近急速に進歩し、最適な性能・規模の回路を実現可能になった FPGA IP で結合する。再構成可能な FPGA IP で演算コア以外の回路を構成することで、専用回路では困難なネットワーク圧縮等への対応を可能にするだけでなく、通常ならば CPU コアによるソフトウェアで実現される部分を FPGA IP によるハードウェアに置き換えることで画期的な高速化、特に低遅延化を実現し、エッジでのリアルタイム応用を可能にする。これにより、ロボティクスを始めとするあらゆるエッジでの深層学習応用で、消費電力を大幅に削減するだけでなく、応答時間の短縮を実現し、深層学習の応用範囲を飛躍的に広げる。

現在の AI エッジコンピューティングの障害は、ソフトウェアによる柔軟な処理技術では必要な電力性能が実現できず、一方ハードウェアで柔軟性を実現する FPGA では学習処理に必要な計算精度・計算能力を低コストで実現することが難しいことである。

そのため、高い電力性能を実現する演算コアと、柔軟な処理を実現できる再構成可能回路を融合することが必須であると考えられる。従来はこのようなアプローチは困難であったが、この状況は近年急速に変化し、Achronix、FlexLogix、Menta、QuickLogic などの複数の IP ベンダーが eFPGA と呼ばれるカスタム LSI 向け FPGA IP を提供している。これらを利用し、高効率な演算コアと FPGA IP を融合させることで、エッジでの AI に必要な柔軟かつ高速な処理を実現できる。

我々は、過去に半導体技術としては最先端ではない TSMC の 40nm プロセスを利用して、演算方式やデータパスを深層学習アプリケーション向けの最適化によって同じ TSMC の当時最先端プロセスであった 12FFC で製造された NVIDIA Xavier とほぼ同等の電力性能を実現した実績がある (Xavier が 1Tops/W に対して我々は 0.5Tops/W)。これを 12FFC で実現したならば電力性能が 3 倍程度高いことを意味する。また、我々は、人工知能ソフトウェア技術の開発、実社会・産業への応用事業の中心とする企業であり、自社開発の深層学習フレームワークは世界最高レベルの性能を発揮し、国内外の様々な応用分野で実際に使われている。このため、開発したプロセッサをただちに産業応用でき、大きな波及効果を実現できる。

しかしながら、現時点で我々は NVIDIA や Intel のような巨大企業ではなく、最先端の半導体開発を自己資金だけで行うのは困難である。本研究開発は、多用途に利用でき、世界をリードする深層学習向けプロセッサを開発・実用化することは、我々の事業の発展だけでなく、日本が人工知能応用技術で世界をリードしていくことに大きな貢献ができる、公共性の高い研究開発であると考えられる。

人工ニューラルネットワークの研究開発は 2010 年代にはいって飛躍的な発展をとげ、様々な分野への応用がひろまりつつある。これは、計算機の能力の飛躍的な発展を背景とした、「深層学習」、すなわち、非常に層数も層毎のニューロン数も多い、深層ニューラルネットワークが可能になったこと、そのような深層ニューラルネットワークを学習させるのに十分な教師データが様々な分野で利用可能になったことによっている。言い換えると、現在の深層学習は、非常に大規模なデータを大規模な計算機クラウドで処理することで学習を実現している。しかし、このアプローチには明らかな限界がある。ネットワークのデータ転送能力とクラウドの処理能力である。「超スマート社会」の基盤となるのは AI によって制御されるロボットが社会・産業のあらゆる面の基盤技術となっていくことだが、そのためにはエッジ側の極めて多様な状況に各デバイスが迅速に対応する必要がある。これは、エッジ側での学習によって初めて実現できる。

現状では、深層学習向けプロセッサの研究開発は、クラウド向けの学習までをターゲットにしたデバイスの研究開発と、エッジでの推論だけを対象にしたデバイスに分化している。これは、エッジ側では既に学習済みのネットワークを利用し、学習はデータを集積したクラウドで行うことを前提にしている。しかし、これは、エッジ側の極めて多様な状況に対応することは困難であることを意味する。

このため、自動運転等の応用には、学習にも対応できる柔軟性をもったプロセッサを使うのが主流である。例えば、NVIDIA Xavier は TSMC の最新の 12FFC プロセスを使って INT8 演算で 30TOPS、30W、すなわち 1Tops/W を実現したとしている。Xavier は、Volta と同一設計の演算コアを使っており、推論・学習の双方に対応可能である。一方、Intel が買収した MobileEye は、2020 年頃に 7nm の EyeQ 5 で、2.4 Tops/W を実現するとしている。但し、EyeQ 5 の「Vision processor」の詳細は明らかになっていない。

<https://www.mobileye.com/our-technology/evolution-eyeq-chip/>

<https://newsroom.intel.com/wp-content/uploads/sites/11/2018/06/intel-mobileye-pb.pdf>

これらから、現在のエッジ側の主要な応用での電力性能は 1Tops/W 程度とみることができるといえる。言い換えると、本プロジェクトの目標の基準である。

3.3.2 最終目標と根拠

最終目標(2020年度)としては、開発成果を組み込んだシステムレベルでの検証を行い、エネルギー消費効率あるいは電力効率(単位電力あたり性能)が、事業開始時点における同等の技術と比較し、10 倍以上となることを示すには、最低 10Tops/W 程度の性能を実現する必要がある。

本提案では、主たる応用としてロボティクスを想定している。ロボティクスにおいて重要なタスクになるのは物体検出である。物体検出とは、画像から物体の位置とその物体の分類を行うタスクであり、高度な作業を行うロボットの実現に不可欠である。物体検出の産業用ロボットへの需要の高い応用例として、乱雑に機械部品の入った箱から部品の位置を特定し、その部品を生産ラインに流すタスクが挙げられる。このタスクにおいては物体検出の速度が全体のボトルネックとならないだけの性能を確保する必要がある。家庭用ロボットについても同様に、物体を検出して移動させるタスクの需要は高い。家庭用ロボットが人間とインタラクションすることを考えると、応答速度が使いやすさを左右する。物体検出の高精度なネットワークとして知られる SSD (Single Shot multibox Detector)[1] を Full HD 画像に適用したときの計算量は 0.7Tops である。NVIDIA Xavier はこれを 20ms 程度で計算できる可能性があるが、物体検出のためだけに 350 平方ミリのチップを利用するのは経済的ではない。また、消費電力も、30W は家庭用ロボットには過大である。すなわち、シリコン面積、消費電力ともに、1 桁以上の向上が必須である。

自動運転においても状況は同様である。自動運転の実現には、リアルタイムなセマンティックセグメンテーションを行う能力が必要である。セマンティックセグメンテーションは、入力画像に何が映っているかの分類をピクセルごとに行うタスクである。自動運転車は、車載カメラが撮影した画像に対してセグメンテーションを行い、画像のどこに歩行者が写っているのかといった情報を得る。遠くの物体を認識するため入力画像は高解像度、例えば Full HD になり、画素数に応じた計算能力が必要とされる。フレームレートが 30fps の Full HD 画像に対し、セグメンテーションの高精度なネットワークとして知られる PSPNet を適用すると、必要な演算性能は 140Tops/s となる。つまり、1Tops/W のプロセッサでは 140W もの電力を消費することになる。加えて、路上の危険な状況への迅速な対応のために低遅延でセグメンテーションを行う必要があるため、応答時間を短くするための演算性能も重要である。自動車が供給できる電力には限りがあり、冷却の問題もあることを考えると、現在の技術では自動運転に必要とされるリアルタイムなセグメンテーションの実現は困難である。逆に、140 Tflops 程度の能力を 14 W 程度で実現することが可能となれば自動運転のためのセマンティックセグメンテーションを行うシステムが実現できる。

一方、電力性能の向上は、近年困難になってきている。これは、いわゆるムーアの法則の限界によるものである。半導体の製造技術は、2010 年前後まではほぼ 3 年毎にデザインルールが半分になる指数関数的な進歩を続けてきたが、ここ 10 年の進歩は遅くなっている。これは、22nm 前後からトランジスタの構造がプレナーから FinFET に変わり、またリソグラフィも 7nm では ArF レーザーから EUV に切り替わる等、技術的ハードルが高くなっているためである。

さらに、トランジスタサイズが小さくなくても、動作電圧を下げられなくなっているため、半導体製造技術の進歩が消費電力の低下につながらなくなっている。

すなわち、本事業で目標とされている、現在の 10 倍以上の電力性能の向上、さらにはそれを超えた電力性能を実現するには、半導体の性能向上に期待するだけでは十分ではなく、新しいアプローチで消費電力を削減する必要がある。

これらの困難の解決を目指すアプローチの一つは、再構成可能回路、いわゆる FPGA の利用である。FPGA は、チップ内部の論理素子(通常 SRAM によるルックアップテーブルで実現)とネットワークをプログラム可能にすることで、大量生産できる標準品で任意の論理回路を実現する技術である。FPGA を AI エッジコンピューティングに応用する研究は多数ある。

一方、FPGA 技術の問題は、プログラム可能にするために同じ規模の回路の消費電力とシリコン面積が大幅に増加することである。このため、内蔵されている基本素子を使う実装では、データ表現を 1-2 ビット程度まで落として場合でなければ汎用の GPGPU コアに基づいた回路に比べて有利にならない。ある程度の規模の乗算器を内蔵した、信号処理向けの FPGA チップも存在するが、シリコン面積に対して乗算器ブロックの割合がそれほど大きくないため、必要な性能を実現するには極めて高価かつ消費電力も大きなものになる。

3.3.3 目標の達成度

本研究開発の目標は、エッジにおける深層学習応用のためのプロセッサ、特に、ロボティクス等のリアルタイム処理に十分な能力と学習に対応できる柔軟性をもつプロセッサとその利用のために必要なソフトウェアを開発することである。エッジにおける深層学習の第一の障害は消費電力であり、第二はリアルタイム性能であるため、本研究開発では、現在実際にシリコンが存在していると見られるチップの中では最高の電力性能を実現している NVIDIA Xavier チップの 1Tops/W の 20 倍以上の電力性能を実現することで、深層学習の応用範囲を飛躍的に拡大することを目指し、さらに、演算コア以外の部分を再構成可能論理で構築することで、従来のプロセッサではソフトウェアで実現する必要があった部分をハードウェアで実現すること

を可能にする。プロセッサ開発は順調に推移し、今年度中に設計開発を完了し、試作段階に移る予定である。

さらに、プロセッサ開発だけではなく、このプロセッサのためのソフトウェア開発も並行して進めている。このため、エッジ・クラウドの双方でソフトウェア開発を進められる体制を構築した。このため、このプロセッサに Chainer フレームワークを移植し、これまでのソフトウェアの蓄積を利用して様々な人工知能応用を進めた。

NEDO 先導研究「次世代人工知能・ロボット中核技術開発／(次世代人工知能技術分野)大規模目的基礎研究・先端技術研究開発／超低消費電力深層学習プロセッサおよびソフトウェア層の研究開発」で開発したアーキテクチャをベースに、さらに改良・省電力化を進めたプロセッサアーキテクチャを設計し、まず 2020 年に製造可能になると考えられる TSMC プロセスで 10 Tops/W を実現できるプロセッサを開発する予定である。さらに、2022 年度時点で利用可能なプロセス技術と、アーキテクチャのさらなる改良で、15-20Tops/W を目指す。実効的な性能としては、柔軟なハードウェアによりネットワーク圧縮等を実現することで、さらなる消費電力の減少をめざす。

なお、本研究と直接関係するものではないが、昨年度までの NEDO 先導研究で開発したアーキテクチャの知見を取り入れて PFN/神戸大学で共同開発した MN-Core プロセッサは昨年度完成し、2020 年 6 月期の Green500 リストに置いて、TSMC の 7nm プロセスを利用している NVIDIA 社の GPGPU, A100 を用いたシステムを上回る消費電力あたりの演算速度を HPL というアプリケーションで実現し、世界で最も電力効率のプロセッサであることを実証した。現在設計開発を進めているプロセッサは、ベースとなるアーキテクチャに同じ高効率アーキテクチャを用いているため、本プロジェクトの目標性能を達成し、世界最高効率のプロセッサを構築できることの確実性が増している。

また、これに並行して、実際の産業・社会応用を進めている。

このように、研究開発全体は順調に推移し、着実に目標に向かった設計開発を実現しつつある。

(4) 個別研究開発テーマ

本研究開発では、「FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発」という大局的目的を実現するため、下記 4 項目の個別研究開発テーマを設定した。

- 実施項目 1: 再構成可能低遅延低消費電力 AI コンピューティングアーキテクチャの研究開発
- 実施項目 2: 再構成可能低遅延低消費電力 AI プロセッサチップの研究開発
- 実施項目 3: ソフトウェアフレームワークの開発
- 実施項目 4: ロボティクス応用のためのソフトウェア開発

個別の研究開発テーマにおける概要、目標および進捗状況の詳細は以下に示す。

3.3.4 研究開発サブテーマ「再構成可能低遅延低消費電力 AI コンピューティングアーキテクチャの研究開発」

3.3.4.1 概要

本項目は PFN, 神戸大学および理化学研究所が協力して研究開発を実施する。以下、我々が既に開発したチップを GPFN1 チップと呼ぶ。GPFN1 チップでは、基本的演算を半精度 16x16 の行列と長さ 16 のベクトルとの積とすることで、他の半精度演算をサポートす

るプロセッサ、特に 4x4 の行列ベクトル積をサポートする NVIDIA Volta に比べても同等の半導体技術では優位な電力性能を実現した。GPFN1 チップでは消費電力の大半を実際の演算器である行列ベクトル積ユニットが消費するので、これをさらに 3 倍程度改善させるには、必要な計算精度を維持しつつ行列ベクトル積部分の消費電力を改善する必要がある。このため、以下のような回路構成を検討する。

- (1) 16 ビット浮動小数点数の他、語長 8 ビットのデータ形式をサポートする
- (2) ビット長の長い演算のための付加回路の規模を小さくする
- (3) 演算結果は丸められることを考慮し、精度に影響しない部分は切り捨て等の処理を行う
- (4) 行列乗算器の入力単位では指数を共通にする(ブロック浮動小数点形式)ことで、回路規模・消費電力を下げる
- (5) 対数変換等の非線型変換を用いて乗算回路の回路規模を小さくする
- (6) 外付けメモリ等とのデータ転送においてデータ圧縮等を利用し、必要なデータ転送速度の削減を行う

3.3.4.2 最終目標と根拠

本実施項目の目標は、

- ・同一の半導体技術・動作電圧において、我々が既に開発した 40nm プロセスでの 0.5Tops/W を最低 4 倍程度向上させること
- ・演算コア以外の部分を再構成可能論理 (eFPGA IP) で構成した、スケーラブルなプロセッサアーキテクチャを開発すること

である。この値を実現することは、我々が保有する技術から可能と考えられる一方、世界の他の企業などが実現することは困難であり、適切な目標であると考えられる。

3.3.4.3 目標の達成度

目標が達成できるかどうかの判断は、2021 年に予定されているアクセラレータチップのサンプルの製造で初めて明らかになるものである。しかしながら、本プロジェクト以前に試作を行っている MN-Core チップとそれを用いた MN-3 ディープラーニング用クラスタシステムに置いて、世界最高の演算速度・消費電力比である約 21Gflops/W を達成している。この技術に半導体のスケーリングおよび短精度浮動小数点演算による更なる演算高速化および下記の改良を加えることで、目標は十分に達成できると認識している。

以下、我々が既に開発したチップを GPFN1 チップと呼ぶ。GPFN1 チップでは、基本的演算を半精度 16x16 の行列と長さ 16 のベクトルとの積とすることで、他の半精度演算をサポートするプロセッサ、特に 4x4 の行列ベクトル積をサポートする NVIDIA Volta に比べても同等の半導体技術では優位な電力性能を実現した。GPFN1 チップでは消費電力の大半を実際の演算器である行列ベクトル積ユニットが消費するので、これをさらに 3 倍程度改善させるには、必要な計算精度を維持しつつ行列ベクトル積部分の消費電力を改善する必要がある。このため、以下のような回路構成を検討する。

- (1) 16 ビット浮動小数点数の他、語長 8 ビットのデータ形式をサポートする
- (2) ビット長の長い演算のための付加回路の規模を小さくする
- (3) 演算結果は丸められることを考慮し、精度に影響しない部分は切り捨て等の処理を行う

- (4) 行列乗算器の入力単位では指数を共通にする(ブロック浮動小数点形式)ことで、回路規模・消費電力を下げる
- (5) 対数変換等の非線型変換を用いて乗算回路の回路規模を小さくする
- (6) 外付けメモリ等とのデータ転送においてデータ圧縮等を利用し、必要なデータ転送速度の削減を行う

2022 年度までのより詳細な開発項目とマイルストーン、目標は以下のようになる。

1-1 8ビットデータ形式のサポート

FP16 の他に語長 8 ビットの形式をサポートする。NVIDIA のプロセッサや ARM SVE では INT8 をサポートするが、これは深層学習にとって最適なデータ型であるわけでは必ずしもなく、回路規模も比較的大きい。このため、深層学習に適した、より回路規模及び消費電力の小さい演算方式を検討する。2018 年度には必要なビット長についての知見をまとめ、19 年度中にはビット長所に応じた適切な回路構成を決定する。

1-2 付加回路の規模の縮小

GPFN1 チップでは、同一の乗算回路をモード切換えによって半精度・単精度・倍精度の乗算に適用可能にしているが、共通なのは半精度乗算の Wallace 木のみであり、そのあとの複数項加算部分の多くは共通化されていない。これは回路規模的にも消費電力的にも好ましくないため、入力の並べかえ等によりこれらを可能な限り共通化する。これについては 19 年度中に回路構成を検討し、20 年度に性能評価を行う。

1-3 下位ビットの切り捨て等による乗算器規模の縮小

通常の固定小数点乗算器では、 n ビットの入力 2 つから $2n$ ビットの結果を得るが、浮動小数点乗算ではこの結果を最終的には 上位 n ビットに丸める。このため、下位 n ビットの答は丸め誤差が平均的には正しい程度に求めればよいと考えられる。このことを利用して、下位ビットを Wallace 木での加算前に丸める等で回路規模・消費電力を減らす構成の検討を行う。これについては 19 年度中に回路構成を検討し、20 年度に性能評価を行う。

1-4 ブロック浮動小数点形式

浮動小数点処理において、仮数部が短くなってくると、指数部に関係した処理のオーバーヘッドが無視できなくなる。ここで指数部に関係した処理とは、指数の比較、指数の比較後の仮数のシフト等である。行列ベクトル積を基本演算とする回路では、演算の対象となる行列やベクトルについて、要素の指数をあらかじめ共通化したブロック浮動小数点形式を使うことで、これらのオーバーヘッドを回避し、回路規模を削減することができる。さらに、必要なメモリ帯域幅を減らすこともできる。但し、この方式は 1-3 の下位ビットの切り捨てはできないので、どちらが良いかを検討する必要がある。これについては 19 年度中に回路構成を検討し、20 年度に性能評価を行う。

1-5 対数表現等の利用

通常の浮動小数点表現の代わりに対数表現を用いると、乗算を加算に変換できるため、乗算器自体のサイズは大きく減少する。一方、対数表現のままでは加算は複雑な非線型関数になるため、通常は $\log(1+\exp(x))$ を区分多項式等で近似する。内積演算の場合には、複数項の加算は真数表現で行うことが好ましいので、乗算のあと指数関数を評価す

ること真数に変換する。浮動小数点表現と対数表現の変換は区間 [1,2) について行えばよいので、例えば 8 ビット程度の精度の場合には、区間全体を 2 ないし 3 次多項式近似することで十分な精度が得られる。この場合には、通常の浮動小数点表現に比べて速度的にも回路規模的にも有利になり、電力性能向上の効果も大きいことが期待できる。これについては 18 年度中に回路構成を検討し、19-20 年度に性能評価を行う。

1-6 データ圧縮の利用

エッジでの深層学習の利用では、電力性能だけでなくハードウェアのコスト自体も極めて重要な要素であり、データ転送速度あたりの消費電力は低いものの非常に高価な HBM 等の技術を使うことは現実的ではない。このため、安価な LPDDR5 等のインターフェースをもつ外部メモリで高い性能を実現することが必須と考えられる。このためには、画像とニューラルネットの係数の両方について、外部メモリとのデータ転送時の必要バンド幅を減らす、すなわち、データ圧縮を利用する必要がある。これについては、18-19 年度に方式検討を行う。但し、この部分は再構成可能論理で実装される。

GPFN1 では、チップ内部に複数コアにまたがった総和・放送等を行うネットワークを実装し、深層学習で重要となる畳み込みや行列ベクトル積を効率的に実行することを可能とした。しかしながら、このネットワークをハードウェアで実装したため、多様な処理に柔軟に対応することは難しくなっている。本提案では、より柔軟な回路の変更を可能にするため、演算コアよりも上位のネットワークや制御回路を eFPGA IP、すなわち再構成可能論理で構成する。これについては、18 年度中に速度、面積、消費電力についての予備的な評価を行い、20 年度までに実際の論理設計に基づいた性能評価を行う。

3.3.4.4 成果と意義

本実施項目の目標が実現されると、世界最高レベルのディープラーニング用アクセラレータのエンジンが実現することになる。実世界におけるディープラーニングの応用場面の多くは、計算に使用できる電力に大きな制限があり、適用できる計算量の限界となっている。特にバッテリーを電源とする知的ロボット、自動運転、言語および対話認識システムの今後の発展には超省電力ディープラーニング・アクセラレータは不可欠であり、この分野における弊社および我が国の優位性構築に大きく貢献することが期待される。

3.3.4.5 成果の普及

本プロジェクトは、現在急速に成長している人工知能の社会実装の鍵となるエッジ AI に使われるプロセッサチップ技術を国内で独自に開発するものであり、海外の競合技術に比べて 1 桁ほど高い電力あたり性能を実現する。さらに、クラウドからエッジまでを統合しハードウェア・ソフトウェアシステムを提供する。この技術が国内で世界に先行して普及することにより、国内の AI 関連技術開発全体が世界に対して大きな優位に立てるだけでなく、ハードウェア・ソフトウェアの双方におけるデファクト・スタンダードの確立につながる。

3.3.5 研究開発サブテーマ「再構成可能低遅延低消費電力 AI プロセッサチップの研究開発」

3.3.5.1 概要

本項目の目標は、前記 1.1 項目の成果に基づき、そのアーキテクチャを実装したプロセッサチップの開発を行う。特に、利用可能なテクノロジーと実際のアプリケーションでの効率・電力性能を考慮し、プロジェクトの目標である高い電力性能を実応用で実現できるものにする。

現在の GPGPU 等の汎用性の高いプロセッサの多くは、チップ内の効率的な集団通信サポートをもたない。キャッシュベースのものではどうしてもキャッシュを経由する通信になり、必要なデータだけを指定して送ることはできないため、消費電力や回路規模が大きくなりがちである。また、多数のコアにまたがった放送や縮約が短時間で可能であれば効率的な並列化ができるアルゴリズムは多いが、階層キャッシュ、あるいはメッシュネットワークでのソフトウェアベースの放送・縮約では並列化が困難になることが多い。本実施項目では、放送・縮約をサポートするチップ内ネットワークにより実行効率を大きく向上させる。これらが無い場合に、実際に必要な問題サイズでアプリケーション実行効率を 50% 程度以上とすることは極めて困難である。

本実施項目では、実施項目 1 で実現する技術を用いることで、下記最終目標を達成する計画である。

3.3.5.2 最終目標と根拠

本項目の目標は、実施項目 1 で開発するプロセッサコアの電力効率から、さらに半導体技術の進歩を取り入れて電力性能を向上させ、チップ全体の設計を行うことである。項目で開発するプロセッサでは、コア間ネットワークを再構成可能論理で構成するため、ソフトウェアのオーバーヘッドなしに高効率・低レイテンシな放送・縮約オペレーションを実現でき、結果として高い実行効率を実現できる。実応用において電力性能を理論値の 70% 程度とすることを目標とする。

3.3.5.3 目標の達成度

現在設計を進めているアクセラレータチップは既に第一次デザインリリースが終わり、半導体設計業者が下位設計をすすめていることが現状である。今後第二次デザインリリース、テープアウトと進み最終的なエンジニアリング・サンプルが 2021 年中に完成する予定であり、設計手順として順調に進んでいる。

従って、目標の達成度は設計の進捗状況として高いレベルにあると判断している。

3.3.5.4 成果と意義

本実施項目の目標が実現されると、世界最高レベルのディープラーニング用アクセラレータのエンジンが実現することになる。実世界におけるディープラーニングの応用場面の多くは、計算に使用できる電力に大きな制限があり、適用できる計算量の限界となっている。特にバッテリーを電源とする知的ロボット、自動運転、言語および対話認識システムの今後の発展には超省電力ディープラーニング・アクセラレータは不可欠であり、この分野における弊社および我が国の優位性構築に大きく貢献することが期待され、大きな意義が発揮されると判断している

3.3.5.5 成果の普及

本項目の目標は、実施項目 1 で開発するプロセッサコアの電力効率から、さらに半導体技術の進歩を取り入れて電力性能を向上させ、チップ全体の設計を行うことである。項目で開発するプロセッサでは、コア間ネットワークを再構成可能論理で構成するため、ソフトウェアのオーバーヘッドなしに高効率・低レイテンシな放送・縮約オペレーションを実現でき、結

果として高い実行効率を実現できる。実応用において電力性能を理論値の 70% 程度とすることを目標とする。

このようなエンジニアリング・サンプルが完成すれば、その超省電力性能により、多くのディープラーニング適用機器に組み込むことが可能となり、成果の普及を実現することができる。

3.3.6 研究開発サブテーマ「ソフトウェアフレームワークの開発」

3.3.6.1 概要

本事業内で開発したプロセッサで利用できる深層学習ミドルウェアの研究開発を行う。基本的には、ディープラーニングフレームワーク Chainer / Pytorch から利用可能とすることで、既にあるソフトウェアの蓄積を無駄にすることなく移行することを可能にする。

当初時点でのプロセッサアーキテクチャに基づいて必要なソフトウェア開発項目、ミドルウェアの構成を明確化する。中間目標段階までに、プロセッサで動作するミドルウェアを完成させ、最終目標段階では実際の応用で利用する。

3.3.6.2 最終目標と根拠

本項目の目標は、実施項目 1 で開発するプロセッサコアの電力効率から、さらに半導体技術の進歩を取り入れて電力性能を向上させ、チップ全体の設計を行うことである。項目で開発するプロセッサでは、コア間ネットワークを再構成可能論理で構成するため、ソフトウェアのオーバーヘッドなしに高効率・低レイテンシの放送・縮約オペレーションを実現でき、結果として高い実行効率を実現できる。実応用において電力性能を理論値の 70% 程度とすることを目標とする。

3.3.6.3 目標の達成度

ソフトウェアフレームワークの開発における最終目標は、1.1、1.2 項目で開発するアクセラレータが高効率で動作し、実際のディープラーニングのフレームワークに組み込まれて動作することである。これが最終目標である根拠は、本開発によるソフトウェアフレームワークなしには、全体システムの動作は応用ソフトウェアにはつながらないため、ソフトウェアには最も重要な目標である。

3.3.6.4 成果と意義

本ソフトウェアフレームワークに関しては、今回の開発計画の先行開発によるアクセラレータを対象としてソフトウェアフレームワークは既に完成していて、実機で動作している。本項目はこのソフトウェアフレームワークの拡張および新設計部分への対応が中心であり、これにより本プロジェクトの成果であるディープラーニングアクセラレータが動作する意義がある。

3.3.6.5 成果の普及

本 1.3 項目の成果は、本テーマ内で開発するディープラーニングアクセラレータがアプリケーションソフトウェアを動作させるために必須の項目である。成果の普及に関しては、この特質から、まずディープラーニングアクセラレータの普及が先行するものであり、それによって本項目の成果も普及すると判断している。

3.3.7 研究開発サブテーマ「ロボティクス応用のためのソフトウェア開発」

3.3.7.1 概要

本テーマ内で開発したミドルウェアを利用したロボティクス応用のための物体検出ソフトウェアを開発する。

2018年時点では、物体検出アルゴリズムとして SSD (Single Shot multibox Detector) [1] の採用を想定するが、物体検出手法研究の進展に合わせて随時見直す。

中間目標段階までに、プロセッサで動作する物体検出ソフトウェアを完成させ、最終目標段階では実際のロボティクス応用のための物体検出システムに組み込み、消費電力と応答時間を測定する。

[1]W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed. SSD: single shot multibox detector. CoRR, abs/1512.02325, 2015.

3.3.7.2 最終目標と根拠

ロボティクス応用のためのソフトウェア開発」の最終目標は、パーソナルロボットに超省電力のディープラーニングを組み合わせることで、真に役に立つパーソナルロボットを完成させることである。しかしながら、この最終目標は本プロジェクトの期間内に完成するものではなく、更に長期間を掛けて熟成させることが求められている。

研究開発期間内(～2022)における開発目標は以下の通りである：

研究開発の現ステージでは、エッジでも適用可能な高い電力性能の演算器コアを開発することが目的である。本コアを使ったクラウド利用可能なプロセッサとして第一版を開発している。エッジ用には開発したコアの規模を縮小することにより、他社推論チップより優れた電力性能が得られることを目標としている。

電力性能向上について、アーキテクチャとして演算精度の削減、搭載メモリ量の削減を実施し、電力削減を行う。また物理設計においては物理配置の工夫による配線等の削減、標準セルのカスタマイズなどを実施し電力削減を行う。研究開発の期間中に、開発したディープラーニングアクセラレータをパーソナルロボットに組み込み、パーソナルロボット内でローカルなディープラーニングによる学習処理が実現することを目標とする。

3.3.7.3 目標の達成度

- ・ 物体認識アプリケーションを想定 (ロボティクス・自動運転で必須の要素技術)
- ・ ハードウェア仕様決定のための性能モデル開発を実施
- ・ ローカルメモリサイズの最小化 ⇒ 回路面積あたりの演算器数 1.3 倍
- ・ 低回路面積オーバーヘッドの行列転置方式の提案
- ・ 物体検出モデルに必要なレイヤーの詳細設計を進行中
- ・ 背景技術として、PFN はすでに高性能な物体認識モデルの学習技術を保有
例: 一般物体認識モデル

3.3.7.4 成果と意義

パーソナルロボットの実用化と普及への最大の障壁は、非常に多様性のある家庭内の状況を学習しつつ行動プランニングをするという非常に困難な課題にある。本技術開発により、パーソナルロボットに組み込めるディープラーニング・アクセラレータの実現は、今までと全く異なったレベルでの環境把握と行動プランニングをかのうにするため、実用化に大きな意義を持つ。

3.3.7.5 成果の普及 (公開しない)

3.4 研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」

本研究開発項目では、端末側(エッジ)で必要となる、カメラや各種センサー入力情報の画像・信号処理および、AI(Deep Learning)処理を、ハードウェアのエンジン処理を基本としながら、低消費電力と低コスト化および、柔軟性を両立し、今後拡大する幅広い AI 応用に適応できる組み込み向け AI エッジ基盤を構築するという目標に対し、AI 認識処理と画像処理において汎用 GPU と比較してそれぞれ 10 倍以上の電力効率化を達成、および、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 を達成するなど、成果を上げつつある。以下に研究開発項目を示す。

- ① 仮想エンジンアーキテクチャ開発
- ② アルゴリズム記述ツール開発
- ③ CV/AI 基本ミドルウェア開発
- ④ リアルタイム SLAM 技術開発
- ⑤ AI エッジ LSI 搭載車載 ECU 試作・評価
- ⑥ 量子化 DNN 技術開発
- ⑦ DNN 変換ツール開発
- ⑧ エッジ環境最適化技術開発
- ⑨ AI エッジ LSI 試作開発

個別の研究開発テーマにおける詳細は以下に示す。

3.4.1 研究開発サブテーマ「画像・信号処理 AI 基盤技術 (仮想エンジンアーキテクチャ) 開発」

3.4.1.1 概要

端末側(エッジ)で必要となるカメラや各種センサー入力情報の画像・信号処理および Deep Learning 処理をハードウェアのエンジン処理を基本としながら、ハードウェア・アクセラレータをアルゴリズムに沿ってプログラム記述可能な「仮想エンジンアーキテクチャ」により、低消費電力と低コスト化と柔軟性を両立し、アルゴリズムの進化と、拡大する幅広い AI 応用に適応できる組み込み向け AI エッジ基盤を構築する。

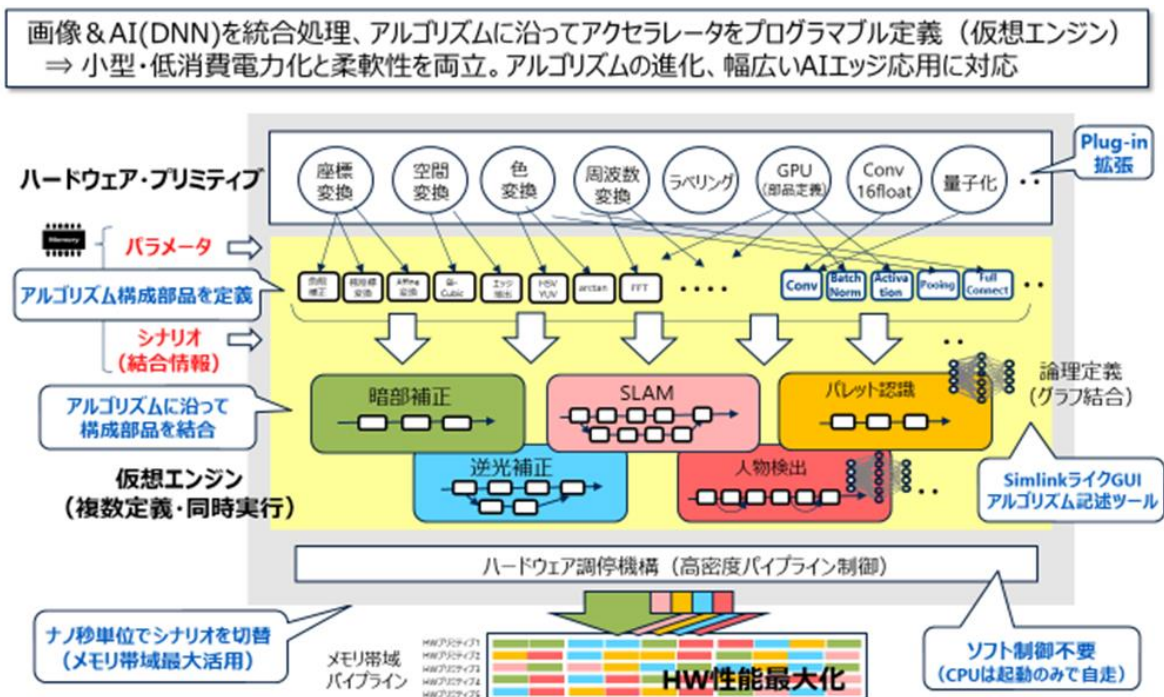


図 3.4.1.1 画像・信号処理 AI 基盤技術(仮想エンジンアーキテクチャ)開発(概念図)

3.4.1.2 最終目標と根拠

「仮想エンジンアーキテクチャ」のハードウェア設計完了および試作 LSI において本開発の全機能のシステム検証、性能評価を完了すること。根拠としては、単にハードウェアの開発を完了するだけでなく、試作 LSI によるシステムに組み込むことによる評価で実証することにより「仮想エンジンアーキテクチャ」の有効性を確認する。

3.4.1.3 目標の達成度

「仮想エンジンアーキテクチャ」のハードウェア設計は、RTL 設計とシミュレーションにおいて各種 EDA ツールを活用した効率的な検証の実施、および FPGA 基板を使用した実機動作検証を行うことにより、2019 年度で開発完了した。また 2020 年度に入り試作 LSI1 搭載の評価ボードによるシステム動作検証、機能評価を行い、さらに試作 LSI2 で実施予定だった LiDAR から Visual SLAM への機能向上検討および量子化演算器の検討も前倒しで終了した。これにより 2020 年 6 月末迄に全ての評価を完了させた。

3.4.1.4 成果と意義

今回の仮想エンジンアーキテクチャは従来の弊社の技術に対して、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柔軟なネット定義を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量子化演算器導入、を実現した。性能評価では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た。またリアルタイム SLAM 処理ライブラリを開発し、高精度な自己位置推定処理において GPU と比較して処理時間が 1/20 に短縮できることを確認した。これにより本来のエッジ向きの最適なデバイスとしての特長が発揮されることとなり大いに意義があることだと言える。

3.4.1.5 成果の普及

成果の普及は仮想エンジンアーキテクチャが AI エッジプロセッサのプラットフォームとして市場に広がることであり、そのためには既存ソリューションとの優位性をデモ等でアピールし、LSI1 が評価完了した後の早期の事業化対応に期待されることである。

3.4.2 研究開発サブテーマ「アルゴリズム記述ツール開発」

3.4.2.1 概要

前述の開発項目仮想エンジンを定義するアルゴリズム記述支援ツールを開発する。これにより、ハードウェア要素プリミティブに対するパラメータ設定およびアルゴリズムに沿った要素部品の連結によるアクセラレータ定義を視覚的に記述可能とする。本ツールにより、各種画像処理アルゴリズムおよび Deep Learning ネット処理を柔軟に定義可能で、今後のアルゴリズムの進化および各種応用に合わせた最適化にも対応可能とする。

3.4.2.2 最終目標と根拠

中間目標として、ツールの設計・開発を α 版として完成させ、機能評価やライブラリの活用評価を行う。最終目標としては、 α 版を進化させて操作性や動作の高速性などの改善検討を行い、 β 版を完成する。

根拠としては、ツールが進化することにより効率的ライブラリ開発が可能になり応用展開も可能になることからソフトウェア開発の効率アップとハードウェアの普及が期待できるからである。

3.4.2.3 目標の達成度

仮想エンジンアーキテクチャの画像処理エンジンに対応したアルゴリズム記述ツールは GUI ツールとして 2019 年度内に α 版の開発を完了した。また AI ライブラリ対応用に仮想エンジンアーキテクチャの汎用 GPU 部に対応した DNN コンパイラ(Glow ベース)の α 版を開発した。また、2021 年 3 月までに、操作性、高速性の改善を行い β 版を完成する。

3.4.2.4 成果と意義

従来の仮想エンジンのソフトウェアは 16 進数のコードを直接記述して作成していたので、内容が分かりにくく記述ミスも起こしやすいという課題があった。今回のアルゴリズム記述ツールにより GUI ツールでは画像処理の流れを視覚的にわかりやすく表すことができ、DNN コンパイラではアセンブラでニーモニックに沿って記述することによって仮想エンジンの 16 進数コードに変換してくれるので、ソフトウェアの開発効率アップという点では大いに意義があることだと言える。

3.4.2.5 成果の普及

成果の普及はまず開発者がソフトウェア開発する際に、今後種々のライブラリ開発をしていく上で有効なことである。さらには共同開発するパートナーへの普及、事業化後のユーザーに対する普及へと広めることが LSI そのものの普及に繋がるので、ツール自身の機能アップ・信頼性向上を図ることが今後注力するポイントである。

3.4.3 研究開発サブテーマ「Computer Vision/AI 基本ミドルウェア開発」

3.4.3.1 概要

前述の開発項目アルゴリズム記述ツールを活用し、業界標準の Computer Vision API(Open VX/Open CV など)および、AI API(TensorFlow など)をアクセラレートするミドルウェア(ライブラリ)を開発する。このミドルウェアは業界標準との整合を最大限図り、応用システムの効率的構築を可能とするものである。

3.4.3.2 最終目標と根拠

最終目標は OpenCV、OpenVX、TensorFlow など業界標準 API の基本関数および拡張を含めた主要関数に対応する。根拠としては標準開発環境対応により応用展開性が向上し、前述のアルゴリズム記述ツールと同様にソフトウェア開発の効率アップとハードウェアの普及が期待できるからである。

3.4.3.3 目標の達成度

2019 年度までに、業界標準 API の中から開発ライブラリと関数の選定が終了した。2020 年度に入り、CV および AI の基本ライブラリ設計、開発に着手し Open VX および Open CV に類似する機能のライブラリを開発した。また AI 関連ライブラリについてはプロセッサベースから専用ハードウェアに対応した開発に着手している。

3.4.3.4 成果と意義

業界標準 API に対応するということは、事業化を行った際のユーザーに対するアピールと信頼性の向上という意味では重要なテーマである。OpenCV や OpenVX は対応する関数の数が非常に多いのでどれだけサポートできるかが大いに意義があることだと言える。

3.4.3.5 成果の普及

事業化で LSI を普及促進するには、業界標準 API のサポートは不可欠である。ユーザーが使い易く、開発し易くするためには、非常に大変ではあるが基本関数および拡張を含めた主要関数への対応を継続的に行うことが重要である。

3.4.4 研究開発サブテーマ「リアルタイム SLAM 技術開発 (SLAM ライブラリ)」

3.4.4.1 概要

現在開発済みの SLAM ソフトウェア搭載の試作装置は FAPC を使用し機能実現しているが、物流ロボットなどの製品搭載にあたり、FAPC のままでは、消費電力、サイズ、重量、コストなど多くの問題がある。車載電子制御ユニット (ECU: Electronic Control Unit) として製品化を検討しているが、車載 ECU 用として通常使用される低いパフォーマンスのマイコンチップでは処理速度が大幅に不足し、実用レベルでの動作が期待できない。

SLAM 処理のリアルタイム性を確保するため、本プロジェクトで開発する画像・信号処理 AI 基盤/AI エッジ LSI 上で高負荷部分を切り出し、ハードウェア処理する事で解決を図る。

3.4.4.2 最終目標と根拠

8km/h でのリアルタイム SLAM 処理の実現。

3.4.4.3 目標の達成度

開発した aIPE 上で並列動作する SLAM ライブラリを試作 LSI1 に実装し、評価中。
8km/h でのリアルタイム SLAM 処理達成見込み。

3.4.4.4 成果と意義

弊社が開発する物流ロボット等に LSI1 として試作した AI エッジ LSI を実装した車載 ECU と共に搭載し、弊社製品の高機能・高性能化を推進する。

3.4.4.5 成果の普及

物流ロボット等に必要の高負荷処理を順次ライブラリとして開発し、製品開発へ適用する。

3.4.5 研究開発サブテーマ「AI エッジ LSI 搭載車載 ECU 試作・評価」

3.4.5.1 概要

本プロジェクトで開発する AI エッジ LSI は、民生応用だけではなく、物流ロボットなどを含む車載 ECU に搭載できる品質の確保が必要である。そのために、早い段階で LSI のパフォーマンスだけでなく、消費電力や EMC、熱評価などの試験を実施し、応用適用時の課題を洗い出し、LSI 論理設計およびチップレイアウトなどにフィードバックする。

3.4.5.2 最終目標と根拠

車載 ECU としての品質確保のため、消費電力、EMC、熱に問題なき事を確認。

3.4.5.3 目標の達成度

試作 LSI1 を搭載した車載 ECU 基板を開発し、LSI 及び ECU 基板の正常動作を確認完了。

消費電力、熱については問題ない見込み。

3.4.5.4 成果と意義

本プロジェクトで開発した AI エッジ LSI をソシオネクストで製品化し、弊社が開発する物流ロボット等に SLAM ライブラリを含めて搭載し弊社製品の高機能・高性能化を推進する。

3.4.5.5 成果の普及

AI エッジ LSI の製品化計画が固まり次第、順次対象技術を車載 ECU に実装し、製品開発を進める。

3.4.6 研究開発サブテーマ「量子化 DNN 技術開発」

3.4.6.1 概要

本項目では、AI による認識処理において、量子化(ビット低減)による認識率の低下を最小限に抑制する最適学習アルゴリズム、最適ハードウェアの開発、評価を行う。本技術により、今後拡大する組み込み向け AI エッジ応用で求められる、より小型で低消費電力な AI 基盤を実現する。

3.4.6.2 最終目標と根拠

試作 LSI での検証・評価、アルゴリズム改善による認識精度向上

(目標: 8bit 処理相当の実用的な認識率)

目標値は他社ベンチマークより設定

3.4.6.3 目標の達成度

試作 LSI1 に 8bit と Ternary(2bit)、Binary(1bit)の量子化 DNN エンジンのハードウェア実装を完了し、評価を実施した。ResNet50 を Ternary(2bit) 技術にて、FIX8 と同等の認識精度で、LSI による実機電力評価において、量子化 DNN エンジン単体で高消費電力性能値を達成した。

画像処理の実行性能(fps/W)としては、AI 認識処理において、NVIDIA Jetson Nano, Google TPU, Intel MyriadX, NVIDIA Xavier と比較して、最大 10 倍の消費電力性能であることを確認した。

3.4.6.4 成果と意義

車両自動運転、インテリジェント監視・見守り、ロボット、AI 家電など、AI 技術を使った社会基盤の革新に置いて、汎用 GPU での処理は電力効率の面で大きな課題がありました。この課題に対し、本研究開発において、汎用 GPU での AI 認識処理に比べ、電力効率 10 倍以上を達成したことは、エッジ処理基盤を構築する上で大きな成果と考える。

3.4.6.5 成果の普及

2019 年度より実用化に向けた活動を実施している。2019 年度では、顧客/パートナー100 社以上を訪問し、ユーザー要望を抽出しました。また、2019 年 10 月 CEATEC および 2019 年 11 月 ET&IoT Technology にて、FPGA デモを展示。また、2020 年 3 月にはソシオネクストからプレスリリースを実施。4 月には NEDO 含めた 4 社共同のプレスリリースを実施した。

2020 年 6 月現在、試作 LSI1 を外部顧客および関係会社へ提供し、実アプリケーションでの実践的な評価を実施中。

3.4.7 研究開発サブテーマ「DNN 変換ツール開発」

3.4.7.1 概要

TensorFlow など汎用の開発環境で構築した Deep Learning ネットワークを、本研究開発する量子化 DNN の画像認識処理が可能なフォーマットへ変換するツールを開発する。

3.4.7.2 最終目標と根拠

各種、汎用環境で構築したニューラルネットワークを用いた機能検証、性能評価

3.4.7.3 目標の達成度

TensorFlow を使った学習結果から、OpenVX の C ソースコードを生成するツールを開発し、学習結果から容易に量子化 DNN 推論が可能な処理環境を構築した。

3.4.7.4 成果と意義

AI は、世界レベルで急速に進展しており、開発環境の進化も著しい。業界標準との整合を最大限に図り、ターゲットとする組み込み向けシステムに最小限の工数で適用可能とすることで、幅広い応用システム展開と、効率的システム構築が可能となる。本研究で開発した変換ツールにより、最も一般的な学習環境である TensorFlow から電力効率の良い AI 認識処理の実行環境を容易に構築すること可能であり、市場でのエッジプラットフォーム構築への貢献は大きい。

3.4.7.5 成果の普及

2020 年 6 月現在、試作 LSI1 を外部顧客および関係会社社提供し、実アプリケーションでの実践的な評価を実施中。

3.4.8 研究開発サブテーマ「エッジ環境最適化技術開発」

3.4.8.1 概要

AI 認識処理において、クラウドサーバーとエッジ連携での最適役割分担による、エッジ側必要機能の見極めと機能開発および、評価を行う。

3.4.8.2 最終目標と根拠

試作 LSI での検証・評価、エッジ環境最適化方式検討。

3.4.8.3 目標の達成度

今後重要になる MaaS システムにおけるクラウド・エッジ処理分担を定義し、エッジ側で必要となる認識技術を定義する。ユースケースとして、走行車両のカメラ画像認識により、特定の場所における車両混雑、駐車場の空き状況などを「エッジ車両で情報集約 ⇒ クラウド処理 ⇒ Navigation 通知」などのシステムを想定し、エッジ側に必要な処理を定義した。特に重要と考えられる Object Detection(Vehicle Detection/Pedestrian Detection)/Semantic Segmentation を実現する商用ソフトウェアを導入し、試作 LSI1 で評価を実施した。

3.4.8.4 成果と意義

MaaS システムにおけるクラウド・エッジ処理分担を定義し、エッジ側で必要となる認識技術を、商用ソフトウェアを使って実用的な評価を実施することは、Society5.0 の自動運転インフラシステムの実現に大きく貢献する。

3.4.8.5 成果の普及

現状、本研究開発の普及活動を画策中。

3.4.9 研究開発サブテーマ「AI エッジ LSI 試作開発」

3.4.9.1 概要

研究開発テーマのアウトプットとして、アルゴリズムの進化への対応、幅広い応用への適合、エッジ環境最適化に対応可能な、組み込み向けの進化型・低消費電力 AI エッジ LSI の試作開発を行う。

3.4.9.2 最終目標と根拠

試作 LSI のシステム検証・評価完了。目標性能:NVIDIA Jetson に対して、1/10 の消費電力、1/5 のチップサイズ

3.4.9.3 目標の達成度

現在の試作 LSI は、各種評価やデモができるように 2018 年 10 月当初に計画した SoC 仕様に対し、CA53 コア数増加、画像出力用 HDMI 搭載、外部 I/F(USB3.0/2.0,PCIe)等を追加しており、チップサイズとしては増加する結果となった。また、消費電力については目標達成を確認した。

3.4.9.4 成果と意義

仮想エンジニアークテクチャ技術、量子化 DNN エンジン技術を搭載した試作 LSI を開発し、汎用 GPU と比較して、AI 認識処理、画像処理で処理効率 10 倍、SLAM 処理時間 1/20 を達成した。

これらの成果は、市販の AI 処理 LSI と比較しても十分競争力があり、研究開発として大きな成果があったと考える。

3.4.9.5 成果の普及

2020 年 6 月現在、試作 LSI1 を外部顧客および関係会社へ提供し、実アプリケーションでの実践的な評価を実施中。

これらの評価のフィードバックより、研究開発終了後は直ちに実用化が可能と考える。

また、展示会、学会等での発表も計画している。

3.5 研究開発テーマ「ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発」

3.5.1 概要

監視カメラなどに使われる撮像素子の高解像度化に伴い、エッジで実施する AI 処理負荷の増大が見込まれている。現状はこの処理に GPU が使われるが、エネルギー効率が低く発熱量が膨大となるため、低解像度に画像をスペックダウンした上で AI 処理を行っているのが実状であり高解像度のメリットを生かした推論の実現が難しい。この非効率性は、AI 処理の膨大な演算量と、メモリとプロセッサ間で発生する膨大なデータ転送量に起因しており、この両者を大幅に削減することが消費エネルギー効率を高めるために重要となる。

この課題を解決するため、本研究開発テーマでは、メモリとプロセッサ内部でのデータ再利用効率を高め省エネルギー化する「ソフトテンソルプロセッサ技術」、高効率かつコンパクトにモデルを圧縮し推論時の演算負荷を削減する「高効率モデリング技術」、4K/8K などの超高精細画像に基づき遠隔まで細かく広範囲の状況を少ないカメラ台数で把握するモデルを実現する「広範囲大規模認識技術」の 3 層の研究項目を構成し、車・船舶などの移動体や高所等に設置されたカメラから超広範囲かつ高精細なセンシングを極めて低い消費電力で実行できる AI エッジ技術を目指している。

最終年度には、電力効率 20 倍以上(現状の GPU の電力効率を 30GOP/J と想定)の改善を目標としている。

項目① ソフトテンソルプロセッサ技術の研究開発

本項目では、エネルギー効率の高いディープニューラルネットワーク(Deep Neural Network、以下 DNN)推論処理を実現する FPGA 向けのプロセッサ IP であるソフトテンソルプロセッサ技術の研究開発を行っている。これまで、3 次元テンソル型に配置された Tensor Processing Element(以下、TPE)が複数連結された 4 次元テンソル型の構造をもつソフトテンソルプロセッサにより、畳込み演算の高速化に取り組んできた。TPE が計算した 2 次元畳込み演算結果を加算するリダクション処理方法を改良し、より少ない TPE でより高速な畳込み演算を実現する方式を開発した。また、低ビットの DNN 高速化とエネルギー高効率化を狙い、DNN の畳込み層のパラメータのゼロビットに関する演算をスキップするゼロビットスキップを取り入れ、可変ビット長の入力の内積演算と入力ビットが大きい場合のクロック数の低減を実現した。また、DNN 記述を入力しソフトテンソルプロセッサ用の制御コードなどを出力するモデル変換技術の開発にも取り組み、ダミーデータを用いて推論を実行しながら DNN 記述の解析を行うことで柔軟な層解析が可能な方式を開発し、項目②の高効率モデリング技術を用いて軽量化したモデルの動的解析にも対応した。これらは計画通り進捗しており、最終目標達成の見込みである。

項目② 高効率モデリング技術の研究開発

本項目では、DNN の高い認識精度を維持しながら、畳込み層のチャンネル削減と演算精度の低ビット量子化により、メモリ使用量と演算量の両面を軽量化する技術の研究開発を行っている。チャンネル削減では、学習済みニューラルネットワークの各層間にアテンション・モジュールを挿入し、各層の冗長なチャンネルを推論誤差最小化に基づき特定して削減する技術を実現した。また、この技術を近年の多様な分岐・合流経路を含むネットワークにも対応可能な方式に改良した。更に、このチャンネル削減技術と演算精度の低ビット量子化を実施可能な高効率化ツールを開発し、高効率化対象の DNN モデルに対し、量子化とチャンネル削減を適用した軽量化モデルの出力機能を実現した。これらは計画通り進捗しており、最終目標達成の見込みである。

項目③ 広範囲大規模認識技術の研究開発

項目③-1 行動認識技術(行動検出)

本項目では、高解像度のカメラ画像に対し人物の行動を認識する技術の研究開発を行っている。行動認識対象として万引き行為に着目し、周辺警戒のために顔を左右に振る動作を検知するため顔を検知する方式に加え、顔がカメラ側を向いていない場合の対応として骨格情報から肩などの状態変化も用いる方式を開発した。その結果、実験環境での演技映像において2019年度の目標性能達成を確認できた。これらは計画通り進捗しており、最終目標達成の見込みである。

項目③-2 広域状況認識技術(渋滞検出、火災検出)

本項目では、高解像度のカメラ画像に対し高所に設置したカメラ画像からの「火災」と「渋滞」といった異常検知を行う広域認識技術の研究開発を行っている。「火災」では、濃い黒煙に限らず煙を検出可能となる技術開発を行い、実火災映像のデータセットにおいて2019年度の目標精度の達成を確認した。また、「渋滞」では、車両密度と流量からなる渋滞曲線に基づく臨界車両密度を超えた状態を渋滞と定義し検知する技術を開発した。車両検知情報をもとに渋滞と判定する道路領域を自動抽出した渋滞判定を実現し、高所カメラから撮影した実道路映像からなるデータセットにおいて2019年度の目標性能達成を確認できた。これらは計画通り進捗しており、最終目標達成の見込みである。

項目③-3 実時間物体検出技術

本項目では、移動する高精細カメラ画像から広範囲の物体を検出する技術の研究開発を行っている。小物体と隠蔽に頑強な検出方式の開発を行った。小物体対策として Upsampling 方式および Subnetwork 方式の開発を実施した。また、隠蔽対策として Soft-NMS の方式の開発を実施した。更に、学習データの物体種別ごとのサンプル数の不均衡に対応可能なロス関数を導入し、2019年度の目標達成を確認した。これらは計画通り進捗しており、最終目標達成の見込みである。

項目③-4 実時間物体追跡技術

本項目では、移動する高精細カメラ画像から物体を追跡する技術の研究開発を行っている。複数の物体が存在し頻繁に隠蔽が発生する場合への対応として、検出結果の対応付けによる Tracking-by-Detection アプローチにおいて追跡軌跡の再同定を用いる方式の開発を行った。更に、ロス関数の改善と部位特徴表現による学習を行い、2019年度の目標達成を確認した。更に、車載カメラから歩行者の表情までも検出・追跡できる場合を想定した状況認識の可能性検討、及び、項目③-3 と共同で車載撮影システムによる撮影を実施した。これらは計画通り進捗しており、最終目標達成の見込みである。

項目③-5 長距離物体検出技術

本項目では、カメラ画像を用いた主に船舶周囲の監視実現に向け、高解像度カメラによる海上で長距離遠方の物体検出を行う画像認識技術の研究開発を行っている。長距離遠方の学習データの収集と、船舶のように細長い物体の検出に頑強な画像認識アルゴリズムの開発を行い、長距離遠方の物体についても精度を大きく落とすことなく認識可能とし、2019年度の目標達成を確認した。これらは計画通り進捗しており、最終目標達成の見込みである。

3.5.2 最終目標と根拠

(公開しない)

3.5.3 成果と意義

項目① ソフトテンソルプロセッサ技術の研究開発

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目② 高効率モデリング技術の研究開発

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目③

項目③-1 行動認識技術(行動検出)

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目③-2 広域状況認識技術(渋滞検出、火災検出)

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目③-3 実時間物体検出技術

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目③-4 実時間物体追跡技術

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

項目③-5 長距離物体検出技術

2018 年度および 2019 年度の目標を達成しており、今年度の目標も現状の研究開発を進展させて達成する見込みである。

3.5.4 成果と意義

本研究開発の事業範囲である、広域監視カメラシステム(固定、車載)、及び、船舶用監視カメラシステムの 2037 年時点での温室効果ガス排出状況について、事業を実施しなかった場合とした場合を比較したところ、

温室効果ガス排出削減効果がある試算結果となり、省エネルギー化に貢献する見込みである。

3.5.5 成果の普及

本研究開発の成果の普及については、本研究開発は、各領域でトップクラスのシェアを有する実事業会社自身による事業検討と研究開発を一貫通貫した産学連携体制で推進しており、学会発表・論文、特許出願、新聞雑誌 Web などのメディア掲載、プレスリリース、展示会や講演にも積極的に取り組んでいる。

現時点で、下記の実績を挙げており、最終年度である今年度の成果も追加することで、さらに増加する予定で進めている。

現時点での実績は下記のとおりである。

- ・ 特許は、出願済・出願準備中を併せ 17 件、今後も増加見込みである。
- ・ 論文の実績は 9 件であり、今後も増加見込みである。
- ・ 学会発表・講演の実績は 19 件であり、今後も増加見込みである。
また、本研究開発の共同研究機関の共著も 1 件あり、今後も共著などにも取り組む予定である。
- ・ 新聞・雑誌、プレスリリースの実績は、13 件であり、今後も増加見込みである。
- ・ 展示会への出典の実績は、8 件であり、今後も増加見込みである。

例として、展示会への出典、講演などでは、CEATEC のような大規模展示会への出典、発表や ET & IoT Technology など、今後、本研究開発成果が適用される分野の展示会において講演することで、本研究開発成果についての普及につながると考えている。

研究開発開始時点において、「NEDO の革新的 AI エッジコンピューティングをテーマとする技術開発プロジェクトを開始」(2018 年 9 月 25 日)を、OKI、JMU、ALSOK、会津大による 4 機関の共同プレスリリースを行い、プレスリリースによる本研究開発の社会への認知度向上の活動を行っている。

2 年目には、NEDOおよびOKIで「ディープラーニングモデルの新たな軽量化技術を開発」(2019 年 9 月 9 日)の共同プレスリリースを行い、16 件の記事掲載(新聞 5 雑誌 1 Web 10)があり、注目される活動も行っている。本研究開発の成果を今後適宜公表する予定である。

3.6 研究開発テーマ「5G 時代を見据えた高度自律的学習機能搭載のための AI エッジコンピューティング技術の研究開発」

本研究開発項目では、複数深層学習の BaseNetwork 層共通化及び FPGA 実装 (BaseNetwork 層共通化及び FPGA 化効果による、消費電力(対 GPU)従来比 1/10)という目標に対し、消費電力 1/11 程度まで達成できることが確認できているとともに、当該開発技術は汎用圧縮ツールとして商用化が実現できている。また 5G エッジ・クラウド統合環境の構築のもとでの実現性評価も行っており、成果を上げつつある。

社会的ニーズの高い AI エッジコンピューティングの適用領域で且つ Society5.0 の実現を支える有力な技術の一つである「ドローン」においては、各種センサー群や飛行制御機構等による限られたリソース(容量・電力)のもとで、今後の高度な自律化のために、深層学習等のアルゴリズムを複数種類同時に実装しなければならない状況が確実である。こうした背景を踏まえ、本研究開発では、リソース制約の大きいエッジにおける高度自律的学習機能及びドローンへの搭載を含む実装技術の開発を行うとともに、5G エッジ・クラウド協調環境のもとでの社会実装・事業化に資する検証を目的としている。

2020 年度現在において、複数深層学習の BaseNetwork 共通化開発技術の FPGA 実装による電力対性能(GPU 比)目標値 1/10 以上及びネットワーク圧縮ツール商用化を達成しているとともに、深層強化学習の BaseNetwork 層共通化及び FPGA 実装、高度自律的学習のための人工意識ソフトウェア開発及び(対深層強化学習との)有効性評価、AI エッジ統合制御システム開発及び5G 通信エッジ・クラウド連携システム環境のもとでの実証を進めている。

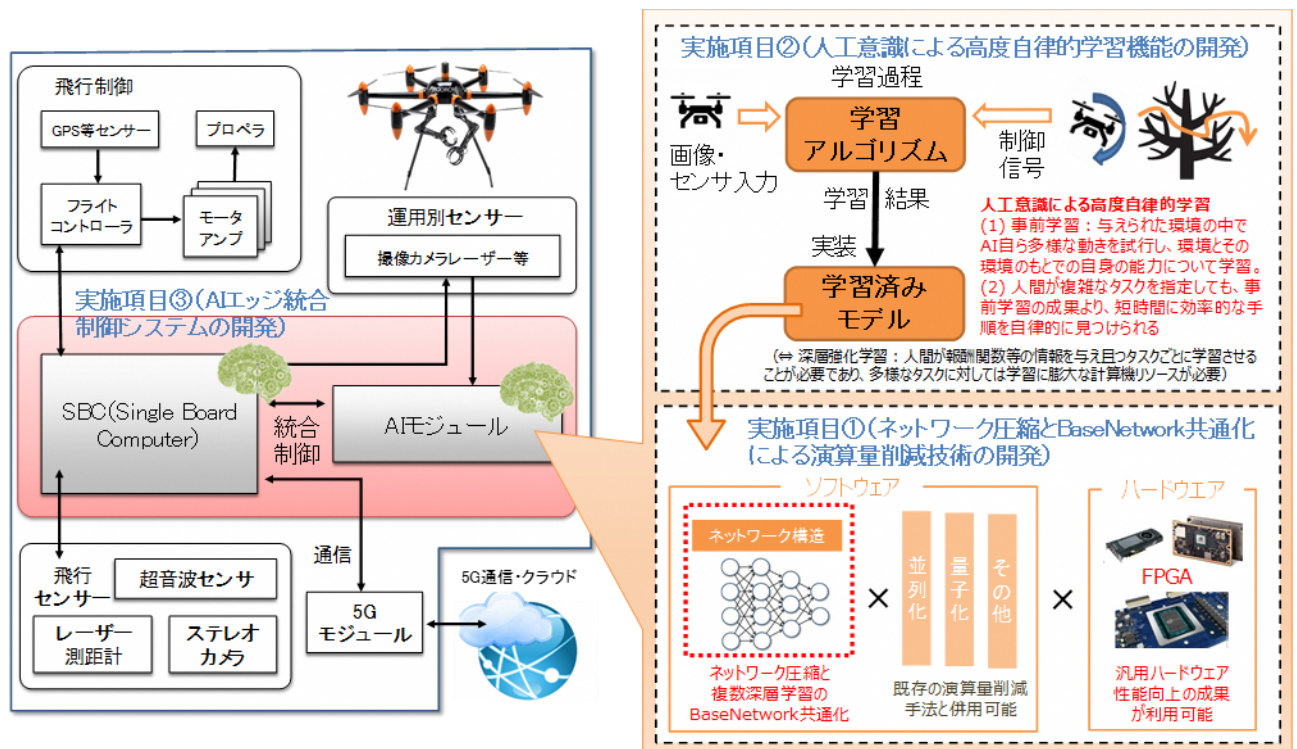


図 3.6 研究開発の全体像 (概念図)

3.6.1 研究開発サブテーマ「ネットワーク圧縮と BaseNetwork 共通化による演算量削減技術の開発」

3.6.1.1 概要

エッジに搭載可能な(本研究開発ではドローンへの搭載を対象)、学習・推論演算量低減・低消費電力のためのアルゴリズム及び実装技術の開発を行い、実機検証のために主に FPGA への実装を行う。

(1) 深層学習に対するネットワーク圧縮技術の開発

深層学習の構成に対するネットワーク圧縮技術に基づき、1.3 で開発する AI エッジ統合制御システムに組み込む FPGA 実装可能なアルゴリズムを開発する。具体的には、エッジへの複数種の深層学習の実装を目的として、深層学習の下位層においてフィルタリングによる特徴抽出等を司る Base Network 層を共通化する技術を開発する。

また、任意の深層学習のネットワーク構造に対してある演算量の制限を付けた上で各層の設計諸元(カーネルのサイズ、入力チャンネルのサイズ、出力チャンネルのサイズ等)を自動的に振り、性能及びネットワーク圧縮率を確認しながら探索し最適化する深層学習のネットワーク自動圧縮設計ツールを開発する。一方、一般に深層強化学習は周辺的环境に応じて適切な制御(例えばドローン飛行制御)を獲得するために、多様な環境をシミュレータ等で用意し学習を実施する必要がある。そのため、環境シミュレータと連動した学習機能を持つ深層強化学習用のネットワーク自動圧縮設計ツールを開発する。

(2) AI エッジ学習環境シミュレータの開発

深層強化学習等のための学習環境開発を行う。具体的には、フィールドでの実空間を模擬できる仮想シミュレータ(AI エッジ学習環境シミュレータと呼ぶ。)を開発する。

(3) 実装方式の開発

(1)及び(2)で開発のアルゴリズムについて、1-3 で用いる AI エッジ統合制御システムに組み込むための AI モジュール開発を行う。最適化されたネットワーク構造のサイズ等の検証をまず GPU ベースの基板上に実装し確認の上で、深層学習、深層強化学習の FPGA への実装開発を行う。

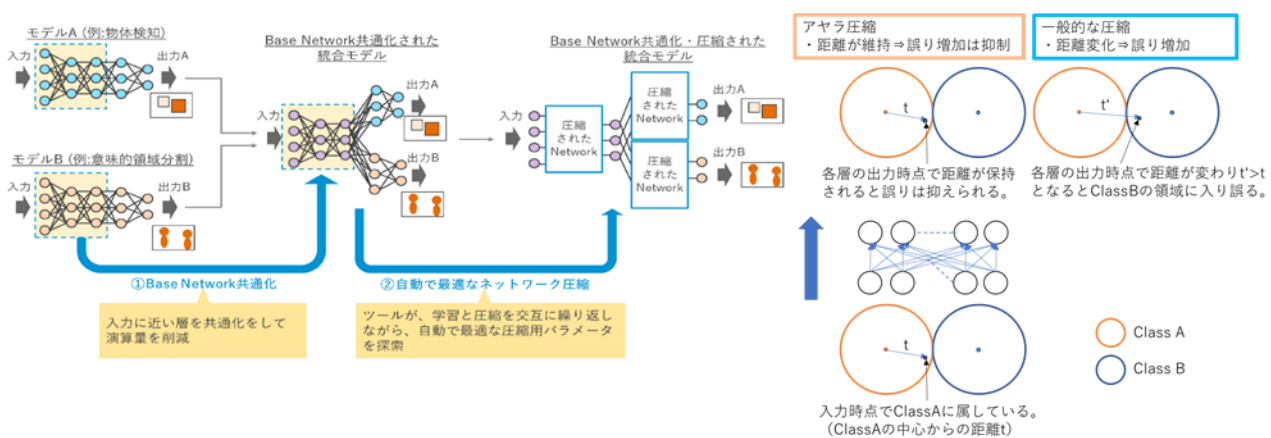


図 3.6.1.1 ネットワーク圧縮と Base Network 層共通化技術 (概念図)

3.6.1.2最終目標と根拠

年度毎及び最終目標は以下の通りである。

- ・2018年度：複数深層学習の BaseNetwork 共通化及び GPU 実装(FPGA 実装に向けた事前評価)
(BaseNetwork 層共通化効果による、計算量従来比 1/10 目標)
- ・2019年度：複数深層学習の BaseNetwork 層共通化及び FPGA 実装
(BaseNetwork 層共通化及び FPGA 化効果による、消費電力(対 GPU) 従来比 1/10 目標)
- ・2020年度：深層強化学習の BaseNetwork 層共通化及び FPGA 実装
(BaseNetwork 層共通化及び FPGA 化効果による、消費電力(対 GPU) 従来比 1/10 目標)

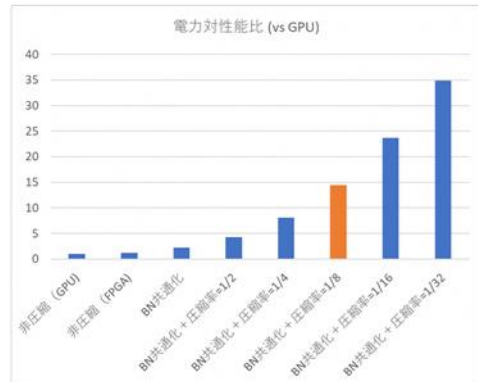
3.6.1.3目標の達成度

2020 年度現在において、複数深層学習の BaseNetwork 層共通化及び FPGA 実装に関する電力対性能を評価し、目標値である電力対性能(GPU 比)1/10 に対して、1/11 程度まで削減できることを確認している(モデルサイズは圧縮により元サイズの 3%以下。精度劣化を抑えながら演算量削減)。下図に電力対性能(GPU 比)評価結果を示す。

GPU/FPGAの電力・推論時間の見積り※1※2※3

入力：256 x 256のRGB画像 モデル：VGG16ベースの物体検出モデルとセマンティックセグメンテーションモデル

デバイス	圧縮操作	一回の推論時間 [msec]	消費電力 [W]	モデルサイズ [byte]	フレームレート [fps]	電力対性能 [fps/W]	電力対性能比 (vs GPU)
GPU(K80※4)	非圧縮	70.06	138	202,099,200	14.3	0.104	1.00
	非圧縮	769.67	10	50,524,800	1.3	0.130	1.25
FPGA (Arria 10※5)	ベースネット共通化	430.65	10	27,589,568	2.3	0.232	2.23
	ベースネット共通化+圧縮率=1/2	222.61	10	14,384,736	4.5	0.449	4.32
	ベースネット共通化+圧縮率=1/4	118.58	10	7,782,320	8.4	0.843	8.11
	ベースネット共通化+圧縮率=1/8	66.57	10	4,481,112	15.0	1.502	14.44
	ベースネット共通化+圧縮率=1/16	40.57	10	2,830,508	24.6	2.465	23.70
	ベースネット共通化+圧縮率=1/32	27.57	10	2,005,206	36.3	3.628	34.88



※1 FPGAの消費電力はコンパイラによる予測値
 ※2 GPUの消費電力・推論時間は、レイヤー単位の計測値から予測
 ※3 モデルサイズはパラメータ数から換算(パラメータ一つにつきGPUは32bit、FPGAは8bitで計算)
 ※4 NVIDIA Corporation GK210GL [Tesla K80] (rev a1)
 ※5 インテル Arria 10GX

※BN共通化は、ベースネット共通化の意

図 3.6.1.3.1 ネットワーク圧縮と Base Network 層共通化技術による電力対性能(GPU 比)

また、2020 年度最終目標に向けて、特定領域内人物追尾のためのドローン視点のカメラの制御を検証ユースとして、複数深層学習／深層強化学習の BaseNetwork 層共通化及び FPGA 実装による消費電力(対 GPU) 従来比 1/10 目標に向けて開発・検証を進めている。

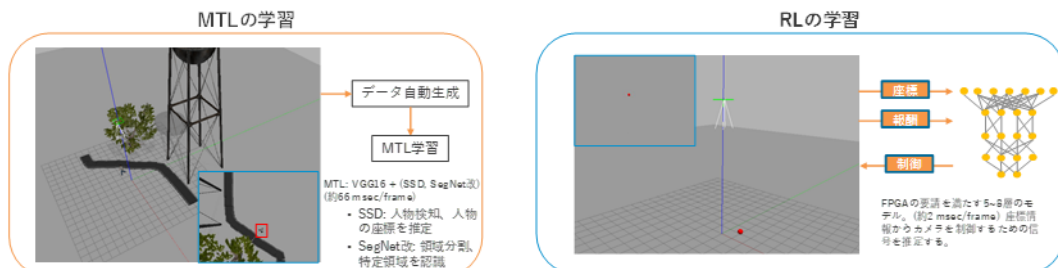


図 3.6.1.3.2 ネットワーク圧縮と Base Network 層共通化技術による電力対性能(GPU 比)

3.6.1.4 成果と意義

本開発技術はプルーニング等の既存の圧縮技術と併用でき、それら既存圧縮手法を含み一式ワンストップで取り揃えた汎用自動ネットワーク圧縮ツール(Pressai)を2020年3月にリリースした。また、本商用化に伴い、今年度に圧縮ツールの深層強化学習対応を追加で開発実施することで実施計画書の見直しも行った。既存 AI プラットフォームとの対比による自動ネットワーク圧縮ツールの優位性を下図に示す。

圧縮ツール		モデル提供	学習			圧縮				推奨実行ファイル		Edgeデバイス	
ツール名	開発元	DLモデル	学習	蒸留	転移学習	アラヤ圧縮	Prune*1	量子化	行列分解	CPU最適化	HDL記述言語	Deploy	Monitor
自動ネットワーク圧縮ツール(仮名)	株式会社アラヤ	-	○	○	-	○	○	○	-	○	○	○*2	-
Cloud AutoML Vision	Google LLC	○	○	-	-	-	-	-	-	-	-	-	-
Cloud AutoML Vision Edge	Google LLC	○	-	-	-	-	-	-	-	-	-	○	-
TensorFlow Lite	Google LLC	-	-	-	-	-	○	○	-	○	-	-	-
Optimization tool kit	Google LLC	-	-	-	-	-	○	○	-	-	-	-	-
QNNPACK	Facebook	-	-	-	-	-	-	○	-	-	-	-	-
SageMaker+AWS market place	Amazon	○(300以上)	○	-	-	-	-	-	-	○	-	○	-
Azure IoT Edge	Microsoft	○(数十個)	○	-	-	-	-	-	-	-	-	○	○
Transfer Learning Tool Kit	NVIDIA	-	○	-	○	-	-	○	-	-	-	-	-
TensorRT	NVIDIA	-(AWS連携)	-	-	-	-	-	○	-	○	-	○	-
Neuralnetwork Distiller	Intel	-(AWS連携)	○	○	-	-	-	○	○	-	-	-	-
PocketFlow	Tencent	-	-	○	-	-	-	○	○	-	-	-	-
keras_compressor	ドワンゴメディアビレッジ	-	○	-	-	-	-	-	○	-	-	-	-
Blueoil	LeapMind	-	○	-	-	-	-	○	-	-	-	-	-
DeLTA-Lite	LeapMind	○	○	?	-	-	-	?	?	-	-	-	-
DeLTA-Kit	LeapMind	-	-	-	-	-	-	-	-	-	-	○	-

図 3.6.1.4 開発・商用化した自動ネットワーク圧縮ツールと既存 AI プラットフォームとの対比

3.6.1.5 成果の普及

エッジ AI を実現するための汎用自動ネットワーク圧縮ツール(Pressai)を2020年3月にリリースしている。

- 2019年11月18日 NEDO と共同プレスリリース「AI の深層学習用ニューラルネットワークモデル向けの自動圧縮・実装ツールを開発」
- 2019年11月21日『ET / IoT Technology Award 2019』 Edge Technology 優秀賞



図 3.6.1.5 汎用自動ネットワーク圧縮ツール(Pressai)

顧客となるエッジ機器メーカーに本ツールを提供し、そのメーカー内部で圧縮・実装を可能とすることで、顧客は自身が開発したネットワークモデルやデータセットの情報を外部に提供することなく情報を保護したままネットワークモデルを圧縮しエッジ機器に搭載することが可能となり設計効率も向上する。エッジ機器の台数ベースのライセンスビジネスを展開中である。

また、本汎用自動ネットワーク圧縮ツールの強化学習への拡張も、市場動向も鑑みて開発を進めることで Pressai と同様なライセンスビジネス展開も検討する。

3.6.2 研究開発サブテーマ「人工意識による高度自律的学習機能の開発の研究開発」

3.6.2.1 概要

エッジにおける高度な自律制御実現を目的として、人工意識による高度自律的学習機能の開発を行う。人工意識とは、汎用人工知能の実現を指向する中で、既存の知識(=内部モデルと呼ぶ)を再利用し、新しい問題を少ない入力情報で解く機能の一つと捉えることができる。少ないデータ量で学習して新規なタスクを解くことでデータ効率性を向上し、自律飛行ドローンやロボット等のエッジ AI アプリケーションにおけるデータ収集に要するコスト・消費電力の削減を目指す。

少ないデータ量で学習して新規なタスクを解くプロセスでは次の3つのデータ効率性が考えられる。

- (1) 内部モデル構築データ効率性：エージェントの内部モデルを構築するときに必要なデータの効率性
- (2) 新規タスク求解データ効率性：内部モデルを活用し、新規タスクを学習し解けるようになるまでに必要なデータの効率性
- (3) 推論時データ効率性：少ない推論回数で新規タスクを解けるための、アテンション（どこを・そのような順番で見るべきか等）に関する方策学習におけるデータ効率性

中間年度となる2020年度までは(1)の内部モデルアーキテクチャ設計・実装、及び、(2)の効率性実現のための初期開発・評価を実施し、ステージゲート以降となる2021・2022年度は(2)の高効率化、及び、(3)の効率性実現を行う。

3.6.2.2 最終目標と根拠

年度毎及び最終目標は以下の通りである。

- ・ 2018年度：人工意識アーキテクチャ設計（機能ブロック構成と機能ブロック間インターフェースの設計）
- ・ 2019年度：人工意識アルゴリズム確立（機能詳細及びアルゴリズムの設計・開発）
- ・ 2020年度：人工意識ソフトウェア開発及び（対深層強化学習との）有効性評価（学習量・計算量削減効果による、計算量（対深層強化学習）従来比1/10目標）
- ・ 2021～22年度：人工意識のシミュレータ学習及びFPGA実装（学習量・計算量削減効果による、消費電力（対GPU）従来比1/10目標）

3.6.2.3 目標の達成度

高度自律的学習機能を実現するアーキテクチャを構築し（下図左）、環境予測とポリシー生成を担うモジュールの具体的な実装形態として、WorldModelベースのアルゴリズム及びシミュレータを開発した（下図右）。本アルゴリズムは、タスクを鑑みた環境の圧縮表現（特徴①-1）と、合成や再利用による新規ポリシー生成が可能であること（特徴①-2）を特徴とすることで、深層強化学習と比較して、新しいタスクが与えられたときに少ない試行回数で高速に学習を行うことが可能となる。

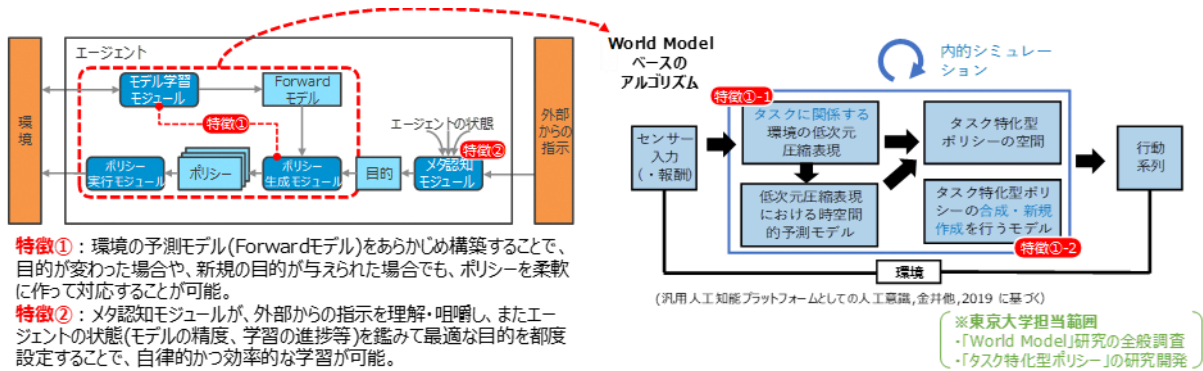
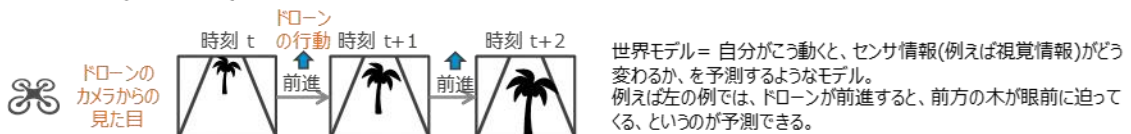


図 3.6.2.3.1 人工意識による高度自律的学習機能

World Model(世界モデル)とは



ドローンシミュレータを用いた内的シミュレーションの結果



図 3.6.2.3.2 WorldModel とシミュレーション結果

WorldModel ベースのアルゴリズムでは、現在の状態とエージェントの行動から、内的シミュレーションにより、将来の状態を予測することができる。(Dream と呼ばれる) 将来の状態の予測が可能となることで、実際の環境上で動くことなく、新しいタスクの解決方法を学習や、事前に計画した上で動くことが可能となる。SLAM 等のセンサー制御技術に比べて環境変化や新環境対応において優位と想定している。

現在、ドローンタスクを想定した 3D 迷路シミュレータ環境での評価において、WorldModel を利用した想像に基づく学習(Dream 学習)及び推論により、学習時において 25~30%のステップ数削減、また、推論時において 50%のステップ削減となることを確認している。

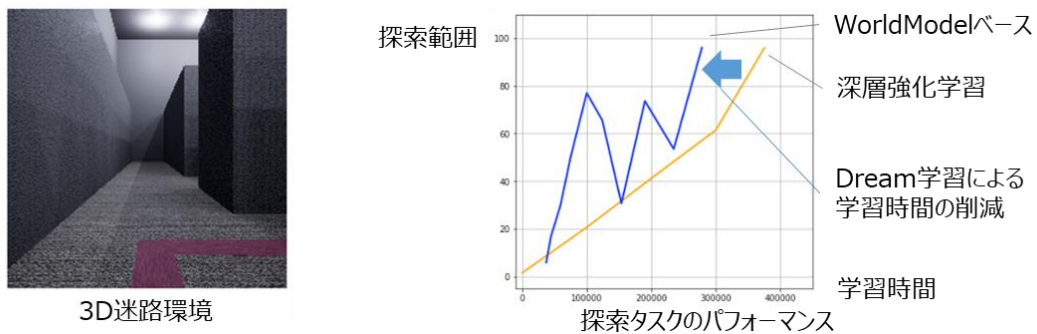


図 3.6.2.3.3 WorldModel ベースアルゴリズムのシミュレーション評価結果

また、Predictive Coding ベースのアルゴリズムを用いて、ロボットアームを想定した制御タスクとして Meta-World と呼ばれるベンチマークセットを用いた評価を実施した。

Predictive Coding ベースのアルゴリズムは、目標指向型の行動計画のためのアーキテクチャであり、センサデータや画像データの入力に対して、事前の少量データでの訓練により近未来予測だけでなく遠未来予測が可能であり、また、一般的な強化学習手法とは異なり、人間にとって解釈可能な目的を明示的に与えることが可能である。

Meta-World を用いた 7 自由度の汎用ロボットアームを想定した reach タスクで評価した結果、Predictive Coding ベースのアルゴリズムは、既存の強化学習アルゴリズムに対して学習時において 1~10 万倍のステップ削減、また、推論時において数十から 200 倍のステップ削減しつつ、より高い成功率となることを確認している。

	Predictive Coding ベース	既存の著名なメタ RL アルゴリズム
学習ステップ数 (環境とインタラクション必要な量)	377	数百万~数千万
推論時に必要なステップ数	1	数十~200
タスクの成功率	93%	40% ~ 50%

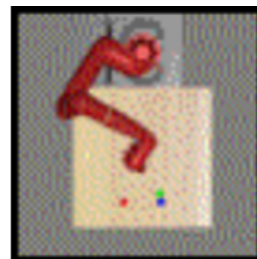


図 3.6.2.3.4 Predictive Coding ベースアルゴリズムのシミュレーション評価結果

現在、中間年度（2020 年度）目標である、人工意識ソフトウェア開発及び（対深層強化学習との）有効性評価（学習量・計算量削減効果による、計算量（対深層強化学習）従来比 1/10 目標）に向けて、性能改善や他のタスクでの開発及び評価を継続している。そして、2020 年度までのドローン適用を想定したナビゲーションタスク、また、ロボットアームを想定したリーチングタスクにおいて得られた成果を踏まえ、ステージゲート以降となる 2021・2022 年度では、事業化出口を睨み、「データ効率性を実現する高度自律的学習機能」の開発を進める。ここでは産業ニーズの高いドローンや UVG での移動タスクや、ロボットアームでのマニピュレーションタスクを対象として高い学習データ効率性の実現を目指す。人工意識による検討対象タスクの概念下図赤枠に示す。

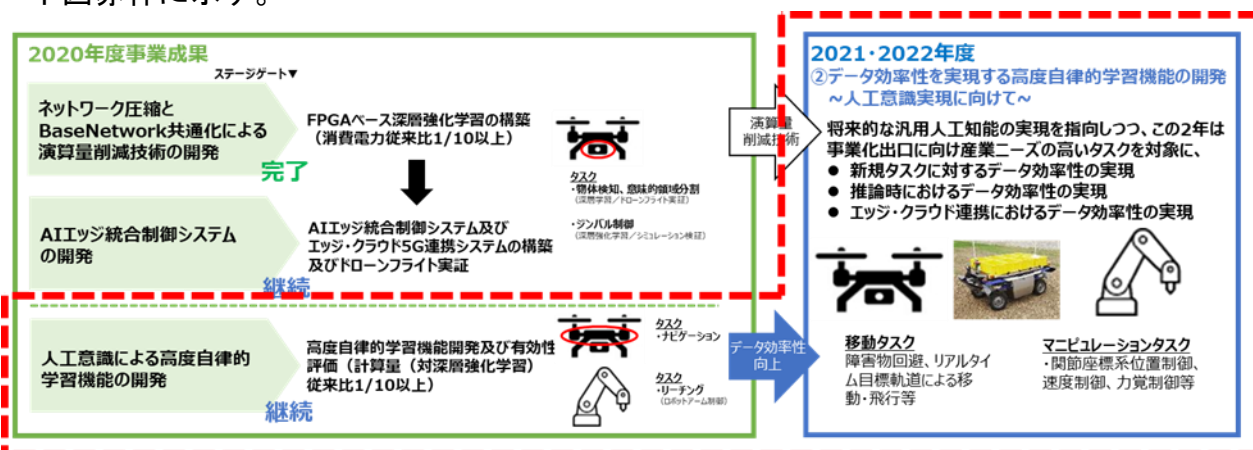


図 3.6.2.3.5 人工意識の想定対象タスク

3.6.2.4 成果と意義

2020年度まではシミュレーション評価であるが、2021～22年度のシステム実装評価を踏まえ、ドローンにおける主に人の認識に基づく自律的な飛行制御の実現により、市場化に向けた第一次顧客としてはまず、警察、消防、海上保安庁等の官公庁や、警備会社や保険会社（損害保険会社）等の民間企業が考えられる。また警備会社のエンドユーザとしては各種施設、個人等の局所的な警備から、農場や森林、海域等の広範にわたって状況把握が必要な警備などが包含される。



図 3.6.2.4.1 自律的学習機能の適用が期待される領域と想定規模

具体的な警備分野での活用方策としては、以下が例として挙げられる。

- ・ 施設警備・監視システムの構築
- ・ 農場や森林等の広域警備、防災システムの構築
- ・ 船舶救助や遭難者捜索等、海上保安への対応
- ・ 山岳救助も念頭に置いた山小屋等での警備・監視システム構築

また、同じく本研究開発で想定している、非 GPS 環境を含めた構造物周辺・内部等における作業代替（人間作業員による対象物への作業の代替：ドローンによるロボットアームを用いた対象物へのアクチュエーション）を想定した自律的な作業制御の実現により、飛行・点検・監視に付随する多様な作業（設備メンテナンス等）の取り込みと市場拡大が想定される。また以下のように政府から示されている「空の産業革命に向けたロードマップ 2019」においても 2022 年度以降の有人地帯での目視外飛行の実現に向けては運航管理システムでの空域のシステム管理と並行して機体の自律化・知能化が技術開発工程として明言されており、レベル 4 実現に向けて業界ニーズとして継続開発が必要とされていることがわかる。

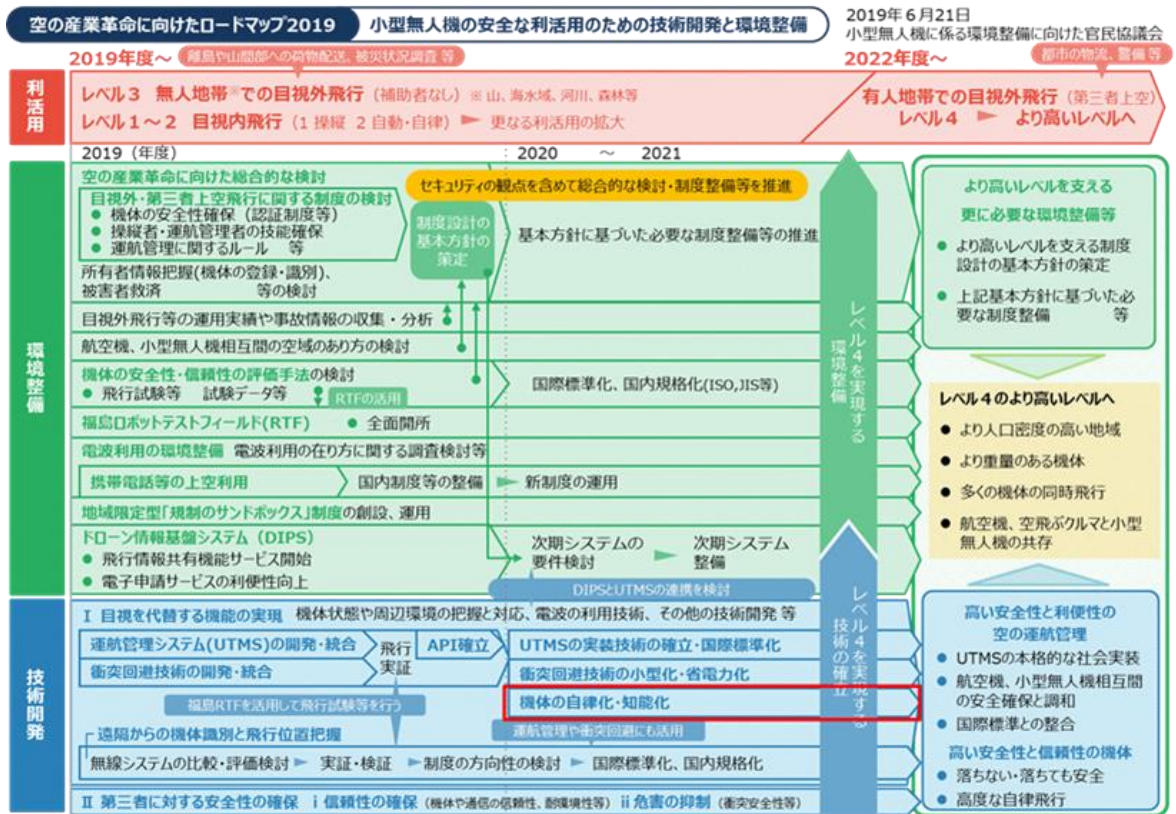


図 3.6.2.4.2 空の産業革命に向けたロードマップ 2019

3.6.2.5 成果の普及

AI エッジ統合制御システム開発の人工意識アーキテクチャ対応を、KDDI との協業体制で行うとともに、自律的学習機構が求められる、建設・自動車・電子機器等広範な産業分野の主要事業者及びベンダーとの協業体制で成果の普及を進めることを想定しており、既に先行的に一部業界における協業体制構築も進めている。一例として、建設機械分野において、株式会社電通国際情報サービスとの協業体制において、自動化支援サービスのローンチや、国内大手プラントメーカーと共同で、ドローンによるプラント内自律飛行点検の実証実験を実施しており、同様な協業を様々な産業分野で進めることで成果の普及を図る。

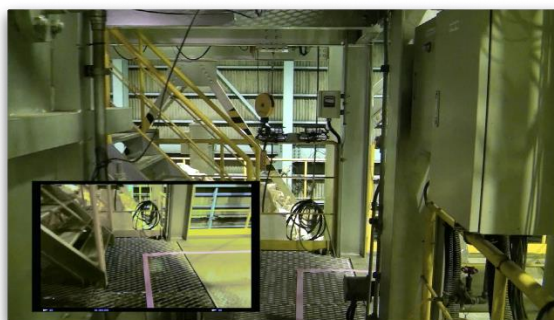


図 3.6.2.5 成果普及の例(人工意識)

3.6.3 研究開発サブテーマ「AI エッジ統合制御システムの開発」

3.6.3.1 概要

3.6.1 及び 3.6.2 で開発した AI モジュールを組み込む AI エッジ統合制御システムの開発、及び、AI エッジコンピューティング／5G 連携環境構築のもとでの実証を行う。

(1) AI エッジ統合制御システムの開発

ドローンの自律制御の実現に求められる機能性能要件を規定して 1.1 及び 1.2 の開発にフィードバックするとともに、1.1 及び 1.2 で開発した AI モジュールを組み込む(通信・センサー・CPU に関わる)AI エッジ統合制御システムの要件定義、開発、及び、ドローンへの実装と機体結合検証を行う。

(2) AI エッジコンピューティング／5G 連携のドローンフライト実証

AI エッジコンピューティング／5G 連携環境のための、通信要件(データレート、遅延等)、及び、エッジ・クラウド要件(ドローン・5G エッジネットワーク・コアネットワークにおける通信制御と AI 機能の配置等)の定義、検証シナリオ策定、(1)の開発と連動する形での検証環境準備、及び実証を行う。

なお、具体的な実証環境は、5G の社会実装スケジュール(まず 2020 年前後に超高速通信を実現し、続けて超低遅延及び多数同時接続を実装)と整合する形で準備する。

3.6.3.2 最終目標と根拠

年度毎及び最終目標は以下の通りである。

2018 年度:

- ・通信・センサー・CPU に関わる AI エッジ統合制御システムの要件定義及びモジュール開発
- ・深層学習に関するモジュール開発(センサー、CPU)
(計算量従来比 1/10 の複数深層学習実証)

2019 年度: AI エッジ統合制御システム開発

(消費電力従来比 1/10 の複数深層学習のシステム実証)

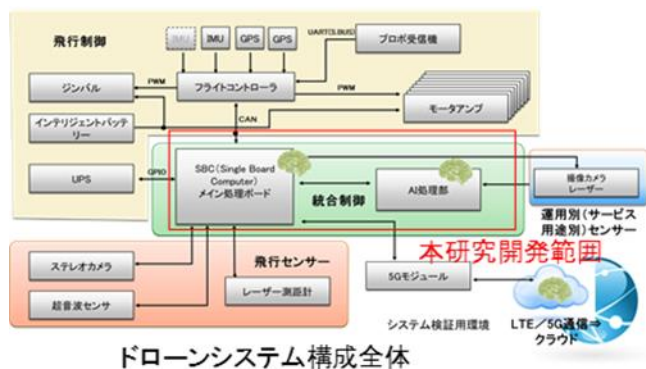
2020 年度: AI エッジコンピューティング/5G 連携の実証

(消費電力従来比 1/10 の複数深層学習・深層強化学習に関する実証)

2021～22 年度: 人工意識に関する AI エッジ統合制御システム開発及びドローンフライト実証 (消費電力従来比 1/10 の人工意識に関するドローンフライト実証)

3.6.3.3 目標の達成度

要件定義、及び、1.1 の成果である圧縮技術及び AI エッジ統合制御システムを開発した。また、ドローン監視ユースケースにおけるドローン空撮映像からの人物検出タスクの評価を行い、消費電力当たりの検知性能(フレームレート)で従来比 10 倍の複数深層学習のシステム構築が実現できた。



ドローン搭載可能なAIエッジ統合制御システム

図 3.6.3.3.1 AI エッジ統合制御システム

ドローン搭載可能な5G 通信モジュールとモバイルエッジコンピューティング環境上に前述と同様の AI アルゴリズムを構築済み。2020 年度内に5G エッジ-MEC 連携システムの基礎検証を実施予定。

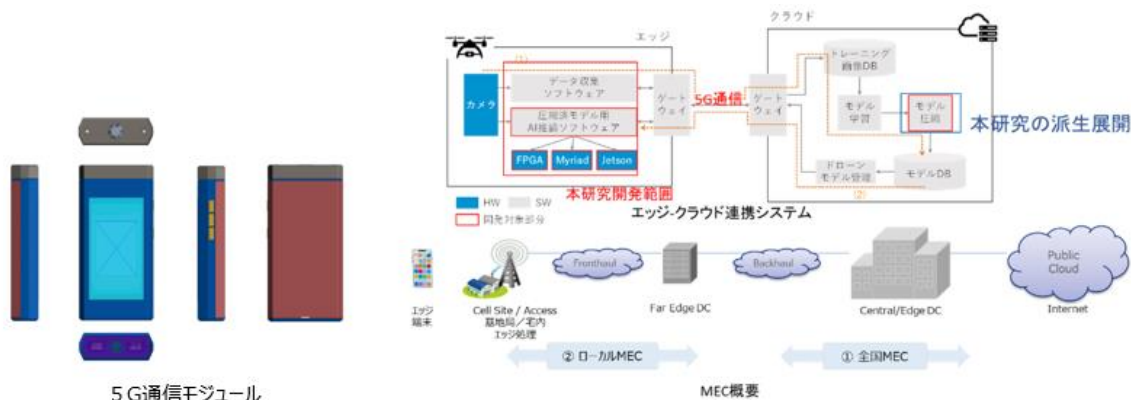


図 3.6.3.3.2 開発 AI エッジ統合制御システムと 5G エッジクラウド連携システム設計

3.6.3.4 成果と意義

当該 AI エッジ統合制御システム及び 5G エッジクラウド連携システムプラットフォームは、具体業種事業者のユースケースニーズを踏まえ、検証データとしても利用しながら実装を進めているところである。具体的なサービス提供としては、まず、ドローンの空撮動画からエッジの画像認識 AI で人物をリアルタイムに検知しユーザに通知するユースケースより実装を進めている。

• ユースケース分類

分類	特徴	分析を行う場所	分析処理の重さ
リアルタイム分析	ドローンの空撮イメージを撮影した直後に分析し、その結果を制御者に通知したり機体制御に反映する	エッジ	軽い
非リアルタイム分析	ドローンの空撮イメージを分析する。分析のタイミングはフライト後などでよい	クラウド	重い

• ソリューション毎の分析ユースケース需要

		ソリューション(お客様例)		
		構造物点検 (電力会社)	広域監視 (鉄道会社)	物流 (物流サービス提供者)
映像からのリアルタイム物体検知	ドローンの空撮映像をリアルタイムに分析し、人物や車両など、特定の物体を検出する	○	◎	◎
映像の解析によるリアルタイム機体制御	ドローンの空撮映像をリアルタイムに分析し、障害物などが検知された場合は自動で回避などの機体制御を行う	◎	◎	◎
写真からの不良検出(非リアルタイム分析)	ドローンが空撮した静止画を分析し、構造物の不良(サビ・ヒビ等)を検出する	◎	-	-

図 3.6.3.4.1 ドローンソリューション毎の分析ユースケース需要

項目	説明
ユースケース概要	車や船による物資輸送をドローンによって代替する
AI活用概要	着陸時等にドローンの近くにいる人物を検出し、注意喚起のアラートを発出する
タスク種類	物体検知
検知対象	人物(将来的にはより多様な障害物)

図 3.6.3.4.2 サービス化ユースケース(ドローン物流 AI エッジ画像認識)

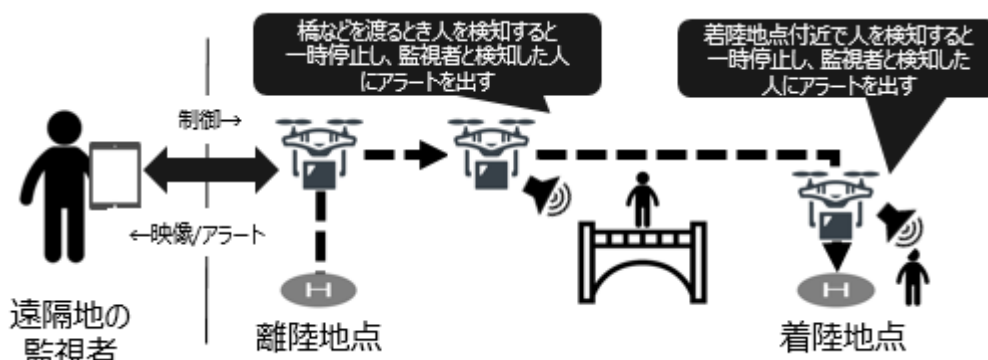


図 3.6.3.4.3 サービス化ユースケース(ドローン AI エッジ画像認識)

3.6.3.5 成果の普及

当該スマートドローンプラットフォームは、まず STEP1 として、個別ソリューションにて 2019 年 6 月に役務提供を開始しており、STEP2 として、2020 年 2Q には「KDDI スマートドローン」お客様運用メニューとしてシステムの本格提供を開始している。本研究開発成果である AI エッジ統合制御システム及び 5G エッジクラウド連携システムを含めた仕組みは、Society5.0 実現に向けた 5G・エッジ活用の重要技術として 2020 年度に對外訴求する。

3.7 研究開発テーマ「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」

本研究開発項目では、エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発という目標に対し、下記の成果を上げている。

1. 軽量化 DNN(Deep Neural Network)モデル「SSS-Net」を開発し、従来の DNN モデルと比較して、演算量及びパラメータ量を削減しつつほぼ等価な認識精度を達成した
2. ドメイン固有言語コンパイラの Xilinx 社 FPGA 向けのバックエンドを開発した。カメラパイプライン処理やフィルタ処理などのエッジビジョン AI 処理において、手動実装と比較して、ほぼ同等な性能と消費リソース量を保ちつつ、LoC(Lines of Code)を 1/10 以下に削減することを達成した。
3. ドメイン固有言語コンパイラからベンダ非依存の高位合成基盤を経由して、様々なベンダの FPGA 向けの回路を生成するための検討と設計、プロトタイプ開発が終了した。現在本開発中であり、2020 年度終了時点で基本テストのカバレッジが 100%となる見込みとなっている。

個別の研究開発テーマにおける詳細は以下に示す。

3.7.1 概要

現在、AI を活用した様々なアプリケーションの実装や実用化が世界中で取り込まれている。

グローバルで、AI フレームワークの実装・実用化が本格化している。AI アプリケーションの中でも、最も早く実用化されるのが画像(ビジョン)の AI 処理であり、大量の画像をリアルタイムで処理するためには、エッジコンピューティング上でビジョン AI 処理(エッジビジョン AI)を実行可能にする技術が重要である。自動運転や Factory Automation 等におけるエッジビジョン AI の実用化に向けた技術開発が進む中、高画素化及び時間当たりの処理イメージ数の増加により、デバイスの負荷は増加傾向にあり、実用化が本格化した時を見据えて、エッジビジョン AI の軽量化技術の早急な開発は喫緊の課題である。

本事業では、ディープラーニングのパラメータだけでなくアクティベーションも量子化・枝刈り等することで、メモリ消費量を大幅に効率化するだけでなく、圧倒的な処理の高速化を目標としている。

開発したエッジビジョン AI の実装対象は、様々なプログラムの搭載を可能とするリコンフィギュラブルデバイス(FPGA)を主とし、エッジビジョン AI を容易に実装可能な開発プラットフォームを整備することで、より高性能なエッジビジョン AI アプリケーションを広く普及させ、日本がエッジビジョン AI のリーダーとなることを目的としている。

エッジビジョン AI の実用化に向けては、従来に比べて遥かに複雑な DNN モデルを消費電力と性能のバランスをとりつつなるべく小さい消費リソースで実現する技術が必要である。エッジビジョン AI を構成するコンポーネントの中でも、特に計算量が多い DNN においては、モデルレベルでの軽量化や専用ハードウェアアーキテクチャによる高速化が不可欠である。

本研究では、モデルレベルでの軽量化アプローチの一環として、畳み込み層や全結合層の重みやアクティベーションを量子化・枝刈り等することで、計算量・メモリ消費量を大幅に効率化する手法を確立する。これにより、ハードウェアへの実装を現実的なものとし、かつ圧倒的な処理の高速化を行う。当然ながら、このモデル圧縮はネットワークの計算精度に影響を与えるため、圧縮・推論・テストを反復して精度検証を行う必要がある。本研究では、効率の良い圧縮モデル

を素早く構築するために、既存のディープラーニングフレームワークとシームレスに連携可能な形でモデル圧縮技術の実装を行うものとする。

現在、深層学習においては学習・推論問わず NVIDIA の GPU を用いるのが主流となっている。これはディープニューラルネットワークの計算には大量の浮動小数点演算が必要であったためだが、シンプルなビット演算だけでニューラルネットの計算が実現できるのであれば、専用に設計されたハードウェアで、GPU より高速に、効率よく計算できる可能性があり、結果として実現されるハードウェアの回路規模も大幅に小さくなりうる。しかし、専用チップの製造には年単位の時間と相応のコストが掛かり、量子化の方法も進化があることを考えれば、回路自体を再構成可能な FPGA の上にハードウェアとして実装するというアプローチも解になりうる。

今後の深層学習を行うデバイスについては、CPU、GPU、FPGA、あるいは専用チップかの4択がありうる。それぞれの選択肢は排反ではなく、例えば CPU が深層学習用命令を実装することで実質的に専用チップの機能を取り込む、という可能性もある。また、学習と推論で、例えば、学習は GPU、推論は FPGA のように、別々のデバイスがドミナントとなる可能性もある。

現在、FPGA で動作するアプリケーションを開発する際には、Verilog HDL などのハードウェア記述言語か、C/C++のようなプログラミング言語を拡張した高位合成言語で記述する必要があり、多くのエンジニアが簡単に活用できるレベルにはない。また、高位合成言語については、ベンダ固有の拡張を使用して最適化を行う必要があり、これに数ヶ月～1年といった期間を要するため、FPGA は GPU に比べて処理速度や消費電力の点で優れている面があるにも関わらず、その性能に見合うほど普及が進んでいない。これに対し NVIDIA は、並列プログラミング言語、コンパイラ、デバッグ、チューニングされた多数のライブラリからなる、GPU 開発環境である CUDA[1]を提供することで、開発者の裾野を広げ GPU によるエコシステムの構築に成功した。こうした開発環境の整備が FPGA の普及には重要である。

一方、自動運転などのアプリケーション側の視点から見れば、グローバル市場で AI ソリューションをテコにしてシェアを取っていくには、センサー技術や AI フレームワークが日進月歩で実用化が進む中、その時々最適なマルチプロセッサの環境に短期間で実装することが重要となる。現時点では、FPGA に短期に簡単に実装できることができれば、競合製品よりも早く高性能のアプリケーションを市場に出すことができ、グローバル市場を取る上で大きなアドバンテージとなる。

本研究では、ドメイン固有言語とそのコンパイラ、そしてベンダ非依存の高位合成技術を組み合わせ、FPGA を主軸としつつ、多様なハードウェア環境でも動作可能なエッジビジョン AI の開発プラットフォームの実現を目指す。

消費電力・処理能力に制約のあるエッジコンピューティングでは、アプリケーションの高速化自体が大きな差別化要因となりうるため、エッジビジョン AI を超軽量化し短 TAT で簡易に開発・実装できる、一貫通貫な開発プラットフォームを作ることで、世界的に先進なエッジビジョン AI アプリケーションを世に先駆けて実現することを狙う。動作環境としては、GPU のみならず、消費電力性能比が高く再構成可能な FPGA を主対象にする。

プロジェクトの進め方として、一貫通貫する最低限の機能の実現を優先し、イテレーションにより機能を拡充していく適応的なスタイルをとる。たとえば、ディープラーニングフレームワークとの統合に関しては、現在一定のシェアを取っているもの(TensorFlow[2], PyTorch[3]等)のうち一つを選定し、FPGA に関してもユーザーの多い大手ベンダ製品(Xilinx 製 FPGA, Intel 製 FPGA 等)から適用を進め、徐々に多様なフレームワークやハードウェアに適用していく。

また、最終的な想定ユーザ企業を巻き込み、開発プラットフォームとアプリケーションの評価を行うことで、継続的に利活用されるような高い事業性を確認する一連の流れを重視する。

以上の通り、技術と事業の開発を行うことで、エッジビジョン AI が Early Majority のピークに達する頃には、エッジビジョン AI の実装を、ソフト面・ハード面からグローバルでリードできていることを期待している。

3.7.2 最終目標と根拠

本研究開発では、最終目標として、ドメイン固有言語とそのコンパイラ、そしてベンダ非依存の高位合成技術を組み合わせ、FPGA を主軸としつつ、多様なハードウェア環境でも動作可能なエッジビジョン AI の開発プラットフォームの実現を目指す。

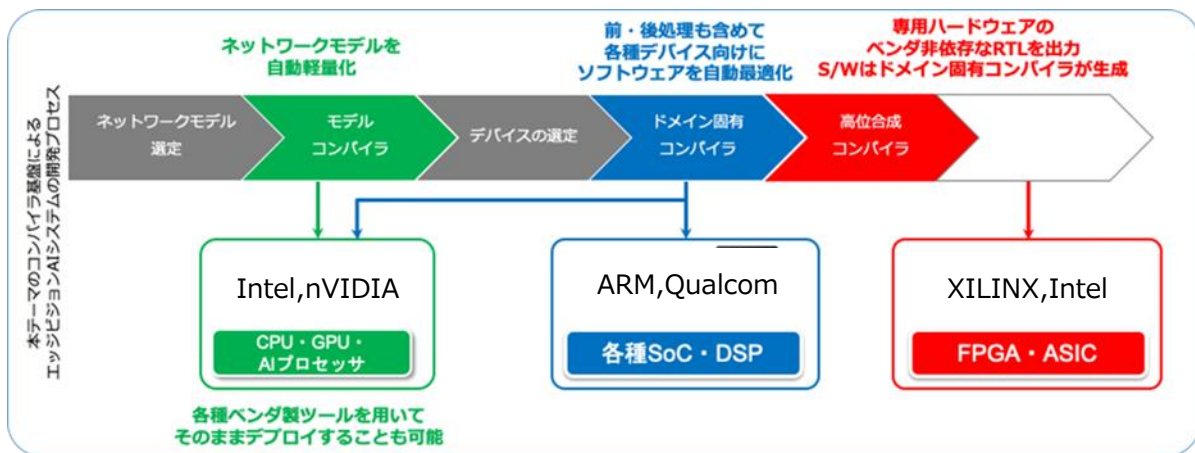


図 3.7.2.1 本テーマのコンパイラ基盤を用いた、エッジビジョン AI システムの開発プロセス

実施項目ごとの最終目標は以下の通りである。

実施項目1：DNN(Deep Neural Network) の超軽量化技術・実装技術の開発

汎用的に利用されている DNN ネットワークモデルの多くは、コンシューマやワークステーションなどの比較的高性能なコンピュータ上で学習・推論されることを前提に設計されている。そのため、それらのネットワークモデルをそのままエッジデバイス上でリアルタイムに動作させることは、性能や必要メモリ量の観点から現実的でないことが多い。

また汎用的な DNN フレームワーク上で実装されるネットワークは、様々な認識タスクやデータセット、クラス数、精度の要求を満たせるように設計されていることが多い。このような性質を持つため、特定のタスク向けの学習を完了した後、推論にのみ利用する場合には、学習パラメータやビット精度を圧縮出来る可能性が存在する。

更に、DNN の学習や推論には GPGPU が利用されることが多いため、GPGPU 上で実行した場合に高い計算効率やメモリ帯域が出る処理が利用されることが多い。このため、同じ DNN ネットワークモデルを CPU や FPGA 上で動作させる場合は、そのプロセッサアーキテクチャの特性の違いから必ずしも高い効率が出せるとは限らず、それぞれのプロセッサアーキテクチャに応じたネットワーク構成や実装の最適化を行う必要がある。

これらの課題に対処するため、利用するタスクの精度や、実行対象のエッジデバイスの性能や搭載メモリ量、プロセッサアーキテクチャなどの特徴に応じて、DNN ネットワークモデルを設計したり、モデルの圧縮や推論処理の近似的あるいは等価的な変換を行う必要がある。このようなモチベーションから、本実施項目では、DNN フレームワーク上で実装されたネットワークモデルに対して、エッジデバイスで実行するための軽量の DNN ネットワークモデルの開発や、量子化や枝刈りなどの技術を利用したモデル圧縮技術についての研究開発を行う。また、汎用的な DNN フレームワークで設計されたモデルに対して、先述の研究開発で得られたモデル圧縮技術を自動で適用するフレームワークの開発を行う。

数値目標として、TensorFlow と PyTorch の DNN フレームワーク上で実装されたネットワークモデルを FPGA 向けにコンパイル可能とし、更に汎用的かつモデル圧縮を何も適用していない同等精度のネットワークモデルと比較して、1/100 の計算量および使用メモリ量の軽量化を目指す。

実施項目2:ドメイン固有言語コンパイラの開発

演算性能やメモリ帯域量の制約が厳しいエッジデバイス上でのリアルタイム処理を実現するためには、動作対象のプロセッサ上での実行効率を最大限に発揮出来るように、プログラムを実装する必要がある。プログラム高速化のためのプログラム変換手法として、マルチコアや SIMD 演算器、スーパースカラなどのプロセッサ内に並列に配置された演算ユニットの活用、ループ変形などによるプログラム内の命令の実行順序の変更、変数や配列などのメモリ配置方法の変更によるキャッシュメモリやスクラッチパッドメモリに代表される高速低容量の高効率利用などが挙げられる。しかしながら、これらのプログラム変換の適用戦略は、対象とするプログラムのアルゴリズムやプロセッサアーキテクチャ、搭載メモリ量などといった各種メトリクスと、ユーザの要求性能に応じて変化する。また、本研究開発においてターゲットデバイスの主軸としている FPGA では、デバイスに搭載されている各種ハードウェアリソースの使用量と性能のスループットの間トレードオフが存在するため、プログラム変換におけるパラメータ空間が、CPU や GPGPU などの汎用プロセッサ向けのソフトウェア実装の場合と比較して、更に増加する背景が存在する。

これらのプログラム変換におけるパラメータの無数の組み合わせに対して、プログラマが個別にエッジビジョン AI のアプリケーションの性能最適化を個別に行うことは、工数の観点から非現実的である。したがって、コンパイラ最適化に代表されるようなプログラム自動最適化技術を積極的に用いる必要がある。

一方で、C/C++などに代表される汎用的なプログラミング言語では、入力とするプログラムの表現能力の高さや曖昧さから、実装されたソースコードとプログラマが意図している仕様や制約との間にギャップが生じている場合が多く存在する。このために汎用言語のコンパイラはプログラマが意図している詳細な制約をソースコードから読み切ることが出来ず、保守的なプログラム最適化を適用せざるを得ない場合がしばしば存在する。

これらの課題を解決するために、本実施項目ではエッジビジョン AI アプリケーションに特化したドメイン固有言語で実装されたプログラムを、FPGA を始めとした様々なターゲットデバイス向けに最適化し、コード生成を行うコンパイラを開発を行う。ドメイン固有言語を採用することで、プログラマは汎用言語と比較してより少ない工数でアプリケーションを実装することが可能となり、コンパイラは制約された言語仕様から得られる情報を用いてより積極的なプログラム最適化を適用することが可能となる。しかしながら、ドメイン固有言語とコンパイラをスクラッチから開発することは工数上現実的ではないため、画像処理向けドメイン固有言語の OSS である Halide[4][5]をベースとして利用し、これらに独自の FPGA 向けのバックエンドを追加実装することで実現することとする。

本実施項目で開発したドメイン固有言語コンパイラによって、2022 年度時点での最先端の研究結果における高位合成を用いて手動で記述した実装と比較して、匹敵する程度の性能及び消費リソース量を、1/10 以下の実装工数で実現出来ることを数値目標とする。

実施項目3:ベンダ非依存のオープンソース高位合成基盤の開発

FPGA や ASIC などに搭載されるハードウェア回路を実装する場合、従来では RTL(Register Transfer Level)で記述されるハードウェア記述言語を利用することが主流であった。しかしながら、RTL 記述での実装では、演算処理のパイプラインスケジューリング、演算ノードに対する演算リソースの割当、FSM(Finite State Machine)などの様々なロジックの実装が必要となるた

め、多大な開発工数とハードウェアに対する理解の深い開発者の存在が不可欠となっていた。

そのような背景から、C/C++言語などの動作レベルのソースコード記述から回路を自動生成する高位合成技術が誕生し、現在では実用的に利用されるようになってきた。各種 FPGA ベンダも、Xilinx 社の VivadoHLS[6]、Intel 社の IntelHLS[7]などといった高位合成ツールを提供している。しかしながら、これらの各種ベンダが提供している高位合成ツールは、しばしば各ベンダに依存した IP を用いた回路を生成するため、事実上特定のベンダの FPGA 向けの回路しか生成することが出来ない。またこれらの高位合成ツール間では、入力となる C/C++言語のサポートや生成されるハードウェア回路のポリシーが、ハードウェア回路の各種最適化の指定に利用される指定子の仕様などが異なるため、同一の C/C++コードを各種高位合成ツールに入力して、各 FPGA ベンダ向けの回路生成をすることは不可能となっている。また、既存では FPGA を利用していた回路を ASIC などの専用回路に実装する際も、それらの高位合成ツールを前提にした再実装が必要となる。

この課題を解決するため、本実施項目ではベンダに非依存な高位合成基盤の開発を行い、様々なベンダの FPGA 及び ASIC 向けの回路の生成の実現を目指す。この高位合成基盤は、東京大学高前田研究室で開発されている Veriloggen[8]をベースとして開発を行う。また、実施項目 2 において開発するドメイン固有言語コンパイラのバックエンドと接続し、エッジビジョン向けのアプリケーションを様々なベンダの FPGA 上で動作可能とすることを目指す。本実施項目で開発した高位合成基盤によって大手 FPGA ベンダである Xilinx 及び Intel 社の FPGA 向けの回路の生成を実現することを目指す。更に Xilinx の FPGA に対しては、ドメイン固有コンパイラの既存の Xilinx 向け FPGA のバックエンドと比較して、同等の性能、アーキテクチャ、リソース量を持つ回路が生成できることを数値目標とする。

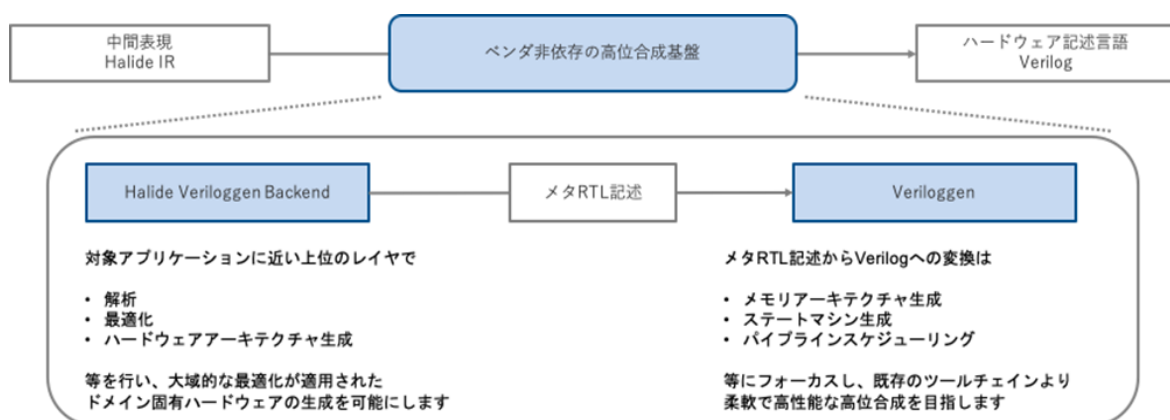


図 3.7.2.2 本実施項目で開発する、ベンダ非依存高位合成基盤を利用したコンパイルフロー図

研究開発目標の設定根拠

エッジビジョン AI については、技術の有効性が検証され、現在そのアプリケーションと実装環境が日進月歩で進化している。処理データ量が増える中、リアルタイム環境での推論が求められるようになると、モデルを軽量化することが重要になるが、実用化に耐えうる軽量化にはまだ課題がある。エッジビジョン AI で世界的に競争力を持つためには、精度高く超軽量化を行う技術と、時間を稼ぐという意味で短 TAT で実装できる開発環境の整備が不可欠となる。

本研究開発は4つの点において、国内外で優位性があると考えている。

- ① AI フレームワークも日々進化しているが、エッジビジョン AI を想定して実用化に耐えうる軽量化による精度が保っているケースは少なく、これが実現できれば、今後増えうるセンサーフュージョンをベースとした IoT の世界において大きな競争優位となる。

- ② エッジビジョン AI はその処理速度と消費電力の制約、そしてリコンフィギュラブル性から FPGA が主流になると見られているが、FPGA の趨勢も今後変わりうる。その中、新しいリコンフィギュラブル・ハードウェアに対応し、移植・新規開発を行える開発環境を整備することが競争優位となる。FPGA メーカーが提供する高位合成ツールは、将来的に発展し、エッジビジョン AI にも対応していく可能性もあるが、こうしたメーカーが競合製品を含めたヘテロジニアスなシステム構成に対応することは、FPGA メーカーがハードウェア販売を事業の中心としている以上、考えにくく、システム構成の柔軟性は将来的にもないと考える。この点、本研究開発は、将来的にハードウェアが変更になっても、短 TAT で適用することを主眼においており、FPGA メーカーのツールよりも汎用性の点で優れている。
- ③ FPGA 等のハードウェアの構造は各社各様であり、これらを効率的に実装するには、数多くの実装経験が製品の完成度を高める上で重要となるが、日本ではフィックスターズ社がこの分野では最も実績のある企業であり、本研究開発で最新技術を研究する大学研究機関とユーザ企業のノウハウを併せることで、海外にも通用する製品作りができる。
- ④ エッジビジョン AI を適用するユーザ企業としては、自動車・産業機械・医療機器などが想定されるが、この分野において日本には世界シェアを持つ企業が未だ多数ある。今回、ユーザ企業として、こうした企業を巻き込んだ研究開発プロジェクトとすることで、加速的に普及を図ることが可能である。

現在想定する競合との競争優位性の比較については、以下のとおりである。

技術名称	技術保有者	年月	性能① 軽量化	性能② 汎用性	品質・機能 等の強み	エコシステム、 ビジネスモデル の強み	コスト	全体 市場 規模	獲得市 場規模 と市場 シェア	総合評 価 (LD、 DH、RA)
提案技術 (軽量化し短 TAT で実装する技 術)	本技術(現状)	2018/6	△	-	軽量度	フレーム ワーク・チッ プの移行を 促す広域の エコシステム	○	-	-	RA
	本技術(事業終了時)	2020/3	○	○	高速性軽量 度汎用性		○	△	低	LD
	本技術(実用化時点)	2022/4	◎	○	高速性軽量 度汎用性		○	○	日本 高 世界 低	LD
	成果普及段階	2027/4	◎	◎	高速性軽量 度汎用性		○	◎	日本 高 世界 中	LD
NVIDIA 社 AI 実装技術 (TensorRT)	本技術(現状)	2018/6	×	×	一貫性	自社エコシ ステムでの 囲い込み	○	◎	高	LD
	本技術(事業終了時)	2020/3	△	×	一貫性		○	◎	高	LD
	本技術(実用化時点)	2022/4	△	×	一貫性		○	○	高	DH
	成果普及段階	2027/4	△	×	一貫性		○	○	高	DH
ThinCI 社 AI 実装技術 (reVision)	本技術(現状)	2018/6	○	×	一貫性	自社エコシ ステムでの 囲い込み	○	△	低	DH
	本技術(事業終了時)	2020/3	○	×	一貫性		○	○	中	DH
	本技術(実用化時点)	2022/4	△	×	一貫性		○	○	中	DH
	成果普及段階	2027/4	△	×	一貫性		○	○	中	DH
DeepPHI 社 AI 実装技術 (DNNC)	本技術(現状)	2018/6	△	×	自社 AI ボード実装	中国市場で の製造工場 との連携	△	×	低	RA
	本技術(事業終了時)	2020/3	△	×	自社 AI ボード実装		△	×	低	RA
	本技術(実用化時点)	2022/4	△	×	自社 AI ボード実装		△	×	低	RA

	成果普及段階	2027/4	△	×	自社 AI ボード実装		△	×	低	RA
--	--------	--------	---	---	----------------	--	---	---	---	----

- ◎・・・業界トップ水準
- ・・・競合よりも優位なレベル
- △・・・競合と同レベル
- ×・・・競合よりも劣るレベル
- LD・・・Leading
- DH・・・Dead Heat
- RA・・・Run After

3.7.3 目標の達成度

当初提案での中間目標に掲げた通り、エッジビジョン AI を初期対象フレームワークで実用化に耐える精度で超軽量化し、汎用 FPGA に短 TAT で実装する技術のベースを確立しつつある。実施項目ごとの達成状況は以下の通りである。

実施項目 1 : DNN(Deep Neural Network) の超軽量化技術・実装技術の開発

DNN の推論処理で頻用される畳み込み演算を、図 3.7.3.1 のようにチャンネルシフト演算と 1x1 畳み込み演算に分解し、従来の畳み込み演算と等価に演算する方法を開発した。従来の DNN モデルでは、3x3 や 5x5 の畳み込み処理が利用されることが多く、大量の積和演算回数とメモリアクセスを必要とするため、計算量とメモリ使用量の両面で支配的となる処理であった。本手法では畳み込み演算の範囲を 1x1 で行うことが出来るようになるため、計算量とメモリ使用量を大幅に削減することが可能となる。一方で、従来の畳み込み演算から新たに必要となるチャンネルシフト演算は、ハードウェア化と親和性が高いため、対象ハードウェアが FPGA となる場合には非常に小さいコストで実装することが可能となる。

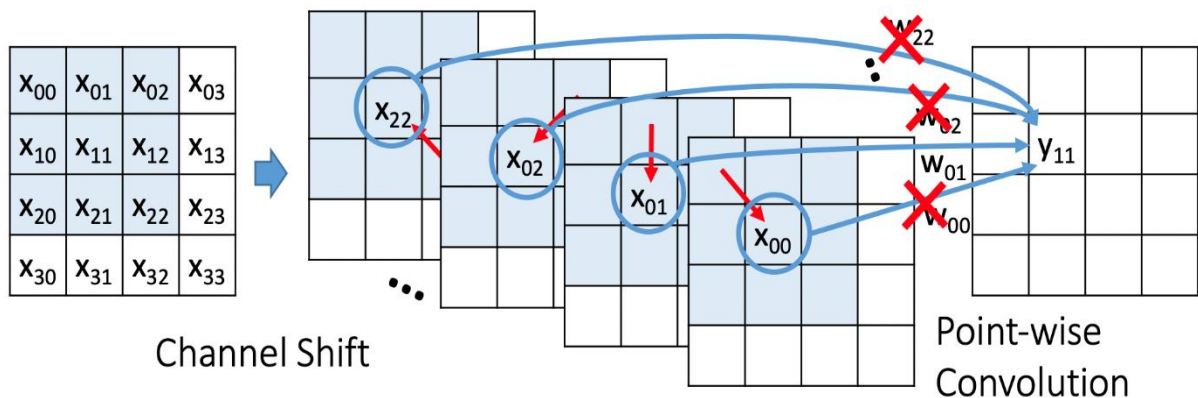


図 3.7.3.1 畳み込み演算から、チャンネルシフト演算と 1x1 畳み込み演算に分解するイメージ

また、本提案手法の畳み込み演算を組み込んだ、軽量化 DNN モデル「SSS-Net」を開発した。表 1 は SSS-Net と既存の DNN モデルの間で、ImageNet2012 データセットのクラス分類タスクに対して、パラメータ量や積和演算量、認識精度を比較した結果である。この結果により、既存の AI モデルと比較してパラメータと計算量を削減しつつ、ほぼ等価な認識精度を達成することが確認された。

表 3.7.3.1. SSS-Net と既存の DNN モデルとの比較結果 (ImageNet2012 Dataset を使用)

Model	#Layers	Params [MB]	GMACs	Top-1 Acc.	Top-5 Acc.
AlexNet	8	60.97	7.270	56.5	79.1
VGG16	16	138.36	154.700	71.6	90.4
ResNet50	50	25.56	3.870	73.2	91.3
MobileNetv2	18	3.47	0.300	71.8	91.0
SSS-Net×0.5 (Ours)	18	0.13	0.008	64.2	85.9
SSS-Net×1.0 (Ours)	18	0.80	0.032	68.0	88.2
SSS-Net×2.0 (Ours)	18	14.31	0.128	70.0	89.9

実施項目2:ドメイン固有言語コンパイラの開発

従来のドメイン固有言語コンパイラの FPGA 向けバックエンドでは、生成するハードウェアリソース量をコンパイル時に確定しなければならなかったため、固定解像度の入出力データに対してのみコンパイル可能であったという制約が存在した。これに対して、より広範なアプリケーションやユースケースに対応するために、可変解像度の入出力データに対応する改善実装を行った。可変解像度版では、入出力データの最小最大サイズ情報を Halide のスケジューリング指定として入力し、その制約のもとで必要最大限なハードウェアリソースを生成することによって、コンパイル時にハードウェアリソース量を確定させた。

本機能の性能評価のために、6 種類の処理からなる撮像処理アプリケーションを対象とし、従来の固定解像度版と、新たに開発した可変解像度版の間で、ベンチマーク比較を行った。ベンチマーク結果を表 3.7.3.2 に示す。ベンチマークの結果、固定解像度版と比較して、僅かなりソース使用量の増加で、同一性能のハードウェアが生成できることが確認できた。したがって、本開発によって、より実用的なアプリケーションやユースケースに対応することが実現できた。

表 3.7.3.2. 撮像アプリケーションに対する、固定解像度版と可変解像度版でのベンチマーク比較結果

		固定解像度版 (Old)	可変解像度版 (This Work)
消費FPGA リソース	BRAM18K	24	32
	DSP48E	19	18
	FF	9,593	9,796
	LUT	10,477	11,397
周波数		150MHz	150MHz
性能		FullHD 60FPS	FullHD 60FPS

また、DNN の前処理などで行われることの多い、画像のフィルタ処理やリサイズ処理についても、本研究開発によって開発されたドメイン固有言語コンパイラによって生成された FPGA 向け生成コードと、手動で FPGA 向けに最適化された OpenCV 実装[9][10]のベンチマーク比較を、同一のデバイス、性能及び周波数のもとで行った。その結果、LUT、FF、DSP などのリソース量は手動実装時より多いリソースを消費量となってしまうが、BRAM は手動実装に比べて 76%のリソース消費量が達成できていることが確認された。また、処理時間のレイテンシは、すべてのテストケースにおいて、手動実装より低く優れたデザインが生成できていることが確認できた。このようなリソースと性能を達成しつつ、プログラマが実装すべきソースコード量は約 1/10 に削減できていることも確認された。

これらの結果により、本プロジェクトで開発しているドメイン固有言語コンパイラは、手動で最適化されたデザインと比べて、わずかにリソース消費量が多い傾向はあるものの、より低いレイテンシのデザインを、遥かに少ない工数で達成することが実現出来ていると言える。以降の研究開発でリソース消費量を更に最適化することにより、2021 年度終了時点で、手動最適化実装と比較して同等のリソース消費量を達成可能な見込みとなっている。

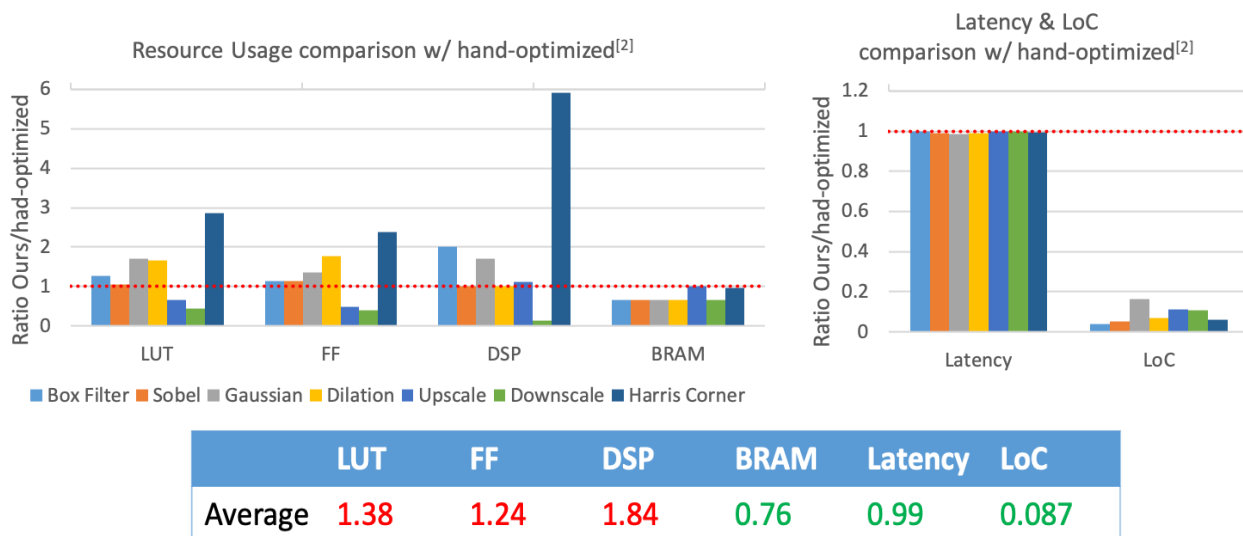


図 3.7.3.2 ドメイン固有言語コンパイラの FPGA バックエンドが生成したハードウェア実装と、手動最適化されたハードウェア実装の、リソース使用量、性能、ソースコード量の比

実施項目3:ベンダ非依存のオープンソース高位合成基盤の開発

実施項目 2 で開発中のドメイン固有言語コンパイラに対して、Veriloggen 向けのバックエンドのプロトタイプ実装とテストを通して、Veriloggen に対して新たに必要な機能の洗い出しと、コンパイルパスにおける変換戦略の検討を行った。

その結果、Veriloggen に対しては大きく下記の 3 つの機能が新たに必要ということが分かった。

1. データ再利用用の中間バッファ (ScratchPad, RingBuffer)
2. 外部モジュールとの接続 (FromExtern/ToExtern)
3. AXI4-Stream 及び FPGA 内部の FIFO バッファとの接続の対応

上記のうち、1 と 2 に関しては対応が完了し、現在 3 の機能追加の検討及び実装を進めている。

また、Veriloggen 対応のために必要なドメイン固有コンパイラ側のコンパイル機能の洗い出しも一通り完了した。現在は Veriloggen 向けのバックエンドの本実装を開始しており、全テストモジュールのうち、33%のテストモジュールが本コンパイルフローを、ベンダ非依存な FPGA 向けにコンパイル出来ることが確認できている。2020 年度終了時点で基本テストの全モジュールが Veriloggen バックエンドを通じてコンパイルし、ベンダ非依存な FPGA 向けの回路生成が達成可能な見込みとなっている。

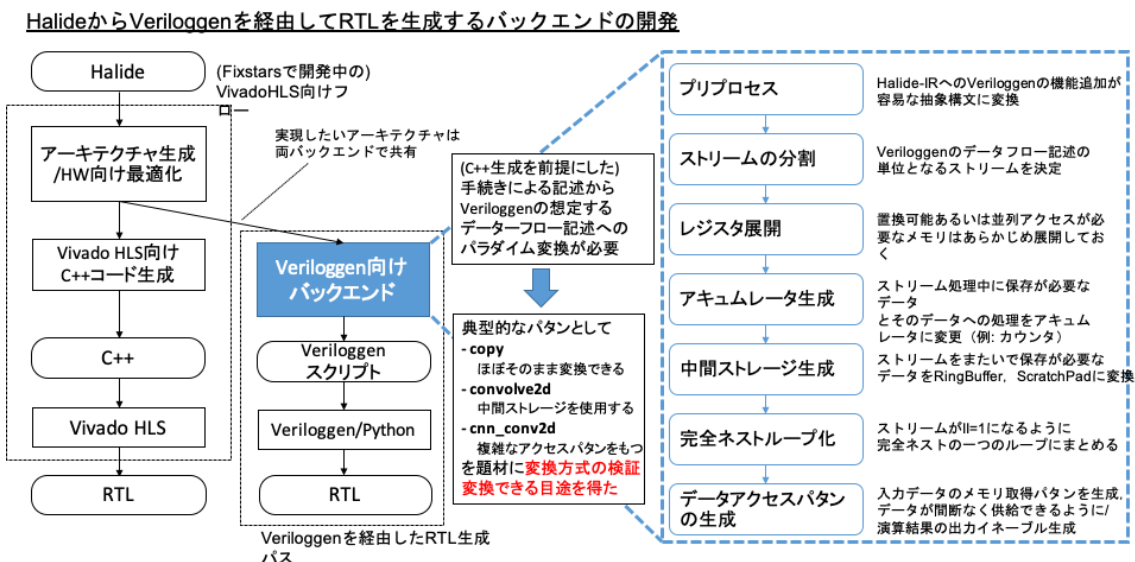


図 3.7.3.3 Halide Veriloggen バックエンドのコンパイルフロー設計図とプロトタイプ結果

3.7.4 成果と意義

当技術を利用することで、限られたハードウェアリソースの中で、より情報量が多くより複雑な機能を有するエッジビジョン AI を実現することができ、また、開発にかかる期間を大幅に短縮化し、極めて短期間でハードウェア実装まで行えるようになる。

現在の AI 処理では、豊富な開発環境を用意している欧米の巨大チップメーカー製の汎用的チップの利用が主流だが、当技術を活用することでエッジビジョン AI を実装した専用チップを極めて短期間で開発できるようになり、高性能な専用チップの開発・利用促進につながる。

エッジビジョン AI の活躍の場は広く、自動運転や Factory Automation、ドローンといった成長著しい分野での利用が見込まれている。

最終目標に基づく成果及びその社会的意義は以下の通りである。

◆技術レベル

エッジビジョン AI を超軽量化し、短 TAT で実装する環境プラットフォームについて主要 AI フレームワークから主要なヘテロジニアスなリコンフィギュラブルデバイスに実装する技術を確立する。

◆エネルギー消費効率/電力効率

現状のエッジビジョン AI に比べてエネルギー消費効率が 100 倍となる実装開発環境プラットフォームの利用が始まることで、2037 年の CO2 削減目標の実現に向けて CO2 削減が進んでいる。

◆技術・性能に係る目標

エッジビジョン AI を FPGA に実装すると 10 倍以上のエネルギー消費効率となるが、その実装開発環境プラットフォームが利用されている。

◆2037 年温室効果ガス排出削減効果への貢献

◆その他のアウトカム

エッジビジョン AI が超軽量化でき、短 TAT で容易にリコンフィギュラブルなハードウェアに実装できることにより、日本のプロセッサ及びアプリケーションの競争力が向上する効果が期待できる。

こうした環境プラットフォームを整備することで、リコンフィギュラブルなハードウェア (FPGA 等) の処理スピードの高速化並びに処理規模の増大化においても消費電力並びにコストを抑える事が現される事により以下の波及効果がある。

①発熱量低減

1-1 IoT エッジデバイス

熱源による機能劣化するデバイス(例:CMOS イメージセンサ等のセンサー)の近くに配置可能、サイズのコンパクト化によるアプリケーションの柔軟性拡大

1-2 IoT クラウドデバイス

放熱を必要としない事によるデータサーバ等での高額な放熱機構を必要としない為、データセンターの小面積化や電力削減によるCO2削減にも大きく寄与できる。

② 高速処理

非ノイマン型アーキテクチャーによりニューロコンピューティング、並列演算用メモリプロセッサの実現化の可能性もあり、将来のIoT、ビッグデータ社会に役立てられる。

◆我が国の経済再生への貢献

日本の輸出(年約70兆円)を品目別で見ると、電気機器は、輸送用機器、一般機械、に次ぐ12兆円の規模があり、半導体製品はそのうち4兆円弱と、テレビ・液晶産業が台湾・韓国勢に覇権を取られた今、依然中心的な製品となっている。その半導体も、NANDを中心としたメモリが主体で、かつて一世風靡したLSI・論理回路は、ここ20年でWindow/Intel、最近ではNVIDIAなどの米国企業やファブレスでデザインのみを行うアジア企業にシェアを取られ、日本企業のグローバルでのシェアは年々低くなり、Logic ICは2017年には生産が1500億円(生産動態統計)にまで大幅に減少している。

こうした中、本研究開発の成果は、①産業のサービス部分の価値が高まる中、超軽量化した最新エッジビジョンAI技術を短TATで実装した自動車、産業機械、医療機器等でのアプリケーションが国際競争力を増して輸出増加に繋がることへの貢献、②本研究開発に統合された日本のASIC等の開発環境が向上し、グローバルで利用されることによる半導体の輸出増加が期待できる。

半導体の主要応用製品分野は、JEITAによると全体で製造業の44%を占めている。製造業は日本のGDPの21%(109兆円)で、半導体の主要応用製品分野は日本のGDPの約10%を占めており、本研究開発により製品開発を先進的にできれば、その波及効果は大きいと考えている。

3.7.5 成果の普及

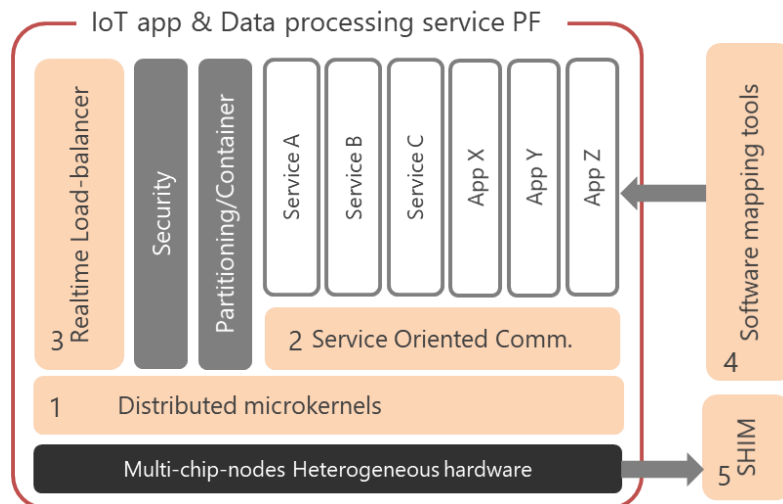
株式会社フィックスターズでは、2章にて後述するエッジビジョンAIシステムのプロトタイプを行うSaaSサービスを通して、本研究開発で研究開発されたDNNモデルのアルゴリズム及び、コンパイラによって生成されたコードをユーザへ提供を計画しており、現在は本研究開発で得られたコンパイラ技術のサービスへの導入作業を進めている。

3.8 研究開発テーマ「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」

本研究開発テーマ「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」では、「自動運転をはじめとするインテリジェント化するエッジシステムに不可欠となるソフトウェアプラットフォームを開発し、ヘテロジニアス・マルチ・メニーコアを最大限に活用できる高効率で安全なソフトウェアの開発を可能にし、欧米に先行されているソフトウェア基盤技術領域において世界トップクラスの成果を目指す。」という目標に対し、各実施項目の目標を全て達成し、成果を上げつつある。

【実施項目】

1. 分散マイクロカーネル OS の研究開発(分散 MK OS)
2. 分散 MK OS 内蔵 SOC の研究開発
3. Hybrid-scheduling/Load-balancing アルゴリズムの研究開発
4. ソフトウェアマッピング支援ツールの研究開発
5. ヘテロジニアス・マルチ・メニーコア向けハードウェア記述標準モデル SHIM の研究開発と国際規格化



実施項目 1~5
 図 3.8.1 本研究開発テーマ全体の取り組み（概念図）

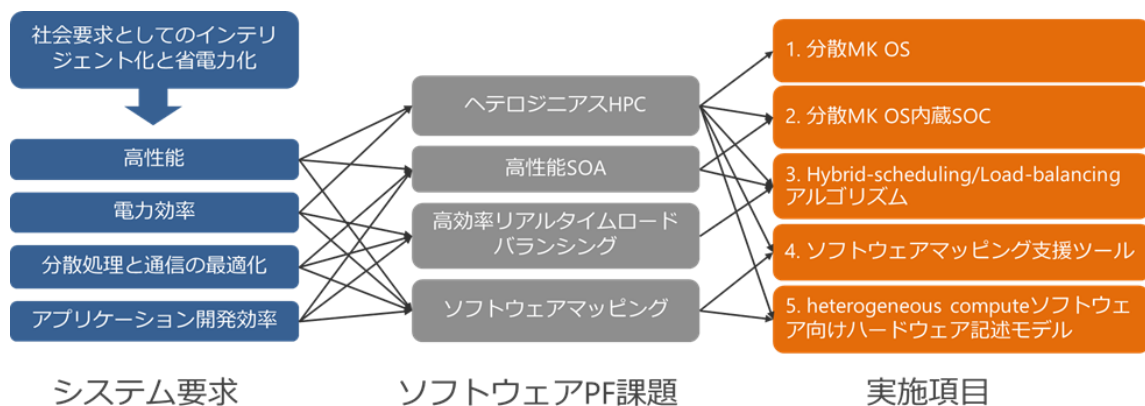


図 3.8.2 実施項目と課題の関係

個別の研究開発テーマにおける詳細は以下に示す。なお、実施項目1と2はその密接な関係上、「3.8.1」項にまとめて記載する。

3.8.1 研究開発サブテーマ①「分散マイクロカーネル OS の研究開発 (分散 MK OS)」、及び、サブテーマ②「分散 MK OS 内蔵 SOC の研究開発」

3.8.1.1 概要

分散マイクロカーネル OS(分散 MK OS)アーキテクチャにより従来シングルカーネル技術のヘテロジニアス・マルチ・メニーコアにおける課題の解決を図る。また、分散 MK OS にメッセージパッシングを統合した SOC(Service Oriented Communication)により、サービスのディスカバリ、ヘテロジニアス・マルチ・メニーコアにおける異種コア、マルチチップ間のメッセージパッシングを分散 MK OS に内蔵し、ミドルウェアによる従来技術の多重構造の通信を大幅に効率化する。

3.8.1.2 最終目標と根拠

ヘテロジニアス HPC:OS 自体が高い並列性を持ち、ヘテロジニアス HPC 向けに根本から再設計された各プロセッサ毎に独立した MK(Microkernel)を配する分散マイクロカーネルアーキテクチャを用いた分散 MK OS の開発することである。高効率ヘテロジニアスコンピューティング対応のために二つの異なるプロセッサアーキテクチャを持つマルチコアチップに跨る分散 MK を構成し、QoS 制御機構を開発し MK メッセージパッシングにおいてスレッド毎に通信帯域の割り当て機能を持つ、チップ内/チップ間/ノード間の統合的動的通信最適化機構を開発し、従来のミドルウェア型メッセージパッシングに対して 10 倍のメッセージパッシング効率を実現する。

本目標の根拠について以下に述べる。今日、一般道で公道実験している多くの自動運転車はトランクに汎用 OS である Linux を搭載した大型ワークステーションコンピュータを積んでおり、その消費電力は 100~1000W 程度であるが、これを 100W 以下に引下げた上で、性能は同等とし、かつ一般道の高度な認識、判断処理を必要とされる 10Hz 以下で処理するには Linux に比べて約 10 倍のリアルタイム性が必要なためである。

3.8.1.3 目標の達成度

2018 年度目標である「高効率ヘテロジニアスコンピューティング対応」は達成済みである。高度自動運転技術開発向けプラットフォームである UltraZ AD に分散 MK OS を移植。また分散 MK の各種 POSIX 関連機能対応、大規模システムで安全性確保のため時間保護機能およびコンテナ機能を開発した。また、UltraZ AD と同様の PCIe インタフェースを持つ NXP 社の S32V と、メニーコアプロセッサである KALRAY 社の MPPA とに分散 MK OS を移植。双方のプロセッサ間で OS 内蔵 SOC が有効に動作することを確認済みである。

2019 年度目標である「QoS 制御機構を開発し MK メッセージパッシングにおいてスレッド毎に通信帯域の割り当て機能のプロトタイプ動作」についても達成済みである。これにより、メッセージ量が多い場合の性能向上に加え、スレッド毎のメッセージバッファの確保と管理方法を開発し、クリティカルなスレッドが他のスレッドのメッセージング状況に影響を受けずメッセージングを可能にする事ができた。

2020 年度目標である「チップ内/チップ間/ノード間の統合的動的通信 最適化機構を開発し、従来のミドルウェア型メッセージパッシングに対して 10 倍のメッセージパッシング効率を実現」に関しては、2018 年、2019 年度を土台として順調に研究開発が進んでいる。特に、分

分散 MK の特徴を最大限に活用しつつ、POSIX や AUTOSAR などの高次元の OS 標準インタフェースを実現するために、Lock-free アルゴリズムを最大限に活用し Memory Allocator や mfuture (multi-kernel 版 future) などの仕組みも加えて開発することで、最先端 OS である Linux をリアルタイム性のみならずスループットにおいても上回る性能を示している。

3.8.1.4 成果と意義

分散 MK OS については、MIT (FOS)、ETHZ (Barrelfish) など海外でマルチカーネルの研究は行われているがまだ実用化には至っていない。またミドルウェア層としての SOC は既存だが MK レベルでのメッセージパッシングは前述の FOS などで研究段階であり実用化されていない。本研究開発によって、世界で初となるヘテロジニアス・マルチ・メニーコアに対応したトップクラスの性能を持つ分散マイクロカーネル OS が実現することになり、特にハイパフォーマンスかつ低消費電力を必要とし、かつ高いスケラビリティが要求される今後のエッジコンピューティングシステムにとって、基盤となる技術の一つを確立することが出来ると考えている。

3.8.1.5 成果の普及

本研究開発の成果はプロジェクト終了後速やかに製品化され事業化される計画である。

3.8.2 研究開発サブテーマ「Hybrid-scheduling/Load-balancing アルゴリズムの研究開発」

3.8.2.1 概要

特許化済みの Semi-priority based scheduling をベースに異なるプロセッサコアを持つヘテロジニアス・マルチ・メニーコアに対応させ、各プロセッサの性能特性、メモリ特性、スレッドメモリアクセスプロファイル、メッセージパッシング利用プロファイルをロードバランスの計算に取り込むことで高効率なスケジューリングを実現する。

3.8.2.2 最終目標と根拠

異なるプロセッサ/メモリの処理及び電力あたりの通信性能、更に各 Service/アプリケーションの通信プロファイルをパラメータに含み、かつ高速な割込み処理やディスパッチなどリアルタイム制約を考慮した Hybrid-scheduling/Load-balancing アルゴリズムを実現する。

通信プロファイリングを用いた、演算/メモリ/通信の 3 つのプロファイルデータによる統合ロードバランシング機構に時間保証機構を追加し、ロードバランシングによるアプリケーション処理効率において 10 倍の性能を実現する。

本目標の根拠を述べる。自動運転システムなどで要求されるシステムに対するダイナミックな付加変動に対してリアルタイムかつ高効率にハードウェアを稼働させるため、動的な負荷分散を実現するロードバランシングアルゴリズムは不可欠である。ヘテロジニアス・マルチ・メニーコアに対応したアルゴリズムを実現することで、非対応のロードバランサに対比して 10 倍以上の効率を実現することで、プロジェクトゴールでもある 10 倍の電力効率の実現に寄与するためである。

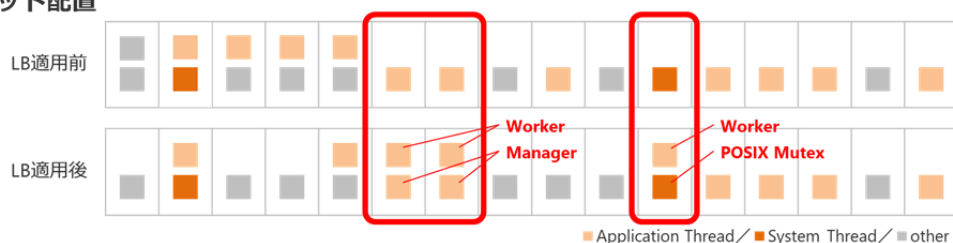
3.8.2.3 目標の達成度

異なるプロセッサ/メモリの処理及び電力あたりの通信性能、更に各 Service/アプリケーションの通信プロファイルパラメータを含み、かつ高速な割込み処理やディスパッチなどリアルタイム制約を考慮した Hybrid-scheduling/Load-balancing アルゴリズムを実現する。

通信プロファイリングを用いた、演算/メモリ/通信の 3 つのプロファイルデータによる統合ロードバランシング機構に時間保証機構を追加し、ロードバランシングによるアプリケーション処理効率において 10 倍の性能を実現する。

本目標の根拠を述べる。自動運転システムなどで要求されるシステムに対するダイナミックな付加変動に対してリアルタイムかつ高効率にハードウェアを稼働させるため、動的な負荷分散を実現するロードバランシングアルゴリズムは不可欠である。ヘテロジニアス・マルチ・メニーコアに対応したアルゴリズムを実現することで、非対応のロードバランサに對比して 10 倍以上の効率を実現することで、プロジェクトゴールでもある 10 倍の電力効率の実現に寄与するためである。

・スレッド配置



・性能向上効果

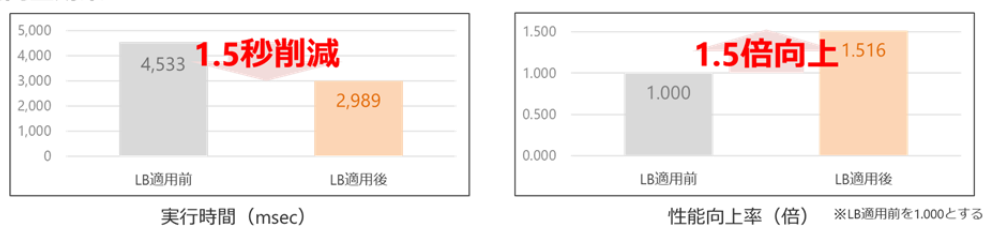


図 3.8.2.3 スレッド配置と性能向上効果

2019 年度目標である「演算/メモリ/通信の 3 つのプロファイルデータによる統合ロードバランシング機構のプロトタイプ動作」についても達成済みである。

2020 年度目標である「統合ロードバランシングの時間保証機構を追加しロードバランシングによるアプリケーション処理効率において 10 倍の性能を実現」については、予定通り研究開発が進んでいる。

3.8.2.4 成果と意義

英ブリストル大、CMU など類似の研究が始まっているがまだ実用化には遠い。本研究開発で Hybrid-scheduling/Load-balancing アルゴリズムを実現することで、統合ロードバランシングを実現するスケジューラを実現することは、エッジコンピューティングのみならず、サーバサイドを含むコンピューティング全体の技術トレンドである SOA (Service Oriented Architecture) で重要となってくる処理負荷のダイナミックなスケールアップのための基盤を実現していくことであり、様々なアプリケーションにおいて、高い性能とスケーラビリティを実現することに大きく寄与すると考える。

3.8.2.5 成果の普及

本研究開発の成果はプロジェクト終了後速やかに製品化され事業化される計画である。

3.8.3 研究開発サブテーマ「ソフトウェアマッピング支援ツールの研究開発」

3.8.3.1 概要

本サブテーマの目的は、サービス指向アーキテクチャを構成する各種ソフトウェアモジュールを、ヘテロジニアス・マルチ・メニーコア上にマッピングするための支援ツール(図3.8.4 参照)を研究開発することである。

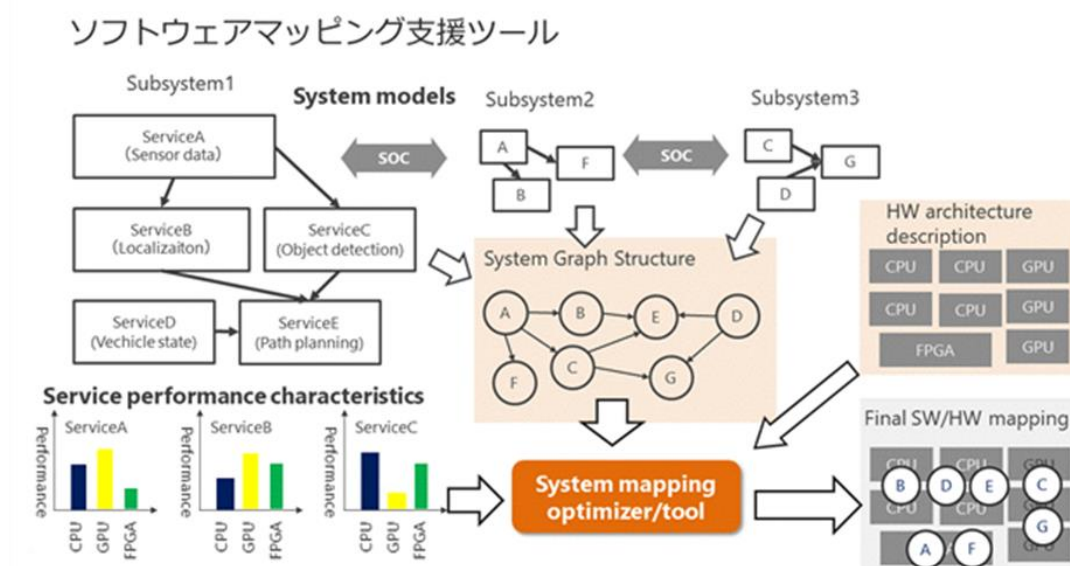


図 3.8.3.1 ソフトウェアマッピング支援ツール

3.8.3.2 最終目標と根拠

本実施項目の目的を達成するために、事業開始時点での当該分野技術水準として、2種類の課題を解決する必要があった。第一に、ヘテロジニアスなプロセッサ環境においてソフトウェアモジュールを最適配置するアルゴリズムの研究開発である。事業開始時点では効率よいアルゴリズムは知られておらず、研究開発の必要があった。第二に、サービス指向アーキテクチャを表現したシステムモデルを入力とするための連携機構の研究開発である。サービス指向アーキテクチャ自体が新規技術分野であるため、連携機構についても新規に研究開発する必要があった。さらに、事業全体で実現するソフトウェアプラットフォームの一部として、これらの研究開発技術をプロトタイプ動作させることを考えた。そのため、本実施項目の最終目標は、「ヘテロジニアスなプロセッサ環境におけるソフトウェアモジュールの最適配置アルゴリズム」および「最適配置アルゴリズムのシステムモデル記述との連携機構」のプロトタイプ動作とした。

3.8.3.3 目標の達成度

2020年6月時点の状況において「ヘテロジニアスなプロセッサ環境におけるソフトウェアモジュールの最適配置アルゴリズム」および「最適配置アルゴリズムのシステムモデル記述

との連携機構」について、それぞれ第一版の単体プロトタイプは完了しており、現在評価、改良中である。また、本事業開始前から名古屋大学資産であったモデルベース並列化(MBP)およびシステムレベル設計環境(SystemBuilder)を拡張し、上述の研究成果を加え、全体を統合したヘテロジニアス・マルチコア向けモデルベース並列化(HS-MBP)設計環境(図 3.8.3.3 参照)を構築中である。さらに現在、最終的な実証に向け、期間終了までに実証完了予定である。そのため、進捗としては順調であり、最終目標に向けた現在の達成度は 80%程度と考えている。

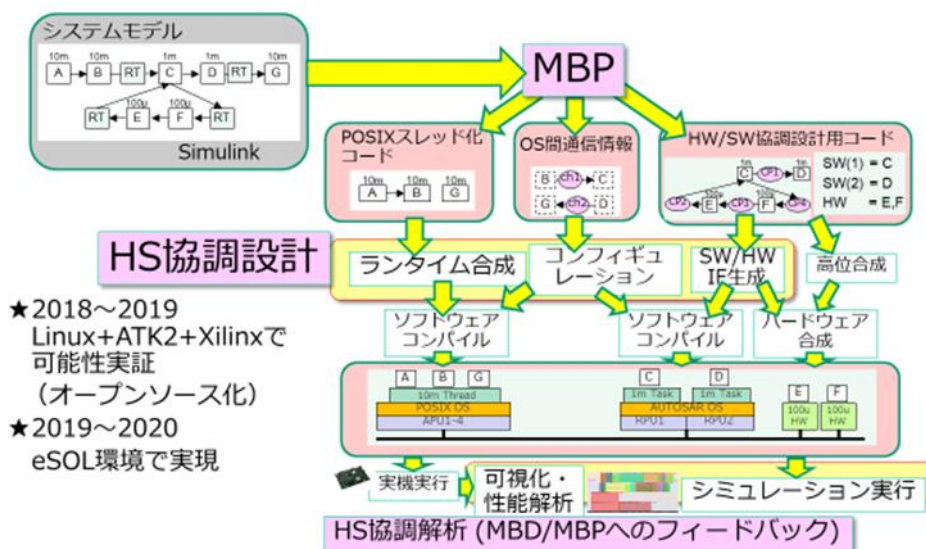


図 3.8.3.3 ヘテロジニアス・マルチコア向けモデルベース並列化(HS-MBP)設計環境

3.8.3.4 成果と意義

ヘテロジニアスなプロセッサ環境におけるソフトウェアモジュールの最適配置アルゴリズム」については、エッジ HPC の特性を活かしたアルゴリズムを提案、国際会議および国際学術論文誌において発表した。この論文の主旨は以下の様なものである。ヘテロジニアス・マルチコア向けソフトウェアマッピングを一般的に定式化した場合、非線形問題となるところ、アクセラレータにおいてのみ高い性能を示す処理のみをマッピング候補とすることで混合整数線形計画問題として定式化し、実用的な時間で求解することができるようになった。

「最適配置アルゴリズムのシステムモデル記述との連携機構」については、システムモデル記述として AMALTHEA データモデルを選定、ソフトウェアマッピングツール MBP の内部データ構造である BLXML に変換するツールを開発、自動運転関連システムの AMALTHEA サンプルモデルを入力としたヘテロジニアス・マルチコア向けソフトウェアマッピングに成功した。AMALTHEA は今後、欧州組込み産業、特に自動車産業、においてベンダ間で共通利用する設計標準記述として使用されることが想定されている。

ヘテロジニアス・マルチコア向けモデルベース並列化(HS-MBP)設計環境では、MBP 出力から、ターゲットとするヘテロジニアス・マルチコア環境である Xilinx Zynq UltraScale+ MPSoC ZCU102 上で動作させる設計フローを実証した。現在 eSOL 社 OS と組み合わせ、システムモデルを入力とし、ヘテロジニアス・マルチコア環境で動作させるソフトウェア設計フローを実証予定である。

3.8.3.5 成果の普及

ソフトウェアマッピングを含めた一部分はイーソル社より実用化(詳細は後述)、他はオープンソース化の予定である。

3.8.4 研究開発サブテーマ「ヘテロジニアス・マルチ・メニーコア向けハードウェア記述標準モデル SHIM の研究開発と国際規格化」

3.8.4.1 概要

SHIM のヘテロジニアス・マルチ・メニーコアへの適用に関する基本項目検討および性能評価を行う。実施項目1～4とのインターフェース、ユーザインタフェース、見える化機能などの研究開発、および標準化を実施する。

3.8.4.2 最終目標と根拠

IEEE/IEC での国際標準化を実現する。本研究開発で実現を目指すソフトウェアプラットフォームが迅速に新たなハードウェアに対応できることを確実にするため、本プラットフォームが対応する標準ハードウェア記述モデルを国際標準化することが有効なためである。

3.8.4.3 目標の達成度

2018 年度目標「見積精度向上」に関しては仕様としては MCA SHIM 2.0 仕様に盛り込み済みである。2019 年度目標「ヘテロジニアスアーキテクチャ対応」では仕様面では IEEE SHIM Standard draft に盛り込み済みである。ツール機能面では実施項目 4 のソフトウェアマッピングツールの開発と併せて SHIM2.0 による見積精度の向上のための開発を実施中である。

2020 年度目標である IEEE/IEC 国際標準化においては、2019 年に IEEE/Computer Society/Design Automation/Software-Hardware Interface for Multi-manycore (IEEE/C/DA/SHIM)として P2804 SHIM WG を設置、その後の WG 活動と各種レビューを経て、無事 2020 年 1 月に IEEE Std. 2804-2019 として標準が IEEE より発行された。

3.8.4.4 成果と意義

コンピュータ技術領域において IEEE 標準は国際的に広く信頼され、多くの実績を持つ。その IEEE 標準にて、本ハードウェア記述モデルを新たな IEEE 標準として成立させ、発行出来たことは、今後の本プロジェクト成果の普及において、国際的に大きな意味を持つ。

3.9 研究開発テーマ「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」

本研究開発項目では、「AI をエッジシステムで利用する際に必要となる、ニューラルネットワークおよびその前後処理、ルールベース AI の補完処理、時分割処理にも対応可能な、高速かつ高効率・低消費電力なエコシステムをハードウェア、ソフトウェア、開発環境も含めて実現する」という目標に対し、これまでの研究開発により基本アーキテクチャを決定し上記協調設計を進めており、競争力のある電力効率目標を達成する見込みを得ており目覚ましい成果を上げつつある。個別の研究開発テーマにおける詳細は以下に示す。

3.9.1 研究開発サブテーマ「多分岐結合型推論プロセッサの研究開発」

3.9.1.1 概要

本研究では、グラフ構造で表される処理対象をコンパイラで多数のスレッドに分解、整列した後、それらのスレッドをプロセッサの資源状況とスレッドの処理優先度を動的に判断して実行制御できる並列プロセッサの基礎検討を行い、今後 20 年間の使用に耐えるフレキシブルかつスケラブルな計算機構造を実現できる高効率プロセッサを定義する。

3.9.1.2 最終目標と根拠

多分岐結合型推論プロセッサ(以下 DILP)ハードウェア・ソフトウェアアーキテクチャを定義し、そのアーキテクチャに従いハードウェアおよびソフトウェアを実装する。世界最高レベルの電力効率:15TOPS/W、面積効率:1mm²/TOPS を目標として、コスト競争の厳しい分野、電力制約の厳しい分野もターゲットとする。

3.9.1.3 目標の達成度

15TOPS/W を実現可能なアーキテクチャ開発を完了。DILP のターゲット分野である組込み用途に向け、処理単位が小さくても高処理効率を実現できるアーキテクチャを検討し、ベクトルユニット共有型 MIMD マルチスレッドプロセッサとすることを決定した。基本クラスタのハードウェア設計は、CPU 及び周辺の実装を完了した。ソフトウェア開発環境として、主にフレームワークライブラリ、シミュレータの開発を実施した。

3.9.1.4 成果と意義

AI をエッジシステムで利用する際のキー技術となる高速かつ高効率・低消費電力なプロセッサおよびソフトウェア開発環境を日本で実現することは、深層学習を核として世界レベルで技術開発と多方面への実用化競争が起きている中で、日本がそれらの技術、製品を使うだけの立場だけではなく、日本の産業に貢献できる技術、製品を戦略的に育てていく重要な意義を持つと考えている。

3.9.1.5 成果の普及

本研究で提案する並列プロセッサ技術の普及にはエコシステムの開発が重要であり、基本クラスタ構成のハードウェアが完了後 PoC 等の途中研究成果を活用し、エコシステムの醸造に重要な標準規格への提案を開始し、主にユーザの立場となるアドバイザー企業からのアドバイスも反映しつつ、量産開発想定時期である 2022 年以降に標準

に沿った形での開発方式・環境としてエコシステムが成立できるように活動を実施していく。

3.9.2 研究開発サブテーマ「多様な AI プロセッシングアーキテクチャの評価と性能比較」

3.9.2.1 概要

様々な AI プロセッサアーキテクチャの方式比較、および様々なアプリケーションによる性能評価。DILP とのグラフィックプロセッシングユニット (GPU) を並列科学アプリケーションカーネル及び DILP の機械学習フレームワーク (Tensorflow, PyTorch, Chainer) での画像認識の実装と性能評価を行い、フレームワークの比較検討、および各種フレームワーク間で重みデータ等を交換する為のオープンニューラルネットワークエクステンション (ONNX) フォーマットについて調査を実施した。

3.9.2.2 成果と意義

これまでの結果により、今後進める予定の DILP の性能評価の為の環境 (OpenCL ベースカーネル) と DILP での機械学習アプリケーション評価のための準備ができた。今後 DILP シミュレータ等を利用してアプリケーションの性能評価をすすめ、GPU 等の他のアーキテクチャとの性能比較を実施する。

3.9.3 研究開発サブテーマ「ヘテロジニアス向けコンパイラの研究開発」

3.9.3.1 概要

ヘテロジニアス向けコンパイラの研究開発の環境整備と課題抽出を実施。ヘテロジニアス環境では、①アクセラレータでの処理の切り出し方式、②処理の一部をアクセラレータに切り出すことに伴う整合性確保、の 2 点が大きな課題である。OSCAR 自動並列化コンパイラの技術を活用・拡張することでこれらの課題を解決し、動的多分岐・結合トレース型 AI プロセッサのエコシステムの一部として事業化を目指す。

3.9.3.2 最終目標と根拠

既存の OSCAR 自動並列化コンパイラに、アクセラレータでの実行時間の予測機能と、ユーザとコンパイラとのインタラクティブなインタフェースを構築することにより①の切り出し方式を実現する。さらに、自動並列化コンパイラの強力な解析能力を拡張することで②の整合性確保を目指す。

3.9.3.3 目標の達成度

アクセラレータでの実行時間の予測機能に関して、実行時間を予測するための研究・実装を行い、アクセラレータで実施すべき処理の判別手段としての有効性が確認できた。

また、自動並列化コンパイラに対し、アクセラレータ実行部分をユーザが指示する機能を開発した。さらに、インタラクティブなユーザインタフェースを持つと同時にシステムの整合性を確保することのできるコンパイラの拡張機能として必要なループ並列化技術をまとめ、製品化に向けた機能開発を実施した。

3.9.3.4 成果の普及

成果の普及には、実行時間の予測技術を他のアクセラレータ向けに拡張し、ユーザインタフェースを現実に即したものとするために、実アプリの評価からのフィードバックを行ってヘテロジニアス向けコンパイラの実現を目指す。

3.9.4 研究開発サブテーマ「OSの抽象化とミドルウェア最適化技術ならびに統合開発環境の研究開発」

3.9.4.1 概要

- ① OSの抽象化による汎用性担保と最適化の研究開発
ヘテロジニアスマルチコア環境下で複数かつアーキテクチャが異なるコアを効率的かつ並列実行可能なOS(ハイパーバイザ)の研究開発を実施する。
- ② 制御系へ向けたミドルウェア最適化の研究開発
画像処理、画像解析の為にミドルウェアであるOpenCL,OpenVXについて、制御系に向けて最適化と上記OSへの実装を行い、ユーザへのインタフェースを提供する。
- ③ 複合コンパイル環境に対応した統合開発環境の研究開発
シームレスにシステム設計と開発が行える統合開発環境の研究開発を実施する。

3.9.4.2 最終目標と根拠

- ① OSの抽象化による汎用性担保と最適化の研究開発
現状の市販ハイパーバイザの通常スケジューリング時オーバーヘッドは約3%あるのに対して今回のハイパーバイザでは2%のオーバーヘッドを目標とする。
- ② 制御系へ向けたミドルウェア最適化の研究開発
現状のLinuxベースOpenCL,OpenVXの処理時間に対し20%の処理時間短縮を行う。
- ③ 複合コンパイル環境に対応した統合開発環境の研究開発
開発エンジニアがハイパーバイザ設定から上位のパーティション内の設定、ビルド、デバッグをシームレスに実行できる環境を目指す。

3.9.4.3 目標の達成度

- ① OSの抽象化による汎用性担保と最適化の研究開発
現状、ハイパーバイザ上のパーティション内では、ネイティブタスクまたは、OSSのリアルタイムOSが動作するまでの実装を完了した。
- ② 制御系へ向けたミドルウェア最適化の研究開発
現状、制御系リファレンスRTOSに仮実装中。ハイパーバイザ上のRTOSパーティションが実装され次第、移植を開始する。達成度としては65%程度と考える。
- ③ 複合コンパイル環境に対応した統合開発環境の研究開発
プラットフォーム、設定項目の整理を実施している。

3.9.4.4 成果と意義

制御系に向けたハイパーバイザ、ミドルウェア、開発環境が整備され、部分的にオープンソース化されることにより、多くの企業でヘテロジニアスマルチコア環境を利用してもらうこと、及び日本からオープンソースのプラットフォームを発信、貢献が可能。

3.9.4.5 成果の普及

- ① 本プロジェクト参加企業、学術やアドバイザー企業への早期のリリース
- ② オープンソースコミュニティからの普及

3.9.5 研究開発サブテーマ「リアルタイム空間理解、判断アルゴリズムの研究開発」

3.9.5.1概要

ロボットや無人搬送車(AGV)が、周囲の物体や環境をリアルタイムに認識するアルゴリズム、また、認識結果にもとづいて、行動計画をリアルタイムに決定するアルゴリズム、および、それらの高速ハードウェア実装技術を開発。

3.9.5.2最終目標と根拠

開発したリアルタイム認識、判断アルゴリズムを DILP 上に実装し、ロボットに搭載して、従来比 10 倍速の複雑形状物品のロボットピッキングを実証する。これは、物流倉庫や工場内の作業自動化に必須の課題である。

3.9.5.3目標の達成度

グラフ適用の新規認識アルゴリズムとデータフロー演算(FPGA 実装)により、物体認識速度 0.7 秒(中間目標:3 秒以下)を達成。IEEE 論文誌採択。

新規ロボット軌道計画アルゴリズムとデータフロー演算(FPGA 実装)により、軌道計画時間 0.5 秒(中間目標:3 秒以下)を達成。フラグシップ国際学会採択。
現在、上記各アルゴリズムの DILP 実装検討を推進中。

3.9.5.4成果と意義

リアルタイム認識、判断アルゴリズム、および、その高速ハードウェア実装技術。
産業分野における様々な複雑な作業の完全自動化を実現し、労働者、熟練者不足の社会課題を解決するとともに、競争の激しい AI 分野での国際的プレゼンスを強化できる。

3.9.5.5成果の普及

本委託事業にて開発したアルゴリズム、ハードウェアをコアとするロボット SI 事業をグローバルに展開予定。

3.9.6 研究開発サブテーマ「多分岐結合型推論プロセッサ性能評価用ベンチマークの研究開発」

3.9.6.1 概要

本プロジェクト開発のDILPが対象とする重要分野の代表的アプリケーションベンチマーク群を機械学習系、信号処理系と、制御系から選定し、これらベンチマークが第三者によって検証可能(ソフトウェア実行可能)な形式(C/C++プログラム)とする。ベンチマークリファレンスソースコードは、東京工業大学で開発された C2RTL ツール(ハードウェア構造C/C++記述からハードウェア回路記述を自動生成)によって、ベンチマーク処理を実行する専用回路を直接的に生成できる形式として開発することで、専用回路実装における潜在的計算量、メモリ帯域・容量、IO 帯域、計算並列性などのベンチマーク固有ハードウェア特性をプラットフォームに非依存な客観的数値として提供可能な形式であることを特徴とする。このことで、各種 AI プラットフォームの各ベンチマーク性能傾向の相関性の本質的理解を促進し、各プラットフォームアーキテクチャやソフトウェア実装の設計途上段階での設計改良のための重要な知見を提供することで、市場競争力の高いエッジシステムプラットフォームの短期間の開発を可能にする。

3.9.6.2 最終目標と根拠

機械学習系(画像認識 CNN, 時系列予測 RNN)、信号処理系(FFT, Pitch Detection)、ロボット制御(姿勢推定 NDT Scan Matching, 経路計画 Path Planning)のベンチマークリファレンスソースコードを開発し、本プロジェクト開発の DILP の性能評価(計算資源稼働率、メモリ帯域・容量占有率、バス占有率、消費電力)を行うとともに、これらベンチマークのハードウェア実装における潜在的計算性能、回路規模、メモリ帯域・容量、消費電力、等のベンチマーク固有ハードウェア特性数値を計測する。

3.9.6.3 目標の達成度

2020 年度末時点では、上述のベンチマークリファレンスソースコードの開発を終了する予定であり、その後の2カ年で、DILP シミュレータの性能プロファイル機能の実装と、DILP 性能評価、ベンチマーク固有ハードウェア特性計測を実施するため、概ね予定通りに研究目標が達成できている。

3.9.6.4 成果と意義

既に、画像認識 CNN ベンチマークでは、高精度認識が可能な Resnet モデル(34 層)の C++コードから C2RTL ツールによって4千万ゲート規模で21.9TOPS の処理性能を有するハードウェアを自動合成することに成功しており、これほどの大規模処理がソフトウェアとハードウェアとで両方実装可能であるような C++ベンチマークは前代未聞であり、産業上・学術上で大変大きな意義がある。

3.9.6.5 成果の普及

本研究テーマに関連した国際会議発表を3件行っており、また C2RTL ツール自体の産業移転活動も進めており、C++によるソフトウェア・ハードウェアベンチマークを最終的には GitHub 等で公開することを検討している。

3.10 研究開発テーマ「完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発」

本研究開発項目では、完全自動運転に向けたシステムオンチップ(ハードウェア)のみならず、オペレーティングシステム(OS)・ミドルウェア・アプリケーション技術を創出するという目標に対し、各組織下記の研究開発を達成し、成果を上げつつある。

- I. 完全自動運転に向けた SoC の研究開発 (株式会社アクセル)
- II. 完全自動運転に向けたコンパイラ・OS の研究開発 (国立大学法人東京大学)
- III. 完全自動運転に向けたミドルウェアの研究開発 (国立大学法人埼玉大学)
- IV. 完全自動運転に向けたアプリケーションの研究開発 (株式会社ティアフォー)

個別の研究開発テーマにおける詳細は以下に示す。

3.10.1 研究開発サブテーマ「完全自動運転に向けた SoC の研究開発」

3.10.1.1 概要

本研究開発テーマでは、完全自動運転に向けた SoC(System on Chip)の研究開発、および試作チップの製造を行う。完全自動運転では、「認知」「判断」「操作」に必要な大量の処理を、決められた時刻までに処理する必要がある(リアルタイム処理)。一方、今後量産製品においてシステムが運用される環境は、ボンネット内やトランク内などの限られた空間および放熱環境、バッテリーからの限られた電力という厳しいものと想定される。本プロジェクト開始時点で、多くの自動運転システムは、高性能の CPU や GPU を組み合わせたもので、処理能力、消費電力、システムのサイズ、放熱性能などの点で多くの課題がある。

これらの課題を解決するためには、面積、電力当りの性能向上が必要不可欠であり、ハードウェアのアーキテクチャレベルからの最適化が有効なアプローチの一つと考えられる。そこで本研究開発テーマでは、様々な特性を持った計算資源をうまく組み合わせたヘテロジニアスコンピューティングが有効であると考え、完全自動運転に特化した独自のアクセラレータおよびメニーコア混在のヘテロジニアス SoC の研究開発、および試作チップの製造を行う。

3.10.1.2 最終目標と根拠

最終目標として、2021 年度に試作チップの製造を完了し、2022 年に全体を結合した実証実験を実施し試作チップによる省電力・演算性能達成見込み達成を確認する。省電力・演算性能達成見込みは以下の 3 点である。

1. 開発成果を組み込んだシステムレベルでの施策を行い、想定用途やシステムによる実用性を検証すること。
2. 事業開始時に広く普及している技術と比較して、エネルギー消費効率あるいは電力効率 10 倍以上を達成すること。
3. AI エッジコンピューティングにおける限られた計算資源でのリアルタイム性の保証と従来比 10 倍以上の高速なデータ処理を達成する。

最終目標1においては、実用化を見据えた研究開発の場合、必須のものと考え、チップを用い、全体を結合した実証実験を実施することで達成を確認する。

最終目標2および3においては、プロジェクト開始時において使われている自動運転システムの多くが、100W を超える消費電力である。先述した今後量産製品においてシステムが運用される環境での使用を想定すると、使用可能な電力や放熱環境から、3~5 倍程度の電力対性能の向上が求められると考える。プロジェクト終了時までの競合技術の進

化も踏まえ、事業化に向けたアドバンテージの必要性も考慮し、10 倍という目標を設定した。

本研究開発テーマにより、完全自動運転に特化したハードウェアアクセラレータを実装しつつ、メニーコアをベースとした SoC に組み込むことによって、他の汎用を目指すプロセッサとの圧倒的優位性を確立し、早期に完全自動運転におけるデファクトスタンダードを達成し、事業化へつなげることを目標とする。

3.10.1.3 目標の達成度

- ・ 2019 年度：FPGA プロトタイピングの立ち上げ完了・評価開始

達成度：100%

- ・ 2020 年度(中間目標)：FPGA を使った SoC プロトタイピング実装および評価

達成度：20%(2020 年 6 月時点)

※2020 年度において、新型コロナウイルスの影響は限定的である。一部の研究開発において、機材を使用した作業が滞っているが、他の作業と再スケジュールすることで、影響を最小限にするよう務めている。

3.10.1.4 成果と意義

成果：

2019 年度および 2020 年度前半迄の成果としては、自動運転システムに必要な要求仕様を洗い出すと主に、全体アーキテクチャ検討を行い、LSI 全体設計を行った。また、L 拠点間での共同研究開発に必要な開発環境の整備や設計データ群の整理を完了した。本研究開発テーマの1つの主要な内容である、自動運転システムで使用される処理のハードウェアアクセラレータの研究開発のうち、CNN アクセラレータの研究開発を完了した。同様に、もう1つの主要な内容である、リアルタイムコア(メニーコア)においても、同様に開発を完了し、FPGA での評価へ移行した。試作 LSI の開発および製造においては、ASIC ベンダの選定を完了し、開発を開始した。ASIC ベンダの選定においては、事業化まで一貫して対応可能なベンダを候補とし、選定した。

意義：

FPGA プロトタイピングによる評価を行うことで、早期にハードウェアアクセラレータの効果測定、性能改善を図る。以降、実際に LSI として製造することで、戦略的に実用化・事業化し、市場への展開を目指す。委託先となる ASIC ベンダにおいても、事業化まで一貫して対応可能なベンダを選定候補とし、早期の実用化・事業化を可能とする。また、本研究開発テーマで対象とする自動運転プラットフォームの Autoware は、研究用途の Autoware.Ai のみならず、実用を目指した自動運転プラットフォームである Autoware.Auto も対応しており、実用化・事業化を強く意識した研究開発内容である。

3.10.1.5 成果の普及

先述の CNN アクセラレータの研究開発において、2019 年度で研究開発を完了し、事業化へシフトした。事業化自体は、本研究開発テーマにおける 1 つの項目を横展開した形である。本 CNN アクセラレータの事業化では、まずは FPGA や ASIC 向けの IP、および FPGA ボードによるシステム製品などとして展開する。その後、ASIC 化を含めた製品化を目指し、現在具体的な検討を進めている。

上述した CNN アクセラレータの事業化について、プレスリリースとして発表を行った。

リリースページ：https://www.axell.co.jp/ir/pdf/AXELL_PR20200623.pdf

3. 10. 2 研究開発サブテーマ「全自動運転に向けたコンパイラ・OS の研究開発」

3. 10. 2. 1 概要

ヘテロジニアス SoC チップ向けコンパイラ・OS の研究開発を実施する。本コンパイラ・OS はメニーコア、ハードウェアアクセラレータ、FPGA などを高効率且つ容易に利活用可能とすることを旨とし、各共通計算アルゴリズム向けに最適化されたライブラリや、ハードウェアアクセラレータを高効率に動作するためのドライバを開発する。

3. 10. 2. 2 最終目標と根拠

2021 年度に試作チップ向け開発を行う。その後 2022 年度まで実証実験による省電力・演算性能達成見込みの確認を実施する。

ソフトウェア開発効率と計算資源利用効率の両面で高効率にヘテロジニアス SoC を利活用できる手段が必要不可欠であるため、アプリケーション開発者が容易にヘテロジニアス SoC の恩恵を授受出来ることを目指して本コンパイラ・OS の開発を行う。

3. 10. 2. 3 目標の達成度

2019 年度:コンパイラ・OS の実装完了

達成度:100%

・2020 年度(中間目標): FPGA による性能評価実施

達成度:15%(2020 年 6 月時点)

3. 10. 2. 4 成果と意義

成果:

アクセラ開発のメニーコアと ARM コア間的高速通信及び、アプリケーションの移植性を考慮した API を実装し、従来の INET ソケット通信より高速な Unix Domain Socket の API の実装を行った。さらにメニーコアを N*N クラスタ並列処理を実現するための API や、Mcube Kernel における C 言語を用いた簡易 Publisher/Subscriber プログラムを実装したが、必要なライブラリである libcpp のビルドに関するエラーが多数発生した。これらを解析し、アサーション及び浮動小数点コンパイルオプションの調整を行うことで回避可能であることを確認した。

意義:

本コンパイラ・OS により、メニーコア、ハードウェアアクセラレータ、FPGA などを容易に行うことができ、アプリケーション開発が行い易くなる。さらなるユーザの利便性、性能、効率などの向上のために継続的な評価も行うことにより、実用に則したコンパイラ・OS が提供できる。例えば、研究用途の Autoware.Ai のみならず、実用を目指した自動運転プラットフォームである Autoware.Auto への対応が可能になる。

3. 10. 2. 5 成果の普及

4.10 を参照。4 組織合同の成果・普及について記載している。

3.10.3 研究開発サブテーマ「完全自動運転に向けたミドルウェアの研究開発」

3.10.3.1 概要

ROS(Robot Operating System)をベースとしたヘテロジニアスな環境(メニーコア、マルチコア、ハードウェアアクセラレータ)に完全自動運転アプリケーションを分散して処理できる環境を研究項目 2 で提供される API を用いて構築することを目標とする。さらに、性能評価を効率的に行うための分析ミドルウェアの研究開発も実施する。

3.10.3.2 最終目標と根拠

自動運転アプリケーションがヘテロジニアスな環境(メニーコア、マルチコア、アクセラレータ)に分散して処理できるような環境を構築することを最終目標とする。ROS にはデータ収集に関する便利なツール「ROSBAG」が付属しており、収集したデータを様々なメッセージタイプに変換してノード間で通信させることができる。さらに、本研究項目では、2020 年度に予定している FPGA での評価を効率的に行うための、分析ミドルウェアの研究開発も実施する。分析ミドルウェアとして動作確認ツール(実行速度測定ツール、トレーサ)や解析ツールを開発し性能向上を妨げているボトルネック解析や、分散処理効果を確認するために用いる。

3.10.3.3 目標の達成度

- ・2019 年度:動作確認ツールの開発を完了
達成度:100%
- ・2020 年度(中間目標):解析ツールの開発完了、FPGA プロトタイプ向けミドルウェアの実装・評価
達成度:30%(2020 年 6 月時点)

3.10.3.4 成果と意義

成果:

OS とミドルウェアとの協調設計に向けて OS・ミドルウェアで必要とするシステムコール(OS が提供する関数)のリスト化、コンパイラの必要要件の洗い出し、アクセラのハードウェアの策定のため、自動運転アプリケーションのメモリ使用量の調査を行った。

実施計画提出時点では、ROS 1 をベースに研究開発を行うことを計画していたが、各委託先との打合せの結果、リアルタイム性や機能安全など実用性を考慮し、分散処理ソフトウェアプラットフォーム(ミドルウェア)を ROS 2 ベースで再設計を行うことにした。再設計に向けて ROS 2 のクライアントライブラリ(RCLC/RCLCPP/RMW)についても調査を行った。メニーコアの疑似環境として Kalray MPP256 環境にて動作検証することとし移植を開始し、RCLC(C 言語のクライアントライブラリ)、RCLCPP(C++言語のクライアントライブラリ)に関して移植が完了した。さらに、ROS 2 で使用する DDS(Data Distribution Service)については、オープンソースで提供されている FastRTPS 及び Eclipse Cyclone DDS を対象に調査を行った。まずは、ROS2 の標準搭載の DDS である FastRTPS から移植作業を開始した。並行して、もう一つの DDS の候補である Eclipse Cyclone DDS を PC 上で動作確認を行い、FastRTPS との性能評価(リアルタイム性)比較を行った。

動作確認ツールとしては、入力データと出力データを比較するツールの研究開発を行った。これより、今後、本ミドルウェアで開発を行うアプリケーションの動作確認・検証を容易に行うことができる。

2020年6月時点では、中間目標のFPGAプロトタイプ実装評価に向け、設計・実装を進めている。さらに、新設計に適した、解析ツールの設計を開始した。

意義:

提案ミドルウェアにより、メニーコアでの複雑な開発(多くのコアでの配置や通信の制御等)を容易に行うことで、アプリケーション開発が行い易くなる。さらに、ROS 2 ベースの新設計に変更することにより、リアルタイム性や機能安全を考慮した実用に適したミドルウェアを提供できる。例えば、研究用途ではなく、実用を目指した自動運転プラットフォームである Autoware.Auto への対応が可能になる。

。

3.10.3.5 成果の普及

4.10を参照。4組織合同の成果・普及について記載している。

3.10.3.6 その他

本ミドルウェアを含む Autoware on Many-core Platform: NoC ベース組込みメニーコアプロセッサ向け自動運転プラットフォームを、組込みシステムの査読ありシンポジウムである組込みシステムシンポジウム(ESS2019)で発表し、優秀論文賞(ベストペーパー)受賞した。

3.10.4 研究開発サブテーマ「完全自動運転に向けたアプリケーションの研究開発」

3.10.4.1 概要

完全自動運転に向けた SoC、OS、ミドルウェア上で動作する完全自動運転に向けたアプリケーションの研究開発を実施する。車載での実用的な価格、消費電力、発熱量の制約下では、高性能であっても従来のベストエフォート型のプラットフォームで危険(Risk)をコントロールすることは困難である。AI エッジコンピューティングを用いて、車載を意識した危険をコントロールして社会に受け入れられる完全自動運転に向けたアプリケーションを構築する。

3.10.4.2 最終目標と根拠

車載向けの限られた計算資源の下で、完全自動運転に要求される認知・判断・操作について、安全性の担保のために実時間性の保証を最終目標とする。

車載向けの限られた計算資源の下で、完全自動運転に求められるタスクの分析を行い、ヘテロジニアス SoC の協調設計、および必要に応じてアルゴリズムの車載向け再設計を含めたアプリケーション再構築を行い、実時間性の保証という観点でヘテロジニアス SoC からアプリケーションまでの全体最適なシステムを構築する。

3.10.4.3 目標の達成度

・2019年度:タスクの選定と分析および協調設計によるヘテロジニアス SoC への初期フィードバックを完了

達成度:100%

状況:初期目標として EKF(拡張カルマンフィルタ)アクセラレータ実装を目標として、FPGA での動作及び性能を確認完了

・2020 年度(中間目標):FPGA 評価を実施し、演算性能を算出する

達成度:30%(2020 年 6 月時点)

状況:EKF の検証および、NDT(Normal Distribution Transform)アクセラレータの初期設計を実施中

3.10.4.4 成果と意義

成果:

完全自動運転機能のリアルタイム制御向けの開発においては、ハードウェア/ソフトウェアの協調設計と、リアルタイム性を確保できなかった場合に車両を安全に停車させる機能を開発した。協調設計においては、IP(Intellectual Property)の流用性も考慮して複数のデータ並列処理へ転用可能な Application Specific SIMT アーキテクチャを採用し、完全自動運転の中でも安全性への影響が大きい EKF(Extended Kalman Filter:自己位置推定に使用するフィルタ処理)の開発を完了した。

完全自動運転ソフトウェアの実車両適応評価では、今期購入した JapanTaxi 車両 2 台とタジマ製小型 EV 車両 1 台の評価を行った。完全自動運転ソフトウェアから車両を制御するための車両インターフェースを開発して自動運転制御可能とし、直進、停止、右左折といった基本動作から、低速域制御や緊急停止制御の改善および検証を通して、同一車種の別車両や複数車種に対して適応できていることを確認した。

意義:

Autoware を効率的かつ安全に動作可能なハードウェアとアクセラレータ連携のシステムを構築し、AS-SIMT アーキテクチャを採用した SoC システムの研究を進めた。これにより、従来までの高性能 PC を用いずに安全性・リアルタイム性を確保して Autoware を動作させられるシステムの目途を立てることが可能になった。

また、車両実験も進めており、JapanTaxi 車両での Autoware 実動作確認で実証データの蓄積及び性能解析を進めている。将来的には SoC で車両を動作させる事を目的としている。

3.10.4.5 成果の普及

4.10 を参照。4 組織合同の成果・普及について記載している。

3.10.4.6 その他

Autoware を中心とした自動運転エコシステムの拡大も進めており、特に本研究プロジェクトと関係の深い協業先として ARM Ltd.が挙げられる。

本研究では Autoware の低負荷な処理をホストコンピュータ汎用 CPU で処理し、高負荷処理をアクセラレータにオフロードする形式を取っている。ここで、ホストコンピュータ上のプロセッサを ARM 製の低消費電力 CPU を活用する事で、さらなる低消費電力化を図っている。

SoC 化に関しては他にも、Renesas、NXP、ADLINK、QuantaComputer から連携の引き合いがあり、将来的には自動運転 IP コア展開事業の促進・広範化に寄与できると見込んでいる。

3.11 研究開発テーマ「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」の研究開発成果

3.11.1 概要

本研究開発では、今後必要性が急増する「AI エッジデバイス用セキュリティのホワイトボックス化」の要請に応えるため、RISC-V オープンアーキテクチャをベースとして RISC-V に不足しているセキュリティ基盤技術を開発する。この基盤技術はトラスト実行環境(TEE。ハードウェア信頼起点のアシスト付き)を軸として産業分野への実応用を容易にするための鍵管理技術を埋め込み、これらのセキュリティ基盤を適用したセキュア産業 IoT の PoC システム設計を含めた研究開発を実施する。

セキュリティシステムはシリコンハードウェア(SoC)から社会実装に至る深い階層構造を必要とするため、本研究開発では実施項目を

実施項目① セキュア MCU アーキテクチャ

実施項目② トラスト実行環境(TEE)

実施項目③ アーキテクチャの産業用途即応化

実施項目④ セキュリティ社会実装 PoC 設計

の四階層で構成し、これらを連携させて研究開発を進める。

3.11.2 最終目標と根拠

実施項目① セキュア MCU アーキテクチャ

「セキュア MCU アーキテクチャ」では、AI チップでサイバーセキュリティを確保するために必須となる、セキュア MCU のシステムアーキテクチャ主要概念を開発する。

チップのセキュリティでは安全な鍵注入や暗号技術の安全な管理がユーザから求められている。これに答えるために RISC-V 64bit とは独立に鍵管理、真性乱数生成を行うセキュア MCU の仕様検討を行う。パソコンでは TPM、スマートフォン等は GlobalPlatform の Secure Element が使われていることが多いが、機能が限られている問題がある。本開発では今後活用が進むチップ認証などを含んだ証明書の検証なども行えるように MCU にする。

セキュア MCU は RISC-V 64bit と同一 SoC 上に組み込まれることを想定し、RISC-V 32bit で構成して低消費電力、物理的耐性を持ち、機能性、安全性を確保する。また、チップそれぞれ固有に注入された鍵は、情報を秘匿するために利用できるだけでなく、そのチップとその上で動作するソフトウェアの真正性の検証やチップの認証にも利用できる。更に、それをベースにするハードウェア、ソフトウェアそれぞれのサプライチェーンのために活用できるようにする。

そのため、チップ利用者がチップの認証などに利用できるような鍵管理のエコシステムの一部となる固有鍵サーバを構築し、セキュア MCU の固有鍵を管理する。

実施項目②-1 RISC-V 版 TEE 仕様策定およびハードウェア開発

「トラスト実行環境(TEE)」では、TEE の RISC-V 64bit 上での実装方式の開発と、RISC-V 基金における標準化活動を推進する。通常の計算とは独立して安全に計算する環境 TEE は非営利団体である GlobalPlatform において仕様が決まっておらず、通常の OS が走る Normal World と Trusted OS が走る Secure World は物理的に分離されている。この機能はスマートフォンの認証システムやセットトップボックスの DRM 実装で活用されているが、RISC-V では TEE の実装についてまだ規格が決まっておらず、ワーキンググループ(WG)での協議段階である。従来の研究開発プロジェクトでは新しい方式を提案しても規格策定で採用に至らない事が多かった。そのため本プロジェクトでは技術開発と並行して、RISC-V 基金

のWGにも参加して標準化の議論も進める。また、本プロジェクトで開発する TEE の Trusted OS や Trusted Application は AI エッジ・産業用途に供する。

実施項目②-2 TEE ソフトウェア

最終年度の目標は Trusted OS を含む TEE に関するソフトウェアが安全にリモートから更新できる技術 Over the Air の策定を予定している。既に Secure MCU を信頼の基点とするデバイスの真正性検証は 2020 年度末に開発予定であり、これを発展させる形で実装する。このために IETF で規格されている TEEP (Trusted Execution Environment Provisioning) のプロトコル実装を RISC-V 上で行ってきたが、これに外部から真正性を確認できる Remote Attestation を付加して、製品に組み込みやすい形にする。Over the Air については自動車会社で活用されているが、規格に沿った実装ではなく、広く活用できるものではない。本プロジェクトの方式は IETF 準拠であり、仕様が公開されている RISC-V で実装することで広く活用されるようにする。

なお、このような Over the Air を実現するためには、デバイスや Trusted Application (TA) の真正性を検証するための鍵と、そのデバイスを使ったサービスを実施するための鍵を連携して運用ができるサーバが必要となるため、実施項目①の固有鍵サーバや実施項目④の鍵管理サーバとの連携の実現を目指す。また、Over the Air で配布するアプリケーションを管理するサーバも考慮したものとする。

また、RISC-V の TEE の実用性を示すために脆弱性の評価を行い、商用に耐えうることを確認する。近年、各種団体が主導する認証制度が進んでいるが、それに適用する実装を行うことで、活用を目指す。

実施項目③-1 産業用途 RISC-V

当初の最終目標とした「産業用途で求められる CPU の研究開発(マイクロアーキテクチャ検討 等)」は RISC-V 基金における追加アーキテクチャ検討の決定等の状況を鑑みて 2020 年度以降その実施を見合わせるものとし、「広範な市場ニーズを把握するためのユーザコミュニティ組織の形成」を当初計画の 2022 年度より前倒して実施することにより、探索したニーズを他の実施項目に追加反映することに注力する。

実施項目③-2 ソフトウェア移植容易性

本実施項目では、ユーザコミュニティを通じて RISC-V の産業応用におけるソフトウェア移植への要請事項を調査し、その移植容易性向上の効果を評価する。ユーザニーズの調査を通じて、ソフトウェア移植を容易化するのに必要な要件を網羅的に抽出する。

実施項目④-1 セキュア産業 IoT 模擬環境での実証

近年、エッジデバイスの急速な増加に伴い、サービス全体の信頼性の担保が喫緊の課題となっている。この課題を解決すべく、本実施項目では、サプライチェーンにおけるステークホルダーである製造会社・販売会社や、運用開始後のステークホルダーである保守会社などのシステムコンポーネントを無条件で信用しないことを基本とし検証を求めるゼロトラストネットワークの構築をめざす。具体的には、本プロジェクトで開発したオープンセキュアチップを用いて、エッジデバイスの物理的なセキュリティ確保と管理を効率的に行うエッジシステム及び鍵管理サーバを PoC システムとして構築し、生産工場や次世代モビリティシステム等の具体的な事業フィールドを想定した模擬環境で実証する。

実施項目④-2 セキュア AI エッジ応用ライフサイクルの社会実装

RISC-V TEE のサプライチェーン信頼性保証と運用環境における鍵管理等の生成、交換、修理から廃棄までの統合技術に関する有効性確認を実施する。有効性確認では、実フィールドにおけるユースケースに対応した RISC-V TEE システムを用いた応用ライフサイクルの仮説検証を行なうことで、セキュア AI エッジが提供する価値のソリューションアイデアを複数提示し、ユーザの用途に応じて使い分けられるよう出口戦略にする。

3.11.3 目標の達成度

実施項目① セキュア MCU アーキテクチャ

「セキュア MCU アーキテクチャ」を実現するためのハードウェアプラットフォーム並びにその上で動作するセキュア OS の実装まで完了している。現状ではセキュア MCU 単体での動作が可能である。セキュア MCU 内部に搭載している RISC-V 32bit コアはオリジナルであり、製品レベルの品質を保持しているところも優れた点である。

今年度は暗号化処理に関わる部分はソフトウェアにて実装するが、来年度以降、システム全体のスループット向上のために、ハードウェア実装の暗号化エンジンを同システムに搭載する予定で進める。

直近の課題として、実施項目②で開発予定の RISC-V 64bit アプリケーションコア部との回路接続とその回路を制御するためのソフトウェア(デバイスドライバ、API)があるがこれらは 2021 年 3 月までに実装完了予定である。

また、セキュア MCU それぞれのための鍵を管理する固有鍵サーバならびに、セキュア MCU の鍵を処理するセキュア OS 用アプリケーションを開発中であり、2020 年度末までには固有鍵サーバとセキュア MCU が連携可能となる予定である。

実施項目②-1 RISC-V 版 TEE 仕様策定およびハードウェア開発

「トラスト実行環境(TEE)」で使われる RISC-V 64bit アプリケーションコア部のベースとなる RISC-V マルチコアアーキテクチャの FPGA 化まで実現できた。さらにセキュア MCU との通信を実現する回路の実装まで実施できており、セキュア MCU 含めたシステム全体としての動作が実施できる段階まで到達できた。また、RISC-V 64bit アプリケーションコアの起動はセキュア MCU で管理され、ブートローダがセキュア MCU の検証を完了しないと実行できない仕組みを加えている。これは Apple の iBoot と似た仕組みであるが、ブートローダの更新などを第三者機関が管理できるようにすることで広く産業用途に活用できることを目指す。

実施項目②-2 TEE ソフトウェア

当初の目標通りに Trusted OS を実施項目②-1 で開発される TEE ハードウェア上に作成する目処は立っている。既に GlobalPlatform が規定する TEE 用の API は開発しており、アプリケーション(Trusted Application:TA)の開発ができるようになっている。この RISC-V での GlobalPlatform の API 実装は RISC-V Security Committee および GlobalPlatform とともに定期的な会議を行っており、産総研がその一翼を担っている。

実施項目①-1 で開発される Secure MCU を信頼の基点として、デバイスを認証するための証明や機密情報が保存されることが想定され、TA からこの証明書や機密情報が活用できるようにする。TA と Secure MCU の通信、および API の実装は産総研とセコムで行っており、2020 年度末までには完成する予定である。

TA をリモートからインストール/アップデート/デリートをリモートから管理する TEEP(Trusted Execution Environment Provisioning)プロトコルは IETF で議論されているが、その Hackathon

に産総研・セコムが参加し、その仕様に準拠した実装を行っている。また、TEEP の提案書には産総研の研究者が提案者の一人となり、規格策定に貢献している。

外部から TA の実行を承認するためにデバイスの真正性、および意図した TA が確実に実行されることを確認する Remote Attestation の実装も進めている。2020 年度末までには、Secure MCU および TEE ハードウェアを FPGA 上で動作させることを予定している。

実施項目③-2 ソフトウェア移植容易性

2018 年度は、既存の制御用ソフトウェアを調査し、移植工数の削減に寄与する要因を抽出した。抽出した要因に基づいて評価対象ソフトウェアを選定する基準を策定した。2019 年度は、2018 年度に策定した選定基準に従い、評価対象とする制御用アプリケーションソフトウェアを選定した。選定したソフトウェアを机上で解析し、移植容易性を評価した。2020 年度は、ユーザコミュニティを通じて産業応用におけるソフトウェア移植への要請事項を調査する。また、ニーズ調査に並行して、RISC-V Foundation の動向をリファレンスとして調査する。

実施項目④-1 セキュア産業 IoT 模擬環境での実証

2020 年度は、昨年度に引き続き産業 IoT における将来動向、ニーズ調査を実施する。様々な産業 IoT の内で「生産工場・物流倉庫」と「モビリティ」を主として、現状システムおよび将来システムを想定し、脅威分析を行うことで脅威シナリオを作成する。また、作成した脅威シナリオの対策として必要となるセキュリティ機能要件を抽出し、オープンセキュアチップの有用性を示すモデルケースの策定を行う。

生産工場・物流倉庫分野では、アセットシェアリングにおける脅威シナリオを示し、顧客と事業者の双方に利する柔軟なリソース配分方式を提案し、TEE による実行環境分離及びアセット管理サーバを用いたシステム構成を具体化する。

モビリティ分野では、普及が進む MaaS 向けテレマティクス保険において、脅威シナリオに加えて車両データ分析を行うサーバ処理の肥大化が課題であることを示し、エッジ側で安全に運用・管理するためのシステムアーキテクチャ、データフロー、ステークホルダーマップから必要となる機能構成を明らかにする。

これまで、生産工場・物流倉庫分野ではリソース配分方式の提案と、モビリティ分野ではエッジ側アーキテクチャ、データフロー、ステークホルダーマップの作成が完了しており、本年度目標に対して計画通りに進められている。

実施項目④-2 セキュア AI エッジ応用ライフサイクルの社会実装

2020 度はセキュア AI エッジ応用ライフサイクルの内外動向調査結果報告書と RISC-V TEE システムを用いて実現すべきエコシステムの仮説立案を計画している。新型コロナウイルスの影響により現時点で国内外動向継続調査の進捗は遅れているが、電話会議等の代替手段を用いて調査を実施し、有効なユースケース仮説とソリューション検討を実施する。

3. 11. 4 成果と意義

実施項目① セキュア MCU アーキテクチャ

RISC-V 32bit ベースでのセキュア MCU を FPGA 上で動作する事ができ、かつその上で動作するセキュア OS まで開発できたことにより、RISC-V 64bit アプリケーションコア部との接続が可能となった。さらにチップそれぞれに固有鍵の注入できる機能の搭載により、セキュリティ性能向上と使いやすさの両立ができています。

これらの成果より、実施項目②への成果の適用ができ、システム全体としての動作が可能となり、社会実装 PoC への適用が可能となる。

実施項目②-1 RISC-V 版 TEE 仕様策定およびハードウェア開発

TEE を実行する RISC-V 64bit アプリケーションコアと低消費電力の RISC-V 32bit セキュア MCU 含めたシステム全体としての動作が可能になったことにより、TEE を搭載し、さらに Trusted Application の実装が実現できるため、社会実装 PoC への適用が可能となる。RISC-V 64bit アプリケーションコア部の起動を検証するセキュア MCU、およびその検証がネットワークを介した第三者機関で行えるようにすることで、ユーザに対して信頼のできる実行環境となる。また、ハードウェアベンダーのみならず、その鍵管理ベンダーや組み込みソフトウェアベンダーへ活用領域を広げることで国内産業振興に使えるようにする。

実施項目②-2 TEE ソフトウェア

TEE は Arm TrustZone を使ったスマートフォンや Intel SGX を使ったサーバで利用が出ているが、ハードウェアベースの信頼の基点をベースにして、デバイスや TEE 内で実行されるコードの真正性を行う Remote Attestation を含むものは少ない。Intel では自社で提供する SDK に Remote Attestation を使える機能を提供するが、そのためには Intel が用意するサーバを信用しなければならない。また、Over The Air についても多くの実装があるが、IETF のような規格に沿うものはほとんどなく、認証制度自体が作成中で模索段階である。本開発では規格策定自体に貢献しつつ、仕様が公開されている RISC-V で実装することで、ハードウェア・ソフトウェア双方のセキュリティが確保できる技術を開発している。また、本開発では協調領域と競争領域を明確に分離し、TEE の仕様や API などは広く公開してセキュリティを確保しつつ、TEE 実装については競争領域として国内産業振興に使えること目指している。

実施項目③-2 ソフトウェア移植容易性

産業分野のシステム製品のコストパフォーマンスとセキュリティの向上のためには、ライセンス料が廉価でホワイトボックス化が可能な CPU を使用することが好ましい。この条件を満たす CPU の中では、主に海外で実用化が進んでいる RISC-V が最有力候補であるが、産業向け機能に関する議論は行われていない。RISC-V に必要な産業向け機能を明らかにし、半導体メーカーによる産業向け RISC-V の製品化を促進することで、産業分野のシステム製品のコストパフォーマンスとセキュリティの向上を図ることができる。

実施項目④-1 セキュア産業 IoT 模擬環境での実証

本プロジェクトで開発するセキュアオープンアーキテクチャを標準技術として普及させるためには、適切なテストケースにおける適用事例を示す必要がある。本実施項目では、生産工場・物流倉庫分野とモビリティ分野をテストケースとして選定し、適用に向けた具体化を進めている。これらのテストケースは、セキュリティに関する標準的な国際規格 (IEC62443, ISO21434(WP29)) に将来的にアラインするものであり、テストケースとして適していると考えられる。したがって、本プロジェクトで開発するセキュアオープンアーキテクチャを標準技術として普及させる上で、本実施項目の成果は意義がある。

実施項目④-2 セキュア AI エッジ応用ライフサイクルの社会実装

セキュリティ計算基盤が支える IoT やモビリティのイノベーション価値向上を達成するためには、研究開発するセキュア AI エッジがグローバルに見てどのような位置に属するのか明確な指標とそれを実現する方策が必要である。さらに、Society5.0 等のエコシステムを想定したユースケースの仮説を立案し、提供する価値のソリューションアイデアを複数提示することも重要であり、そのためには、国内外の動向調査に基づいた出口戦略策定と、その動向調査の

フィードバックを受けた RISC-V TEE システムを適用したセキュア AI エッジ応用ライフサイクルのエコシステムを検討することが必要不可欠である。

3.11.5 成果の普及

- ・市場向け活動の第一歩として 2020 年 8 月に第一回オープンフォーラムを実施済み。TRASIO が研究開発する技術に対する認知を高めて、オープンコミュニティへの発展の期待を喚起した。
- ・オープンコミュニティはこの実績を更に発展させる活動として、2021 年度初めまでに活動開始のアナウンスを予定。
- ・オープンコミュニティの開設目的
 - TRASIO 研究成果(協調領域の技術基盤)に対するユーザの認知を高める
 - 研究開発成果の紹介(プレゼンテーションの実施、ハンドアウト資料の提供)
 - 研究開発成果であるセキュリティ基盤技術のハンズオン体験(ハードウェア、ソフトウェアの動作に関する実体験)の提供
 - セキュリティ技術、RISC-V 技術に対する広範なユーザニーズの探索
 - 探索したニーズの研究開発実施項目へのフィードバックを推進することにより、市場ニーズを積極的に取り込んだオープンシステムを開発して、研究開発成果の普及を促進する。

3.12 研究開発テーマ「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」

3.12.1 研究開発サブテーマ「実施項目 0：AI エッジで必要となるセキュリティ評価分析基盤の研究開発」

本研究開発では、AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術を確立させるという目標に対し、以下のような成果を上げつつある。

- AI エッジデバイスなどの対象物が、評価済みあるいは登録済みであることを確認する際に必要となる個体管理用識別子のセキュリティ評価方法について、日本からの ISO TC292 WG4 に対しての国際標準化提案に貢献し、審議開始が承認された。
- AI エッジデバイスの入出力セキュリティを評価するためのシミュレータの MILS (Model In the Loop Simulation: モデル動作)レベルモジュールの構築が完了し、公開可能な部分の外部発表を行った。
- AI エッジデバイス内の重要な情報を保護するための特殊パッケージを試作し、その特性の基礎評価を完了させた。
- AI エッジデバイス内のファームウェア(バイナリ)などを解析するツールを試作し、DARPA CGC (Cyber Grand Challenge)の課題の不具合箇所にも自動適用可能なことを確認した。
- AI エッジデバイスなどの個体管理用識別子の読取、照合・識別、貼付を評価するための装置やシステムの構築が進み、公開可能な部分の外部発表を行った。

個別の研究開発テーマは、AI エッジデバイスの入出力に関わる部分を実施項目 1、AI エッジデバイスの内部実装保護に関わる部分を実施項目 2、AI エッジデバイスの個体管理に関わる部分を実施項目 3、全体を俯瞰し新たにセキュリティ評価分析基盤として整備すべき部分を実施項目 0 が担っており、それらの詳細を以下に示す。

3.12.1.1 概要

AI エッジに対して横断的なセキュリティ評価を行う際に必要となる評価項目を明らかにし、それらの内、既存の仕組みでは対応できない評価項目に対して新たな評価分析カテゴリを確立する。その際、必要とされるセキュリティ要求仕様やセキュリティレベルは応用分野毎に異なるため、その応用分野における主要なステークホルダー(製品やサービスの提供者、調達者、セキュリティ評価機関、セキュリティ認証機関など)とも対話しながら新たな評価分析カテゴリのセキュリティ要求仕様やセキュリティ評価分析手法の策定や整備を行う。これらにより、社会一般に受け入れられ易いセキュリティ評価分析基盤の確立を目指す。

3.12.1.2 最終目標と根拠

最終目標は、2022 年度末までに AI エッジを対象に含む新たな評価分析カテゴリを確立させることである。その際、全体のセキュリティは一番セキュリティの弱い箇所で決まるため、全体を俯瞰しながらセキュリティ要求やセキュリティ評価分析手法が定まっていない箇所を特定し、その部分を強化することが肝要となり、また、社会一般に受け入れられ易いセキュリティ評価分析基盤とするために、応用分野毎の主要なステークホルダー(製品やサービスの提供者、調達者、セキュリティ評価機関、セキュリティ認証機関など)とも対話しながら検討を進めることが重要となる。

3.12.1.3 目標の達成度

2018年度には、AI エッジ向けのセキュリティ評価項目全体を洗い出し、既存の評価方法などで対応可能なものと今後整備が必要な項目とに分類し、更に今後強化すべき箇所を明らかにするという目標を達成した。上記分類の結果、後者については各実施項目の中で重点的に取り組み、2019年度には2018年度で明らかにされた箇所の内、対象がセキュリティ評価認証済みまたは登録済みであることを確認するための個体管理用識別子のセキュリティ評価分析基盤に焦点をあて、そのセキュリティ要求仕様とセキュリティ評価分析手法を策定するという目標を達成した。2020年度は、応用分野として経済産業省が定めた自動走行システムにおけるサイバーセキュリティ対策検討体制(工程表)(図 3.12.1.3) 右下の評価・認証体制に対する評価方法/評価環境/体制整備のために、AI エッジが組み込まれる部分のセキュリティ要求仕様とセキュリティ評価分析手法の策定に取り組んでおり、その目標は達成可能であると見込んでいる。

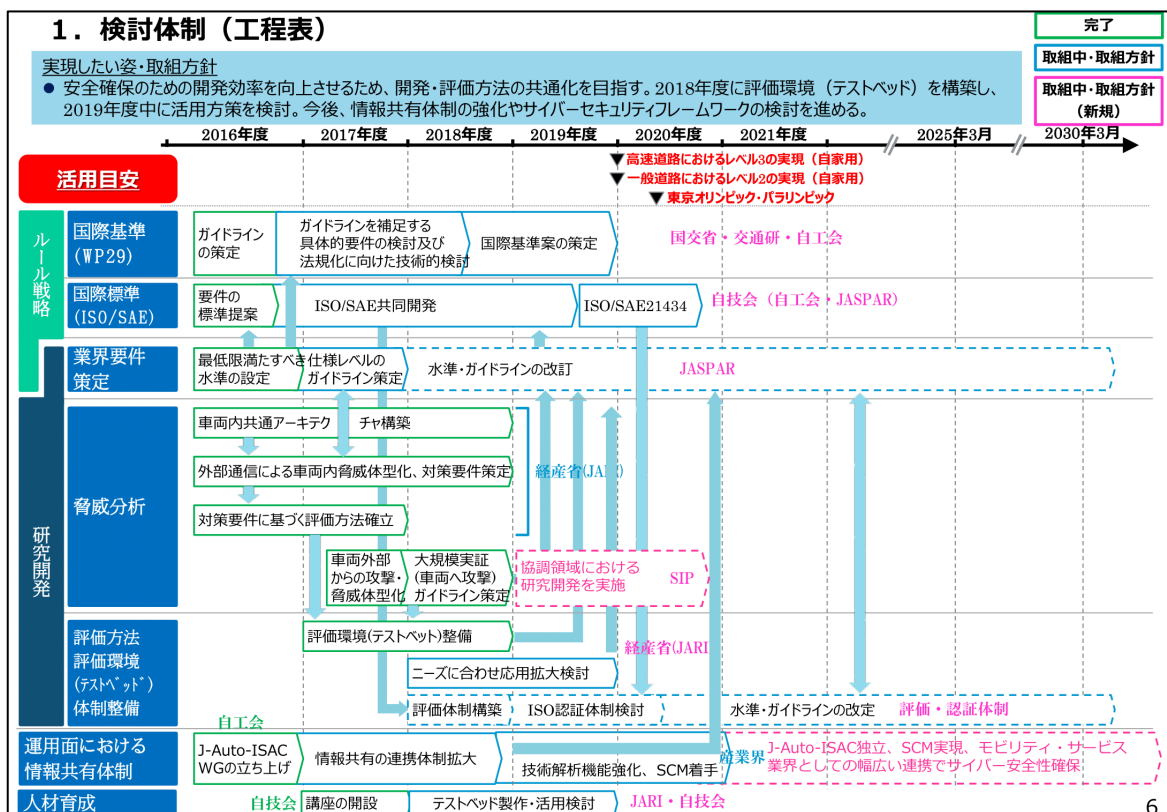


図 3.12.1.3 経済産業省の自動走行システムにおけるサイバーセキュリティ対策検討体制(工程表)

3.12.1.4 成果と意義

2018年度の成果であるAIエッジ向けセキュリティ評価項目の洗い出しに関しては、それにより、今後ニーズが高まる分野とその分野で不足しているセキュリティ評価項目を明確にすることができ、リソースをよりニーズの高い方に向けることができたという意味で、その意義は大きい。2019年度の成果である個体管理用識別子(人工物メトリクス)のセキュリティ要求仕様とセキュリティ評価分析手法は、AI エッジデバイスやその部品などの対象物が評価済みであることを確認する際の識別子のセキュリティ評価に利用できる。人工物メトリクスは、日本発の技術であるが、標準的な評価分析手法が定まっていなかったため、その国際標準化提案を日本からISO TC292 (Security and resilience) WG4 (Authenticity, integrity and trust for products and documents)に対して行い、国際標準化に向けた審議の開始にも繋がった。意義に関しては図 3.12.1.4 が示すとおり、政府模倣品・海賊版対策総合窓口 に寄せられて

いる相談対象商品の 1/3 程度は AI エッジデバイスとも関連の深い機器や機械が占めており、その件数も増加していることから、AI エッジデバイスやその部品に対する将来的な模倣品・海賊版増加リスクに対して布石を打つこととして意義も大きい。

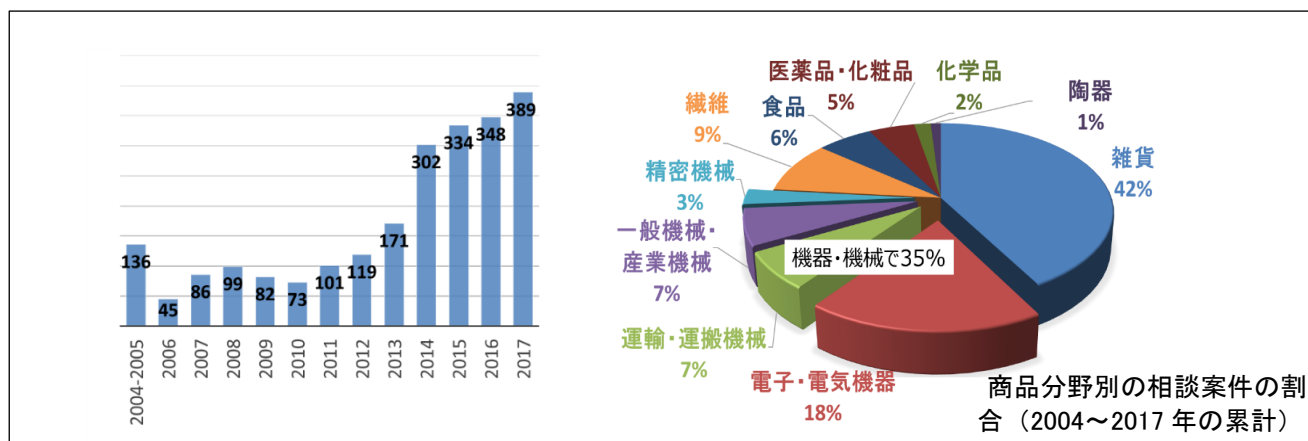


図 3.12.1.4 政府模倣品・海賊版対策総合窓口に寄せられている相談件数の推移(左)と商品分野別の相談案件の割合(右)

3.12.1.5 成果の普及

求められるセキュリティ要求やレベルは応用分野により異なるため、AI エッジ応用分野のステークホルダー(製品やサービスの提供者、調達者、セキュリティ評価機関、セキュリティ認証機関など)の意見も伺いながら、社会に受け入れられ易いものになっている。個体管理のための識別子(人工物メトリクス)に対しては、関連する企業群、横国大、産総研、経産省模倣品対策室などで『人工物メトリクスタスクフォース』を作り、成果の主要部分の国際標準化を進めている。これにより、成果が世界に波及し、関連市場がセキュリティ上の脅威により阻害されることを防止する。AI エッジ入出力セキュリティ評価シミュレータの自動走行システムへの応用に関しては、そこで必要となるセキュリティ要求仕様と評価分析手法を ISO/SAE 21434 (Road vehicles - Cybersecurity Engineering) に準拠させると共に、従来から重視されてきている機能安全(セーフティ)等への悪影響を防止するために、電気電子システムの(自然発生的なエラーによる)誤動作リスクを低減するための ISO26262 (Road vehicles - Functional safety(機能安全))、誤操作などのリスクの低減も考慮に入れた ISO/PAS 21448 (Road vehicles - Safety of the intended functionality (SOTIF))などとも互換性を保ちながら、また、内閣府 SIP プロジェクト第2期 自動運転(システムとサービスの拡張)の「仮想空間での自動走行評価環境整備手法の開発」で取り組まれているセーフティの取り組みとも連携する体制を取ることで、セキュリティ対策や評価が、従来から重視されている機能安全(セーフティ)へ悪影響を与えることを避け、本分野においてより普及し易い方向を見据えながら取り組みを進めている。

[1]経済産業省

「自動走行システムにおけるサイバーセキュリティ対策自動走行ビジネス検討会 2019年6月
https://www.meti.go.jp/shingikai/mono_info_service/jido_soko/pdf/sanko_03.pdf

[2]政府模倣品・海賊版対策総合窓口「模倣品・海賊版対策の相談業務に関する年次報告」
 2018年6月

¹ 2004年8月に企業等からの要望を受けて経済産業省 製造産業局 模倣品対策・通商室に設置された政府の一元的な相談窓口。

² 2020年6月時点では DIS (Draft International Standard)。

3.12.2 研究開発サブテーマ「実施項目1：AI エッジ入出力セキュリティ評価シミュレータの開発」

3.12.2.1 概要

AI エッジデバイスへの入力である物理世界からのセンシングデータや、それらに基づく出力である機器への制御信号に対するセキュリティリスクが世界的に危惧され始めていることから、攻撃の実現可能性や影響、セキュリティ強化策の効果などをシミュレーション、あるいは、シミュレータと実機により評価できるよう「AI エッジ入出力セキュリティ評価シミュレータ」の構築を行う。構築されたシミュレータはシステムレベルでの検証が可能となるものを目指す。以下に詳細を示す。

- (1)「AI エッジ入力攻撃モデル」では、攻撃によって発生するシステムへの入力や状態変化をシミュレートすることを目的とし、センサやAIへの攻撃をMILSの設計フローにおいて利用可能な形式でモデリングする。攻撃によって発生するセンサ群やAIエッジ、制御機器などシステムの構成要素における振る舞いを模擬するために、システム構成要素のモデルに組み込むエージェント型のモデルを構築する。AI エッジ入力攻撃モデルは、以下3つのモデルに細分化される。
 - (1-1)「センサ入力攻撃モデル」: AI エッジ入力として接続されるセンサに対する成りすまし攻撃や、センシングを妨害する攻撃をモデリングする。
 - (1-2)「AI エッジインターフェース攻撃モデル」: AI エッジがセンサ及び制御対象機器と接続する通信路などAIエッジが持つインターフェースに対する攻撃をモデリングする。
 - (1-3)「AI エッジハードウェア攻撃モデル」: システムの構成要素に対し物理的にアクセス可能なことを想定した攻撃をモデリングする。AIエッジにおけるアナログ情報の改竄・漏洩やアナログ回路およびデジタル回路へのフォールト注入攻撃をモデリングする。
- (2)「AI エッジ出力影響評価モデル」では、攻撃がシステムに与える影響をシミュレートすることを目的とし、システム構成要素のモデルにおける内部状態とシステムで定義される故障や異常状態などの達成すべきセキュリティ基準とを紐付けるモデリングを行う。AI エッジ入力攻撃モデルとの連動したシミュレーションにより、攻撃によって発生したシステム構成要素の内部状態が、システム全体で規定されるセキュリティ基準を満たすか判定することができる。AI エッジ出力影響評価モデルは以下2つの構成要素に細分化される。
 - (2-1)「セキュリティ基準判定機能」: システムの機能要件から定義される異常状態、故障状態、セキュリティ上攻撃が成立している状態、すなわちセキュリティ基準が未達の状態を定義し、シミュレータからの出力結果からセキュリティ基準に達しているかを判断する機能を提供する。この機能を実現するためには、セキュリティ基準が未達の状態(あるいは達成している状態)をシミュレータと接続可能なフォーマットでモデル化する必要がある。システムの機能要件の例としてはISO26262におけるSafety goalやユーロNCAPが挙げられる。シミュレータと接続可能なフォーマットとしてはXCCDF(eXtensible Configuration Checklist Description Format)が挙げられる。シミュレータから受け取る情報からXCCDFで記述されたSafety goalが侵害されていないかを判断し、判定結果を出力する。

- (2-2)「内部状態抽出機能」:セキュリティ基準評価モデルが判定するために必要となるシステム構成要素の内部状態を抽出する機能である。例としては、ISO26262 においてハザード分析の結果「アクチュエータ X は速度 Y 以上でアクティベートされてはならない」という要件があった場合、この要件はセキュリティ基準判定機能として XCCDF で記述される。内部状態抽出機能では、速度及びアクチュエータのアクティベート状況を抽出する。
- (3)「AI エッジ処理セキュリティ強化モデル」では、対策の効果を設計段階で定量的に評価することを目的とし、センサフュージョンや認識処理、オブザーバを用いた状態フィードバック制御など高度な処理において、攻撃の影響を緩和するあるいは攻撃を検出する機能のモデリングを行う。AIエッジ入力攻撃モデル、AIエッジ出力影響評価モデルとの連動したシミュレーションにより、セキュリティ強化技術の導入によって攻撃によって発生したシステム構成要素の内部状態が、システム全体で規定されるセキュリティ基準を満たすかどうかを判定することができる。AI エッジ処理セキュリティ強化モデルは、対策の実装レイヤに応じて以下3つに細分化して検討する。
- (3-1)「セキュアセンサフュージョン」:マルチモーダルセンシングを前提とし複数のセンサ情報をエッジで AI 処理する過程において攻撃対策あるいは攻撃検出を行うアルゴリズムの検討と、そのシミュレーションモデリングを開発する。
- (3-2)「セキュアフィードバック制御」:オブザーバやカルマンフィルタなどの状態フィードバック制御をセキュリティの視点から改良を検討し、攻撃検出や攻撃の影響緩和が可能なアルゴリズムを開発する。また、そのシミュレーションモデルや HILS を想定した実機シミュレータを開発する。
- (3-3)「セキュア AD/DA 変換回路」:セキュア AD/DA 変換回路はセンサやアクチュエータのアナログ物理量と AI エッジにおけるデジタル情報の双方向変換機能を狙ったアナログ情報の改竄や漏洩に対して防御ならびに攻撃の検知と回避を具現する半導体集積回路レベルのアナログ・セキュリティ強化技術を確立する。

3.12.2.2 最終目標と根拠

2022 年度末において、実機接続可能な(HILS)セキュリティ評価シミュレータとしての開発を行い、統合の完成を最終目標とする。根拠として、2020 年度末までに、(1)攻撃実験の物理モデル仕様検討、物理モデル設計、物理モデル実装、統合評価、(2)出力影響評価シナリオの見直し、再設計、再実装、統合評価、(3)セキュアセンサフュージョン、セキュアフィードバック制御について、設計、実装、単体評価、統合評価を各々行い、MILS(モデルでの動作)技術によるモデル化を行い、次いで、HILS 技術によるハードウェア接続に進む。

3.12.2.3 目標の達成度

攻撃実験の動作仕様検討、動作モデル設計、動作モデル実装、統合評価を行い、MILS レベルモジュールの構築を行った。AI エッジ入出力セキュリティ評価シミュレータとしては、七つのモデル(1. 外界環境 2. センサ:周囲計測 3. 攻撃 4. 認知・判断 5. コントローラ 6. システム 7. 状態計測)から構成され、攻撃モデル以外は各ドメインでのモデルベースデザインにおけるモデルを利用することを想定し、各モデルの入出力はシミュレーション実行

時にログ出力可能な構成とすることにより評価の数値化を可能にした。センサフュージョンとしては、自動運転で主流となるレーダ、LiDAR、カメラによる物体検知や測距、測角を想定したシミュレータのアーキテクチャを検討し、カメラ画像の入力としては、Unreal と MATLAB/Simulink の連動による実現を検討し、フィジビリティスタディを完了した。

各ドメインの具体的なモデルベースデザインは、以下のとおりである。カメラセンサシステムの認知・判断に対応する部分に関しては、車線検出に対してカメラ映像の色調が影響を与えることを見出し、カメラセンサの色調を変化させる環境を構築した。また、Adversarial Patch (以下 AP と呼ぶ) と呼ばれる AI が誤動作する画像を用いた攻撃を想定し、AP を生成する環境を構築することで、車両が AP によって人の検出に失敗し AEB (Autonomous Emergency Braking: 衝突被害軽減ブレーキ) が動作しない攻撃シナリオを模擬することを可能とした。本想定においては、(独)自動車事故対策機構の自動車アセスメントである JNCAP (Japan New Car Assessment Program) やヨーロッパ新車アセスメントプログラムである Euro NCAP (European New Car Assessment Program) の運転シナリオをもとに動作を規定し、センサ攻撃がその動作のアセスメントプログラムに及ぼす影響を調べるためのシミュレータ(図 3.12.3.1)を開発した。

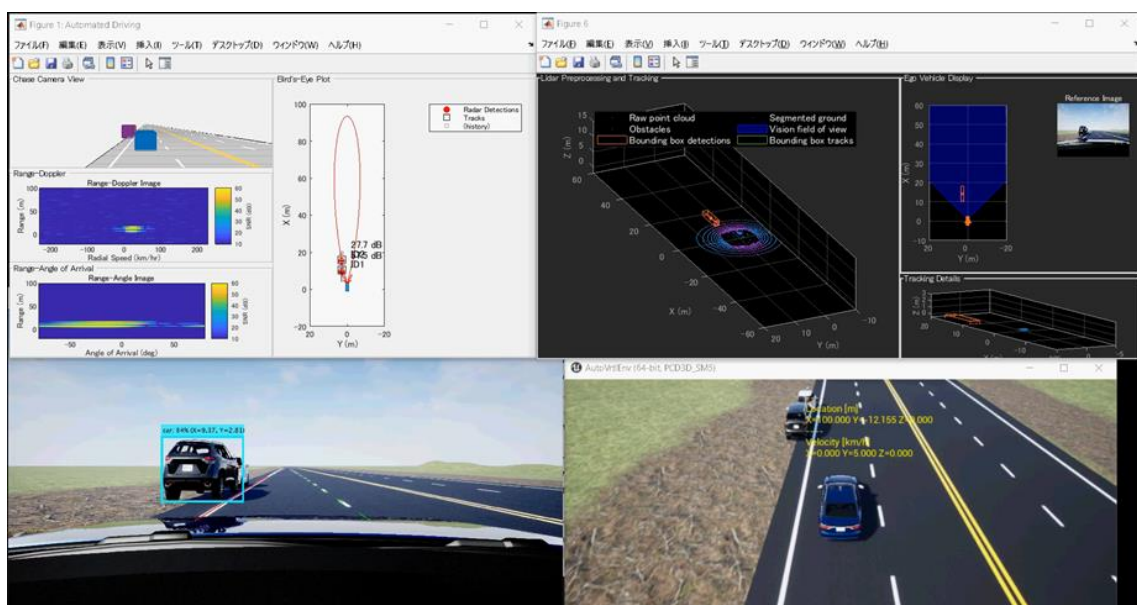


図 3.12.2.3.1 AI エッジ入出力セキュリティ評価シミュレータの画面例

カメラのイメージセンサに対しては、セキュアセンシングとして ToF (Time of Flight)型において、環境・他のセンサからの干渉、さらには意図的な攻撃からセンサデータの正当性保障のために、攻撃モデルを構築すると共に耐攻撃性を有する設計を行った。また、コモンモード・電化注入型検波方式を用いたイメージセンサにおいては、従来方式と同程度のダイナミックレンジ及び背景光除去性能、周波数選択性に加え、特定符号の選択性を有していること、及び従来方式と比較し、より高い検波周波数での動作を実現することを示した。レーダに対しては、FMCW (Frequency Modulated Continuous Wave : 周波数連続変調) への攻撃の影響を評価することを目的とし、コントローラが受け取るレーダの対象検出結果に着目して攻撃の影響を評価可能とするモデル化を実施した。FMCW レーダは、高性能なファスト FMCW 方式を採用し、MIMO (Multi Input Multi Output) …技術に基づく現実的なレーダをモデル化した。超音波センサについては音響レベルでのモデリングを行い、攻撃を再現できる MILS レベルモジュールの構築を行った(図 3.12.2.3.2)。

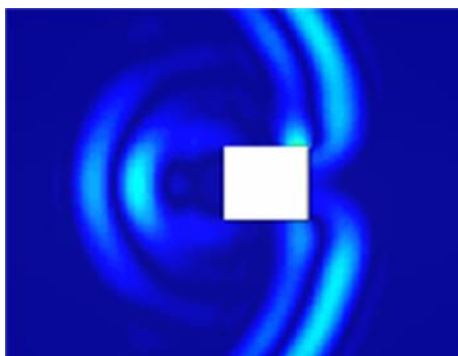


図 3.12.2.3.2 センサに対する攻撃の影響を評価するための MILS レベルモジュールでの物理モデルシミュレーションの例

アナログ情報の改竄・漏洩やアナログ回路及びデジタル回路へのフォールト注入攻撃のモデリングとしては、電磁波・電圧サージ・グリッチ等による外乱注入の表現手法を探索し、セキュアなアナログデジタル変換機能に向けた要素回路テストチップを CMOS デバイス技術にて設計し、アナログデジタル変換器（時間領域 ADC）のベーシックモデルを MATLAB/Simulink 上に構築(図 3.12.2.3.3)することによりセキュア AD/DA 変換回路の設計を完了した。

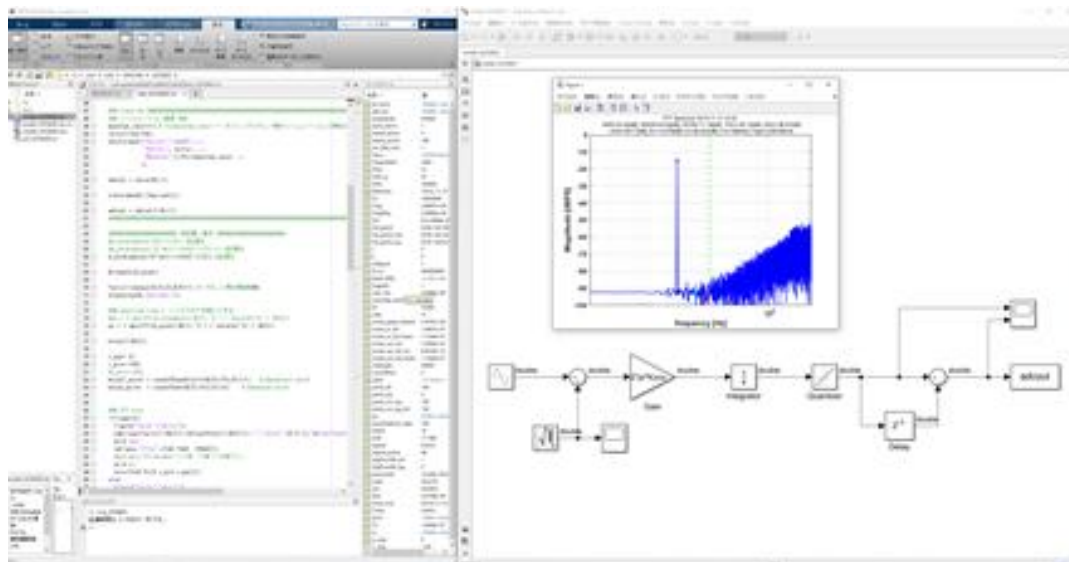


図 3.12.2.3.3 電磁波・電圧サージ・グリッチ等による外乱注入評価モデルモジュール

3.12.2.4 成果と意義

シミュレーションによるセキュリティ評価が実施できない現状においては、製品の市場出荷後に研究者など第三者からの脆弱性報告によってリコールがなされ、修正が行われる。これに対し、攻撃と影響のシミュレーションモデルによる評価によってセキュリティリスクの検出率を 10 倍以上引き上げ、設計段階で脆弱性を塞ぐことにより、リコール率を 1/10 以下に抑えることができる。この結果、製品のセキュリティ設計/開発に必要なエネルギー消費効率/電力効率は 10 倍以上となる見込みである。

3.12.2.5 成果の普及

得られた成果の内、公開可能なものは学術的な信憑性を高めるために、添付資料(特許論文等リスト)記載のとおり、外部発表を行っている。

3.12.3 研究開発サブテーマ「実施項目2：AI エッジ内部実装保護技術の研究開発」

3.12.3.1 概要

電子デバイスを解析して、そこに封入した回路情報や回路に流れる信号情報が解読されると、設計情報が復元されて製品を容易に模倣・改造される。事業者が技術開発に投じた金銭的・時間的投資を回収する前に、模倣品が市場を席巻するという状況は長らく改善されず、すでに看過できない段階に達している。安価に改造・複製された製品が想定外の利用のされ方をして、二次被害をもたらす危険もある。技術開発国の知的財産が、模倣品製造国に搾取される問題とその影響は、国益を守るために緊急に解決すべき課題である。

本実施項目では、電子デバイスに搭載した回路情報、電子デバイスに流れる信号等の秘匿したい情報を保護するための実装技術、及びその評価技術を開発する。具体的には、電子回路(以下、IC)に封入した情報を侵襲解析困難にするIC保護技術(以下、IC保護技術)の開発、および、IC保護等の加エプロセスを経た電子デバイスの品質保証技術(以下、品質保証技術)の開発を行う。

本実施項目で開発するIC保護技術は、ユースケースに応じたコストコンシャスな(cost-conscious)技術であることを前提とする。サンプル出荷のための少量生産品と市場拡大のための大量生産品では、セキュリティリスク低減として選択可能な手段が異なる。リバースエンジニアリングの脅威は、技術の存在が明らかになった時点から生じるため、サンプル品や少量生産品についても、リバースエンジニアリング対策は必要である。今後、各種センサや通信モジュール等のエッジデバイスが社会インフラとして活用される状況は加速し、多品種・少量生産の電子デバイスが次々と市場投入される。リバースエンジニアリング対策に十分な費用をかけられない少量生産品やサンプル品は、類似品を開発する事業者が技術開発のコストを削減するために、情報抜き取りの対象にされる(図3.12.3.1)。

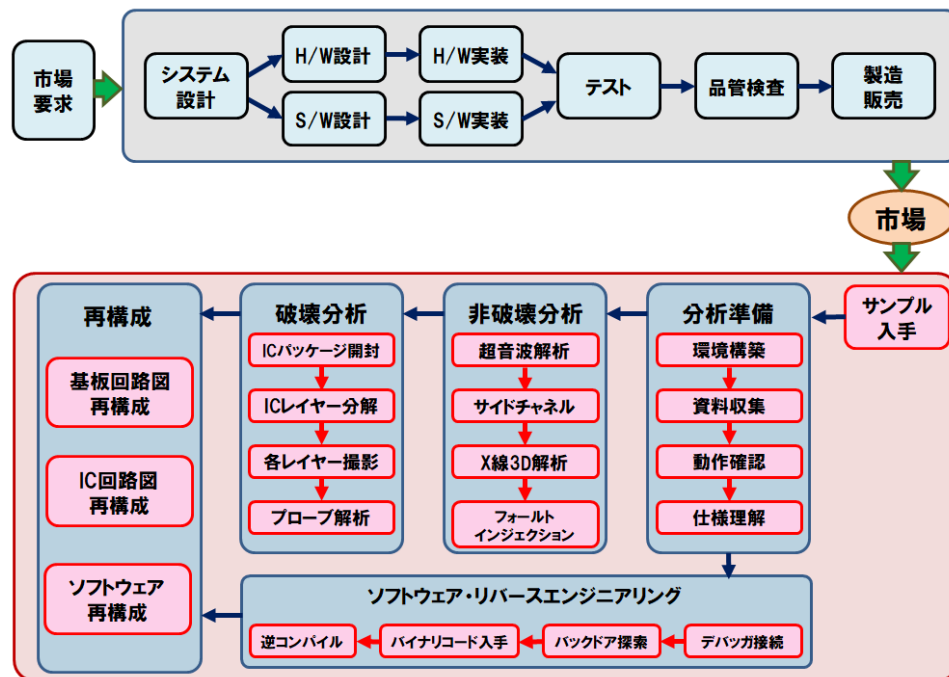


図3.12.3.1 デバイスの市場投入とその後の情報抜き取りフローの例

3.12.3.2 最終目標と根拠

本実施項目の IC 保護技術は、カスタム IC、特に FPGA 等の SoC(system on chip)を搭載した電子基板について、物理解析による情報の漏洩を遅延させるための対策技術と位置付ける。具体的には、サンプル出荷から技術評価までの情報漏洩抑止等、特定の想定下での侵襲攻撃による情報抜き取りから電子デバイスを保護し、情報の漏洩を遅延させることを目的として、低コストで運用できる耐リバーズエンジニアリング技術を開発する。

IC 保護技術の開発(以下、実施項目 2-1)では、電子デバイスに封入した情報について、換装やデキャップ等による侵襲的な情報搾取への対策技術を開発する。実施項目 2-1 で開発した技術をスムーズに社会実装し、デファクトスタンダード(事実上の標準)の地位を獲得するためには、本事業で開発する IC 保護技術の運用コストが低く、安定して運用できる技術であることが望ましい。

一般に、半導体製品の製造の前期工程では、高精度の設備・複雑な形成プロセスを実現するための高額な設備投資が必要である。実施項目 2-1 では、半導体製造の後期工程(半導体の形成後プロセス)で適用するための IC パッケージの加工技術を開発する。具体的には、(1)セキュリティシールド形成技術、(2)フィルタ実装技術、(3)基盤ソフトウェア(ファームウェア)による電子デバイスの制御技術の開発を行う。技術の詳細は、セキュリティ保証に係るため公開しない。実施項目 2-1 の(1)セキュリティシールド形成技術とは、フレキシブル基板上にセキュリティコアを形成する技術である。薬液等による樹脂開封時の樹脂溶解とともに配線溶解する材料の選定と、基板を湾曲させる実装工法に適した材料構成を行う。微弱な信号や微細な特性を扱うと品質劣化による動作不具合が生じるため、セキュリティコアをロバスト化する設計も行う。(2)セキュリティシールド実装技術では、電子デバイス上の専用回路(ASIC)に接続させるセキュリティコアをフレキシブル基板上に形成して、研磨等により樹脂開封されるとセキュリティコア切断される(回復困難な断電状態になる)仕組みを実現する。(3)基盤ソフトウェア(ファームウェア)による電子デバイスの制御技術では、実装したセキュリティコアの特性を組み込むことにより、電子デバイスの出荷時の状態を識別可能にする。電子デバイスの制御技術を用いると、出荷時にセキュリティコアを有効化した電子デバイスが出荷後にセキュリティコアを無効化された場合、その電子デバイスでは、以降、不可逆的にプログラムを起動不能にする。

品質保証技術の開発(以下、実施項目 2-2)では、(1)IC パッケージの加工により保証が滅失した部位の品質の再保証(品質保証)のための技術、(2)加工によりアドオンされたセキュリティ機能の保証(セキュリティ保証)のための技術、(3)IC パッケージに IC 保護技術を施す加工プロセスの保証(プロセス保証)のための技術を開発する。

実施項目 2-2 の(1)品質保証では、放電、衝撃、振動等の外的要因に対して、IC 保護技術を施した部位が、規定の温度・湿度の範囲内では異常を生じないことを確認する。具体的には、物理的ストレス環境下での、加速劣化試験(accelerated aging test)や温度サイクル試験(temperature cycle test)を実施する。本実施項目では、評価対象の物理系から発生したパルス(電流、電圧、電磁波等)を観測して、パルスの時間変位・空間変位の情報から、対象に起きている異常やその原因を特定する技術を開発する。具体的には、重畳波形(合成波形)から、素波形を分離し、素波形に混入する異常波形の有無を検知し、検知した異常波形を特定して抽出し、抽出した異常波形から分析対象の物理系に起きている異常を推定するための技術開発である。また、階層的な分析プロセスをなす波形分析では、各プロセスの分析パラメータが他の分析プロセスに影響を与えるため、分析対象の物理系から発生したパルスの時間変化に合わせて、分析パラメータを自動設定する技術を開発する。

実施項目 2-2 の(2)セキュリティ保証では、電子デバイスから取り外した IC を再装着できないこと、パッケージ開封された IC は動作しないことを確認する。前者は、技術仕様と加工

プロセスを非公開とすることにより、加工の不可逆性を保証する。後者は、薬液による開封試験、精密研磨による開封試験によりセキュリティ機能の十分性を評価する。本実施項目では、開封試験のための評価環境を事業者(産業技術総合研究所)の施設内に構築して、評価基準と評価手順を開発する。

実施項目 2-2 の(3)プロセス保証では、加工のための材料調達から加工品出荷までの加工プロセスについて、独自の基準を設けて自己適合評価し、外部評価機関により評価結果の追認を受ける。

セキュリティ評価とプロセス評価で用いる評価基準は、外部評価機関の助言を得て開発することにより、基準の健全性を確保する。評価基準の開発のために、大量の参照基準文書を分析する技術を開発して、文書の構造解析、文章の類似度解析、文書の照合作業のための情報分析を行う。

実施項目 2-3 の「ファームウェア解析ツールの開発」では、AI エッジデバイスを制御する基盤ソフトウェアであるファームウェアに、機器乗っ取りにつながるプログラムの不具合やトロイの木馬が存在しないことを検証するためのプログラム解析ツールを開発する。ファームウェアのソースコードが提供されることはほとんど期待できないため、実行プログラムのバイナリーコードを直接処理し、制御やデータの流れを記号的に再構成できる自動化ツールを開発する。並行して、トロイの木馬に特有な悪意ある振舞いを検知する手法の開発も行う。

実施項目 2-4 の「脆弱性自動検知技術の開発」では、実施項目 2-3 のファームウェア解析ツールの解析結果をもとに、サイバー攻撃による機器乗っ取りにつながる脆弱性(プログラムの不具合)の存在を自動検出する手法の開発を行う。全ての脆弱性を自動検知することは原理的に不可能であるため、危険性が高く組み込み機器にまだ多数存在するケースに絞った自動化ツールを開発する。

実施項目 2-5 の「ソースコードに依らない制御フロー整合基盤の開発」では、たとえファームウェアに脆弱性があったとしても、制御の流れをチェックするコードを事前に挿入しておくことで、実際の機器の乗っ取りは生じさせない技術の開発を行う。ファームウェアのソースコードが利用できない状況でも、サイバー攻撃による被害を未然に防ぐ技術の開発を行う。

3.12.3.3 目標の達成度

IC 保護技術の開発では、集積回路(以下、IC)に実装された機密情報を保護するため後付け樹脂加工によるフィルタの付加を安価に実現できることは社会実装として有効なことである。2022 年度末にはフィルタの後付け加工ビジネスを開始できるよう実装技術、量産化技術を確立させることを最終目標とする。

品質保証技術の開発では、品質保証・セキュリティ保証・プロセス保証の評価基準と評価プロセス(評価手順)を開発し、評価に必要な技術の開発と評価環境の構築を行うことを目標とする。評価基準の開発では、外部評価機関による助言をもとに基準を開発し、査読を受けることにより、評価基準の客観性と十分性が第三者に確認可能であることを目標とする。

評価手順の開発では、評価受入から評価完了までを2週間で実施できることを目標とする。品質評価は、本実施項目で開発する技術により、電流、電圧、電磁波等外的要因による異常発生の有無の確認と評価結果の生成までを自動で行う。従来の人手による逐次的な分析作業に代えて、階層的な自動分析プロセスを導入して人手による作業を極力排除することにより、分析に係る時間を大幅に圧縮する。

セキュリティ評価技術の開発では、薬液による開封試験、精密研磨による開封試験の評価環境の構築と評価事業で運用するための体制づくりを行う。また、評価事業のサービス化に向けて、パイロット評価による PoC(Proof of Concept; 概念実証)を行い、評価基準と評価プロセスの事業化の確実性を高める。

実施項目 2-3、2-4、2-5 のファームウェア保護・分析技術の開発においては、各項目で開発された技術が統合され、実際の AI エッジデバイスのファームウェアを対象に、脆弱性やトロイの木馬の振舞いがプログラム実行させずに自動検知できること、および、脆弱性があつたとしてもサイバー攻撃による機器乗っ取りを未然に防ぐこと、を可能とすることを最終目標とする。こうした技術は、現状では自動化できておらず、また、検知等の精度や処理効率も実用レベルに達していない。ここに掲げた最終目標を達成することで、AI エッジデバイスのセキュリティを大幅に向上させることが可能となる。

3.12.3.4 成果と意義

IC保護技術の開発では、ハードウェアによる対策としてIC防護技術のために、リバースエンジニアリングが困難となるフィルタの仕様に係る回路、材料の基礎調査、検討を行っている。回路基板と配線材料の選定のための調査結果を踏まえて、回路機能性を維持したまま樹脂開封

薬品等の化学的侵襲解析への耐性を有するセキュリティコア基材と配線材を特定するために、セキュリティコア仕様や実験ステップについて検討を行い、モック品を製作し樹脂開封薬品等の化学的侵襲解析評価を行った。半導体封止樹脂を溶解させる酸系薬品にセキュリティコア材料、配線回路材料ともに溶解し化学的侵襲解析への耐性があることを確認した。本年度は物理的侵襲解析への耐性技術の基本評価を完了させるため、セキュリティコアの湾曲加工工法の確立を行うとともに、FPGAへの後付け加工並びに機能・品質評価を進めている。

物理的侵襲解析への耐性技術については、セキュリティコアを湾曲形状に加工して実装することを進めている。形状加工については目途が立ち、耐性ならびに品質評価を行っている。

セキュリティコア特性については、基礎的な識別を可能とするレベルまで固有になることを確認している。

セキュリティコア特性、固有値そのものの環境変化耐性(品質評価)についても本年度に終える予定である。

品質保証技術の開発では、放電、衝撃、振動等の物理的な外的要因に対して、IC保護技術を施した部位が、規定の温度・湿度の範囲内では異常を生じないことを確認するための試験装置の整備を完了した。現在、評価分析対象の物理系から発生したパルス(電流、電圧、電磁波等)を観測して、パルスの時間変位・空間変位の情報から、対象に起きている異常やその原因を特定する技術を開発して、技術評価を実施している。下図は、試験装置の一例である。左下図は、ICチップの局所的な異常発熱を赤外線像として捉える装置、右下図は、動作温度・動作湿度を変化させて規定の温度・湿度の範囲内では異常を生じないことを確認する装置である。

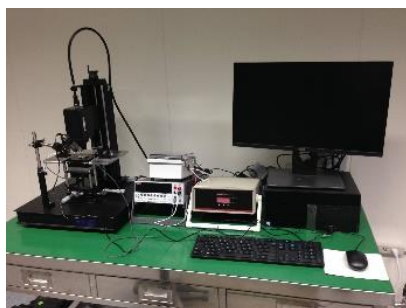


図 3.12.3.4.1 評価装置

評価プロセスの開発では、波形分析の各要素技術の開発を完了した。具体的には、独立成分分析 ICA (independent component analysis)、波形特徴抽出 SSF (scale-space filtering)、クラスタリング分析 CA (clustering analysis) を層状に組み合わせた波形分析技術を試作開発して、技術評価を実施中である。

セキュリティ評価技術の開発では、薬液による開封試験、精密研磨による開封試験の評価環境を構築した(図 3.12.3.4.2)。評価環境の一部は、品質評価技術の開発のために使用する。

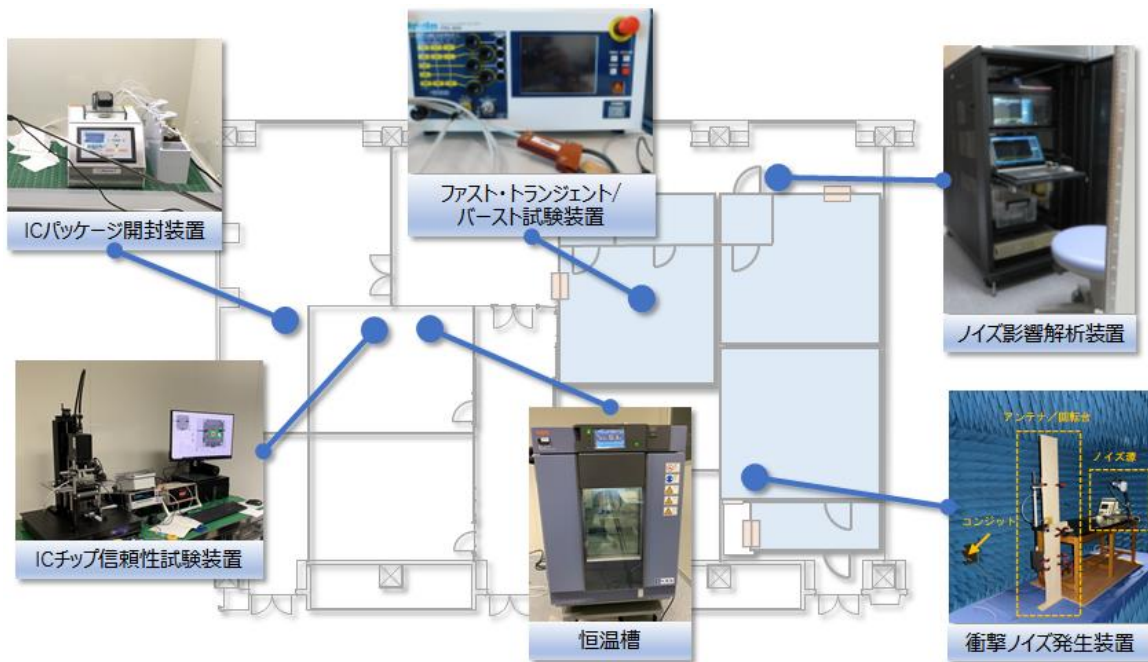


図 3.12.3.4.2 評価環境

また、評価基準開発のための基準規約文書の分析技術の開発では、階層的クラスタリング技術・自然言語処理技術・確率的潜在的機械学習技術(トピックモデル技術)を組み合わせた文書分析ツールを試作開発した。現在、分析精度を向上させるために、ローレンツ曲線による分布モデルによる分析設定の最適化の技術を開発して、技術評価を実施中である。

ファームウェア保護・分析技術の開発においては、まず、ファームウェアのバイナリ解析ツールを用いて、Cyber Grand Challenge (CGC)の関連する課題の大半を、自動解析できることを確認するとともに、ツール解析結果から、トロイの木馬に特有の振舞いを抽出し同定する手法を開発した。また、ツール解析結果からスタックオーバーフロー脆弱性により、関数の戻り番地が上書きされる可能性があるかどうかを自動判定する手法を開発した。そして、プログラムが事前に定められた制御フローから逸脱しないことを保証する制御フロー整合基盤技術をソースコードなしに実現する技術の検討を進め、脆弱性があった場合でも、サイバー攻撃による機器乗っ取りが生じないよう予防することができることを確認した。以上により当初に設定した中間目標を達成しており、今後いくつかの困難が予想されるものの、現状では概ね計画どおりに研究開発が進捗している。

3.12.3.5 成果の普及

本実施項目の背景にあるリバースエンジニアリング技術は、IC への侵襲的な解析技術と、非侵襲的な解析技術に分類できるが、前者の実態については公表されている情報が極めて少なく、また、科学的/学術的な根拠に基づいた評価技術が未確立である。そこで本実施項目では、電子デバイスの IC を換装・パッケージ開封して、情報を直接抜き取ろうとする破壊的な侵襲に対して無防備なサンプル品や少量生産品に対するコストコンシャスな耐リバースエンジニアリング技術とその評価技術を開発する。

意義としては、近年、下図に示すとおりアジア太平洋州での半導体生産が急速に伸びており、その背景に日本や欧米からの半導体製造装置の輸出等による技術流出とともに、アジア太平洋州企業の（模倣を含む）技術の向上が見込まれている。電子機器については、リバースエンジニアリングに無防備な電子デバイスが、IC の換装やパッケージ開封による情報抜き取りにより、事業者が技術開発に投じた金銭的・時間的投資を回収する前に、安価な模倣品が市場を席卷する状況が看過できないレベルに達している。

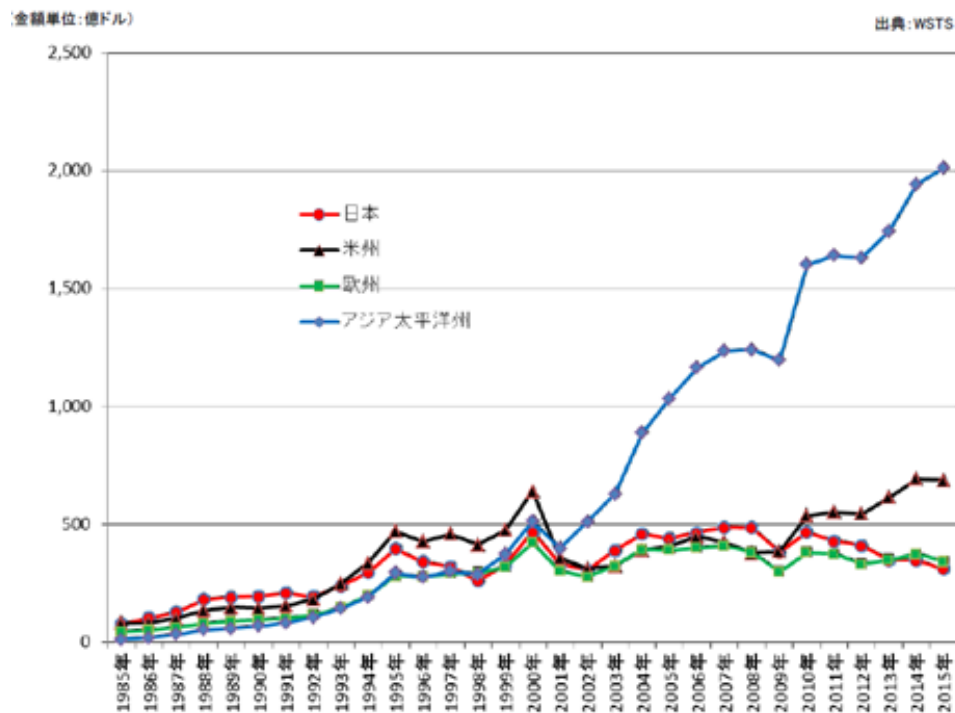


図 3.12.3.5 地域別にみる世界の半導体市場の変遷

本実施項目で取り組むコストコンシャスなリバースエンジニアリング対策技術とその評価技術は、リバースエンジニアリングに無防備な電子デバイスを対象に、安易な模倣行為への対策として有効である。具体的には、IC 保護技術により後付けしたフィルタを IC 起動のための固有値として活用することにより、セキュリティコアへの侵襲攻撃を防ぐ。また、セキュリティコアという固有値が IC 毎に異なることを活用して換装を防ぐ。今後、化学的侵襲解析と物理的侵襲解析に対する耐性評価を行う。化学的侵襲への耐性評価は、材料の選定、セキュリティコアの構成の適正性に還元できる。また、物理的侵襲への耐性評価は、セキュリティコアの形状（湾曲形状）に還元できる。各耐性評価の結果は、セキュリティ評価の基準開発にも用いられる。

ファームウェア保護・分析技術についても、近年、サプライチェーンにおけるセキュリティリスクが指摘されており、AI エッジデバイスのファームウェアについても、脆弱性やトロイの木馬に代表されるリスクが顕在化しつつある。こうしたリスクを、ソースコードの利用を前提とし

ない自動化技術で軽減することができれば、AI エッジデバイスの安心・安全な利活用を大きく促進することにつながり、これは大きな意義を持つ。

3.12.3.6 事業化までのマイルストーン

リバースエンジニアリングの糸口を与えないために、IC 保護技術に関する情報は非公開とし、広く一般には普及はさせてはならない。品質評価技術については、技術開発に注力しており、現在は、本格的な成果の普及は実施していないが、将来的に公開可能な部分については、口頭発表による成果の普及も予定している。ファームウェア保護・分析技術についても、現在は、手法の確立とツール実装を通じた性能確認・向上に集中しているが、公開可能な部分について一般向けの技術解説を発表している。

3.12.4 研究開発サブテーマ「実施項目 3：AI エッジの個体管理を支えるための人工物メトリクスの研究開発」

3.12.4.1 概要

付加価値の高い AI エッジデバイスやその部品に対する偽造や改変などの攻撃に対応するため、人工物メトリクスを用いた個体管理の研究開発を行っている。

人工物メトリクスの評価対象として、ランダム微細凹凸表面構造を有し評価の難しいナノ人工物メトリクスチップ（以下 NAM チップと呼ぶ）を選択し、このランダム微細凹凸表面を光学読取で得た画像を個体管理に使用するための評価に必要な装置およびシステムを構築している。具体的には、評価対象用 NAM チップ製造、画像照合、画像読取、NAM チップ貼付に関して進めてきた。評価対象用 NAM チップ製造では、2021 年 3 月迄に光学読取に対応したランダム微細凹凸構造の最適化、2022 年 3 月までにパイロット評価向けの個体管理仕様を策定することになっている。この NAM チップの製造は、ナノスケールランダム微細凹凸形状の形成技術を保有している会社に依頼しているが、将来の標準化を考慮して複数の試作ルートの確保に着手している。

光学的に読取った照合画像には、ランダム微細凹凸表面構造情報を含む白色干渉画像を使用している。白色干渉画像の特徴を把握し、それを反映したアルゴリズムを用いた画像照合技術とその評価技術の研究開発を行っている。また、本プロジェクトの利用では 1:1 認証が見込まれるので、ID 情報を記録する媒体（外枠、QR コード、ソフトウェア秘密鍵 etc）によるシステムを想定し、そのセキュリティ評価の検討も行っている。

白色干渉画像の取得では標準化の為に必須である、小型かつ低コスト製造が可能な画像取得装置の構成に関する研究開発および白色干渉画像の読取安定性の評価も行っている。

AI エッジデバイスの個体管理に NAM チップを用いるためには、NAM チップを AI エッジデバイス表面に、画像取得ができる状態で、かつセキュリティ要件を満足する実装が必要である。本研究開発では、AI エッジデバイスとして半導体デバイスやその部品を想定し、半導体パッケージ表面への実装に求められる要件の検討に必要な評価環境の構築を行っている。

一方、光学的読出に適さない用途への展開、あるいは既存 PUF との比較を通しての評価・分析・レベル分け等に必要な情報の取得を目的に、NAM の電氣的読取技術の研究開発も行っている。具体的には、ナノスケールのランダム微細凹凸構造を形成した Si ウエハ上に表面形状に敏感な MOS-FET のゲートを形成し、ゲート直下の電荷分布に反映された構造情報を、ドレイン電流特性の変化から読み出す方式に関する研究開発も並行して行っている。

3.12.4.2 最終目標と根拠

偽造困難な識別子の評価用サンプルを試作すると共に、そのサンプルを用いた個体識別子のセキュリティ評価分析装置やシステムの構築を行い、個体管理を支えるための人工物メトリクス評価技術を確立させる。具体的な最終目標は次の通りである。

- 「照合・識別評価」: 複数の製造メーカーが対応可能な汎用的な技術を用いて光学読取用のランダム微細凹凸構造の最適化を行うと共に、提供可能な評価用サンプル数をこれまでの数百個程度から、10,000 個以上にし、それらを用いて、FMR (False Match Rate) と FNMR (False Non Match Rate) を 10^{-8} 以下とする。
- 「読取評価」: 照合・識別に用いる干渉画像取得において必要なセキュリティ性能を示す最適装置構成を決定し、小型機を試作する。なお現有装置価格は～20M¥程度であり、目標コストは 5M¥以下とした。また、偽造困難な識別子を電氣的に読み出す際に達積可能な精度を明らかにする。
- 「貼付評価」: 偽造困難な識別子を部品等に十分な強度を持って貼り付け可能であることを評価するため、1,000 個以上の評価用サンプル試作し、また、それらの評価を可能とする環境を構築する。

3.12.4.3 目標の達成度

目標の達成度は次の通りで、最終目標に向かって計画通りに進行中である。

- 「照合・識別評価」: 12,600 個の評価用 NAM チップを搭載したウエハ試作に成功し、現在その白色干渉画像取得作業を進行中である。また、コスト削減のため複数社による試作に着手した。評価の結果、画像照合初期サンプルで、FMR と FNMR を 10^{-8} 以下とする見通しを得た。現在、12,600 個の干渉画像を用いた評価作業に着手している。
- 「読取評価」: 小型専用読み取り装置を試作し、白色干渉画像の取得と取得画像に対する照合・識別評価が可能であることを確認した。また、NAM 構造上への MOS-FET 試作プロセスを構築し、単一凹凸上への MOS-FET 試作を電氣的に読取り、シミュレーション結果をおおよそ再現した結果が得られている。
- 「貼付評価」: 半導体パッケージとして最も標準的である QFP-PKG を選択し、その表面への NAM チップの貼り付けとその評価が可能であることを実証した。1,000 個以上のサンプル試作が可能な試作装置の基本仕様を策定した。

3.12.4.4 成果と意義

人工物メトリクスの市場を広げるためには、目に見え難く実感し難いセキュリティを科学的/学術的根拠に基づき評価可能とする必要がある。本研究開発テーマでは、それを実現するための装置やシステムの構築をコストも視野に入れながら行っている。将来的な市場のセキュアで健全な拡大に寄与するものであり、その意義は大きい。

3.12.4.5 成果の普及

人得られた成果の内、公開可能なものは学術的な信憑性を高めるために学会等で積極的に発表すると共に、実施項目 0 を経由して国際標準化での議論に活用している。また、構築した評価装置やシステムを用いたセキュリティ評価受託を開始している。

4. 実用化・事業化に向けての見通し及び取り組みについて

4.1 「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」の研究開発における実用化・事業化の見通し（日本電気株式会社）

4.1.1 概要

本事業の実用化・事業化は 2019 年 9 月に設立したナノブリッジ・セミコンダクター株式会社(NanoBridge Semiconductor Inc., 略称 NBS)が行う。事業化のための、原子スイッチ関わる権利および日本電気株式会社との対外的な契約は新会社が承継する準備を進めている。

4.1.2 実用化・事業化への課題と対応策

本研究開発での ASSoC の製造は、異なるファウンドリーの 2 つのファブ利用する予定である。国内において技術ノード 28nm のトランジスタを製造するファブがないことから、海外のファブにおいてトランジスタ層および下層配線層を形成し、国内のファブにおいて原子スイッチおよび上層配線層を形成することで ASSoC が形成された 300mm ウエハを製造する。大量生産や生産管理のしやすさやの観点からは単一のファブで 300mm ウエハを製造することが望ましい。

また、ユーザーを広めるためには無償もしくは安価な ASSoC の設計ツールが必要である。現状の ASFPGA 設計ツールでは、論理合成を行うために商用の高額なツールライセンスが必要である。高額ゆえにライセンスを保有しているユーザーは限られているため、ユーザーを増やすためには安価にライセンスを提供する必要がある。また ASFPGA の設計ツールにおいてもデバック環境の提供、ユーザーインターフェースの改良等を行う必要がある。これらの設計ツールに関してツールベンダーとの協業を検討している。

4.1.3 実用化・事業化の体制

本事業の実用化・事業化は新会社(NBS)が担当する。事業化の当初は、顧客先での原理実証(PoC)、少量チップの製造受託事業となる。下図(a)は IT 機器を製造する顧客からの製造受託事業の商流を示したものである。新会社が顧客からチップの製造委託を受け、ファウンドリーに製造を発注しウエハを受け取る。新会社において後工程・テストを行い、チップを顧客に納入し、委託費を受け取る。事業化の当初は、少量生産であり、また顧客の数も限られると考えられたため、本商流となる。

事業を拡大させるためには、大量生産や顧客の数の増大への対応が必要となる。下図(b)は、事業が拡大した場合のライセンス事業を説明したものである。新会社は ASFPGA/ASSoC の対価としてファウンドリーよりロイヤリティを受け取る。ファウンドリーに ASFPGA/ASSoC を回路 IP ライブラリとして登録され、顧客(ファブレス・LSI 設計会社)の要望により、ファウンドリーがチップを製造する。ファウンドリーが回路 IP の対価を徴収し、新会社にロイヤリティとして支払う。

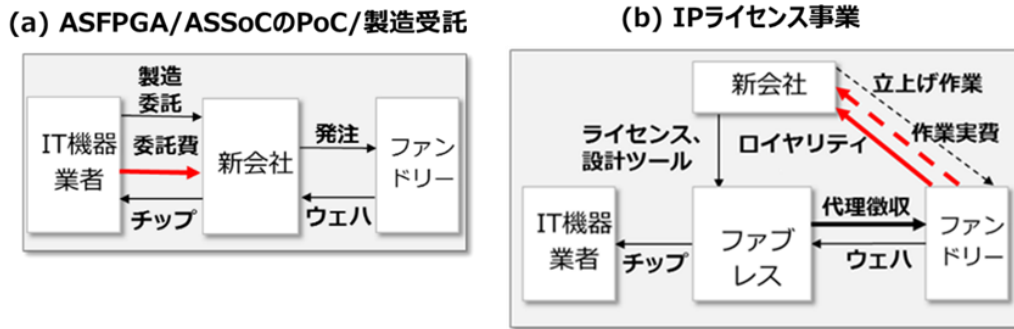


図 4.1.3 実用化・事業化における体制

4.1.4 市場規模と経済効果

全世界の FPGA 市場は 7,500 億円である。近年は、FPGA の市場の伸びは半導体 LSI の伸び(2.4%)を上回り、年率 8%程度で成長を続けている。この高い伸びは、FPGA が他の半導体 LSI 市場である ASIC/ASSP 等を置換えていることによる。

低電力・高性能化・耐環境性能を備えた ASSoC/ASFPGA は、従来 FPGA 置換え、さらなるロジック LSI である ASIC/ASSP の置換えが可能である。

2023 年度における FPGA 市場におけるターゲット市場は約 5,500 億(内 1,000 億が ASSoC)と考えられる。本市場は、技術ノード 14nm 以降の高性能・大規模 FPGA を必要とする市場を除いたものである。5,500 億円の内、ASSoC および ASFPGA の市場占有率を 12%、その内 IP ビジネスでロイヤリティ 2-5%が売上となる。

4.1.5 ベンチマーク

本技術と競合する技術である不揮発 FPGA を提供する競合ベンダーはマイクロセミ(米国)である。同社の不揮発素子技術はフラッシュメモリまたはアンチヒューズである。フラッシュメモリは書き換え可能だが、放射線エラーが存在する。アンチヒューズは放射線エラーがない一方で、書き換えが不可である。原子スイッチは放射線エラーがなく、書き換えも可能である。待機時に電源を切り、低電力化できる点は同じである。

既存の FPGA メーカーによる SoC-FPGA との比較では、インテル(米国)、ザイリンクス(米国)などの FPGA メーカーが販売する SoC-FPGA があり、搭載 FPGA コアは SRAM 型である。ASFPGA コアは、面積を小さく出来る分、従来の SRAM 型 FPGA コアより電力効率(4 倍)が良く、動作時の低電力化も可能である。

FPGA コア IP 事業モデル採用企業間の比較を行う。エッジ AI 用 SoC に向けて、アプリごとに最適な FPGA コア IP の提供という意味での先行ベンダーは、米国ベンチャー企業フレックスロジック社である。彼らの FPGA コアは揮発性であるため取り扱いが難しい。原子スイッチ FPGA コアは不揮発性であるため、回路書き込み後の取り扱いが容易で、ユーザーの使い勝手が良い。

4.1.6 事業化までのマイルストーン

2019年9月、技術ノードが40または65nm世代の原子スイッチの製造を国内ファブで行う新会社を立ち上げ活動中である。本開発の再委託先であるソリトンシステムズや、東京大学稲葉研究室からのベンチャー企業、自動車関連会社をはじめとするエンドユーザーにおいてASFPGAやASSoCの価値実証を行う。

2022年3月、28nm世代のASSoCの技術が本テーマで完成する。

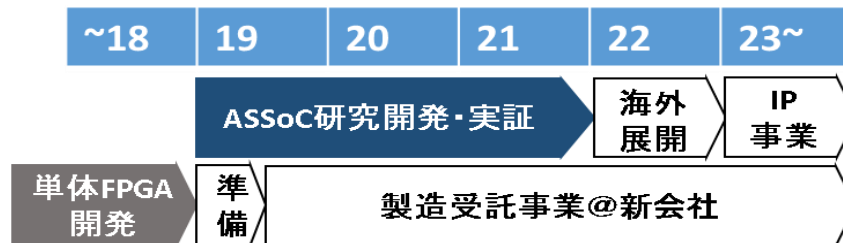


図 4.1.6 実用化・事業化におけるマイルストーン

4.2 動的再構成技術を活用した組込み AI システムの研究開発における実用化・事業化の見通し

4.2.1 動的再構成技術を活用した組込み AI システムの研究開発における実用化・事業化の見通し（ルネサスエレクトロニクス株式会社）

4.2.1.1 概要

動的再構成組込み AI チップに、AI フレームワークを含む開発ツールまでを垂直連携することでえられる従来技術比10倍以上の電力性能の効果を、高効率・高速 AI チップ製品・サービスとして提供。

また、製品適用分野としては以下の市場を想定している。

表4.2.1.1 AI搭載が期待される事業例

市場	顧客	提供ソリューション案
ファクトリーオートメーション	機器メーカー、SIer	自律制御、予知保全、異常検出、検査、動作安定向上
ビルディングオートメーション	機器・住設メーカー、SIer	異常検出、メンテナンス支援
コグニティブカメラ	機器メーカー、SIer	物体認識、人物・行動認識、予測、個人認証
家電/ヘルスケア	機器メーカー	物体認識、人物・行動認識、疾病予防、個人認証
自律型モビリティ	機器メーカー、SIer	自動運転、物体認識、自律制御
サービスロボット	機器メーカー	自律制御、物体認識、人物・行動認識

4.2.1.2 実用化・事業化への課題と対応策

本研究成果である AI 電力性能は、組込み機器製造販売事業者だけでなく、AI 応用の特徴からエンドマーケットでの利用者を対象としたサービス提供事業者および機器・サービス利用者自身まで広範囲な活用が期待できる。そのため、これら多様化するビジネス形態に対し、製品・サービスの提供と、その対価獲得のための仕組み構築が必要となり、この実現が実用化・事業化の課題である。

本研究では、研究成果を「開発ツールおよび AI ライブラリを含む AI チップに最適化された開発環境」と、「高効率・高速処理を実現する動的再構成 AI チップ」とで構成し、これらをルネサスエレクトロニクスが MCU・MPU として提供中の生産・品質管理された半導体ソリューションの形態にて汎用的に市場に提供することを想定している。

4.2.1.3 実用化・事業化の体制

「開発ツールおよび AI ライブラリを含む AI チップに最適化された開発環境」と、「高効率・高速処理を実現する動的再構成 AI チップ」の製品・サービスは、ルネサスエレクトロニクスが提供する半導体ソリューションの形態にて提供する。

4.2.1.4 市場規模と経済効果

(公開しない)

4.2.1.5ベンチマーク

エンドポイント学習に必要な AI チップの電力性能は、本研究の設定目標である 10TOPS/W に変更はない。また、現状においてエンドポイントでの学習に対応した競合はほとんど存在せず、ほぼすべての競合 AI チップは推論専用であり、これについても研究当初の想定通りである。目標達成時には市場をリードする AI チップを実現可能である。

4.2.1.6事業化までのマイルストーン

本成果をルネサスエレクトロニクスにて、市場提供可能な品質を有する半導体ソリューションとし、MCU・MPU を中心に構築済みの営業体制および販売チャネルを活用し、製品・サービスとしての提供を開始する。

(時期は公開しない)

4.2.2 競争学習機構による汎用・超軽量エンドポイント学習技術の開発」における実用化・事業化の見通し（SOINN 社）

4.2.2.1概要

マーケットニーズが顕在化している、①川崎重工社“Successor”向け事業、②“SWITCH VISION”事業を起点に、そこで培った技術やノウハウを生かし、研究成果を多面的・多角的に市場に投入する。SOINN on DRP 開発の最終ゴールは、産業向け、民間向け用途問わず広く社会実装を進め、日本発の技術で日本が世界をリードできる事業、産業を創出することにある。

4.2.2.2実用化・事業化への課題と対応策

①川崎重工社“Successor”向け事業とその発展

国際ロボット連盟の統計では、日本のロボット密度（製造業1万人あたりの産業用ロボット利用台数）は約3%である。この数値は日本以外のロボット化が進む国々でも同様で、高齢化による労働力不足が深刻化する中、世界主要国の製造ラインの94~99%を人が担っていることを示している。

ロボット化が困難なタスクの、主な理由・課題を以下に示す。

<課題>

・技能：人の感覚や経験に大きく依存する工程や、作業のばらつきが大きい工程がある。

・プログラミング：ロボット導入時のプログラミングが煩雑、もしくは事実上不可能な工程がある。

・コスト：多様なセンサ、頻繁な設備改善、多品種少量生産への対応が必要である。



<課題への対応策>

- ・“Successor”にはベテラン職人の「技」をリアルタイムに計測する機能がある。
- ・SOINN は多様な時系列・マルチモーダルデータの学習と、学習済モジュールを使った高速推論(予測)が可能である。
- ・上記双方のメリットを融合し、職人のばらつきのある作業工程をダイレクトに SOINN に学習させ、再現させる。これにより、職人技能のダイレクト学習、プログラミングレス、大幅コストダウンを同時に実現する。

②“SWITCH VISION”とその発展事業

“SWITCH VISION”は OK 品のみから現場エッジ側で教示学習して運用でき、再学習や追加学習も容易(CPU で数分～1時間程度)である。さらに、何をどこまで学習し、判定基準がどこにあるかの表示・確認・微調整も可能など、他に類を見ない特長がある。これをさらに発展させ、SOINN on DRP を活用した“SWITCH VISION”は、名刺サイズで省電力、省演算、低コストの AI デバイスとしてプラントの各所に多数配置され、歩留まり率の改善や、設備の異常検知に大きく寄与すると考える。

<課題>

- ・SWITCH VISION としての適用案件数を増やし、インターフェース設計も含めた様々な現場ニーズへの対応の推進。
- ・プリセットするモジュール機能やインターフェース機能の拡充。

<課題への対応策>

- ・外部販売パートナーと連携して適用案件数を増やすとともに、現場からのフィードバックを分析して改良を重ねる。

4. 2. 2. 3 実用化・事業化の体制

多くの外部企業と協業の上、取り組むことを予定している。

4. 2. 2. 4 市場規模と経済効果

(公開しない)

4. 2. 2. 5 ベンチマーク

(公開しない)

4. 2. 3 事業化までのマイルストーン

(公開しない)

4.3 FPGA IP と可変精度演算コアの融合による超低消費電力エッジヘビーコンピューティング向け SoC の研究開発」の研究開発における実用化・事業化の見通し

4.3.1 概要

近年、クラウドにおける深層学習・機械学習の需要は極めて急速に拡大している。その市場規模としては、2015 年に 3.6 兆円、2025 年には 50 兆円まで増加すると見込んでいる。現在、この市場は NVIDIA が独占している状態にあるが、この研究で開発したサーバー・クラスターを事業化し、エッジ、ロボティクスに応用することで、長期の収益、および経済の活性化が期待できる。

4.3.2 実用化・事業化への課題と対応策

近年、クラウドにおける深層学習・機械学習の需要は極めて急速に拡大している。その市場規模としては、2015 年に 3.6 兆円、2025 年には 50 兆円まで増加すると見込んでいる。現在、この市場は NVIDIA が独占している状態にあるが、この研究で開発したサーバー・クラスターを事業化し、エッジ、ロボティクスに応用することで、長期の収益、および経済の活性化が期待できる。

4.3.3 実用化・事業化の体制

多くの外部企業と協業の上、取り組むことを予定している。

4.3.4 市場規模と経済効果

(公開しない)

4.3.5 ベンチマーク

(公開しない)

4.3.6 事業化までのマイルストーン

(公開しない)

4.4 研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」の研究開発における実用化・事業化の見通し

4.4.1 低進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し (ArchiTek 株式会社、株式会社ソシオネクスト)

4.4.1.1 概要

(ArchiTek 株式会社)

本委託事業において試作 LSI1 により仮想エンジンアーキテクチャを回路実証することができた。今後はこの仮想エンジンアーキテクチャを AI エッジプラットフォームとして展開すべく IP としての実用化およびより小型で低消費電力の LSI を開発し事業化したいと考えている。

(株式会社ソシオネクスト)

ソシオネクストでは、Factory Automation、物流ロボット、セキュリティ・見守り、車載カメラ・運転支援の市場に向けて、本成果を搭載した ASIC、ASSP のビジネスを展開したいと考えている。

4.4.1.2 実用化・事業化への課題と対応策

(ArchiTek 株式会社)

仮想エンジンアーキテクチャを AI エッジプラットフォームとして展開すべく IP としての実用化および、より小型で低消費電力の自社 LSI 提供を検討

(株式会社ソシオネクスト)

本成果を搭載した ASIC、ASSP のビジネスを展開していく。

4.4.1.3 市場規模と経済効果

(公開しない)

4.4.1.4 ベンチマーク

(公開しない)

4.4.1.5 事業化までのマイルストーン

(公開しない)

4.4.2 進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し（株式会社ソシオネクスト）

4.4.2.1 概要

ソシオネクストでは、Factory Automation、物流ロボット、セキュリティ・見守り、車載カメラ・運転支援の市場に向けて、本成果を搭載した ASIC、ASSP のビジネスを展開したいと考えている。特に、量子化 DNN 技術/DNN 変換ツールは、AI 処理向け IP として ASIC へ適用。また、AI エッジ LSI は別途量産用 LSI を開発し、ASSP 事業を展開する予定。

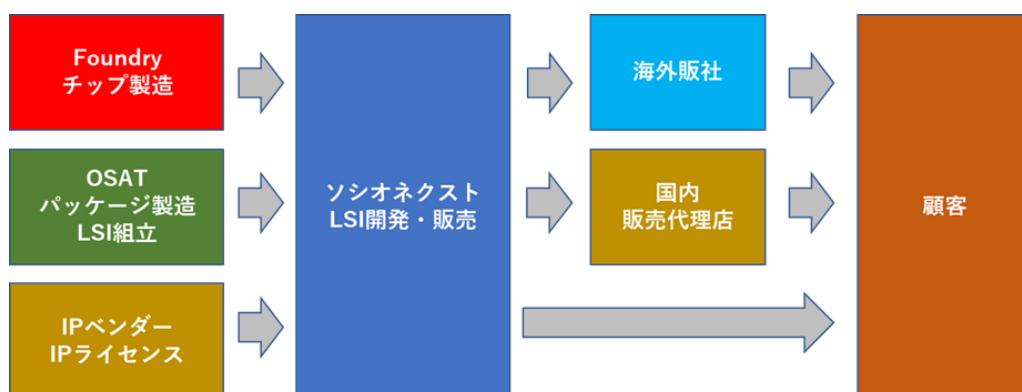
4.4.2.2 実用化・事業化への課題と対応策

AI 認識処理の実用化には、顧客環境・実アプリケーションを使った、より実践的な評価、改善が重要と考えている。

本課題に対しては、2020 年 3 月より外部会社評価および関係会社へ評価を依頼し、実施中。これら評価結果を反映することにより、より実用的なシステムとなるように改善する予定。

4.4.2.3 実用化・事業化の体制

実用化・事業化の体制は、下記のような体制を予定している。



4.4.2.4 市場規模と経済効果

（公開しない）

4.4.2.5 ベンチマーク

（公開しない）

4.4.2.6 事業化までのマイルストーン

（公開しない）

4.4.3 進化型・低消費電力 AI エッジ LSI の研究開発における実用化・事業化の見通し（株式会社豊田自動織機）

4.4.3.1 概要

本プロジェクトで開発した AI エッジ LSI をソシオネクスト社で製品化し、弊社に供給いただく事を前提で実用化・事業化を計画している。弊社は、製品版 AI エッジ LSI を搭載した車載 ECU の製品版を開発し、この ECU に本プロジェクトで開発したリアルタイム SLAM 技術を実装し、各種物流ロボット等弊社製品の高機能・高性能化を計画中。

4.4.3.2 実用化・事業化への課題と対応策

実用化・事業化の課題は、本テーマで開発・試作した AI エッジ LSI の製品化である。弊社としては、AI エッジ LSI の製品化計画が固まり次第、本 LSI を使用した車載 ECU の製品版を開発し、この ECU で SLAM 処理を行う製品開発を推進する。

4.4.3.3 市場規模と経済効果

- ・AI エッジ LSI の製品版供給：ソシオネクスト社
- ・SLAM ライブラリ共同開発：ArchiTek 社
- ・AI エッジ LSI 製品版を搭載した製品開発：豊田自動織機

4.4.3.4 ベンチマーク

（公開しない）

4.4.3.5 事業化までのマイルストーン

AI エッジ LSI の製品化計画が固まり次第、本 LSI を搭載した車載 ECU を使用し、リアルタイム SLAM 処理を行う自律走行台車の製品開発を推進する。

4.5 ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発の研究開発における実用化・事業化の見通し（沖電気工業株式会社、総合警備保障株式会社、ジャパンマリンユナイテッド株式会社）

4.5.1 概要

本研究開発の成果の普及については、本研究開発は、各領域でトップクラスのシェアを有する実事業会社自身による事業検討と研究開発を一気通貫した産学連携体制で推進しており、学会発表・論文、特許出願、新聞雑誌 Web などのメディア掲載、プレスリリース、展示会や講演にも積極的に取り組んでいる。

現時点で、下記の実績を挙げており、最終年度である今年度の成果も追加することで、さらに増加する予定で進めている。

現時点での実績は下記のとおりである。

- ・特許は、出願済・出願準備中を併せ 17 件、今後も増加見込みである。
- ・論文の実績は 9 件であり、今後も増加見込みである。
- ・学会発表・講演の実績は 19 件であり、今後も増加見込みである。

また、本研究開発の共同研究機関の共著も 1 件あり、今後も取り組む予定である。

- ・新聞・雑誌、プレスリリースの実績は、13 件であり、今後も増加見込みである。
- ・展示会への出典の実績は、8 件であり、今後も増加見込みである。

例として、展示会への出典、講演などでは、CEATEC のような大規模展示会への出典、発表や ET & IoT Technology など、今後、本研究開発成果が適用される分野の展示会において講演することで、本研究開発成果についての普及につながると考えている。

研究開発開始時点において、「NEDO の革新的 AI エッジコンピューティングをテーマとする技術開発プロジェクトを開始」(2018 年 9 月 25 日)を、OKI、JMU、ALSOK、会津大による 4 機関の共同プレスリリースを行い、プレスリリースによる本研究開発の社会への認知度向上の活動を行っている。

2 年目には、NEDOおよびOKIで「ディープラーニングモデルの新たな軽量化技術を開発」(2019 年 9 月 9 日)の共同プレスリリースを行い、16 件の記事掲載(新聞 5 雑誌 1 Web 10)があり、注目される活動も行っている。本研究開発の成果を今後適宜公表する予定である。

4.5.2 実用化・事業化への課題と対応策

(公開しない)

4.5.3 実用化・事業化の体制

OKI、ALSOK、JMU の 3 社は、事業会社であり、本研究開発を行う部門と事業部門が連携して実用化・事業化に向けて取り組みを行っている。

4.5.4 市場規模と経済効果

(公開しない)

4.5.5 ベンチマーク

(公開しない)

4.5.6 事業化までのマイルストーン

(公開しない)

4.6 5G時代を見据えた高度自律的学習機能搭載のためのAIエッジコンピューティング技術の研究開発における実用化・事業化の見通し（株式会社アラヤ、KDDI株式会社）

4.6.1 「ネットワーク圧縮とBaseNetwork 共通化による演算量削減技術の開発」における実用化・事業化の見通し（株式会社アラヤ）

4.6.1.1 概要

Base Network 層共通化とアラヤの圧縮技術を組み合わせた複数深層学習に対する圧縮技術は、汎用自動ネットワーク圧縮ツール(Pressai)として2020年3月に商用ツールとしてリリースし実用化した。本圧縮ツールの開発ライセンス料、および、本圧縮技術を搭載した量産製品に対するライセンス料に基づく事業を進めつつ、更なる事業拡大施策を講じる。

4.6.1.2 実用化・事業化への課題と対応策

Pressaiを中核とする実用化・事業化拡大として下記対応を進める。

・対応FPGAの拡大

FPGAはそもそも量産適用となるデバイス数が少ない課題がある。20年度までに開発したIntel社FPGA向け実装を量産適用数が多いLowレンジFPGAへの最適化し、現在ARM社CPUが主流のLowレンジハードウェアを用いたAI市場へのFPGA参入を進める。また、Intel社製FPGAが主要シェアを通信やサーバ領域への適用を進める。更には、車載向けFPGA市場でシェアを占めるXilinx社製FPGA対応開発を進め自動車セットメーカー・サプライヤへの採用を図る。

・ASIC、AIチップ向けIP技術提供

FPGAは単価が高いデバイスであり事業化の観点ではFPGA搭載機器の出荷台数伸長には課題がある。対応策として、20年度までの研究開発成果に基づきASIC/AIチップ向けIP開発を開発する。現在協業するセットメーカーで検討している複数のAIユースケースに適用可能な複数深層学習対応ASIC・AIチップとして性能改善、量産化を目指す。

・エッジAI受託開発の推進

汎用自動ネットワーク圧縮ツールの単体販売では、アルゴリズム開発を必要とするユーザにリーチできない課題がある。対応策として、これまで研究開発で得た圧縮技術の知見に基づきエッジAI実装の開発・導入全般を支援するコンサルティング事業を行う。実装ターゲットハードウェアをこれまで開発してきたMidレンジFPGAに加えて、LowレンジFPGAやCPU、GPU等のエッジデバイスに最適化しつつ、アルゴリズム開発からネットワーク圧縮、エッジAI実装まで一気通貫で行うコンサルティング事業を進める。

4.6.1.3 実用化・事業化の体制

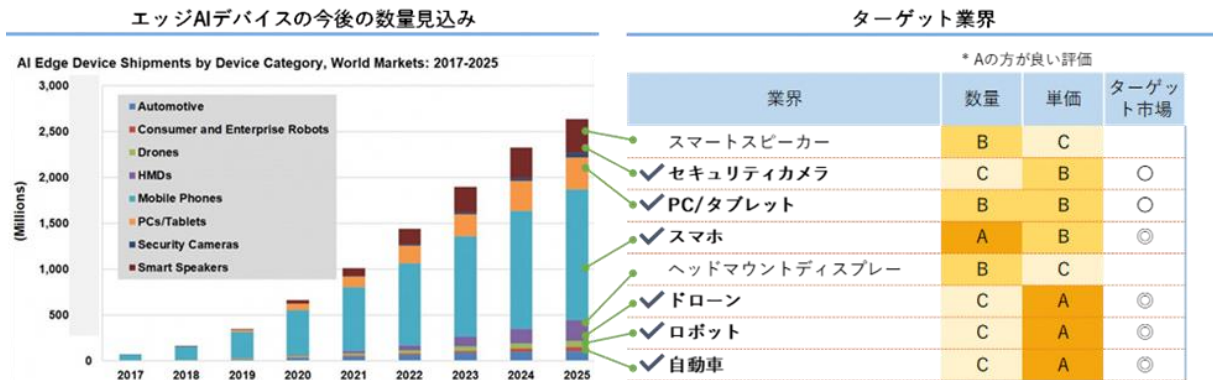
本研究においてKDDIが開発を進めるAIエッジ統合制御システムを通じてクラウドサーバで用いられるFPGAへの適用を進める中でサーバ事業者との協業体制確立を進める。また、車載向けFPGA、AIチップへの技術適用を進める中でユーザニーズに基づく圧縮技術の改良を進めつつセットメーカー及びサプライヤとの協業体制構築を図る。

4.6.1.4 市場規模と経済効果

対応 FGPA の拡大、ASIC・AI チップへの対応、エッジ AI 受託開発の推進により、深層学習用チップセット世界市場における目標シェアを、ターゲット市場(下表の市場分析より):スマホ、自動車、ロボット、ドローン、PC、セキュリティカメラをメインとし、競争環境も鑑み、以下と考える。

深層学習用チップセットの世界市場と目標シェア(1台平均 1000 円と概算)	
市場規模(世界市場)	目標シェア
2022 年(1300 万台)	13 億円
2025 年(4000 万台)	50 億円

市場規模算出の根拠:市場調査会社(Tractica)の市場予測資料を参照



出典: <https://www.tractica.com/newsroom/press-releases/deep-learning-chipset-shipments-to-reach-41-2-million-units-annually-by-2025> より作成

4.6.1.5 ベンチマーク

本開発技術はプルーニング等の既存の圧縮技術と併用でき、それら既存圧縮手法を含み一式ワンストップで取り揃えたツール(Pressai)として提供している。既存 AI プラットフォームとの対比による自動ネットワーク圧縮ツールの優位性を下図に示す。

圧縮ツール		モデル提供			学習				圧縮				推論実行ファイル		Edgeデバイス	
ツール名	開発元	DLモデル	学習	蒸留	転移学習	アラヤ圧縮	Prune*1	量子化	行列分解	CPU最適化	HDL記述言語	Deploy	Monitor			
自動ネットワーク圧縮ツール(仮名)	株式会社アラヤ	-	○	○	-	○	○	○	-	○	○	○*2	-			
Cloud AutoML Vision	Google LLC	○	○	-	-	-	-	-	-	-	-	-	-			
Cloud AutoML Vision Edge	Google LLC	○	-	-	-	-	-	-	-	-	-	○	-			
TensorFlow Lite	Google LLC	-	-	-	-	-	○	○	-	○	-	-	-			
Optimization tool kit	Google LLC	-	-	-	-	-	○	○	-	-	-	-	-			
QNNPACK	Facebook	-	-	-	-	-	-	○	-	-	-	-	-			
SageMaker+AWS market place	Amazon	○(300以上)	○	-	-	-	-	-	-	○	-	○	-			
Azure IoT Edge	Microsoft	○(数十個)	○	-	-	-	-	-	-	-	-	○	○			
Transfer Learning Tool Kit	NVIDIA	-	○	-	○	-	○	-	-	-	-	-	-			
TensorRT	NVIDIA	-(AWS連携)	-	-	-	-	-	○	-	○	-	○	-			
Neuralnetwork Distiller	Intel	-(AWS連携)	○	○	-	-	○	○	-	-	-	-	-			
PocketFlow	Tencent	-	○	○	-	-	○	○	-	-	-	-	-			
keras_compressor	ドワンゴメディアビレッジ	-	○	-	-	-	-	-	○	-	-	-	-			
Blueoil	LeapMind	-	○	-	-	-	-	○	-	-	-	-	-			
DeLTA-Lite	LeapMind	○	○	?	-	-	?	?	-	-	-	-	-			
DeLTA-Kit	LeapMind	-	-	-	-	-	-	-	-	-	-	○	-			

図 4.6.1.5.1 開発・商用化した自動ネットワーク圧縮ツールと既存 AI プラットフォームとの対比

アラヤ独自圧縮の圧縮性能について、マグニチュードベースプルーニング(重みの大きさに基づく枝刈)との比較を下図に示す。圧縮率を高くする(compression_rate を下げる)と、ア

ラヤ圧縮のいても精度低下はあるが(精度:0.7→0.5)圧縮率に応じて実行時間は短縮する。一方で、マグニチュードベースプルーニングも圧縮率に応じて精度低下するが、実行時間は短縮しない。これは、アラヤ独自圧縮が、モデルを構成する各層を枝刈りしつつ、残った非ゼロ重みのみをFPGA実装で効率よく演算可能であることを示している。

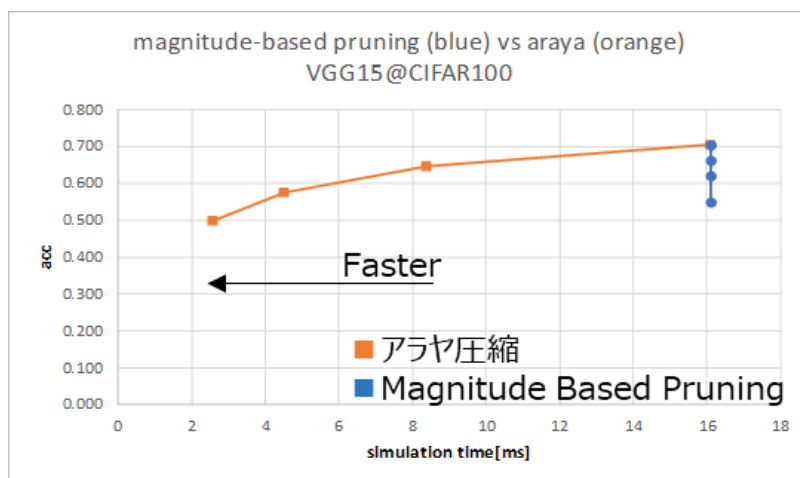


図 4.6.1.5.2 アラヤ独自圧縮性能ベンチマーク

4.6.1.6 事業化までのマイルストーン

実用化・事業化で開発、検討を進める各施策のマイルストーンを下図に示す。20年3月にリリースした Pressai をベースに、対応 FPGA の拡大、ASIC・ASI チップ向け IP 技術開発・提供及びエッジ AI コンサルティング事業推進により、更なる事業化を進める。

	2020年	2021年	2022年	2023年
Pressaiリリース	▼2020年3月			
対応FPGA拡大	Lowレンジ対応▼ Xilinx対応▼	Lowレンジ市場・車載対応		
ASIC、AIチップ向けIP技術提供	ASIC対応IP開発	性能改善	量産対応	
エッジAIコンサルティング事業推進	(エッジAI開発に関する包括的なコンサルティング・受託開発事業を実施)			

図 4.6.1.6 事業化マイルストーン(演算量削減技術)

4.6.1 「人工意識による高度自律的学習機能の開発の研究開発」における実用化・事業化の見通し (株式会社アラヤ/国立大学法人 東京大学)

4.6.1.1 概要

エッジ AI コンピューティングで高度自律的学習機能の実現に向けて、AI 自体が目的設定と自律的な学習を行う、人工意識、と呼ぶアーキテクチャの開発を行う。

・現在の AI アルゴリズムは、学習した技術を蓄積することができないため環境の変化毎に都度大量データを用意して学習する必要がある。この課題を解決するために、環境予測

モデルとメタ認知モジュールを組み込んだメタ学習やモデルベースの強化学習アルゴリズム開発、および、少量データで効率的に学習可能な環境シミュレータの開発を進める。

・ ドローン、ロボットアーム、協業ロボット等の分野におけるタスクに対する実用化を目指しアルゴリズムやシミュレータのライセンス料に基づく事業化を進める。

4.6.1.2 実用化・事業化への課題と対応策

実用化・事業化に向けた取り組みを示す。

- 既存の深層強化学習手法に対して性能や学習効率において優位となる人工意識アーキテクチャに基づくアルゴリズム開発、および、学習環境シミュレータ開発を行う。
- アラヤが既に事業として行なっている、画像認識サービス、建設機械自動化支援サービス、自律飛行ドローンによる点検自動化サービス、建物空調最適化サービスなどへの適用を行うことで、これらサービスの競合優位性を高めていく。
- KDDI の開発する AI エッジ統合制御システム／スマートドローンプラットフォームに、データ効率性を備えた高度自律的学習機能を適用していくことで、KDDI プラットフォームにおけるドローン分野でのライセンスベースの事業化を進める。

上述の取り組みにおける課題として、個別のサービス・事業については、各業界の既存プレイヤーが AI などの先端技術を自社開発することで、スタートアップなどの新興勢力への競争力を強化してきている。対応策として、AI を強みとするアラヤにとっては逆に業界知識の不足によって、実装や効果的な AI の活用が困難となる場面が今後も発生すると考えており、今まで以上に、各業界の知識・ノウハウを持つ企業と連携して、事業を推進していく予定である。

4.6.1.3 実用化・事業化の体制

KDDI との協業体制において、本研究で開発を進める AI エッジ統合制御システムを活用することで実用化・事業化の取り組みを進める体制作りを進める。各ビジネスドメインの事業パートナーのノウハウの提供を受けることで最適な高度自律的学習アルゴリズムに基づく自律エージェントサービスを提供することでエンドユーザへのサービス・価値提供を進める。

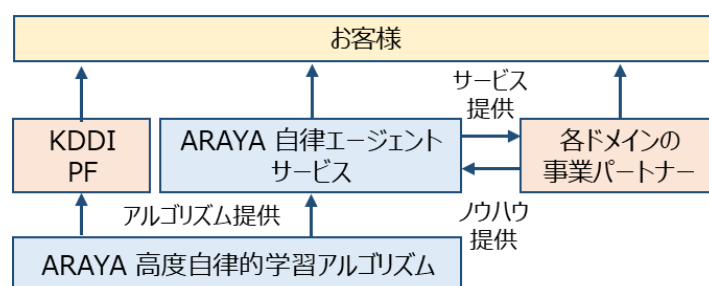


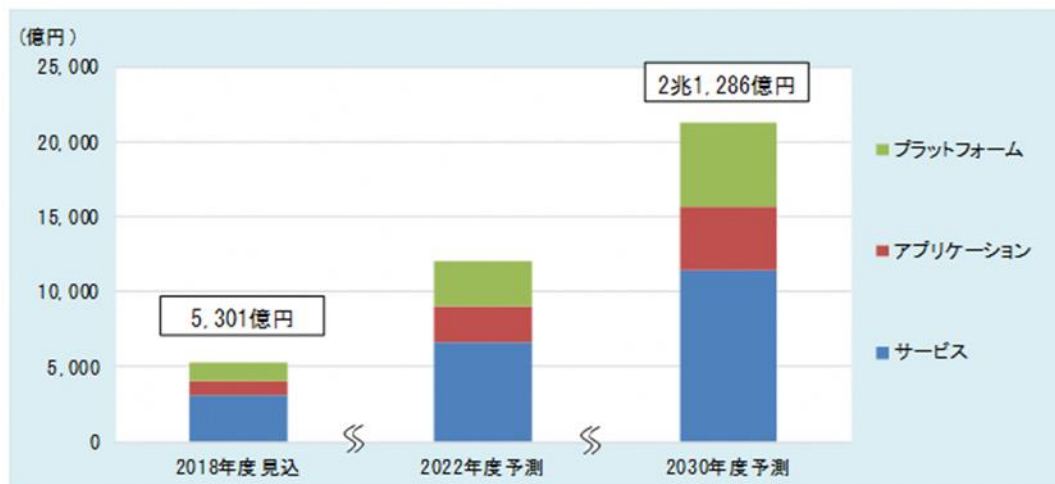
図 4.6.1.3 実用化・事業化の体制(人工意識)

4.6.1.4 市場規模と経済効果

富士経済グループより発表された「2019 人工知能ビジネス総調査」の予測より、AIビジネス国内市場全体 2 兆 1,286 億円において、国内エッジ AI コンピューティング市場は、2018 年度に見込み 110 億円だった市場規模が、2030 年度には 664 億円にまで拡大する見込みとなっている。この市場規模の急激な拡大にあたり、民生機器および産業機器で使用される

学習効率性の改善が大変重要であり、ここにアラヤが提供する自律的学習機能が重要な役割をもち、前述の、5G 及びエッジ・クラウド連携における開発プラットフォームとしての提供及び演算量削減技術とのセットで競争力をもって市場提供できるものと考えている。

メタ学習・模倣学習等の周辺技術動向や上記競争優位の想定を鑑み、国内エッジ AI コンピューティング／プラットフォーム・アプリケーション市場（664 億円の概ね 4 割）の内、2030 年度にはシェア 66.5 億円（25%）程度の貢献を果たす事が可能と考えている。



出典: <https://www.fcr.co.jp/pr/19039.htm>

4.6.1.5 ベンチマーク

既存技術 (SLAM: 自己位置推定/マップ作成、SLAM+深層強化学習) と SLAM+人工意識ありの深層強化学習との比較評価を下図に示す。環境変化や新規環境、新規タスク・目的への対応に対する高効率性と頑健性の実現を目指している。

比較の観点	SLAMのみ	SLAM+深層強化学習	
		SLAM+人工意識なし	SLAM+人工意識あり
方式の概要	マップを作成した上で、マップ上でタスクを実現する飛行経路を計画する。	マップを作成した上で、タスクを実現させるような動作を学習させる。	マップの作成、タスクを実現させるような動作に加え、転移学習によって新規環境への適用性を向上。
環境変化への対応	再度マニュアルで飛行させてマップを作成する必要がある。	マップの再作成の必要はあるが、事前に環境変化を見越したポリシーを学習しておくことで小さな変化には対応可能。	事前に想定した環境変化に対するポリシー以外のポリシーも生成可能なため、より大きな変化に対応可能。
新規環境への対応	同上	再学習が必要。	少ない追加学習で新規環境に適用できる。
新規タスク・目的への対応	作成済のマップ上で新しく飛行経路を計画することで対応可能。	同上	同上
動的な環境への対応	他技術と組み合わせが必要。	事前に動的な環境で学習しておくことで、対応可能。	←
解釈性	全てはマップ上で解釈されるため、視覚的に分かりやすい。	マップ以外の、NNの中での表現は解釈が難しい。	←

図 4.6.1.5.1 既存技術とのベンチマーク

また、学習におけるデータ効率性を指向する先進研究開発として、Google による 2020 年の IEEE International Solid-State Circuits Conference (ISSCC)における研究との対比を示す。Google による、今後の特にクラウドにおける AI(機械学習)の方向性を下記に示す。

- ▶ 少ないデータ量で学習できるようにすることが今後の重要な研究の方向性の一つである(=Few shot learning)。

- アプローチとしては、複数のエキスパートモデルを含んだモデルに対して、少しのデータを与えることで一部のモデルだけをアクティベートさせることでタスクを解く方法 (=sparsely activated models) や、大きいモデルを作成し個別タスクで追加学習する方法等がある。

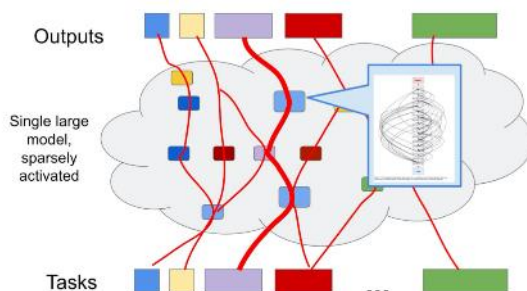


Figure 8: A diagram depicting a design for a large, sparsely activated, multi-task model. Each box in the model represents a component. Models for tasks develop by stitching together components, either using human-specified connection patterns, or automatically learned connectivity. Each component might be running a small architectural search to adapt to the kinds of data which is being routed to it, and routing decisions making components decide which downstream components are best suited for a particular task or example, based on observed behavior.

図 4.6.1.5.2 IEEE ISSCC 2020 での Google による AI(機械学習)の方向性

これに対して、本提案技術では、今後の AI エッジ端末増大に伴い、汎用モデル作成はクラウド上で、追加学習はエッジ上でと役割を分担した上で、データ連携・リソース配分方法の最適化を、5G 及びクラウド・AI エッジ連携の国内初総合的プラットフォームである KDDI プラットフォームにおいて実現することで、システムエネルギー効率性とエッジ高度化の実装システムを先行して社会に提供する。これは世界的に見ても新しい取り組みであり、5G を活用したエッジクラウド連携によるデータ効率性を実現することでこれまでにない新たな価値提供を目指す。

4.6.1.6 事業化までのマイルストーン

本研究開発において高度自律的学習機能・人工意識アーキテクチャの開発を進めることでアーキテクチャの確立と有用性の検証を行う。並行してベースモデルとなる「WorldModel」ベース及び「Predictive Coding」ベースのアルゴリズム開発を進めるためのシミュレータ開発を進め有用性の評価を行う。2021 年度以降はデータ効率性向上に向けた開発と検証を進める。そして 2022 年度末までに AI エッジ統合制御システムを活用したドローン運用案件等の実ユースケースを通じてシステムレベルでの消費電力従来比 1/10 を目指すことを事業化のマイルストーンとする。

また自律型エージェントのサービス展開として、既に深層強化学習及び模倣学習を用いて事業展開を進める建機自動化支援サービスやドローン点検自動化サービス等において、今後本研究にて確立した高度自律的学習機能を適用することで事業化を進める。

	2018年度	2019年度	2020年度	2021年度	2022年度
高度自律的学習機能・人工意識の開発		アーキテクチャ・アルゴリズム確立▼	シミュレーション評価 計算量従来比1/10▼	追加学習及び推論 データ効率の実現▼	システムレベル消費電力 従来比1/10▼
	人工意識アーキテクチャの開発・有用性検証			データ効率性向上に向けた開発・有用性検証	
ベースモデル開発		シミュレータ開発▼	性能評価▼	性能評価▼	性能評価▼
		「WorldModel」ベースアルゴリズム開発		エッジ端末での追加学習・推論効率化に向けた拡張開発・評価	エッジ・クラウド連携、実機評価に向けた拡張開発・評価
			性能評価▼ シミュレータ開発▼		
			「Predictive Coding」ベースアルゴリズム開発		
自律型エージェントのサービス展開	従来アルゴリズムベースのサービス展開		深層強化学習・模倣学習ベースのサービス展開		高度自律化学習機能を用いたサービス展開

図 4.6.1.6 事業化マイルストーン（人工意識）

4.6.2 「AI エッジ統合制御システムの開発」における実用化・事業化の見通し（KDDI 株式会社）

4.6.2.1 概要

「スマートドローンプラットフォーム」は 4G LTE/5G ネットワークに接続するドローン機体、運行管理システム、ソリューション毎のアプリケーションで構成され、機体とアプリケーションの組み合わせによって、構造物点検、広域監視、測量などの様々なサービスを提供できるドローンプラットフォームである。

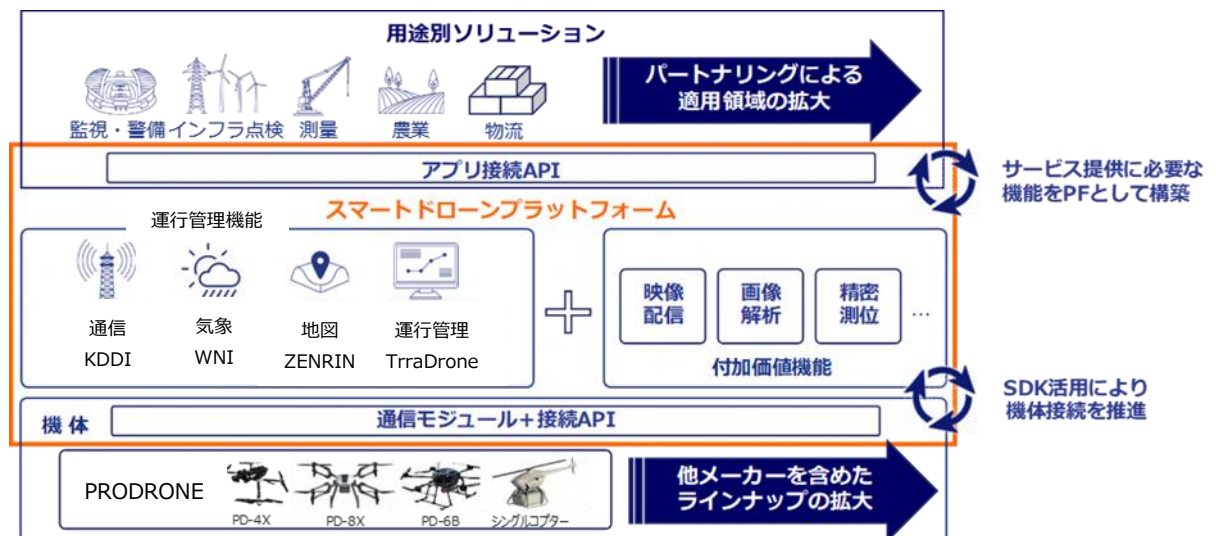


図 4.6.2.1 KDDI スマートドローンプラットフォーム概念図

スマートドローンプラットフォームの付加価値機能として、画像処理技術を中心とした AI 機能を開発中であり、将来的には 5G ネットワークを前提としたエッジ・クラウド連携環境もとで、大規模なデータ収集・解析基盤として機能を拡張、確立する予定である。

4.6.2.2 実用化・事業化への課題と対応策

下図に示す通りドローンは目視外遠隔自動飛行による市場拡大に移行していき、長距離飛行に強みを持つセルラードローン領域が本格化してくる予測であるが、ドローンと地上の無線通信の信頼性およびインフラ投資の観点から構築済みの 4G 携帯回線を活用した運用、

また、これから本格普及する 5G ネットワークと共に爆発的な市場拡大を目指し 3GPP 等の標準化団体を通じて日本のみならず各国で研究が加速している状況である。

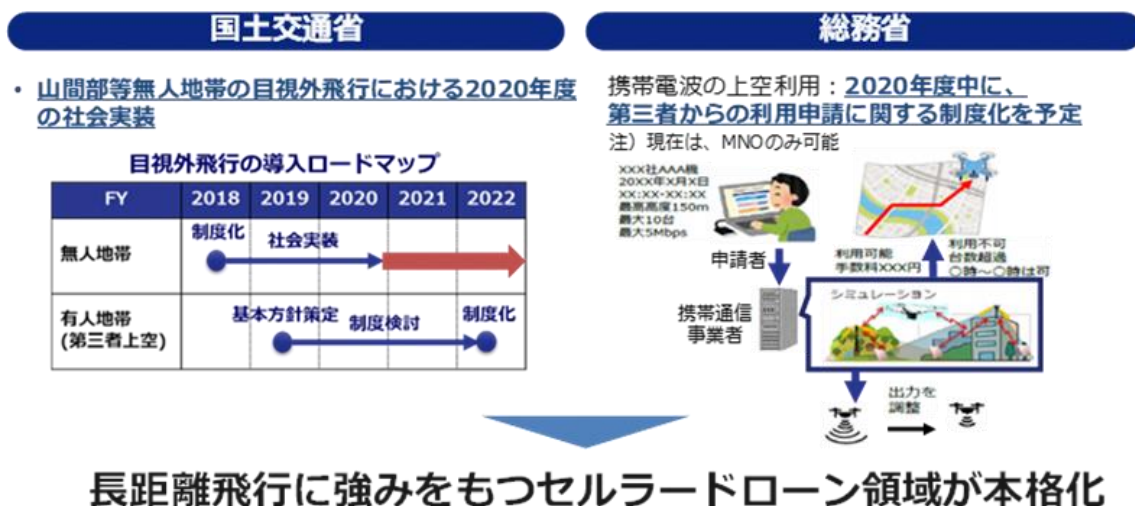


図 4.6.2.2 国内制度の動向

4.6.2.3 実用化・事業化の体制

エッジ(ドローン)での高度な自律処理と 5G との親和性確保のもと、AI エンジンを含め各アプリケーションが接続可能な運航管理システムを構築することで、様々なドローン適用領域への活用を目指している。

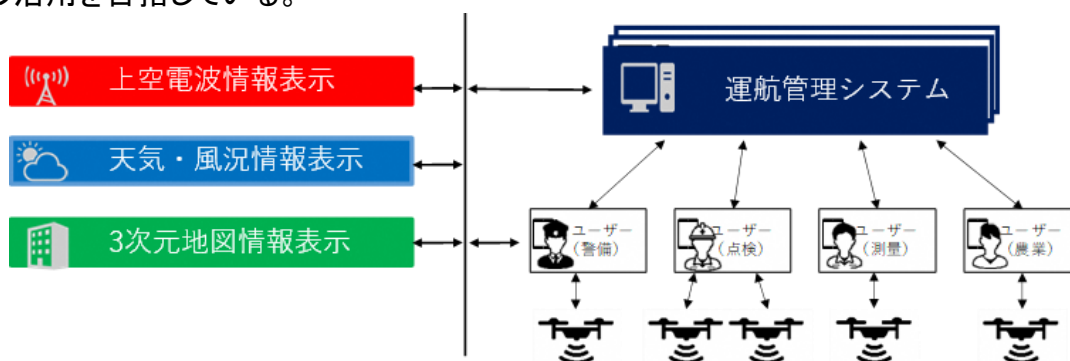


図 4.6.2.3.1 運航管理システムの開発

本研究のドローン搭載イメージを以下に示す。ドローンは大きく制御基板として飛行制御をつかさどるフライトコントローラーとカメラやセンサー制御およびフライトコントローラーや通信機能と連携するコンピューティング機能(下図中 SBC)からなる。

開発成果をコンピューティング機能上に搭載することで飛行制御の自律制御の高度化や点検等のミッションの高度化が実現可能である。

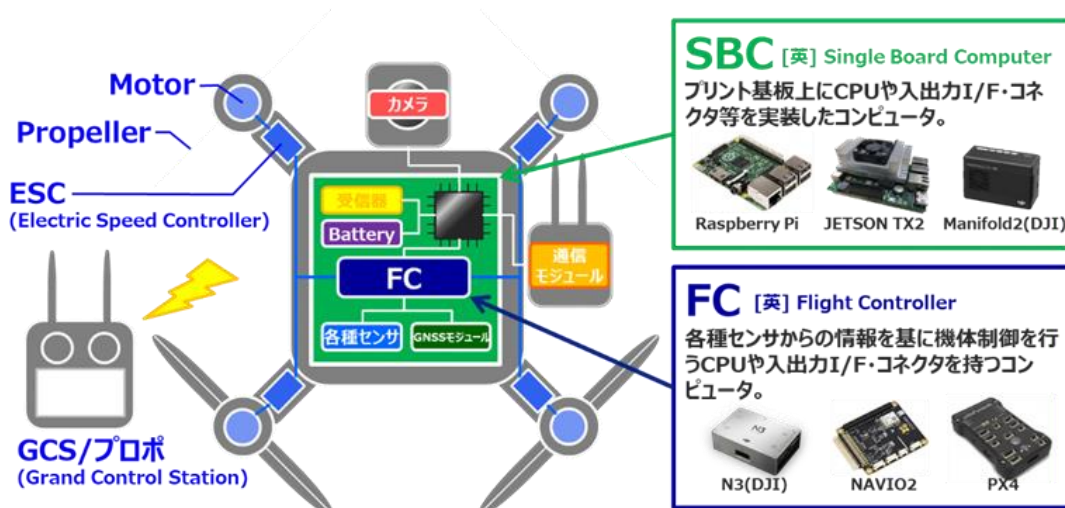


図 4.6.2.3.2 ドローン内部構成イメージ

また、エッジクラウド統合制御システムの構成イメージを以下に示す。飛行制御や通信制御とともに前述のコンピューティング機能部にエッジ AI 機能を搭載し SDK (Software Development Kit) として提供することでハードウェアや組み込み開発が最小限でクラウドとコネクテッドな環境を利用することが可能となる。本開発成果および 2021 年度以降の人工意識の開発を推進することでクラウドとエッジが5G 通信を介して学習データを共有することでエッジ端末がより高度化していくという「集合知」としての環境構築が見込まれる。

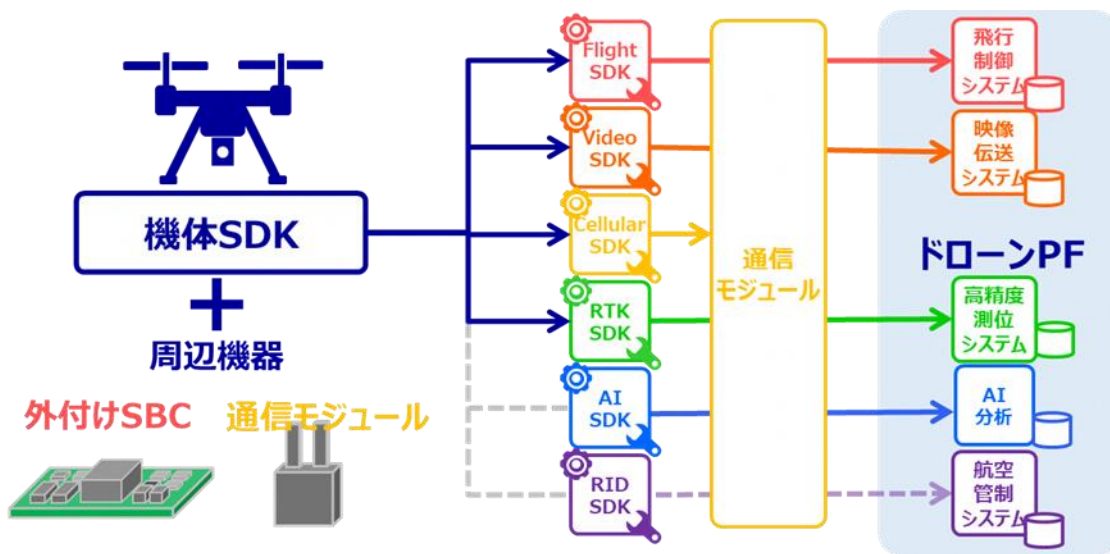


図 4.6.2.3.3 ドローンにおけるエッジクラウド環境 構成イメージ

4.6.2.4 市場規模と経済効果

2020 年度以降、規制緩和などが進み、目視外市場（長距離飛行）が大きく伸びる。インフラ点検・測量の需要増を中心に、各産業へのドローン利用が広がることが見込まれている。

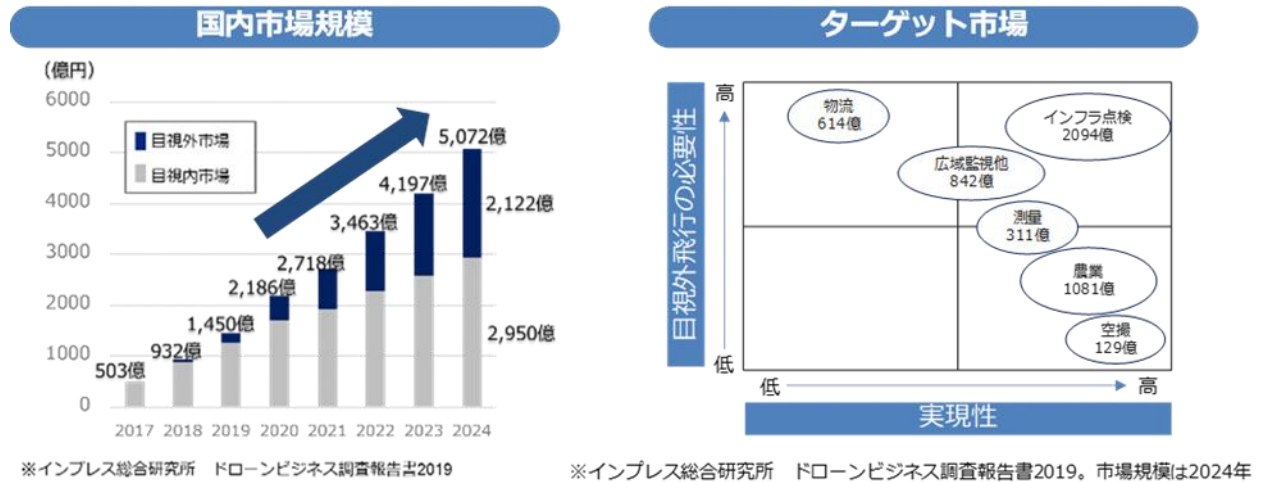


図 4.6.2.4 国内市場の動向

4.6.2.5 ベンチマーク

セルラードローン領域においては、MNO が中心に実証実験等が推進されているところである。特に KDDI および NTT グループが積極的にセルラーを用いたプラットフォーム展開を推進している。今後様々な業種や AI などの分析サービスと連携していくことが予想される。海外の大手 IT 企業も多額の投資をして AI 開発及びデータ分析を実施しているがドローンの対象領域は大手企業のインフラなど通常一般に流出しないデータであり早期に日本として開発基盤を構築することが必要である。また本テーマで提唱しているように今後ドローンのように IOT 機器は通信を介して学習データをクラウドシステムと相互共有する時代になると予測される。本テーマで実証計画予定の 5G 及びクラウド・AI エッジ連携の統合環境は、国内初の総合的なプラットフォーム提供形態となる。

	KDDI	NTT グループ	ソフトバンク	楽天
サービス	スマートドローン インフラ点検、監視、測量等（個ソリ） （2019.6月より開始）	docomo sky ジャパン・インフラ・ウェイマーク 監視、測量、インフラ点検等 （2018.3月より開始） インフラ点検に特化して事業展開 （NTT西100%出資、2019年4月に設立）	SoraSolution 国内ベンチャー（センシロロボティクス）のSL活用し、インフラ点検中心に提供 （2019.11月より開始）	そら楽 配送に特化してサービス提供 （2016.4月より開始）
プラットフォーム	スマートドローンPF →自社構築済 2020年：有人航空管制との接続予定	Air Palette（NTTデータ） →東電PG、日立とドローンの航路を構築する事業組合「グリッドスカイウェイ」を設立	Sora Solution内で提供 （国内通航管理の議論には未参画）	楽天AirMap UTM （北米AirMapとの合併会社）
機体	PRODRONEの他、ラインナップを拡充 （Coretronic(台湾)、DJI、SkyDrive)	DJI、ACSL、Studio(北米) （ドコモより、ドローン搭載用モジュールを提供）	センシロロボティクス （2020年：セルラー対応予定）	ACSL、JD.com(中国) ※モバイル回線は、DCM/KDDIを利用

図 4.6.2.5 ベンチマーク状況

4.6.2.6 事業化までのマイルストーン

- ・STEP1として、個別ソリューションにて2019年6月に役務提供を開始しており、
- ・STEP2として、2020年2Qには「KDDIスマートドローン」お客様運用メニューとしてシステムの本格提供を開始している。
- ・STEP3として、AI分析や飛行制御の高度化を予定しており、本テーマの技術については機体の飛行安全性の観点から前述の官民協議会ロードマップにも示されている通り機体の自律化として2022年度以降の有人地帯の第3者上空飛行実現に必須機能である。

市場は携帯電波の上空利用や運航管理システムなどの社会インフラの整備および航空法等の関連法制度改定に伴い飛行制御関連サービス(現場での仕事の効率化:外業)の高度化が始まり、一方インフラ点検や警備業などでの分析サービス(オフィス内作業の効率化:内業)もAIの発展とともに高度化が予測される。本開発テーマは双方に寄与するテーマであり、また本技術はIOT機器やモビリティ全般に広く普及していく技術にしていく予定である。



図 4.6.2.6 事業化マイルストーン

4.7 「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」における実用化・事業化の見通し

4.7.1 概要

本 PJ では、研究開発で構築したコンパイラ基盤を使用してエッジビジョン AI の開発や評価が完結して行うことができる Web サービス「GENESIS DevEnv」を構築し、1. エッジビジョン AI の応用プロダクトの開発者、2. エッジビジョン AI を構成するソフトウェアの開発者、3. エッジビジョン AI が動作するハードウェアのサプライヤー、の 3 者を媒介するクラウド上のプラットフォームを目指す。

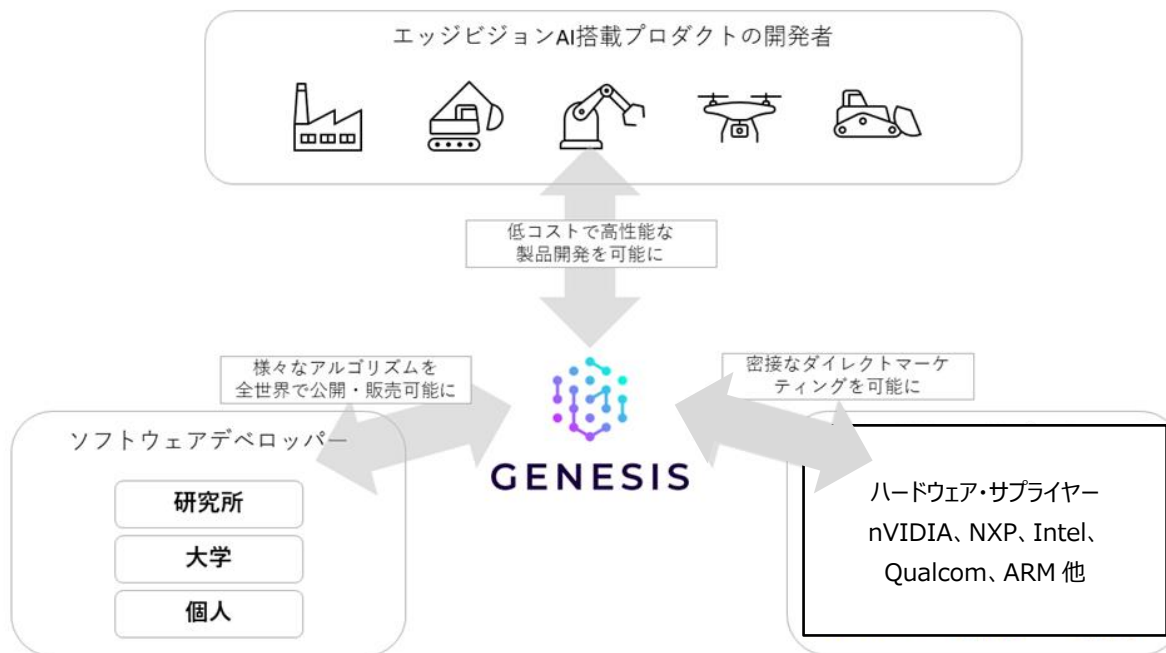


図 4.7.1.1 Web サービス「GENESIS DevEnv」のステークホルダーマップ

来たる Society5.0 社会においては、実社会で生み出される大量かつ多様なデータをリアルタイムかつ自動的に収集・分析・処理することで、これまでとは質的・量的にも異なる次元の価値を生み出すことが期待されている。このような高度な情報化社会を実現するためには、それを支える情報システムにもまた高度な要件が求められる。このようなシステムを実現するにあたり、我々が重要と考える指標は以下の2点である。

1. システムの開発・維持・改善に必要な人的コスト
2. システムの運用に必要なエネルギー効率

Society5.0 社会を支える IoT や人工知能が統合されたシステムは、実社会と密接な関係のもとで実装・運用されなければ価値を持たない。すなわち、実社会の複雑さ・多様さを反映した、多種多様な要件のシステムが必要となることが予想される。このようなシステムを実装するためには、高度なプログラミング技術を修めたエンジニアリング人材が大量に必要である。少子化と高度教育に課題を抱える我が国においては、限られた人的リソースを効率よく運用する手段を考えねばならない。本プロジェクトの研究開発テーマであるドメイン固有のフルスタックコンパイラ基盤は、今まで人手で行っていたエンジニアリング作業を様々なレイヤで自動化することで、システムの開発と維持運用にかかるコストを大幅に削減する。

エネルギー効率、すなわち単位エネルギーあたりのシステムの処理性能は、あらゆる種類、あらゆる規模のシステムで常に問題となる。極論してしまえば、エネルギー効率は高ければ高いほどよい。エネルギー効率の最大化において最も難しいのは、局所的な効率化の組み合わせでは解決できず、系全体での最適化が必要な点にある。つまり、大規模化するシステムでは必須となるモジュール化とは本質的に相性が悪い。しかしながら、本PJのコンパイラ基盤は、適用される問題のドメインをある程度絞ることで複数のレイヤにまたがって大域的な最適化を施すことを可能としており、モジュール化とエネルギー効率を高いレベルで両立することができる。加えて、固定化されたハードウェアで動作するソフトウェアの最適化のみならず、アルゴリズムに特化したハードウェアデザインを自動的に生成できるため、ソフトウェア・ハードウェアの垣根をこえ、システム全体としてエネルギー効率を最大化する構成を自動的に探索・生成することすら可能である。

このように、本PJの研究開発テーマであるコンパイラ基盤は、人的コストとエネルギー効率の両面から、Society5.0社会におけるエッジビジョンAIのシステム開発のあり方を破壊的に進化させる可能性を秘めている。本項では、この研究成果をいかにパッケージングし、持続可能な事業として実社会に届けていくかという点について述べたい。

冒頭でも述べたように、本PJでは、事業モデルとしてSaaSモデルを採用する。短期的には、コンパイラ基盤を開発ツールとして単体で外販する、いわゆるツールビジネスを行うことも可能だが、これは行わない。古典的な開発ツールベンダのように高度な技能を必要とする開発者を対象ユーザーとしてしまうと、市場規模が一定数で飽和してしまい、ツール単価を上げなければ事業が成り立たず、そうするとユーザー数を増やすことが難しいというジレンマに陥ることを避ける、というのが一点目の理由、さらに、ツールを個別に配布してしまうとサポートコストが無視できないほど大きく、事業を小さく始めるということが難しいというのが二点目、さらに、ツールビジネスでは顧客要件に合わせたカスタマイズ対応が必要となることが多く、自社のIPとして価値を集積しにくいという点が三点目である。

本PJでは、研究開発で構築したコンパイラ基盤を基盤技術として使用し、エッジビジョンAIを1.いつでも 2.誰でも 3.どこからでも、容易に実現できる開発環境をWebサービスとして構築し、インターネットを通じて全世界に提供する。



図 4.7.1.2 GENESIS DevEnv のサービス構成図

4.7.2 実用化・事業化への課題と対応策

このモデルには、以下のような利点がある。

1. サブスクリプションビジネスを展開しやすい
2. 潜在的なユーザー数を最大化できる
3. 業界をリードするプラットフォームへと成長する高い可能性がある

サブスクリプションビジネスの本質は、ユーザーとの継続的な関係性を保てることにある。SaaS モデルは、事業者が構築したクラウド環境上にユーザーがアクセスし、その上で様々な活動を行い、結果としてユーザーが利得を得るものである。つまり、ユーザーがどのような活動を行っているか、どのような課題を抱えているか、それをどのように改善したらよいか、という指標を、サービスの運営と並行して得ることができる。サービスが荒削りな段階から、ベータバージョンとしてサービスを公開することで、サービスの価値を徐々に高め、時には方向転換をし、本質的なユーザーの課題を解決するよう、研究開発に対して有効なフィードバックを行う。これは、貴重な国家予算を有効に活用し、真に価値のある研究を推進するとともに、研究結果をサービス価値へと素早く昇華させ、事業の持続性を高めることにもつながる。

次に、Web サービスの特性上、インターネットがつながる場所からブラウザが動作するデバイスさえあればサービスを利用することが可能となるため、潜在的なユーザーは全世界に広がる。市場を独占するプラットフォーマーを目指すならば、サービスへと流入可能な絶対ユーザー数が大変重要であり、そういった観点からインターネット上で Web サービスを提供することは最良の手段と言える。

この構造は、ユーザー数が十分でない立ち上げ期においても、以下のようなシナジー効果により本サービスに関連した売上が見込めることで、事業の持続可能性を高める。本 PJ の主たる事業者である株式会社フィクスターズでは、本サービスが解決しようとしている問題を、高速化・最適化技術を持つソフトウェアエンジニア集団による受託サービスとして解決してきた。自動運転やロボット、ドローンのような次世代の自動モビリティの頭脳、高度化する検査装置機器や医療機器の処理エンジンなど、画像に高度な適用するといった需要は年々向上しており、受託サービスの引き合いは増える一方だが、高度なエンジニアリング技術を密なコミュニケーションの下で提供するという受託サービスの特性上、営業チャネルは事業所の近くに閉じてしまう。すなわち、現時点では、我々の到達可能な市場は日本および北米の拠点周辺に限られてしまっている。しかしながら、本サービスが全世界に対して公開されれば、開発初期のプロトタイピングや性能評価タスクは本サービスの上で完結する。その上で、さらなる最適化や本開発といった巨大案件のリードを取るための営業ツールとして本サービスを利用し、サービス自体のユーザー数拡大と、受託事業とのシナジー効果の双方を得ることができる。

最後に、プラットフォームへと成長する可能性と、その場合の社会的インパクトについて述べる。本サービスは、冒頭でも述べたように、以下の 3 者を媒介する。

1. エッジビジョン AI の応用プロダクトの開発者
2. エッジビジョン AI を構成するソフトウェアのデベロッパー
3. エッジビジョン AI が動作するハードウェアのサプライヤー

前提として、本サービスの主たるユーザーは 1. である。しかしながら、本サービスのコンパイル基盤が、ドメインのアルゴリズムを入力として、特定のチップ上で高速に動作するバイナリもしくはハードウェアデザインそのものを出力するという特性上、2. のソフトウェアデベロッ

パー、および 3.のハードウェアサプライヤーとも密接な関係がある。これらの各ユーザーに対してどのような価値を提供し、本サービスをプラットフォームとして成立させていくかについて述べる。

1.のユーザーに対しては、エッジビジョン AI を応用する複雑なシステムに対して、高い開発効率という価値を提供し、サブスクリプション課金によって対価を得る。これは今まで述べてきた事業モデルから素直な延長上にある。次世代の検査機器や医療機器など、スマートなマシンビジョン、自動運転車、ドローンやロボット等の自動モビリティ分野のユーザーを想定しており、前者は具体的には、ニコン、オリンパス、キヤノン、オムロン等の大手メーカー、後者については、トヨタ、デンソー、ホンダ、スバル、ソフトバンクロボティクス、WHILL、DJI 等である。ここで挙げたメーカーは、全て事業者である株式会社フィクスターズの受託サービスにおける既存顧客、あるいはそれに準ずる近い関係性にあるため、受託サービスの事業部と連携してサービスの普及と利用促進に務める。

一方、2.のユーザーに対しては、本サービス上で利用可能なアルゴリズムを集積するマーケットプレイスを用意し、1.のユーザーのサービス利用料のレベニューシェアを行う。これにより、本サービス上では様々なアルゴリズムが利用可能となり、1.のユーザーの新規流入の増加およびリテンションの維持が達成できる。対象としては、アルゴリズムの研究開発に従事する大学研究者、個人開発者、企業体である。具体的には、現在、日本国内のいくつかの大学と共同研究を行っており、本 PJ のコンパイラ基盤が入力とするドメイン固有言語を用いて、様々なアプリケーション実装し評価するといった研究を行っている。また、Visual SLAM 技術の開発に強みを持つ Kudan 株式会社と事業提携を行っており、Kudan のもつ高度な SLAM アルゴリズムをサービス上で展開し、レベニューシェアによる事業化を進める予定である。

最後に、3.のユーザーに対しては、ハードウェアの販売およびマーケティングチャネルを提供する。マイクロプロセッサのチップベンダーおよび、センサーベンダーがその対象となる。前者は、Intel、NVIDIA、ARM、Qualcomm、などの大手チップベンダから、テキサス・インスツルメンツ、CEVA、Cadence 等の DSP に強みをもつ SoC メーカーおよび IP ベンダ、国内では東芝やデンソーなどの独自の ASIC を製造している企業を想定している。後者は、イメージセンサ最大手のソニーセミコンダクタソリューションズをはじめ、ToF や LiDAR 等の次世代センサのメーカーを幅広く対象とする。1.のユーザーがアルゴリズムの評価結果をもとに専用チップ化を行うといったシナリオを仮定すると、サービス上で構築したハードウェアデザインを流用し、ハードウェアサプライヤーに対して開発を依頼することで、短 TaT での専用チップの製造が可能となる。さらに、ハードウェアサプライヤーにとっては、全世界に広がるプロダクトの応用ユーザーが、どのようなアルゴリズムを、どのようなハードウェアに対して評価しているかといった統計情報を提供することが可能となる。ハードウェアサプライヤーは、この情報をもとに、次期製品開発の方向性、例えばどのような AI 処理を専用ハードウェアとして実装すべきか、メモリバンドと演算リソースのバランスはどれくらいがよいか、シリコンをレジスタとロジックのどちらにどれくらいの比率で割り当てたらよいか、といった、製品特性を決定する重要なパラメータを決定することができる。ダイレクトセールスを重視するハードウェアサプライヤーにとって、これらの情報は製品の市場競争力を高める上で欠かせない情報であり、多くのマーケティング費用を割いて顧客との関係構築および情報取得に努めていることから、十分に市場はあると予想している。

結びに本節の内容をまとめる。本 PJ の研究結果を SaaS としてパッケージングして提供することで、プロダクト・ソフトウェア・ハードウェアの 3 者を媒介するハブとなり、Society 5.0 社会における重点課題であるエッジビジョン AI を応用したシステムの開発を牽引するクラウド上のプラットフォームを目指す。

4.7.3 実用化・事業化の体制

(公開しない)

4.7.4 市場規模と経済効果

(公開しない)

4.7.5 ベンチマーク

(公開しない)

4.7.6 事業化までのマイルストーン

(公開しない)

4.8 「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」における実用化・事業化の見通し（イーソル株式会社）

4.8.1 概要

（公開しない）

4.8.2 実用化・事業化への課題と対応策

（公開しない）

4.8.3 実用化・事業化の体制

（公開しない）

4.8.4 市場規模と経済効果

（公開しない）

4.8.5 ベンチマーク

競合機関との比較では、A 社の組込み OS 技術、B 社のリアルタイム OS 技術があげられる。A 社、B 社ともに、本プロジェクトでの成果目標では、実用化時点ならびに成果普及段階においても、性能／電力効率、リアルタイム性、品質・信頼性、機能等で十分に上回ることができる。OS 以外の実施項目を含め、以下の優位性を持つと考えている。

1. OS: キャッシュコヒーレンス機構を持たない 8 コアを超えるヘテロジニアス・マルチ・メニーコアに対応し、高いリアルタイム性を持つマルチカーネル技術にサービス指向アーキテクチャ(SOA) を統合した世界初の OS - 現時点で Linux に対して 35 倍から 100 倍の性能。
2. ソフトウェアマッピング: 従来のデータ並列に加えて制御アルゴリズムにて多用されるタスク並列を含めたソフトウェアマッピングに対応した世界トップクラスのソフトウェアマッピング技術。ヘテロジニアスアーキテクチャ向けモデルレベルのソフトウェアマッピングは世界初。従来(ヘテロ向け分割と境界部分生成は手動)と比較し、単体で 100 倍を超える生産性向上、全体で 10 倍以上を目指す。
3. ハードウェア記述標準モデル: ソフトウェア視点でヘテロジニアス・マルチ・メニーコアを表現できるハードウェア記述仕様の IEEE における国際標準化による成果普及。

表 4.8.5 分散マイクロカーネル OS と Linux のマルチコア性能比較

API	分散マイクロカーネル OS	Linux
pthread_cond_signal()->pthread_cond_wait()	0.7us	82.8us
pthread_mutex_unlock()->pthread_mutex_lock()	1.5us	54.0us
sem_post()->sem_wait()	1.5us	69.3us
pthread_barrier_wait() ->pthread_barrier_wait()	0.7us	73.8us

4.8.6 事業化までのマイルストーン

（公開しない）

4.9 「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」における実用化・事業化の見通し

4.9.1 多分岐結合型推論プロセッサの研究開発における実用化・事業化の見通し(株式会社エヌエスアイテクス)

4.9.1.1 概要

本研究開発により、DILP の IP 技術と評価結果、これを使うためのコンパイラを含む開発環境が成果として得られる。この成果を利用して、DILP-IP を製品化し、ASSP を販売する半導体製造会社または ASIC を開発する装置メーカーへ販売し、完成した半導体製品が各種装置製造メーカーで利用される。DILP-IP 完成後の 2022 年度(プロジェクト最終年)から自動車車載向け事業化、非車載向け事業化を本格化し、半導体会社とライセンス契約を取り付ける目標である。

4.9.1.2 実用化・事業化への課題と対応策

実用化、事業化のキーポイントとなる世界最高レベルの数値目標としている電力効率(15 TOPS/W)、面積効率(1 mm²/TOPS)の実現が大きな課題である。2020 年度以降、段階的に評価を行い研究内容にフィードバックしていく。

4.9.1.3 実用化・事業化の体制

本研究開発物となる DILP の基本クラスタ部を切り出して、エヌエスアイテクス独自事業として半導体ベンダへの販売、拡販を先行して開始する計画であり、ツールベンダや、ミドルウェア/ソフトウェア開発会社と広く協力関係を築き、実用化・事業化の準備を整える。本研究開発で得られる実証検討結果を用いて、アドバイザー・ボードとも協力し、評価から採用に向けた円滑な活動を行う。特に自動車領域では、大手自動車会社と密接な関係を築きながら、早期の事業化対応を進めていく。

4.9.1.4 市場規模と経済効果

研究開発成果の直接的な製品となるプロセッサ IP 市場(車載、非車載組み込み[スマートフォンは除く])規模として、PJ 完了の 2022 年度で国内 5,800M¥、海外 23,400M¥、2026 年度で国内 7,400M¥、海外 29,600M¥を見込む。2026 年度ではターゲットシェア 15%を目指す。

4.9.1.5 ベンチマーク

本研究開発で目標としている面積効率:1mm²/TOPS は世界最高レベルの数値であり、組み込み制御などコスト競争の厳しい分野でも競争力が高く、電力効率:15TOPS/W も世界最高レベルの数値であり、電力クリティカルな産業分野でも競争力が高く、実用化の期待が大きい。

4.9.1.6 事業化までのマイルストーン

プロジェクト最終年度である 2022 年度より、車載用途では車載半導体ベンダへ拡販活動を開始し、デファクト化を進めて行く。制御系組み込み用途では ASIC 開発メーカーをリードに、組み込み LSI メーカーの ASSP を通してロングテールに展開して普及促進する。

4.9.2 OSの抽象化とミドルウェア最適化技術ならびに統合開発環境の研究開発における実用化・事業化の見通し(株式会社ユーリカ)

4.9.2.1 概要

ハイパーバイザ、ミドルウェアを含む開発環境を自動車、FA、コンシューマなどの幅広いマーケットに提供する。

- ①: ハイパーバイザ、ミドルウェアを含む開発環境をオープンソース化すると同時にデストリビューション(パッケージ化)およびサポートビジネスを実施する。
- ②: 特定顧客向けパッケージの提供

4.9.2.2 実用化・事業化への課題と対応策

課題①: 既存ソフトウェアプラットフォームベンダーとの競合

対応策: ヘテロジニアスマルチコア対応、および仕様策定時から複数のターゲット顧客のアドバイスをもらい、仕様に盛込むことにより、優位性を獲得する。

課題②: メイン顧客の獲得

対応策: ベータ版やリリース候補版を顧客に試用してもらい、フィードバックを反映することにより、顧客獲得を目指す。

課題③: 知名度の向上

対応策: オープンソース化とコミュニティ活動により、新規ユーザの開拓を行うと共に既存ユーザを活性化する。

4.9.2.3 実用化・事業化の体制

内部においては、本製品用のチームの設立、コミュニティ事務局の設立等を実施する。外部とのリレーションについては、本プロジェクトメンバーとの連携を密にし、潜在顧客、ポテンシャルカスタマの開拓をプロジェクト最終年度より実施する。

4.9.2.4 市場規模と経済効果

本プロジェクト成果であるソフトウェアプラットフォームや組込み開発のためのツールの市場規模は、本プロジェクト終了時の予測で、国内市場 約957億円、世界市場 約1.4兆円である。また、国内市場に対しては、0.2%程度の直接的経済効果が発生すると考える。

4.9.2.5 ベンチマーク

(公開しない)

4.9.2.6 事業化までのマイルストーン

(公開しない)

4.9.3 リアルタイム空間理解、判断アルゴリズムの研究開発における 実用化・事業化の見通し（株式会社日立製作所）

4.9.3.1 概要

本委託事業にて開発のリアルタイム認識、判断アルゴリズム、および、その高速ハードウェア実装技術を適用したロボット、無人搬送車（AGV）、検査装置を、既に自動化ニーズの高い物流、生産分野でまず実用化し、さらに、今後自動化ニーズの急速な高まりが予想される小売り分野、三品産業（食品、医薬品、化粧品）へ展開することを狙う。

4.9.3.2 実用化・事業化への課題と対応策

課題①： ハードウェア（DILP や FPGA）の長期間調達、品証体制の確立

対応策： エヌエスアイテクス社や半導体ベンダと共同で体制を構築

課題②： 本開発技術搭載のロボットを活用した、SI 事業の立ち上げ

対応策： 事業部と研究開発グループの連携により、事業化を推進

4.9.3.3 実用化・事業化の体制

・弊社産業ビジネスユニットと研究開発グループの連携体制にて推進

・ハードウェア、ミドルウェアに関して、エヌエスアイテクス社、ユーリカ社と連携

4.9.3.4 市場規模と経済効果

物流、生産、小売り分野の合計の自動化市場は、世界全体で、2030 年時点で 40 兆円弱と予測されており、十分な市場規模であると考ええる。

4.9.3.5 ベンチマーク

本委託事業の目標は、従来比 10 倍速のロボット認識、判断 AI 技術の実証であり、これにより上記市場の獲得を狙う。

4.9.3.6 事業化までのマイルストーン

技術構築完了（本委託事業終了時 2023/3）

ハードウェア調達、品証体制の構築（2024/3）

事業化向け開発推進（2023/4～）

4. 10完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発における実用化・事業化の見通し(株式会社アクセル、株式会社ティアフォー社)

4. 10. 1概要

本研究開発は、完全自動運転に向けたシステムオンチップ (SoC) とソフトウェアプラットフォームの研究開発により、AI エッジコンピューティングにおける限られた計算資源でのリアルタイム性の保証と従来比 10 倍以上の高速なデータ処理を達成する。Proof of Concept (PoC)を提示しメニーコアやアクセラレータが混在するヘテロジニアスな SoC 上でリアルタイム性を保証したコンパイラ・OSとミドルウェア、自動運転アプリケーションの統合基盤を構築する。

研究開発した技術は、適切なシステムインテグレータ (Sier) やサプライヤ (Tier 1/2) を通したライセンス形態で販売するような事業展開を計画しており、自動運転車両展開の促進を見込んでいる。

4. 10. 2実用化・事業化への課題と対応策

現在、自動運転のシステムで使用されているハイエンドな組込みプラットフォームには一長一短があり、汎用性や消費電力の観点から完全自動運転に利用することは難しい。これに代わるプラットフォームを開発するためには、次の課題を解決する必要がある。

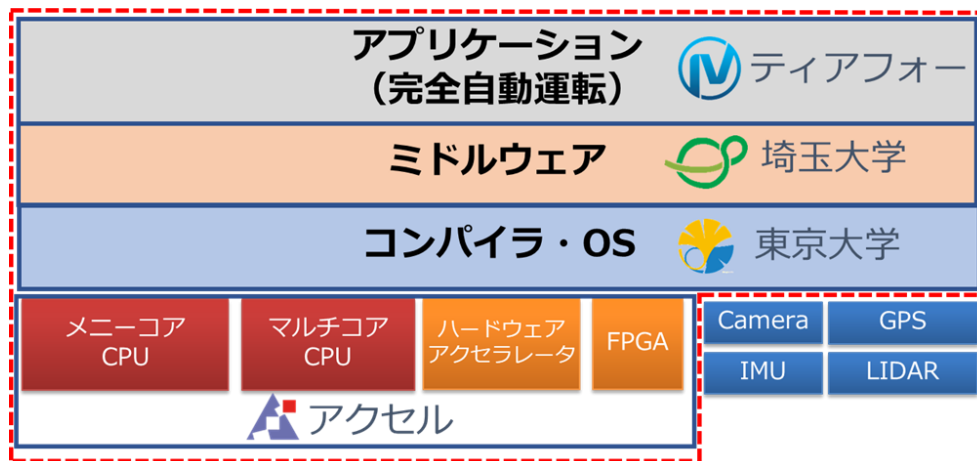
- ・高度な処理を低消費で実現
- ・計算量増大への対応
- ・アルゴリズム構築と最適化

本研究開発ではこれらに対応するため、システム全体に対して幅広い汎用性を求めるのではなく、マルチコアやメニーコアなど極めて汎用性の高いアーキテクチャを除いては、AI の各応用分野に特化した特定処理(あるいはソフトウェアでは遅い、電力効率が悪い処理)はハードウェアアクセラレータで実装するヘテロジニアス SoC (System on a Chip) アーキテクチャを創出し、そのためのコンパイラ、OS やミドルウェア、アプリケーションを含めたソフトウェアプラットフォームの構築を実施中である。

4. 10. 3実用化・事業化の体制

本研究開発の実施体制を図 4.10.3 に示す。汎用のメニーコアと特定のハードウェアアクセラレータを組み合わせた SoC アーキテクチャの研究は、LSI 開発に定評のあるアクセルが担当する。また、東京大学はコンパイラと OS を担当する。埼玉大学は、ヘテロジニアス環境に対応したミドルウェアの研究を進める。アプリケーション開発は、自動運転の公道実験を行っているティアフォーが担当する。

これらの研究成果を組み合わせることで、AI エッジコンピューティング技術を構築することが可能になり、さらに、ディープラーニングを用いた画像認識など特定ハードウェアアクセラレータを入れ替えることにより、完全自動運転の実用化に貢献することができる。本研究開発の実施担当企業のほかに、産業界への波及効果を得るために、民間企業の有志によるオープンイノベーションコンプレックス (OICX) を形成し、AI 製品サービス動向を把握するとともに、研究成果の事業化につながる企業群と定期的 (月 1 回程度) に意見交換していく。内閣府や関連省庁の意見も取り入れながら、我が国に利益をもたらす基盤技術を創出する。



本研究開発の対象

図 4.10.3 実施体制と研究開発項目

4.10.4 市場規模と経済効果

自動運転の市場規模(自動車本体価格を除く)のみを見ても、2025年までに5兆円、2035年までに8兆円に達し、1200万台以上の完全自動運転車が市場に出回ると予想されている(BCGの調査:自動運転車市場の将来予測)。自動運転によるエコドライブ(急発進・無駄なブレーキ等を減らす)により、消費エネルギー(約8%の削減:経済産業省調べ)や渋滞解消することも期待されている。

さらに、本研究の応用分野としては、ロボット、介護、農業、ドローン、製造、防災・減災などAIの様々な分野での活用も見込まれている。特にAI関連製品サービスは、PCやスマートフォンに匹敵する市場規模になることが期待できる。特に完全自動運転車は年間数百万から数千万台規模の出荷が見込まれる。世界に先駆けて完全自動運転に向けたSoCとコンパイラ・OS・ミドルウェア・アプリケーション技術を含むソフトウェアプラットフォームを実用化することは、社会的にも経済的にも、大きなインパクトがある。

4.10.5 ベンチマーク

既存自動運転ハードウェアとの比較:

本研究開発の目的は、完全自動運転に向けたシステムオンチップ(ハードウェア)のみならず、オペレーティングシステム(OS)・ミドルウェア・アプリケーション技術を創出することである。表 4.10.5 に示すように、現在、ハイエンドな組込みプラットフォームには一長一短があり、汎用性や消費電力の観点から完全自動運転に利用することは難しい。

	CPU性能	汎用性	位置推定	物標認識	行動判断	自立制御	消費電力
NVIDIA Drive PX等	△ ARM	△ 画像処理等	△ GPU	○ GPU	△ CPU/GPU	△ CPU	× 100W~
Intel Atom等	○ X86	○ PC一般	× CPU	× CPU	× CPU	△ CPU	△ ~20W
ルネサス R-Car等	△ ARM	× 車載限定	× CPU	× CPU	× CPU	△ CPU	○ ~5W
本提案	△ ARM	△ 自動運転 ソフトウェア制御	○ APU (※1)	○ APU (※1)	○ APU (※1)	○ メニーコア (※2)	△ 20W~

※1 使用頻度の高いワークロードを専用HWによるアクセラレータ(APU)で高速化・省電力化

※2 AIアプリケーションとの協調設計によりワークロードに最適化したヘテロメニーコアアーキテクチャ

表 4.10.5 性能/消費電力比較

本研究開発では、エッジコンピューティングの活用で自動運転システムを構築することで安全性を確保しつつ、低価格・消費電力/発熱・省スペース化実現の両立を目標にしている。既存の自動運転システムの概略を図 4.10.5 に示す。

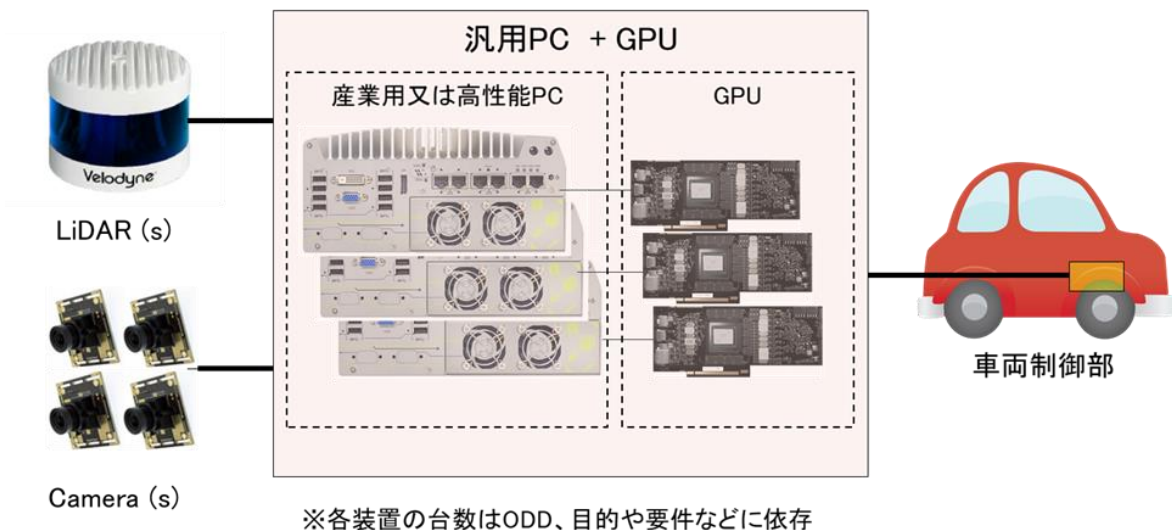


図 4.10.5 既存の一般的な自動運転システム

現在一般的な自動運転のシステムでは高性能 PC 又はそれに近い性質のハードウェアを使用している。特徴は次の通り。

- ・ 消費電力が大きく、高負荷時は 1000W 以上となる可能性もある

- ・ 自動運转向けハードウェアや GPU は総じて高価
- ・ PC 数台でシステムを構築することがあり、車両によってはペイロードの問題が生じる また熱対策などの考慮も必要

本研究ではこれらの問題を解決するために、Autoware の主要動作を行うホストボードを ARM-Cortex 等の低消費汎用 CPU で実行し、従来 GPU を活用していた計算量の多い並列演算処理を専用アクセラレータにオフロードする、エッジ装置を活用したアーキテクチャとしている。また、Autoware を従来システムと同様に安全性を確保したまま新しいシステムで動作させるには、ハードウェアだけではなくミドルウェア/OS/ドライバの対応が必要となっており、これまでに記述した通り 4 組織で連携して研究を進めている。

最終的には次の特徴を持った自動運転システム/IP コアの提供をめざしており、これらを解決出来ている競合は現在の所存在していない。

- ・高負荷処理であっても数 10~100W レベルの消費電力で動作する自動運転(Autoware)システム
- ・SoC 化による小型化達成、量産性向上
- ・ハードウェアレベルの柔軟性、機能拡張性、冗長性の確保

4.10.6 事業化までのマイルストーン(株式会社アクセル)

年度	2021 年度	2022 年度	2023 年度	2024 年度	2025 年度
量産化判断	◇プロトタイプ評価	◇費用・損益	◇サンプル評価	◇ユーザ評価	
プロトタイプ評価	▲プロトタイプ評価				
製品設計	▲フィードバック	▲設計完了			
開発費用投資		▲5 億円			
生産			▲サンプル出荷	▲ユーザ評価	▲量産体制
販売					▲5000 台/月
収益発生					▲1500 万円/月

※量産化判断が必要となるポイント

- プロトタイプ評価段階 : プロトタイプ評価による開発の妥当性判断
- 製品設計段階 : 設計費用投資の可否および損益予測による経営判断
- 生産段階 : サンプル評価による製品の妥当性判断
- 販売段階 : ユーザ評価による採用判断

アクセルにおける事業化戦略として、完全自動運转向けシステムオンチップとプラットフォーム全体の实用化を、2025 年度以降を目指したロングゴールとして想定するとともに、本研究開発プロジェクトにおける個々の成果において、それ単体での实用化・事業化(横展開も含む)を直近のショートゴールとして想定している。これにより、本研究開発プロジェクトの最終目標を目指しながら、直近の収益も達成可能な戦略的な事業化計

画を可能とする。ショートゴールによる短期収益化は、今後の事業化に向けたフィードバックを得るとともに、今後の研究開発費用の捻出、および一定のリスクヘッジとしても機能すると考えている。

ショートゴールの一例として、CNN アクセラレータの事業化が挙げられる。CNN アクセラレータの事業化におけるマイルストーンは、以下のとおりである。

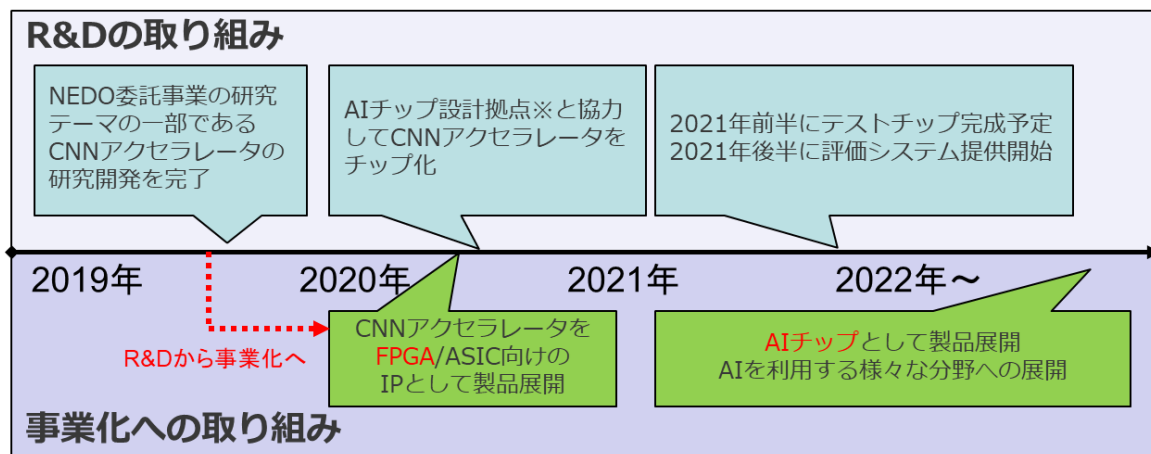


図 4.10.6 事業化におけるマイルストーン(株式会社アクセル)

4.10.7 事業化までのマイルストーン(株式会社ティアフォー)

2022年度の低速自動運転モビリティサービスへの自動運転システムの導入開始に向けて、ティアフォーとしては、チップ搭載車両関連サービス設計を進めるとともに、ソフトウェア分野の人材採用を進める。

自動運転車両を開発する企業に対して、自動運転システムなどを提供するとともに、モビリティサービスを提供するプレイヤーに対して配車管理システムや遠隔監視システムを提供することで、システム使用やライセンス供与に伴う収入を得ることを目指す(初年度(2022年度)は2000万円程度)。更に、2025年度からの中速自動運転サービス開始に向けて、新型 SoC を搭載した車両開発、サービス設計も進めていく。

年度	2021年度	2022年度	2023年度	2024年度	2025年度
サービス設計		▲試作チップ	▲車両向け新 SoC		
ソフト人材採用(費用)		▲5億円(費用)	▲10億円(費用)		
ソフト開発		▲試作チップ向け Autoware 開発	▲新 SoC 車両向け Autoware 開発		
ビジネス立ち上げ			▲低速モビリティサービス	▲新車両モビリティサービス	
収益発生(ライセンス料金/課金)			▲20百万/月		▲60百万/月

4.11 「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し

4.11.1 「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し (NSITEXE)

4.11.1.1 概要

半導体 IP および関連ツールの開発・ライセンス販売、保守メンテナンス、エンジニアリングサービス。

4.11.1.2 実用化・事業化への課題と対応策

今回のプロジェクトで得られた成果を製品化する場合、IP の品質保証の観点から、再度回路検証、ソフトウェアのテスト工数が必要となる。これに対して NSITEXE で保有している回路検証技術ならびにソフトウェアテスト手法を用いる事により、事業化の対応を実現できる。

4.11.1.3 実用化・事業化の体制

NSITEXE は主たる事業として半導体 IP および関連ツールの開発・ライセンス販売を実施しているため、NSITEXE が主体となり、既存製品と同様に今回のプロジェクトの成果を製品化し販売する。

4.11.1.4 市場規模と経済効果

2023 年でプロセッサ IP の国内市場は約 600 億円の規模となる予測で、このうち 20% がセキュアを対象とする領域になる(120 億円)。この市場のうち、車載向け製品 5%、産業用途向け製品 2.5%(車載の半分)獲得する事で、2023 年は 9 億円の売り上げを見込んでいる。その後、市場が成長する事により、2025 年では 31 億円、2027 年では 73 億円の売り上げを見込んでいる。

4.11.1.5 ベンチマーク

半導体 IP 事業における競合を以下にリスト化する。

- 1: SiFive (RISC-V の IP からハードウェア実装までを手掛ける新興企業)
- 2: RAMBUS (セキュリティの IP 会社を買収し、広く事業展開している)

4.11.1.6 事業化までのマイルストーン

事業化開始時期: 2023 年

売上開始時期: 2023 年

4.11.2 「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し(セコム)

4.11.2.1 概要

IoT 機器向け電子証明書を管理(電子証明書の発行や失効など)する電子認証局の運営。

4.11.2.2 実用化・事業化への課題と対応策

課題:

IoT 機器の開発、また、その機器の大規模なサプライチェーンにおいて、実際にどの段階でどのように電子証明書などを組み込むのかは、顧客の状況に依存することとなり、現時点では把握しきれない。

対応策:

実施項目④のセキュリティ社会実装 PoC において運用を試す。オープンコミュニティが開催されたときに実際のユーザから情報収集し分析する。

4.11.2.3 実用化・事業化の体制

本プロジェクトにおいて研究開発を実施しているのはセコム本社の IS 研究所である。一方、実用化・事業化するにはすでに電子認証局を運用しているグループ会社のセコムトラストシステムズ(株)となる想定である。

4.11.2.4 市場規模と経済効果

IoT 向け証明書の市場規模は不明。

IoT 向け電子証明書の発行累計枚数を

2023 年:	10 万枚(100 万円)
2028 年:	20 万枚(1200 万円)
2032 年:	360 万枚(3600 万円)

と想定している。

※()内は証明書単価 10 円で計算

4.11.2.5 ベンチマーク

国内における先行事例

- ・GMO グローバルサイン株式会社: マネージド PKI for IoT
- ・凸版印刷株式会社: Edge Trust
- ・デジサート・ジャパン合同会社: PKI Platform / IoT

4.11.2.6 事業化までのマイルストーン

「4.11.2.4 市場規模と経済効果」に記載したものと同様。

4.11.3 「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」における実用化・事業化の見通し(日立製作所)

4.11.3.1 概要

産業分野では生産工場での無人搬送ロボットの乗っ取り、なりすましを防止する物理セキュリティ担保を顧客価値とする SI 事業およびセキュリティ基盤サービスを想定している。

4.11.3.2 実用化・事業化への課題と対応策

事業化にあたっては、顧客課題の解決に向けたソリューション提案に直結すること示せること、事業拡大に向け特定分野の顧客以外にも適用できることが課題となる。このため、本 PJ では実際の顧客ニーズおよびフィールドを模擬した環境での PoC、複数の事業フィールドに対するユースケース抽出を行う。

4.11.3.3 実用化・事業化の体制

2020 年度は要素技術開発に注力するため、事業化検討を凍結している。2021 年度に実施予定であるセキュリティ社会実装 PoC の成果、オープンコミュニティ参画企業等との連携を通じ、事業化シナリオを検討する。

4.11.3.4 市場規模と経済効果

第 4 次産業革命を背景とした Industry 4.0, Society 5.0 などの政策を受け、IoT セキュリティの世界市場規模は 2023 年までに約 352 億ドルへ拡大すると予測されている。また、ISO21434 発行によりセキュリティ対策が必須となる自動車分野では、車載セキュリティ機器の市場規模が 2025 年までに 780 億ドル、自動車サイバー保険を含むテレマティクス保険市場も 710 億ドルへの成長が予測されている。本 PJ 成果のセキュアオープンアーキテクチャは上記市場成長を支える基盤技術であり、経済効果は大きいと考える。

4.11.3.5 ベンチマーク

産業分野への適用に当たっては OT (Operational Technology) を熟知した上でセキュリティアーキテクチャを構築する必要がある。セキュアプラットフォームを狙うインテル社や、エコシステム構築で先行するサイバートラスト社、RISC-V をベースにオープンセキュリティを進める Google 社などの IT 企業が競合であるが、本 PJ は参画企業で培ってきた OT ナレッジを有するという点でリードしている。

4.11.3.6 事業化までのマイルストーン

2021 年度に生産工場やモビリティシステム等の事業フィールドを想定した模擬環境での PoC 開発を進める。その後、本 PoC による試行を通じて適用可能性について検討する。並行してオープンコミュニティ参画企業等との連携を通じ、事業化シナリオを検討する。

4.12 高速研究開発テーマ「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」における実用化・事業化の見通し

4.12.1 「実施項目0：AI エッジで必要となるセキュリティ評価分析基盤の研究開発」における実用化の見通し（産業技術総合研究所）

4.12.1.1 概要

AI エッジデバイスなどにおいて未整備かつ今後重要となることが予想されるセキュリティ評価項目のセキュリティ評価分析技術を確立し、また、応用分野毎に異なるセキュリティ要求を、ユーザ、ベンダ、評価機関、認証機関等の複数のステークホルダーとも対話しながら、市場に受け入れられ易いセキュリティ要求仕様として策定する。主要な部分は国際標準に準拠もしくは反映させることで、市場に提供される製品・部品などのセキュリティを共通の物差しで評価分析できるようにし、また、調達者が提供するサービスや利用者の用途に応じて適切なレベルのセキュリティ機能を付加価値として提供可能とすることにより、2030年には世界で404.4兆円になると予想されているAI エッジ関連市場のセキュアで健全な発展に貢献する。

4.12.1.1 実用化・事業化への課題と対応策

エッジデバイス等のセキュリティ保護技術やセキュリティ評価分析技術は攻撃者に知られると悪用されるものも多い。そのため、外部発表や国際標準化などで広く公開すべき部分と、関係者のみに共有すべき部分との適切な切り分けが課題となる。この課題に対して海外ではセキュリティ評価分析技術の研究開発やそれらに基づき技術的に深い検討を行うための中核的研究開発拠点を国内に整備し、そこにおいて科学的な根拠とノウハウを蓄積し、関係者との連携を通して必要な情報の浸透と攻撃者側への流出防止の両立を図っている。本事業においても、これに倣い外部との連携を可能とするためのセキュリティ評価分析基盤を整備し、根拠のある最新のセキュリティ評価分析結果や評価手法の創出を可能とすると共に、その応用分野に関係する事業者との連携を通して必要な情報を共有して行くことで、攻撃者側へ重要なセキュリティ情報が漏洩することを防止しつつ、その分野のステークホルダーのみに必要な情報を浸透させる対応策を取る。

4.12.1.2 実用化・事業化の体制

前述のとおり、応用分野毎に求められるセキュリティ要求や市場において事業を展開する事業者が異なることから、攻撃者側に情報が伝わることを防止しつつ各応用分野のステークホルダーのみに適切な保護技術や評価分析技術が伝わる体制を組んでいる。具体的には、エッジデバイスが評価・認証済みまたは登録済みであることを確認するための個体管理用識別子の評価分析に対しては、産業界においてその標準的な評価手法が確立していないことから、学術的／科学的根拠に基づいたセキュリティ要求仕様、評価分析手法を整備すると共に、関連する企業群、横国大、産総研、経産省模倣品対策室などで「人工物メトリクススタスクフォース」を作り公開可能な部分の国際標準化を行うと共に、セキュリティ要求仕様、評価分析手法も関係者により受け入れられ易い方向に修正して行くことで対策や評価分析手法の円滑な普及を進めている。その他の分野においても、研究開発成果の状況に応じた体制の整備を進めている。

4.12.1.3 市場規模と経済効果

本テーマの最終的なアウトカムは、今後拡大が見込まれる AI エッジ関連市場が、セキュリティ上の脅威により阻害されることを防止することにある。電子情報技術産業協会 (JEITA) の報告書[1]によると、AI エッジの基盤となる CPS (Cyber Physical System) と IoT を合算した市場規模は以下の図 4.12.1.3.1 に示すとおり、2016 年に世界で 194.0 兆円、2030 年には世界で 404.4 兆円になると予想されている。仮に不適切なセキュリティアプローチにより市場が 10%縮小した場合、2030 年時点で 40 兆円以上の損失につながる。実際、DX(Digital transformation)が進む中で、セキュリティ上の問題が原因で、多大な損失が生じている事例も出始めている。そのため、AI エッジに対しても適切なセキュリティアプローチを取ることが求められている。

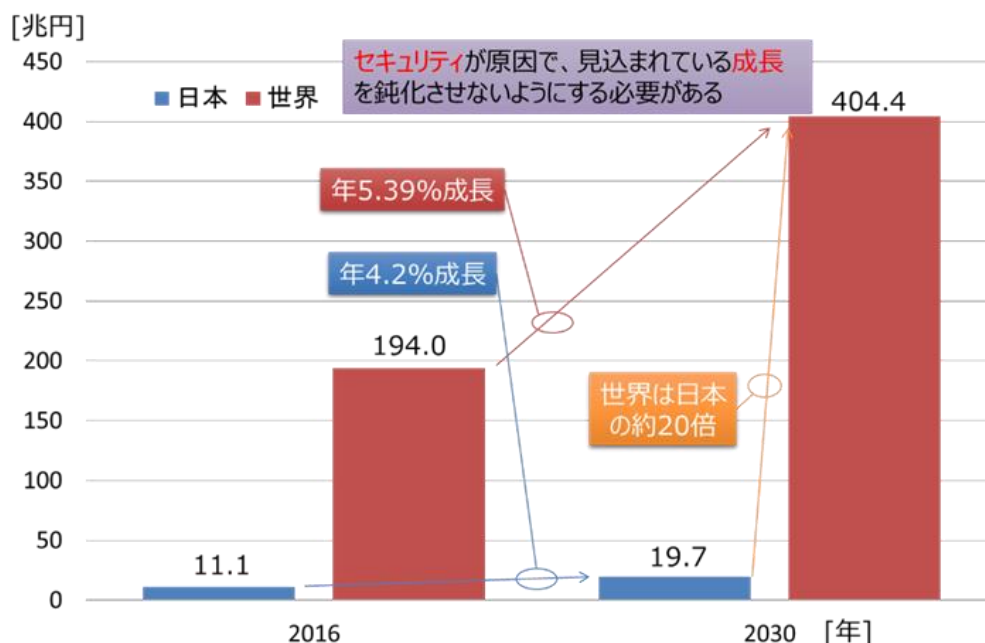


図 4.12.1.3.1 AI エッジ関連の市場規模[1]

また、以下の図 4.12.1.3.2 は、国内の模倣品被害規模推計の被害パターン[2]であり、AI エッジの導入が今後進むと予想される自動車部品・産業機械、電気・電子部品の被害額の割合が大きいことが示されている。

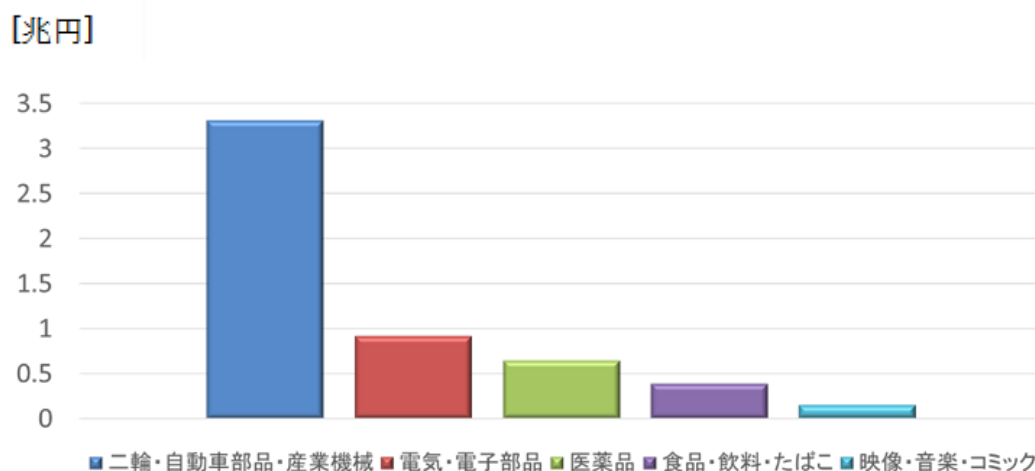


図 4.12.1.3.2 国内の模倣品被害規模(推計)

世界の模倣品等の取引規模は年間 5 千億ドル(55 兆円) [3]であり、セキュリティ評価済みで付加価値の高い AI エッジデバイスに対する模倣品の被害も将来的に大きくなることが予想される。それらの被害額の 10%の縮小に貢献した場合の市場規模は 5.5 兆円 (1US\$=110 円の場合)となる。

[1] JEITA 「CPS/IoT の利活用分野別世界市場調査の発表について」

<https://www.jeita.or.jp/cgi-bin/topics/detail.cgi?n=3455> 2017.12

[2] 経済産業省 「平成 29 年度知的財産権ワーキング・グループ等侵害対策強化事業 我が国模倣品被害の課題分析及び課題解決のための方策検討に関する調査最終報告資料」

https://www.meti.go.jp/medi_lib/report/H29FY/000038.pdf 2017.03

[3] 公益財団法人 日本関税協会 「世界の模倣品等の取引規模、年間 5 千億ドルに迫る—OECD 及び EUIPO の試算発表—(OECD)」 <https://www.kanzei.or.jp/topic/international/2016/for20160526.htm> 2016.04

4.12.1.4 ベンチマーク

前述のとおり、海外では、以下の表に示すようにセキュリティ評価認証制度に関係する機関が、セキュリティ評価において技術的に深い検討が必要となる部分や新たなセキュリティ評価項目を検討する際などにおいて、その国の中核的研究開発拠点と連携する体制が取られている。

表 4.12.1.4 セキュリティ評価認証体制の国内外の状況
(CC : Common Criteria, ISO/IEC 15408 の場合)

	ドイツ	オランダ	フランス	英国	米国	日本
認証機関	BSI	TÜV Rheinland Nederland	ANSSI	NCSC CB	NIAP	IPA JISEC
評価機関	—各国の認定機関により認可された民間企業など—					ECSEC lab. ITSC など
中核的研究開発拠点	フラウン ホー ファー研 究機構	オランダ 応用科学 研究機構 (TNO)	フランス国 立宇宙研究 センター (CNES)	国家サイバー セキュリティ センター (NCSC)	アメリカ国立 標準技術研 究所(NIST)	キャッチ アップ が必須

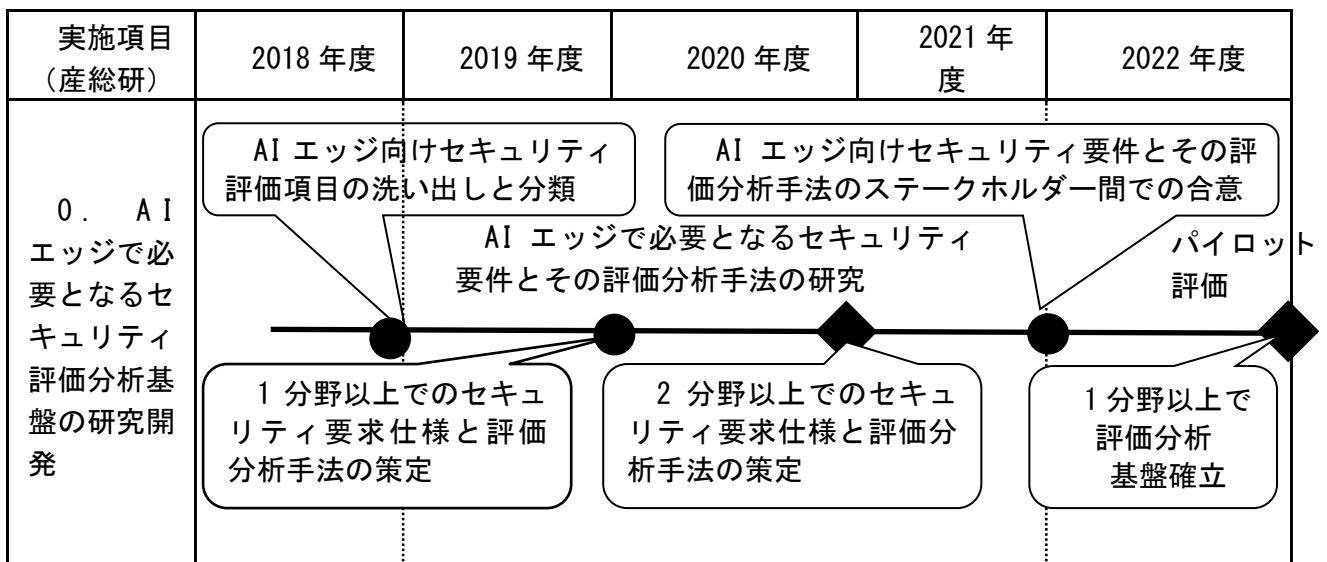
CC の国際承認アレンジメント(C CRA: CC Recognition Arrangement)に参加していない中国においても、2017 年 7 月に施行された中国国家サイバーセキュリティ法 23 条によりセキュリティ専用品の市場投入前の Certification や Test を必須とし、2020 年 5 月からはセキュリティ専用品の強制性国家標準化の議論が開始されている。それらの議論において技術的に深い検討が必要となる部分(評価技術、標準化に係わる理論的/技術的研究、研究成果を活用したセキュリティ技術コンサルテーションなど)において、CNITSEC(China National Information Technology Security Certification Center, 中国信息安全测评中心)と連携する体制が取られている。

これらが示すように、海外の主要な国々では、国内に中核的研究開発拠点を整備しそこで得られた根拠のある最新のセキュリティ評価分析手法や知見を国内での評価に活用しながら

ら、肝となるノウハウが外部に流出することを防いでいる。日本はこの部分において遅れを取っておりキャッチアップが必要となっている。

4.12.1.5 事業化までのマイルストーン

前述のキャッチアップを行うためには、用途に応じたセキュリティ要求仕様の策定と、適切なセキュリティ評価を可能とするための設備、システムの整備と、それらを有効活用するための標準的かつ効果的な評価分析手法の策定が必要となる。さらに、全体のセキュリティレベルは一番セキュリティの弱い箇所で決まるため、全体を俯瞰し脆弱な箇所の強化と評価が必要であると共に、攻撃手法は進化／変化し続けるため、常に最新の対策・評価技術の創出が可能となる基盤の整備が必要となる。これらを実現するためのマイルストーンとして、2020年度末までに2分野以上でのセキュリティ要求仕様と評価分析手法を策定し、さらに、作成した仕様や手法が社会に受け入れられ易くするために、2021年度末までにステークホルダー間で合意可能な内容になるよう改良を行う。また、重要な内容については国際標準に取り込むか国際標準への準拠を行うことにより、国際的な動向とも歩調を合わせる。2023年度からは、セキュリティ評価やそれに伴う技術コンサルテーションなどを開始できるようにすることで、セキュリティ保証スキームの実用化を行う。



4. 12. 2「実施項目 1: AI エッジ入出力セキュリティ評価シミュレータの開発」における実用化・事業化の見通し(電子商取引安全技術研究組合)

4. 12. 2. 1 概要

AI エッジデバイスへの入力である物理世界からのセンシングデータや、それらに基づく出力である機器への制御信号に対するセキュリティリスクが世界的に危惧され始めていることから、攻撃の実現可能性や影響、セキュリティ強化策の効果などをシミュレーション、あるいは、シミュレータと実機により評価できるよう、AI エッジ入出力セキュリティ評価シミュレータを 2020 年度末までに実用化する。また、実用化されたシミュレータはシステムレベルでの検証が可能なものとする。

4. 12. 2. 2 実用化・事業化への課題と対応策

センサに対する攻撃の影響を人の感覚のレベルで評価できることが望ましい。このために、一般的なドライビングシミュレータは実時間と等価な時間内で処理できることが一つの目標となる。現状、攻撃と AI 処理を組み込んだシミュレータは、保有する環境下において実時間に対し 10 倍程度の処理時間が必要である。このため、人が攻撃をどう感じるのかという視点を評価するためにはシミュレータの高速化が必要である。プログラムの最適化や 2021 年度から着手予定としているシミュレータの部分的なハードウェア化を加速して対応する。

また、現状評価指標が Euro NCAP に代表される最終的な車の挙動としての評価基準(人に車が衝突するかという観点)となっており、内部のコントローラやセンサがどうあるべきかという基準にブレイクダウンできていない。セキュリティ分析とモデルベース設計におけるテスト手法を統合し、モデル単位でセキュリティテスト基準を作成するアプローチで解決を試みる。本課題は実施項目 0 と連携して取り組んでいる。

4. 12. 2. 3 実用化・事業化の体制

AI エッジ入出力セキュリティ評価シミュレータは産業技術総合研究所に提供され、同機関において、セキュリティ評価分析技術基盤として活用される予定である。AI 制御システムのセキュリティ強化技術は、知財化後、ECSEC 組合の後継企業により運用され、三菱電機が行うアクチュエータを含む AI 制御システムのセキュリティ強化技術や評価技術は、同社の自社事業のためにも活用、運用される予定である。

4. 12. 2. 4 市場規模と経済効果

本実施項目の成果活用により、現時点の国内企業における当該分野のポテンシャルを確実に維持するとともに、AI エッジコンピューティング分野の高い品質と信頼性の獲得により、ハードウェアからアプリケーションまでを含む生産開発体制の増強、及び雇用促進への寄与が期待できる。具体的な産業貢献の一例を挙げると、緊急時以外は AI が運転の主体となる自動運転レベル 3 の 2030 年の販売台数は 1800 万台近くに上るという調査会社の発表がある。自動運転車の車載機器の開発において、本事業で整備したセキュリティ評価分析技術を基盤とした設計を行うことにより、自動運転の健全な発展を支え、自動車産業における我が国の優位性を保持することが期待できる。

4. 12. 2. 5 ベンチマーク

既存のシミュレータのアプローチでは、“センサへの攻撃”がシステム全体に与える影響の把握と、攻撃への対策設計へのフィードバックの両立が困難であるが、本研究開発によって、モデルベースデザインに組み込み可能なセンサの攻撃モデルを開発することにより、センサ攻撃の影響を最終的な動作シナリオにおいてシミュレート可能としつつ、システム各要素での影響を把握し、各要素および全体での対策効果の検証が可能となる。また、これにより構成要素における“セキュリティ基準”の策定と対策設計へのフィードバックが可能となる。

類似技術とのベンチマーク(比較:強み/弱み)は次のとおりである。

シミュレートのアプローチ	設計へのフィードバック	システム全体の挙動把握
ネットワークシミュレータを用いた攻撃モデル (例: NS-2)	× 振る舞いのモデリングであるため、設計へフィードバックできない。	○ システム全体の現象の把握に優れる。
物理シミュレータベースの攻撃モデル	○ 物理レベルのモデルであるため、原因の特定に優れる。	× センサ単体のモデリングは可能であるが、システム全体の表現は困難。
本事業でのアプローチ: モデルベースデザイン攻撃モデル	○ コード変換可能なレベルでモデリングするため、設計に反映可能	○ システム全体のシミュレートが可能であるため、現象の把握に優れる。

4. 12. 2. 6 事業化までのマイルストーン

2022 年度末までに、実機接続可能な(HILS)セキュリティ評価シミュレータとしての開発を完成させ、事業において活用可能な状態にする。

4. 12. 3 「実施項目 2 : AI エッジ内部実装保護技術の研究開発」における実用化・事業化の見通し (株式会社コネクテックジャパン)

4. 12. 3. 1 概要

IT 化や高速大容量通信の普及が加速するとともにセキュリティへの要求も高まっている。製品例としてはエッジコンピューティングやポスト5G, AI などがあり、アプリケーションとしてはデータセンタやネットワーク、アプリケーションとしてはモビリティや輸送手段などが予想される。つまり模倣やテロに晒されると大きな被害となることが予想される。

本実施項目では、電子デバイスに搭載した回路情報、電子デバイスに流れる信号等の秘匿したい情報を保護するための実装技術、及びその評価技術を開発することであり、電子デバイスそのもの製作過程から作りこむのではなく後付け加工により安価にこれを実現し社会実装に貢献することが目的である。

4. 12. 3. 2 実用化・事業化への課題と対応策

社会実装を実現するためには工法の確立だけではなく、試作、量産が可能な生産技術も構築しておくことが重要となる。本実施項目ではそのための要素技術開発、量産設備開発を 2022 年度末までに完成させる予定である。

4.12.3.3 実用化・事業化の体制

普及させるためのサプライチェーンを形成することが重要となる。すなわち電子デバイスへの後付け加工で機密情報を保護することができることをユーザに伝える、そして加工の試作、特性評価、品質評価、量産対応をターンキーで実現する体制づくりとなる。具体的にはFPGAなどデバイスメーカーとコネクテックジャパンとの協業などを考えている。

4.12.3.4 市場規模と経済効果

概要で述べた商品の傾向として標準品ではなくそれぞれのアプリケーションや使われ方に合わせた電子デバイスの選択、すなわち少品種大量の時代から多品種少量(または変量)の時代となっているといえる。したがって対象として考えられる電子デバイスとして挙げられるのはFPGAといったものである。FPGAの市場について下図に示す。



2020年と2025年のFPGA市場 出典：グローバルインフォメーション

本実施項目のセキュリティ技術が全てにおいて必要とされるものではないが、概要で述べたようにその需要は確実に増えるものと考えられ、当該技術を日本で育成し、事業化を推進することは大きな経済効果を産むものと確信する。

4.12.3.5 ベンチマーク

4.12.4.5を参照のこと。

4.12.3.6 事業化までのマイルストーン

2020年度末までに量産対応できる実装の制御技術開発と要素技術検証のためのユニット試作を開始する。2021年度に実装技術、量産化技術について技術評価を進める。2022年度にはサンプル出荷できる体制を整備する。

4.12.4 「実施項目2：AIエッジ内部実装保護技術の研究開発」における実用化・事業化の見通し（産業技術総合研究所）

4.12.4.1 概要

集積回路(IC)で処理される情報は、サイドチャネル攻撃をはじめとするリバースエンジニアリングにより、情報の一部復元または全部復元が可能である。産業技術総合研究所では、

長年、耐リバースエンジニアリング技術の研究開発を行っている。技術の発展とともに、耐リバースエンジニアリング技術は、対策としての十分性が減失するという特徴がある。株式会社コネクテックジャパンが開発する IC 保護技術は、ユースケースに合わせて、低コストに運用できる耐リバースエンジニアリング技術である。IC に直接加工を施す IC 保護技術であることから、品質の再保証が必要である。また、IC の物理解析による情報漏洩を遅延させるセキュリティ機能の保証も必要となる。いずれの保証についても、低コストで評価できることが求められる。産業技術総合研究所は、IC 保護技術により加工を施した電子デバイスの品質再保証、セキュリティ保証に必要な評価プロセスに機械学習技術を導入して、分析作業と評価作業に係る人的コストと時間的コストを削減する。

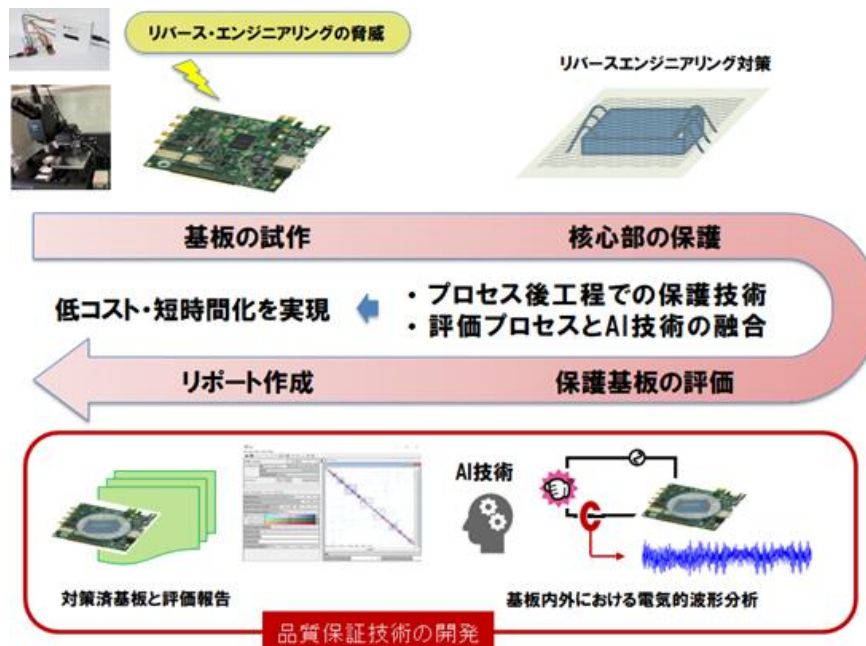


図 4.12.4.1 内部実装保護のための品質保証技術

4.12.4.2 実用化・事業化への課題と対応策

品質評価プロセスでは、放電、衝撃、振動等の物理的な外的要因に対して、IC 保護技術を施した部位が、規定の温度・湿度の範囲内では異常を生じないことを確認する。この評価プロセスで導入する波形分析システムは、分析対象の物理系から発生した電氣的波形を観測して、その時間変位・空間変位の情報から、対象に起きている異常やその原因を特定する。具体的には、観測された重畳波形(合成波形)を成分波形に分離し、素波形に混入する異常波形の有無を検知し、検知した異常波形を特定して抽出し、抽出した異常波形から分析対象の物理系に起きている異常を推定する。具体的な分析の流れを図 4.12.4.2 に示す。

波形分析システムは、従来使われている独立成分分析、波形特徴抽出、クラスタリング分析を組み合わせた波形分析技術である。階層的な分析プロセスをなす波形分析では、各プロセスの分析パラメータが他の分析プロセスに影響を与えるため、分析パラメータの設定変更の玉突き現象が発生し、分析が機能不全に陥ることがある。

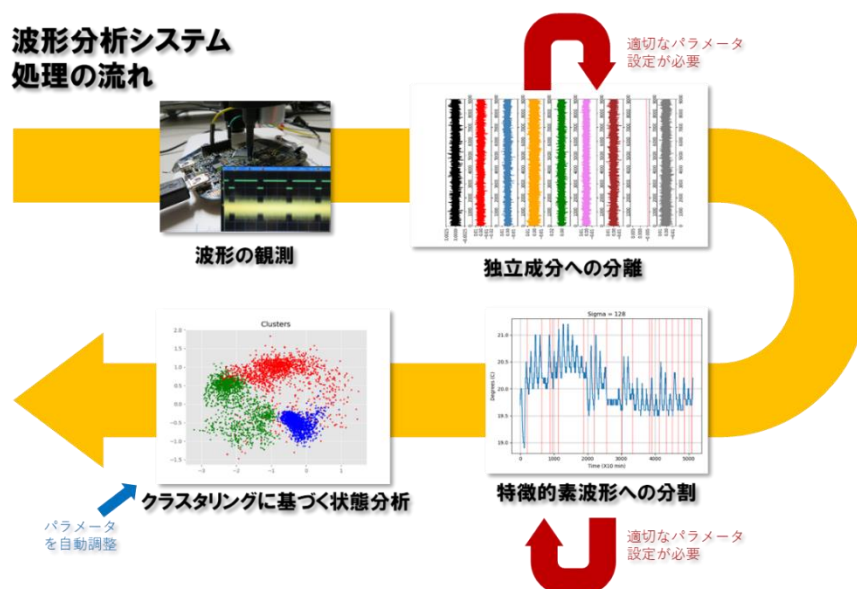


図 4.12.4.2 波形分析システムの処理手順

従来は、熟練の分析者により、経験的に適切なパラメータを組み合わせ評価を行っていた。本実施項目により開発する波形分析システムでは、ノンパラメトリック・ベイズモデルによる分析パラメータの自動設定を行う。完全に自動化できない設定作業については、パラメータ候補を提示することにより、分析作業の手戻りリスクを低減する。さらに、異常が組み込まれた分析対象を除外するために、分析対象としての妥当性を判定する評価システムも開発する。如何に正常と異常を区別するかが技術の実用化の鍵となる。

4.12.4.3 実用化・事業化の体制

産業技術総合研究所は、品質評価(品質再評価)、セキュリティ評価に必要な技術を提供する役割を担う。事業で使用する装置と施設については、装置と施設を共同で運用する可能性がある。評価基準の変更の必要が生じた場合は、事業者と外部評価機関との協議により、客観性と適切性を損なわない範囲で評価基準を改訂する。また、評価基準の見直しの検討は、定期的に行うこととする。

4.12.4.4 市場規模と経済効果

リバースエンジニアリングに関する情報は、その大半が公開されていないことから、状況の積み上げにより市場規模と経済効果を推測する。半導体製品の製造者は、高度な物理的解析技術を有する。半導体製品の不良解析技術は、高度なりバースエンジニアリング技術となり得るため下表の売上額の上位企業は、必要な設備・人員・環境を有している。

1Q18 Top 15 Semiconductor Sales Leaders (\$M, Including Foundries)

1Q18 Rank	1Q17 Rank	Company	Headquarters	1Q17 Tot IC	1Q17 Tot O-S-D	1Q17 Tot Semi	1Q18 Tot IC	1Q18 Tot O-S-D	1Q18 Tot Semi	1Q18/1Q17 % Change
1	2	Samsung	South Korea	12,811	770	13,581	18,581	820	19,401	43%
2	1	Intel	U.S.	14,220	0	14,220	15,832	0	15,832	11%
3	3	TSMC (1)	Taiwan	7,524	0	7,524	8,473	0	8,473	13%
4	4	SK Hynix	South Korea	5,346	109	5,455	8,016	125	8,141	49%
5	5	Micron	U.S.	4,931	0	4,931	7,360	0	7,360	49%
6	6	Broadcom Ltd. (2)	U.S.	3,740	368	4,108	4,160	430	4,590	12%
7	7	Qualcomm (2)	U.S.	3,676	0	3,676	3,897	0	3,897	6%
8	9	Toshiba	Japan	2,747	265	3,012	3,517	310	3,827	27%
9	8	TI	U.S.	2,960	204	3,164	3,339	227	3,566	13%
10	11	Nvidia (2)	U.S.	1,965	0	1,965	3,110	0	3,110	58%
11	15	WD/SanDisk	U.S.	1,795	0	1,795	2,350	0	2,350	31%
12	10	NXP	Europe	1,965	246	2,211	2,017	252	2,269	3%
13	12	Infineon	Europe	1,130	754	1,884	1,360	907	2,267	20%
14	13	ST	Europe	1,378	440	1,818	1,696	518	2,214	22%
15	17	Apple* (2)	U.S.	1,600	0	1,600	1,830	0	1,830	14%
—	—	Top 10 Total		59,920	1,716	61,636	76,285	1,912	78,197	26.9%
—	—	Top 15 Total		67,788	3,156	70,944	85,538	3,589	89,127	25.6%

(1) Foundry (2) Fabless *Custom devices for internal use.
Source: Company reports, IC Insights' Strategic Reviews database.

本実施項目では、少量生産品のデッドコピーや改変を防止することを本来の目的に置いている。この視点から被害状況を見ると、報告の一例は以下の通りである。

2 日本の機械工業における模倣品被害の状況と政府および工業会の模倣品対策

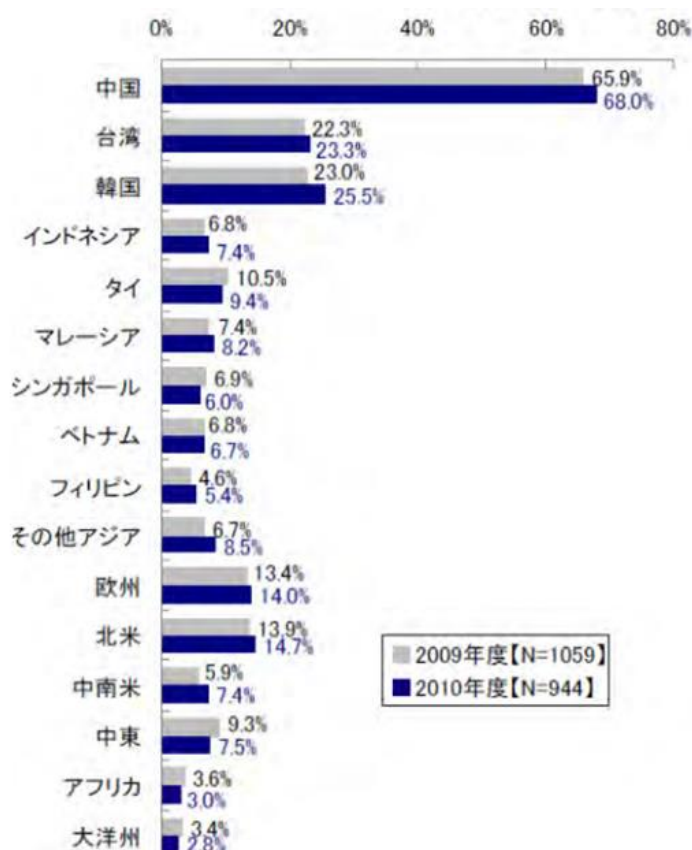
2.1 日本の機械工業における模倣品被害の状況

我が国の模倣被害は深刻さを増している。特に中国は模倣品・海賊版の製造国、消費国、輸出国としての側面を持ち、一般機械・産業機械に限ってみても権利侵害国は中国に集中している。一般機械・産業機械の模倣被害率は全体の被害率が減少している中で増加しており、状況の特徴としては、技術模倣やデッドコピーが多い点にある。模倣品の品質が劣る内容としては「耐久性」が最も多く、外観だけでは判断できないところに模倣品対策の難しさがある。なお、同業他社や業界団体と連携した模倣被害対策への取り組み状況は、全業種平均と比べてやや低調といえる。

2010年度の全商品分野（全業種）における模倣被害総額は、約1,072億円と推計されている。

図 4.12.4.4 日本の機械工業における模倣品被害の状況
(国際連携による模倣品対策調査研究報告書(2016)から抜粋)

こうした状況は、我が国のみにとどまらず、米国、ドイツも大きな損害を被っている。模倣品や改造品は、中国、台湾、韓国が主な製造国となっており(下図)、製造された模倣品・改造品は世界各地に拡散している。



2010年度の調査では、模倣品被害総額は日本国内で1072億円と推計されており、1社あたりの平均被害額は、被害額100億円未満の企業全体で1.7億円、被害額100億円以上の企業も含めると平均被害額は2.7億円になる。こうした傾向は、少量多種のエッジデバイスが市場に投入加速する2020年以降では、被害は急速に拡大すると考えられる。

4.12.4.5 ベンチマーク

類似技術とのベンチマーク(比較:強み/弱み)は次のとおりである。

	ICの取り外し	ICのパッケージ開封	コスト
セキュアブート機能	×機能を喪失しない	×機能を喪失しない	△OSが必要
セキュリティチップ	△本来機能を一部喪失する	○本来機能を喪失する	×高価
本事業のIC保護技術	○本来機能を喪失する	○本来機能を喪失する	○安価

4.12.4.6 事業化までのマイルストーン

2021年度中に評価プロセスに必要な技術開発を完了し、2022年度中に評価プロセスのサービス化に向けて、パイロット評価によるPoC(概念実証)を実施する。主体的に事業を運営する事業者との協議により、装置・施設・基準の運用方法を取り決める。

4.12.5 「実施項目2：AI エッジ内部実装保護技術の研究開発」における実用化・事業化の見通し（IIJ イノベーションインスティテュート）

4.12.5.1 概要

本実施項目で研究開発を行うファームウェア解析ツールは、最終的には、プロジェクトにおいて共同で開発される応用技術と併せ、ファームウェア保護・分析ツールとしてオープンソース化することを想定している。本ツールを活用することで、AI エッジデバイスのファームウェアが、想定外の悪意ある動作を行わないか、あるいはサイバー攻撃によるデバイスの乗っ取りを可能にする脆弱性が存在しないかを一定の範囲で自動解析できると期待される。こうしたツールが広く世の中で使われることで、自社の製品・サービスのみならず、関連するネットワーク AI 製品・サービスのセキュリティレベルの向上につながる。結果として、当該製品・サービスの市場規模の拡大に寄与する。

4.12.5.2 実用化・事業化への課題と対応策

ファームウェア保護・分析技術については、AI エッジデバイスのセキュリティを診断するための自動化ツールとしての位置づけである。したがって、実用化・事業化を進めるには、こうしたデバイスセキュリティ診断ビジネスが国内で進展することが前提となる。一部のベンダによるサービス提供が始まってはいるものの、多くの部分が人手による分析作業に依存しているため、市場規模の拡大が阻害される恐れがある。本プロジェクトでは、ファームウェア解析を自動化することで、この問題を解決することを目指しているが、市場規模の拡大についても、政府における取り組みを通じて、セキュリティ事業者との連携を取り、解析ツールの実用化・事業化の支援を行っていく予定である。

次に、技術的な観点から説明する。本プロジェクトでは、知的財産保護の観点から実施が困難なソースコード提供を前提とせず、実行プログラムのバイナリーコードを直接自動解析する技術の開発に取り組んでいるが、複数の CPU アーキテクチャや OS への個別対応に加えて、解析にかかる処理効率の飛躍的向上が必要となる。さまざまなデバイスのファームウェアを、完全に自動解析して、実用上有益な結果を出力させることは容易ではないため、本プロジェクトでは、実用上最も必要性が高いと思われる脆弱性等に絞り、ツール実行の効率と処理の汎用性を向上させる方針である。

4.12.5.3 実用化・事業化の体制

実用化の初期段階では、本プロジェクトに関連する機関での使用を通じて、機能の拡張等を図っていく。経産省を中心に、企業コンソーシアムを形成して、国内企業の製品・サービスのセキュリティレベル向上に役立てることも考えている。当社の役割は、開発ツールを技術的に正しくメンテナンスすることであると考えているが、需要に応じて製品・サービス化することも排除はしていない。

プロジェクト終了後、3年間程度は、プロジェクト関連機関を中心としてコンソーシアムを形成し、その中での利用に限定し、オープンソース化は行わない。プロジェクト終了後3年を目処に、オープンソース化を進める。一方で、この時点の状況に応じて、オープンソースとせず有償・無償のライセンス供与を行う可能性も検討する。

AI エッジデバイスを考えた場合、ソースコード参照を前提とした対策は現実的ではなく、また、対象となるプロセッサアーキテクチャや OS も多岐に渡る。今回開発するツールは、こうした差異に依存せずに、プログラムの動作を解析するものであり、ひとたび有用性が確立されれば、長期間に渡って利用を継続することが可能なため、有用性が大きい。ただし、個

別アーキテクチャや個別 OS に対応する作業は別途必要であり、こうした作業をコンソーシアム等で分担していくことを想定している。

4.12.5.4 市場規模と経済効果

すでに述べたとおり、国内のデバイスセキュリティ診断ビジネス自体は、まだ始まって間もないが、世界的なサイバーセキュリティ市場は 2016 年に 920 億ドルに達し、2019 年の市場規模は 1160 億ドルと言われている。一方で、2020 年までに、表面化する企業へのサイバー攻撃の 25%以上が IoT 関連になるとみられている。単純計算では、2020 年ころには、IoT 関連のサイバーセキュリティ市場は 300 億ドル規模でなければならないが、現状でそれだけのリソースは割かれていない。また、本プロジェクトに関連するイベントとして、DARPA が 2016 年に開催した Cyber Grand Challenge がある。この実行プログラムの脆弱性を自動的に検知するコンテストの優勝賞金は 200 万ドルであり、DARPA がコンテストの準備にかけた予算は 5,500 万ドルと言われている。本プロジェクトで開発するツールは、オープンソース化による実用化を想定しているが、サプライチェーンセキュリティの観点から、こうしたツールやサービスの重要性は増大しており、今後の市場規模の拡大と、安心して利用できる製品の普及によりもたらされる経済効果が見込まれる。

4.12.5.5 ベンチマーク

ツールの性能のベンチマーク対象としては、2016 に実施された DARPA の Cyber Grand Challenge で上位入賞したツールが挙げられる。こうしたツールは、その後ベンチャー化され、実用化へ向けた取り組みと技術の高度化が進められており、性能評価を引き続き行なっていく予定である。実用的な観点からは、実際に販売されているデバイスのファームウェアを対象とした実験を定期的に行い、未知の脆弱性やトロイの木馬を自動検知できるかチェックしていく方針である。

4.12.5.6 事業化までのマイルストーン

事業化への展望については、当初から変わっていない。以下に、スケジュール線表を再掲する。長期的に見た場合に予想される重大な障害としては、コンソーシアム形成段階におけるツールのスケーラビリティの問題、オープンソース化段階での適切なライセンス条項設定の問題、そして事業化検討段階での利用ノウハウの明文化の問題がある。現段階では問題の深刻度と解決策は明らかでないが、プロジェクト実施中から、対策の検討を進めていく予定である。

年度	2023 年度	2024 年度	2025 年度	2026 年度	2027 年度
コンソーシアム等利用	コンソーシアム形成				
オープンソース化				オープンソース化	
事業化検討			事業化検討		

4.12.6 「実施項目3：AI エッジの個体管理を支えるための人工物メトリクスの研究開発」における実用化・事業化の見通し(産業技術総合研究所)

4.12.6.1 概要

ナノ人工物メトリクス(NAM)をベースとするセキュリティ評価分析基盤を実用化し、プロジェクト終了後は産総研とその再委託先等で個体管理用識別子の読取評価、照合・識別評価、貼付評価を行えるようにする。

4.12.6.2 実用化・事業化への課題と対応策

性能以外の最大の課題は評価装置の開発コストと製造コストであり、それぞれ次のように対応中である。

照合・識別評価技術：

評価用 NAM チップの製造は、現在は1社のみを用いて試作しているが、1社のみでの供給はコスト低減が困難であり、別の会社での試作検討に着手している。

読取評価技術：

現在は既存の白色干渉顕微鏡を改造した装置で読出しを行っている。課題は、装置サイズと装置コストである。現在、卓上サイズで1/5～1/10の製造コストの小型読出し装置を開発中である。

貼付評価技術：

AI エッジデバイスとして最も重要と思われる半導体デバイスの個体管理をターゲットに実装技術を開発している。半導体デバイスは離型剤を内在する封止樹脂で形成されており、その表面への実装(接着)は一般に困難である。この表面への実装技術は汎用性がある。また本プロジェクトでは、少量生産対応の試作装置の開発も予定している。

4.12.6.3 実用化・事業化の体制

産総研とその再委託先等で個体管理用識別子の読取評価、照合・識別評価、貼付評価を行えるようにする。

4.12.6.4 市場規模と経済効果

市場規模および経済効果については、4.12.4.4 節を参照のこと。

4.12.6.5 ベンチマーク

AI エッジデバイス用途での、ナノ人工物メトリクス(NAM)をベースとする個体管理用識別子と、その他の技術をベースとする既存の個体管理用識別子とのベンチマーク(強み、弱み)は次の表のとおりまとめられる。

個体識別子	複製への耐性	部品等への適用	セキュリティ評価分析基盤
シリアル番号、QRコード、電子タグ	×複製可能	△	(複製可能)
PUF (Physically Unclonable Function: 物理複製困難関数)	○複製困難	△ 一部の電子部品にのみ適用可能	ISO/IEC 20897での国際標準化とセキュリティ評価分析基盤の整備が進んでいる。
本事業で開発対象とするナノ人工物メトリクスをベースとする識別子	○複製困難	○容易	存在しておらず、整備が必要。

4.12.6.6 事業化までのマイルストーン

2022 年度末までに個体管理用識別子の読取評価、照合・識別評価、貼付評価を実施可能にする。

●特許論文等リスト（添付資料）

◎研究開発テーマ「不揮発省電力 FPGA コアを用いた低遅延 AI 処理コンピューティング技術の研究開発」

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	伴野 直樹	日本電気株式会社	Three-fold improved set-voltage variability of a Cu atom switch with a split electrode for very-large-scale integration	Japanese Journal of Applied Physics 59, SGGB09 (2020)	有	2020/2

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	伴野 直樹	日本電気株式会社	3x Improved Set-voltage Variability of Cu Atom Switch with Split-electrode for Very Large Scale Integration	Int. Conf on Solid State Devices & Materials (SSDM)	2019/9
2	多田 宗弘	日本電気株式会社	Opportunities and Challenges of Atom Switch for Next AI Hardware	Int. Conf on Solid State Devices & Materials (SSDM)	2019/9
3	丸山 勉	筑波大学	An Implementation of Non-Local Means algorithm on FPGA	The International Conference on Parallel Computing	2019/9
4	宮村 信	日本電気株式会社	革新的 FPGA の軌道上実証結果	第 63 回 宇宙科学技術連合講演会 日本航空宇宙学会	2019/11
5	阪本 利司	日本電気株式会社	原子スイッチが拓く低電力エレクトロニクス	AI チップ設計拠点フォーラム(第 8 回)	2020/1
6	伴野 直樹	日本電気株式会社	Split-electrode による原子スイッチの Set 電圧ばらつき改善	応用物理学会(JSAP) 春季学術講演会	2020/3
7	岡本 浩一郎	日本電気株式会社	ON-state retention of Atom Switch eNVM for IoT/AI Inference Solution	IEEE International Reliability Physics Symposium	2020/4
8	根橋 竜介	日本電気株式会社	A 171k-LUT Nonvolatile Programmable Logic using Cu Atom-Switch Technology in 28nm CMOS	International Conference on Field Programmable Logic and Applications	2020/8
9	阪本 利司	日本電気株式会社	不揮発 FPGA の宇宙応用	応用物理学会(JSAP) 秋季学術講演会シンポジウム	2020/9

番号	発表者	所属	タイトル	会議名	発表年月
10	阪本 利司	日本電気株式会社	Atom-switch FPGA for low-power IoT applications	MEMRISYS Online Mini-Conference 2020	2020/11

◎研究開発テーマ「動的再構成技術を活用した組み込み AI システムの研究開発」

【特許】

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
1	ルネサスエレクトロニクス株式会社	特願 2019-189601	国内	2019/10/16		半導体装置およびその制御方法	藤井
2	ルネサスエレクトロニクス株式会社	(出願準備中)					
3	SOINN 株式会社	(出願準備中)					
4	SOINN 株式会社	(出願準備中)					
5	SOINN 株式会社	(出願準備中)					

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	安藤	東京工業大学	Dither NN: hardware/algorithm co-design for accurate quantized neural networks	IEICE Transactions on Information and Systems、 vol. E102	有	2019/12
2	本村	東京工業大学	深層ニューラルネットワーク向けプロセッサ技術の実例と展望	電子情報通信学会和文論文誌 C、 J103-C (05)	有	2020/05

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	中原	東京工業大学	A Tri-State Weight Convolutional Neural Network for an FPGA: Applied to YOLOv2 Object Detector	The 2018 International Conference on Field-Programmable Technology	2018/12
2	宗形	東京工業大学	雑音畳み込みニューラルネットワークと FPGA 実装	リコンフィギュラブルシステム研究会	2019/1
3	安藤	東京工業大学	Dither NN: 画像処理から着想を得た組み込み向け量子化ニューラルネットワークの精度向上手法	リコンフィギュラブルシステム研究会	2019/5
4	植吉	東京工業大学	無効ニューロン予測による DNN 計算効率化手法	リコンフィギュラブルシステム研究会	2019/5
5	本村	東京工業大学	AI チップ: 世界の研究動向と東工大の研究戦略	東京工業大学研究院公開	2019/10

番号	発表者	所属	タイトル	会議名	発表年月
6	本村	東京工業大学	AI Computing: The Promised Land for Computer Architecture Innovation?	Future Chips Forum 2019	2019/12
7	鈴木	東京工業大学	ProgressiveNN: Achieving Computational Scalability without Network Alteration by MSB-first Accumulative Computation	CANDAR 2020	2020/11

(b) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	SOINN 株式会社		AI・人工知能 EXPO ブース展示	2019/4

◎研究開発テーマ「進化型・低消費電力 AI エッジ LSI の研究開発」

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	岡本 吉史	株式会社ソシオネクスト	革新的なエッジコンピューティング実現を目指す NEDO 委託事業を開始	CEATEC 講演	2018/10
2	岡本 吉史	株式会社ソシオネクスト	NEDO 委託事業 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	CEATEC 講演	2019/10
3	磯野 貴巨	株式会社ソシオネクスト	A 12.1 TOPS/W Mixed-Precision Quantized Deep Convolutional Neural Network Accelerator for Low Power on Edge / Endpoint Device	IEEE Asian Solid-State Circuits Conference (A-SSCC)	2020/11

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1	株式会社ソシオネクスト	「NEDO 委託事業に採択」	ソシオネクストプレスリリース	2018/10/17
2	ArchiTek 株式会社	「AI 画像処理チップ開発」	日刊工業新聞	2020/1/9
3	ArchiTek 株式会社	「隘路抜けるシニア起業家」	日刊工業新聞	2020/1/13
4	株式会社ソシオネクスト	「量子化 DNN エンジン搭載・低消費電力 AI チップを試作」	ソシオネクストプレスリリース	2020/3/17
5	ArchiTek 株式会社	「仮想エンジニアキテックチャ搭載・エッジ AI チップを試作」	ArchiTek プレスリリース	2020/4/3
6	ArchiTek 株式会社	「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」	ArchiTek プレスリリース	2020/6/18
7	株式会社ソシオネクスト	「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」	ソシオネクストプレスリリース	2020/6/18
8	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」	豊田自動織機プレスリリース	2020/6/18

(c) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	ArchiTek 株式会社	「仮想エンジン(aIPE)」の FPGA デモ	CEATEC 2018 (J-Startup 枠)	2018/10
2	ArchiTek 株式会社	「仮想エンジン(aIPE)」の FPGA デモ	イノベーション・リーダーズ・サミット (J-Startup)	2018/10

番号	所属	タイトル	展示会名	発表年月
3	ArchiTek 株式会社	「仮想エンジン(aIPE)」の FPGA デモ	イノベーション・ジャパン 2019 (NEDO 枠)	2019/8
4	ArchiTek 株式会社	「仮想エンジン(aIPE)」の FPGA デモ	イノベーション・リーダーズ・サミット (J-Startup)	2019/10
5	株式会社ソシオネクスト	量子化 DNN FPGA デモ	CEATEC	2019/10
6	株式会社ソシオネクスト	量子化 DNN FPGA デモ	ET&IoT Technology 2019 展示	2019/11
7	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	NEDO 展示ページ ・事業紹介 ・リーフレット「未来社会をもっと便利 に！ 進化型・低消費電力 AI エッジ LSI」	CEATEC 2020 オンライン	2020/10

◎研究開発テーマ「ソフトテンソルプロセッサによる超広範囲センシング AI エッジ技術の研究開発」

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Kohei Yamamoto, Kurato Maeno	沖電気株式会社	PCAS: Pruning Channels with Attention Statistics for Deep Network Compression	arXiv	無	2018/12
2	山本 康平, 橋 素子, 前野 蔵人	沖電気株式会社	ディープラーニングのモデル軽量化技術	OKI テクニカルレビュー第 233 号	有	2019/5
3	Kohei Yamamoto, Motoko Tachibana, Kurato Maeno	沖電気株式会社	Model Pruning Technology for Deep Neural Networks	OKI テクニカルレビュー第 233 号(英語版)	有	2019/7
4	Kohei Yamamoto, Kurato Maeno	沖電気株式会社	PCAS: Pruning Channels with Attention Statistics for Deep Network Compression	arXiv	無	2019/8
5	国定 恭史, 山本 康平, 橋 素子, 前野 蔵人	沖電気株式会社	ニューラルネットワークの枝刈りが感度マップへ及ぼす影響	システム制御情報学会論文誌 33 巻第 5 号	有	2020/5
6	Trong Huy Phan, Kazuma Yamamoto	沖電気株式会社	Resolving Class Imbalance in Object Detection with Weighted Cross Entropy Losses	arXiv	無	2020/6
7	増田 誠, 橋 素子, 山本 一真 他	沖電気株式会社	船舶の自動化、IoT 技術に関する OKI の取り組み	日本マリンエンジニアリング学会学会誌 55 巻 6 号	有	2020/11
8	福井 洋, 天谷 一郎	ジャパンマリンユナイテッド株式会社	自律化船実現に向けた物体検出技術への取り組み	日本マリンエンジニアリング学会学会誌 55 巻 6 号	有	2020/11
9	西村 匡史, ファンチョン フィ, 山本 一真, 増田 誠	沖電気株式会社	周辺特徴抽出と Soft-NMS による隠蔽に強い物体検出の検討	精密工学会誌 86 巻 12 号	有	2020/12

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	国定 恭史, 山本 康平, 橘 素子, 前野 蔵人	沖電気株式会社	ニューラルネットワークの枝刈りが感度マップへ及ぼす影響	第 63 回システム制御情報学会	2019/5
2	Stanislav Sedukhin, Kazuya Matsumoto, Yoichi Tomioka	会津大学	Brain-inspired Co-design of Algorithm/ Architecture for CNN Accelerators	7th International Conference on Smart Computing and Artificial Intelligence (SCAI 2019)	2019/7
3	Kohei Yamamoto, Kurato Maeno	沖電気株式会社	PCAS: Pruning Channels with Attention Statistics for Deep Network Compression	30th British Machine Vision Conference (BMVC2019)	2019/9
4	天谷 一郎, 比留井 仁, 小林 敬明, 石毛 健晴	ジャパンマリンユナイテッド株式会社	深層学習を用いた長距離物体検出技術の開発 (その 1)	日本船舶海洋工学会 令和元年秋季講演会	2019/11
5	ファンジョンフィ, 山本 一真, 増田 誠	沖電気株式会社	小物体に対応した SSD に基づく検出器の開発	ViEW2019 ビジョン技術の実利用ワークショップ	2019/12
6	穂積 和貴, 富岡 洋一	会津大学	高精細映像における連続フレームを用いた効率的なブロック分割物体検出手法	電子情報通信学会 パターン認識・メディア理解研究会 (PRMU 研究会)	2019/12
7	石川 晴也, 林 昌希, ファンジョンフィ, 山本 一真, 増田 誠, 青木 義満	慶應義塾大学, 沖電気株式会社	Robust Multi-Object Tracking with Spatio-Temporal Features	ViEW2020 ビジョン技術の実利用ワークショップ	2019/12
8	Stanislav Sedukhin, Yoichi Tomioka	会津大学	Massively-Parallel Computing of Multi-Channel 2D Convolution	The SIAM Conference on Parallel Processing for Scientific Computing	2020/2
9	川村 聡志, 国定 恭史, 山本 康平, 橘 素子	沖電気株式会社	(2+1)次元畳み込みネットワークの枝刈り効果に関する考察	電子情報通信学会総合大会	2020/3

番号	発表者	所属	タイトル	会議名	発表年月
10	山本 康平, 橋 素子	沖電気株式会社	低ビット量子化と枝刈りの併用による CNNの軽量化	電子情報通信学会総合大会	2020/3
11	西村 匡史, ファンクション フィ, 山本 一真, 増田 誠	沖電気株式会社	周辺特徴抽出と Soft-NMS による 隠蔽に強い物体検出	DIA2020 動的画像処理実利 用化ワークショップ 2020	2020/3
12	磯部 宙, 富岡 洋一	会津大学	ゼロビットスキップ機能付きビットシリアル 内積回路を用いた畳み込み演算 の高速化に関する一検討	電子情報通信学会 VLSI 設 計技術研究会	2020/3
13	江下 尚彦, 森口 拓雄, 徳梅 慎也	総合警備保障株式 会社	広域監視システムにおける渋滞判定 手法	電子情報通信学会 総合大会	2020/3
14	石川 晴也, 林 昌希, ファンクション フィ, 山本 一真, 増田 誠, 青木 義満	慶應義塾大学, 沖 電気株式会社	Part Aware Online Multi- Object Tracking using Attention Mechanism	電気学会 知覚情報/次世代 産業システム合同研究会	2020/3
15	国定 恭史, 前野 蔵人, 橋口 展明, 井下田 吉 男, 富岡 洋一	沖電気株式会社, ジャパンリユニ テッド株式会社, 総合警備保障株式 会社, 会津大学	ニューラルネットワークの推論時にお ける GPU の消費電力の計測	電子情報通信学会 総合大会	2020/3
16	Kazuki Hozumi, Yoichi Tomioka	会津大学	Low-latency Block-wise Object Detection Method using SSD for High Resolution Video	International Conference on Digital Signal Processing (ICDSP 2020)	2020/6
17	徳梅 慎也, 森口 拓雄, 江下 尚彦	総合警備保障株式 会社	U-Net とドメイン適用を用いた煙検 出モデルの作成と評価	第 26 回 画像センシングシンポ ジウム (SSII 2020)	2020/6
18	Sora Isobe, Yoichi Tomioka	会津大学	Low-bit Quantized CNN Acceleration based on Bit- serial Dot Product Unit with Zero-bit Skip	CANDAR 2020: The Eighth International Symposium on Computing and Networking	2020/11
19	ファンクション フィ, 山本 一真	沖電気株式会社	物体検出への階層的なクラス識別の 導入検討	ViEW2020 ビジョン技術の実利 用ワークショップ	2020/12

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1	沖電気株式会社, ジャパンマリユニテッド株式会社, 総合警備保障株式会社, 会津大学	NEDO の革新的 AI エッジコンピューティングをテーマとする技術開発プロジェクトを開始	OKI、JMU、ALSOK、会津大 共同プレスリリース	2018/9/25
2	沖電気株式会社	AI 技術動向の変化と OKI の取り組み	OKI プレミアムフェア 2018	2018/11/16
3	沖電気株式会社	革新的 AI エッジコンピューティング	OKI Web サイト	2018/11/30
4	沖電気株式会社	OKI の技術の今を支え将来を生み出す研究開発センターの挑戦	OKI Web サイト	2018/11 末
5	沖電気株式会社	ディープラーニングモデルの新たな軽量化技術を開発	NEDO、OKI 共同プレスリリース	2019/9/9
6	沖電気株式会社	映像・光・音響センシング技術による船舶 IoT	OKI AI エッジコンピューティング 製品発表会	2019/10/3
7	沖電気株式会社	センシングやネットワークを強みに AI エッジを推進	日経 XTECH Special	2019/10/3
8	沖電気株式会社	OKI レポート 2019	OKI 統合報告書	2019/11/1
9	沖電気株式会社	Interview with Head of Corporate Research & Development Center	OKI Web サイト	2020/2
10	沖電気株式会社	ディープラーニングモデルの軽量化技術を開発	月刊「画像ラボ」5 月号 (日本工業出版(株))	2020/5
11	沖電気株式会社	ディープラーニング軽量化技術による電力消費効率の改善	経団連 チャレンジ・ゼロ (日本経済団体連合会)	2020/6/8
12	沖電気株式会社	OKI、経団連主催の「チャレンジ・ゼロ」へ参加 ～ディープラーニング軽量化により AI の電力消費効率改善技術の開発を促進～	OKI プレスリリース	2020/6/9
13	ジャパンマリユニテッド株式会社	無人化船要素技術：画像認識について	JMU Web サイト	2020/6/30

(c) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	沖電気株式会社	革新的 AI エッジコンピューティング	CEATEC JAPAN 2018、	2018/10
2	ジャパンマリユニテッド株式会社	船舶用周囲監視システム	物流系ユーザー会での技術講演	2018/10
3	沖電気株式会社	AI 技術動向の変化と OKI の取り組み	OKI プレミアムフェア 2018	2018/11
4	沖電気株式会社	映像・光・音響センシング技術による船舶 IoT	OKI AI エッジコンピューティング 製品発表会	2019/10

番号	所属	タイトル	展示会名	発表年月
5	会津大学	ソフトテンソルプロセッサによる超広範囲・高精度にセンシング	IoT ワークショップ「センシング・エッジによる IoT 革新的ビジネスの潮流」	2019/11
6	総合警備保障株式会社	AI・4K・5G で進化する ALSOK の警備サービス ～都市空間セキュリティの実現に向けて～	ET & IoT Technology 2019(組み込み総合技術展 & IoT 総合技術展)カンファレンス 基調講演	2019/11
7	総合警備保障株式会社	不審行動検知 AI、高所からの火災検出システム	ALSOK フューチャーフォーラム	2020/2

◎研究開発テーマ「5G 時代を見据えた高度自律的学習機能搭載のための AI エッジコンピューティング技術の研究開発」

【特許】

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
1	株式会社アラヤ	特願 2020-063139	国内	2020/3/31	出願中	情報処理装置及び情報処理方法	森 俊彰, 水谷永輔, 津田 達也, 大鳥羽 暢彦, 蓮井 樹生, 松本 渉
2	株式会社アラヤ	特願 2020-0711101	国内	2020/4/10	出願中	情報処理装置及び情報処理方法	森 達也, 大鳥羽 暢彦

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	Keiko Fujii, Hoshinori Kanazawa, Yasuo Kuniyoshi	東京大学	Spike Timing Dependent Plasticity Enhances Integrated Information at the EEG Level: A Large-scale Brain Simulation Experiment	Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob 2019),	2019/8
2	杉田 博司	KDDI 株式会社	5G 時代を見据えた AI エッジコンピューティングの開発	マルチメディア推進フォーラム	2019/10
3	金井 良太, 松本 渉	株式会社アラヤ	「エッジ AI を“自動で”実現するアプリケーション 脳科学者が語る AI の未来と、ニューラルネットワークの最先端圧縮技術」	『ET / IoT Technology 2019』エッジテックトラックセミナー	2019/11
4	松本 渉	株式会社アラヤ	「ニューラルネットワークの圧縮技術によるエッジ AI の実現」	『GPU Computing Workshop for Advanced Manufacturing 2019』テクニカルセッション	2019/12
5	玉井 信也	株式会社アラヤ	産業用ドローンシステムへの世界モデルの応用に関する一検討	2020 年度 人工知能学会全国大会 (第 34 回) OS-18 世界モデルと知能	2020/6

(b) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	株式会社アラヤ	出典社ブース	『ET / IoT Technology 2019』	2019/11

◎研究開発テーマ「エッジビジョン AI を超軽量化し短 TAT で実装する技術の研究開発」

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Y. Sada, M. Shimoda, A. Jinguji, H. Nakahara	東京工業大学	"A Dataflow Pipelining Architecture for Tile Segmentation with a Sparse MobileNet on an FPGA,"	International Conference on Field-Programmable Technology (FPT)	有	2019 (採録決定)
2	R. Kuramochi, M. Shimoda, Y. Sada, S. Sato, H. Nakahara	東京工業大学	"FPGA-based Accurate Pedestrian Detection with Thermal Camera for Surveillance System,"	International Conference on Reconfigurable Computing and FPGAs (ReConFig)	有	2019 (採録決定)
3	R. Kuramochi, Y. Sada, M. Shimoda, S. Sato, H. Nakahara	東京工業大学	"Many Universal Convolution Cores for Ensemble Sparse Convolutional Neural Networks,"	13th Int'l Symp. on Embedded Multicore/many-Core Systems-on-Chip (MCSoc2019), pp.93-100.	有	2019
4	M. Shimoda, Y. Sada, R. Kuramochi, H. Nakahara	東京工業大学	"An FPGA implementation of Real-time Object Detection with a Thermal Camera"	FPL, pp.413-414	有	2019
5	A. Jinguji, Y. Sada, H. Nakahara	東京工業大学	"Realtime Object Detection for Many Pedestrian Toward Surveillance Camera"	FPL, pp.424-425	有	2019
6	M. Shimoda, Y. Sada, H. Nakahara	東京工業大学	"Filter-wise Pruning Approach to FPGA Implementation of Fully Convolutional Network for Semantic Segmentation,"	15th International Symposium on Applied Reconfigurable Computing (ARC), pp.371-386	有	2018

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
7	A. Jinguji, T. Fujii, S. Sato, H. Nakahara	東京工業大学	"An FPGA Realization of OpenPose based on a Sparse Weight Convolutional Neural Network,"	International Conference on Field-Programmable Technology (FPT), pp.310-313	有	2018
8	H. Nakahara, M. Shimoda, S. Sato	東京工業大学	"A Demonstration of FPGA-Based You Only Look Once Version2 (YOLOv2),"	FPL, pp.457-458	有	2018
9	M. Shimoda, S. Sato, H. Nakahara	東京工業大学	"Demonstration of Object Detection for Event-Driven Cameras on FPGAs and GPUs,"	FPL, pp.461-462.	有	2018
10	M. Shimoda, S. Sato, H. Nakahara	東京工業大学	"Power Efficient Object Detector with an Event-Driven Camera on an FPGA,"	The 9th International Workshop on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART 2018), pp.1-6	有	2018
11	R. Kuramochi , H. Nakahara	東京工業大学	"An FPGA-Based Low-Latency Accelerator for Randomly Wired Neural Networks,"	FPL, 2020, (accepted)	無	2020
12	H. Nakahara, Q. Zhiqiang, A. Jinguji, W. Luk	東京工業大学, Imperial College London	"R2CNN: Recurrent Residual Convolutional Neural Network on FPGA,"	28th ACM/SIGDA Int'l Symp. on Field-Programmable Gate Arrays~(FPGA), page 319	有	2020/2
13	中原 啓貴	東京工業大学	"畳込みニューラルネットワークのFPGA実装,"	電子情報通信学会誌, Vol. 103, No. 5, pp.501-506	有	2020/5
14	N. Soga, Y. Sada, M. Shimoda, A. Jinguji, S. Sato, H. Nakahara	東京工業大学	"Fast Monocular Depth Estimation on an FPGA,"	IPDPS Workshop (RAW2020), pp.1-4	有	2020/5

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
15	H. Nakahara, Z. Que, W. Luk	東京工業大学, Imperial College London	"High-Throughput Convolutional Neural Network on an FPGA by Customized JPEG Compression,"	The 28th IEEE Int'l Symp. on Field-programmable Custom Computing Machines (FCCM), pp.1-9	有	2020/5
16	Z. Que, H. Nakahara, E. Nuvitadhi, H. Fan, C. Zeng, J. Meng, X. Niu, W. Luk	Imperial College London, 東京工業大学, Intel Corp.	"Optimizing Reconfigurable Recurrent Neural Networks,"	FCCM, pp.1-8	有	2020/5
17	A. Jinguji, S. Sato, H. Nakahara	東京工業大学	"Tiny On-Chip Memory Realization of Weight Sparseness Split-CNNs on Low-end FPGAs,"	FCCM, 2020, page 1	有	2020/5
18	Y. Suzuki, N. Soga, S. Sato, H. Nakahara	東京工業大学	"A Table Look-Up Based Ternary Neural Network Processor,"	The 50th IEEE Int'l Symp. on Multiple-Valued Logic (ISMVL), 2020, pp.1-6,	有	2020/6
19	H. Nakahara	東京工業大学	" 2^{n+1} -valued SSS-Net: Uniform Shift, Channel Sparseness, and Channel Shuffle,"	ISMVL, 2020, pp.1-6,	有	2020/6
20	M. Shimoda, H. Nakahara et al.	東京工業大学	"SENTEI: Filter-wise Pruning with Distillation Towards Efficient Sparse Convolutional Neural Network Accelerators"	IEICE Trans. on Inf.	有	2020/10 (出版予定)

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	丸岡 晃	株式会社フィックスターズ	ドメイン固有言語とコンパイラ技術を用いた専用ハードウェア開発	DA シンポジウム 2019	2019/8
2	山田 貴登	株式会社フィックスターズ	DNN コンパイラと専用 DSL による小型 FPGA デバイスへの深層学習の適用と実装	Design Solution Forum 2019	2019/9

番号	発表者	所属	タイトル	会議名	発表年月
3	山田 貴登	株式会社フィックス ターズ	マルチレイコンパイラ基盤による、エッジ向けディープラーニングの実装と最適化について	Deep Learning Acceleration 勉強会#2	2019/10
4	山田 貴登	株式会社フィックス ターズ	AI チップ戦国時代における深層学習モデルの推論の最適化と実用的な運用を可能にするソフトウェア技術について	日本ソフトウェア科学会 機械学習工学研究会 (MLSE)	2020/7
5	中原 啓貴	東京工業大学	"お題への回答「マーケット」～AI ハードウェア研究者から経営者に Acadexit した者の一感想～"	日本ソフトウェア科学会 機械学習工学研究会 (MLSE)	2020/7
6	佐田 悠生, 下田 将之, 佐藤 真平, 中原 啓貴	東京工業大学	"畳み込みニューラルネットワークを用いた単眼深度推定の FPGA 実装について,"	電子情報通信学会リコンフィギュラブルシステム研究会(慶應大), RECONF2019-56, 2020, pp.73-78.	2020
7	倉持 亮佑, 佐田 悠生, 下田 将之, 佐藤 真平, 中原 啓貴	東京工業大学	"アンサンブル学習を用いたスパース CNN の FPGA 実装に関して,"	電子情報通信学会リコンフィギュラブルシステム研究会(慶應大), RECONF2019-55, 2020, pp.67-72.	2020
8	秋元 宏介, 佐田 悠生, 佐藤 真平, 中原 啓貴	東京工業大学	"ハードウェア実装に適した畳み込みニューラルネットワークのフィルタに関する比較,"	電子情報通信学会リコンフィギュラブルシステム研究会(慶應大), RECONF2019-54, 2020, pp.61-66	2020
9	倉持 亮佑, 佐田 悠生, 下田 将之, 佐藤 真平, 中原 啓貴	東京工業大学	"アンサンブル学習を用いたスパース CNN の FPGA 実装に関して,"	第 33 回多値論理とその応用研究会(神戸), 2020, No.19.	2020
10	鈴木 裕太, 曾我 尚人, 佐藤 真平, 中原 啓貴	東京工業大学	"テーブル参照方式 3 値ニューラルネットワーク推論プロセッサについて,"	第 33 回多値論理とその応用研究会(神戸), 2020, No.20.	2020
11	中原 啓貴	東京工業大学	" 2^{n+1} -valued SSS-Net: Uniform Shift, Channel Sparseness, and Channel Shuffle,"	第 33 回多値論理とその応用研究会(神戸), 2020, No.21.	2020

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1	東京工業大学	「株式会社桜島養魚が AI トラッキング魚体計数機を導入～養殖魚の尾数計数作業の自動化が実現～」	マルハニチロ株式会社 HP ニュース&トピックス	2020/5

番号	所属	発表内容	発表先	発表年月
2	東京工業大学	「マルハニチロ、AIで養殖魚数を管理 自動で計測」	日本経済新聞 2020年5月16日朝刊、5 月15日電子版	2020/5
3	東京工業大学	「マルハニチロ AIが養魚数を自動 計測 桜島養魚でブリ、カンパチ 商 品力向上や経費削減」	みなと新聞 2020年5月19 日 一面、5月18日 電子版	2020/5
4	東京工業大学	「マルハニチロ、ブリなど尾数計数にA I導入」	水産経済新聞 2020年5月 19日 一面、5月19日 電子 版	2020/5
5	株式会社フィクスターズ	フィクスターズ、クラウド上のエッジ AI 開発環境「GENESIS」をベータ公開	株式会社フィクスターズ プレスリリース	2020/10

(c) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	東京工業大学	エッジ AI デザインサービス	Ledge.ai EXPO	2020/4
2	株式会社フィクスターズ	クラウド上で完結する、エッジビジョン AIの開発プラットフォーム	CEATEC	2020/10

◎研究開発テーマ「スケーラブルなエッジ HPC を実現する OS 統合型プラットフォームの研究開発」

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Z. Zhong, M. Edahiro	名古屋大学	Model-Based Parallelization for Simulink Models on Multicore CPUs and GPUs	International Journal of Computers & Technology, Vol. 20, pp.1-13.	有	2020/1

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	権藤 正樹	イーソル株式会社	エッジコンピューティングへの挑戦	eSOL Technology Forum 2019	2019/9
2	Z. Zhong, M. Edahiro	名古屋大学	Model-Based Parallelization for Simulink Models on Multicore CPUs and GPUs	ISOCC2019	2019/10
3	小川真彩 高, 本田 晋也, 枝廣 正人	名古屋大学	ヘテロジニアスマルチコアシステムに対する統合型設計環境	複雑系マイクロシンプジウム (CSMS'20)	2020/3

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1		"Research on highly parallel embedded control system design and implementation method"	Impact https://www.ingentaconnect.com/content/sil/impact/2019/00002019/00000010/art00016	2019/12
2			日本証券新聞	2019/11/14
3			株主手帳 2020 年 2 月号	2020/1
4			イーソルプレスリリース	2020/1

(c) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	名古屋大学	枝廣研究室活動紹介	ET2019	2019

◎研究開発テーマ「動的多分岐・結合トレース型 AI プロセッサのエコシステム開発」

【特許】

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
1	株式会社日立製作所	特願 P2019-218696	国内	2019/12/3	出願	ロボットシステム及びロボットシステムの制御方法	小菅 敦丈, 大島 俊

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Atsutake Kosuge, Keisuke Yamamoto, Yukinori Akamine, Takashi Oshima	株式会社日立製作所	An SoC-FPGA-Based Iterative-CIOSet-Point Accelerator Enabling Faster Picking Robots	IEEE Transactions on Industrial Electronics	有	2020/3

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	Neha Chauhan, Tsuyoshi Ishiki, Dongju Li	東京工業大学	Speaker Recognition using LPC, MFCC, ZCR features with ANN and SVM Classifier for large input Database	ICCCS (International Conference on Computer and Communications Systems) 2019, Singapore	2019/2
2	小菅 敦丈, 大島 俊	株式会社日立製作所	An Object-POST Estimation Acceleration Technique for Picking Robot Applications by Using Graph-Reusing K-NN Search	IEEE Graph computing 2019	2019/9
3	小菅 敦丈, 大島 俊	株式会社日立製作所	ソーティングネットワーク回路によるグラフ型近傍点探索を用いたピッキングロボット向け物体姿勢推定高速化手法	IEICE ソサイエティ大会 2019	2019/9

番号	発表者	所属	タイトル	会議名	発表年月
4	Tomoya Kashimata, Toshiaki Kitamura, Keiji Kimura, Hironori Kasahara	早稲田大学	Cascaded DMA Controller for Speedup of Indirect Memory Access in Irregular Applications	SC19	2019/11
5	Viviana Crescitelli, Atsutake Kosuge, Takashi Oshima	株式会社日立製作所	Towards Activity Recognition under variability of lighting conditions for Human-Robot Collaboration	UPINLBS	2019/11
6	Keiji Kimura, Kazuhi Fujita, Kazuki Yamamoto, Tomoya Hashimata, Toshiaki Kitamura, Hironori Kahasara	早稲田大学	Automatically Parallelizing Compiler Cooperative OSCAR Vector Multicore	IWIA2020	2020/2
7	山本 一貴, 藤田 一輝, 柏俣 智哉, 高橋 健, Boma A. Adhi, 北村 俊明, 川島 慧大, 納富 昭, 森 裕司, 木村 啓二, 笠原 博徳	早稲田大学, オスカーテクノロジー株式会社, 株式会社, 株式会社 エヌエスアイテクス	マルチターゲット自動並列化コンパイラにおけるアクセラレータコスト推定手法の検討	IEICE ETNET	2020/2

番号	発表者	所属	タイトル	会議名	発表年月
8	Boma A. ADHI, Tomoya Kashimata , Ken Takahashi, Keiji Kimura, Hironori Kasahara	早稲田大学	Compiler Software Coherent Control for Embedded High- performance Multicore	IEICE	2020/3
9	Viviana Crescitelli, Atsutake Kosuge, Takashi Oshima	株式会社日立製作 所	A camera fusion approach for Multi Person POSe Estimation in low light environments	2020 IEEE Sensors Applications Symposium	2020/3
10	Viviana Crescitelli, Atsutake Kosuge, Takashi Oshima	株式会社日立製作 所	Towards Real-Time Multi- Person POSe Estimation in low light environments	IEICE 総合大会 2020	2020/3
11	小菅 敦文, 大島 俊	株式会社日立製作 所	A 1200x1200 8-Edges/Vertex FPGA-based Motion-Planning Accelerator for Dual-Arm- Robot Manipulation Systems	2020 Symposia on VLSI Technology and Circuits	2020/6
12	Neha Chauhan, TsuyoshiIs shiki, Dongju Li	東京工業大学	Speaker Recognition using fusion of features with Feedforward Artificial Neural Network and Support Vector Machine	International Conference on Intelligent Engineering and Management (ICIEM 2020)	2020/6
13	Nabilah Shabrina, Dongju Li, Tsuyoshi Isshiki	東京工業大学	Small Area Fingerprint Verification using Deep Convolutional Neural Network	International Conference on Intelligent Engineering and Management (ICIEM 2020)	2020/6

(b) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	株式会社日立製作所	ロボットピッキングシステムデモン レーション	Hitachi Social Innovation Forum 2019	2020/3
2	株式会社日立製作所	ロボットピッキングシステムデモン レーション	国際物流総合展 2020	2020/3

◎研究開発テーマ「完全自動運転に向けたシステムオンチップとソフトウェアプラットフォームの研究開発」

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Seiya Maeda, Yuya Maruyama, Takuya Azumi	埼玉大学, 大阪大学	Static Program Placement Platform for Embedded Multi-Core Multi-Cluster Systems	Demo Session of IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)	有	2019/4
2	Hiroyuki Chishiro, Kazutoshi Suito, Tsutomu Ito, Seiya Maeda, Takuya Azumi, Kenji Funaoka, Shinpei Kato	東京大学, 株式会社アクセル, 埼玉大学, 株式会社ティアフォー	Towards Heterogeneous Computing Platforms for Autonomous Driving	IEEE 15th International Conference on Embedded Software and Systems (ICES2019)	無	2019/6
3	安積 卓也, 丸山 雄也, 前田 青也, 加藤 真平	埼玉大学, 大阪大学, 東京大学	Autoware on Many-core Platform: NoC ベース組込みメニーコアプロセッサ向け自動運転プラットフォーム ※優秀論文賞 (ベストペーパー) 受賞	組込みシステムシンポジウム (ESS2019)	有	2019/9
4	西村 啓佑, 粟本 真一, 千代 浩之, 加藤 真平	東京大学, 東京工業大学	ROOP: ヘアメタルプログラム向けオンラインジャッジプラットフォーム	第 31 回コンピュータシステムシンポジウム (ComSys2019)	有	2019/12
5	Keita Miura, Takuya Azumi	埼玉大学	Converting Driving Scenario Framework for Testing Self-Driving Systems	EUC 2020: International Conference on Embedded and Ubiquitous Computing	有	2020/12
6	Yuqing Yang, Takuya Azumi	埼玉大学	Exploring Real-Time Executor on ROS 2	IEEE International Conference on Embedded Software and Systems (ICES2020)	有	2020/12

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	Seiya Maeda, Yuya Maruyama, Takuya Azumi	Univ. of Saitama/ Univ. of Osaka/ Univ. of Saitama	Static Program Placement Platform for Embedded Multi-Core Multi-Cluster Systems	Demo Session of IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)	2019/4
2	Hiroyuki Chishiro, Kazutoshi Suito, Tsutomu Ito, Seiya Maeda, Takuya Azumi, Kenji Funaoka, Shinpei Kato	東京大学, 株式会社アクセル, 埼玉大学, 株式会社ティアフォー	System on a Chip and Software Platforms for Fully Autonomous Driving	IEEE 15th International Conference on Embedded Software and Systems (ICES2019)	2019/6
3	Hiroyuki Chishiro	東京大学	Towards Heterogeneous Computing Platforms for Autonomous Vehicles	The 2019 International Symposium for Advanced Computing and Information Technology [Keynote]	2019/8
4	安積 卓也, 丸山 雄也, 前田 青也, 加藤 真平	埼玉大学, 大阪大学, 東京大学	Autoware on Many-core Platform: NoC ベース組込みメモリーコプロセッサ向け自動運転プラットフォーム ※優秀論文賞 (ベストペーパー) 受賞	組込みシステムシンポジウム (ESS2019)	2019/9
5	西村 啓佑, 粟本 真一, 千代 浩之, 加藤 真平	東京大学, 東京工業大学	ROOP : ヘアメタルプログラム向けオンラインジャッジプラットフォーム	第 31 回コンピュータシステムシンポジウム (ComSys2019)	2019/12
6	Shinpei Kato	東京大学	Heterogeneous Computing Platforms for Autonomous Vehicles	2020 International Solid-State Circuits Conference (ISSCC2020) [Invited]	2020/2

番号	発表者	所属	タイトル	会議名	発表年月
7	Takuya Azumi, Yuya Maruyama, Shinpei Kato	埼玉大学, 大阪大学, 東京大学	ROS-lite: ROS Framework for NoC-Based Embedded Many-Core Platform	IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020)	2020/10

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1		NEDO「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」事業採択のお知らせ	PRTIMES https://prtimes.jp/main/html/rd/p/000000025.000004053.html	2019/3/29
2		アクセルとティアフォー、AI ハードウェアアクセラレータの製品化に向けて連携を強化	PRTIMES https://prtimes.jp/main/html/rd/p/000000033.000004053.html	2020/6/23
3		西新宿エリアで5Gを活用した自動運転タクシーの実証実験（フェーズII）を始動	PRTIMES https://prtimes.jp/main/html/rd/p/000000022.000050415.html	2020/10/9

(c) 展示会への出典

番号	所属	タイトル	展示会名	発表年月
1	株式会社ティアフォー	Autoare をコアとしたオープンエコシステムの紹介	CES 2020	2020/1
2	株式会社ティアフォー	自動運転仕様の JPN TAXI 車両	第12回 オートモーティブ ワールド –クルマの先端技術展–	2020/1
3	株式会社アクセル, 株式会社ティアフォー	Autoware Ecosystems with ARM Devices	Arm DevSummit 2020 ※オンラインイベント	2020/10
4	株式会社アクセル, 株式会社ティアフォー	完全自動運転社会の実現へ！高速データ処理を実現する SoC 研究開発	CEATEC 2020	2020/10

◎研究開発テーマ「セキュアオープンアーキテクチャ基盤技術とその AI エッジ応用研究開発」

【特許】

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
1	株式会社日立製作所	特願 2020-140295	国内	2020/8/21	出願済	データ処理方法、エッジ装置、および、データ処理システム	新保 他
2	株式会社日立製作所	特願 2020-138965	未定	2020/08/19	出願済	シェアリングコンピューティングシステム、リソースシェアリング方法	木下 他

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	Trong-Thuc Hoang, Ckristian Duran, Duc-Thinh Nguyen, Hoang ; Duc-Hung Le, Akira Tsukamoto, Kuniyasu Suzaki, Cong-Kha Pham	電気通信大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Quick Boot of Trusted Execution Environment With Hardware Accelerators	IEEE Access, 74015 – 74023 (Impact Factor 4.098)	有	2020/4
2	須崎 有康, 佐々木 貴之	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所, 日本電気株式会社, 横浜国立大学	Trusted Execution Environment によるシステムの堅牢化	情報処理学会学会誌, 情報処理	無	2020/05
3	須崎 有康	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所	Trusted Execution Environment の実装とそれを支える技術	電子情報通信学会 基礎・境界サイエティ誌 Fundamental Reviews	無	2020/10

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
4	Trong-Thuc Hoang, Ckristian Duran, Khai-Duy Nguyen, Tuan-Kiet Dang, Quynh Nguyen, Quang Nhu, Phuc Hong Than, Xuan-Tu Tran, Duc-Hung Lee, Akira Tsukamoto, Kuniyasu Suzuki, Cong-Kha Pham	電気通信大学, 産業技術総合研究所, DUT, DTU, VNU-UET, VNU-HCMUS, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Low-power High-performance 32-bit RISC-V Microcontroller on 65-nm Silicon-On-Thin-BOX (SOTB)	IEICE Electronics Express, Vol.VV, No.NN, 1-6, pp.1-6 (Impact Factor 0.788)	有	2020/10

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	須崎 有康	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	TEE を中心とする CPU セキュリティ機能の動向	一般社団法人 組込みシステム技術協会 (JASA)機関紙 Bulletin JASA	2019/4
2	荒川 文男, 池田 誠, 塚本 明, 須崎 有康	東京大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Trusted Execution Environment (TEE) with Open Processor Cores	MPSoC (Multicore and Multiprocessor SoCs) Forum 2019	2019/7
3	濱口, 柏山	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	「セキュアオープンアーキテクチャ技術基盤とその AI エッジ応用研究開発」のプロジェクト紹介	内閣サイバーセキュリティセンター (NISC)へのセキュリティ関連取り組み説明会	2019/9

番号	発表者	所属	タイトル	会議名	発表年月
4	須崎 有康	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	ASPLOS2019 参加報告	情報処理学会学会誌、情報処理	2019/9
5	須崎 有康	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	RISC-V TEE の開発	RISC-V Day Tokyo 2019	2019/9
6	柏山	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	「セキュアオープンアーキテクチャ技術基盤とその AI エッジ応用研究開発」のプロジェクト紹介	モバイルコンピューティング推進コンソーシアム(MCPC) AI&ロボット委員会講演	2019/10
7	須崎有康, 塚本 明, 小島一元, ホアン トロン ツック, 師尾 彬	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所	3 種類の TEE 比較 (Intel SGX, ARM TrustZone, RISC-V Keystone)	電子情報通信学会、ハードウェアセキュリティフォーラム 2019	2019/12
8	柏山	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	CASE クルマの Cyber Security 実装検討	ITS Japan 第 17 回 ITS シンポジウム 2019	2019/12
9	Kuniyasu Suzaki, Kenta Nakajima, Akira Tsukamoto, Tsukasa Oi	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所	WiP: Library Implementation and Performance Analysis of GlobalPlatform TEE Internal API for Intel SGX and RISC-V Keystone	The 19th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom 2020)	2020/12
10	須崎有康, 塚本 明, 小島一元, 中嶋健太, Hoang Trong Thuc, 師尾彬	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 電気通信大学, 東京大学	TEE比較	電子情報通信学会、暗号と情報セキュリティシンポジウム (SCIS2020)	2020/1
11	荒川文男, 池田 誠, 塚本明, 須崎 有康	東京大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	RISC-Vのセキュリティ仕様のローエンドプロセッサへの適用性の検討	電子情報通信学会、暗号と情報セキュリティシンポジウム (SCIS2020)	2020/1

番号	発表者	所属	タイトル	会議名	発表年月
12	Trong-Thuc Hoang, Akira Tsukamoto, Kazumoto Kojima, Kuniyasu Suzaki	電気通信大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Survey on Timing Side-Channel Attacks on Microarchitecture and Countermeasures	電子情報通信学会、暗号と情報セキュリティシンポジウム (SCIS2020)	2020/1
13	宮澤 慎一	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	プラットフォームセキュリティの技術の変遷	電子情報通信学会、暗号と情報セキュリティシンポジウム (SCIS2020)	2020/1
14	柏山	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Malware Cyber-Securityへの量子コンピューティング応用可能性に関する考察	モバイルコンピューティング推進コンソーシアム(MCPC) 技術解説書論文集「拡大する量子コンピューティングその社会実装ポテンシャル」	2020/3
15	Ckristian Duran, Trong-Thuc Hoang, Akira Tsukamoto, Kuniyasu Suzaki, Cong-Kha Pham	電気通信大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	TEE Boot Procedure with Crypto-accelerators in RISC-V Processors	Fourth Workshop on Computer Architecture Research with RISC-V (CARRV 2020)	2020/5
16	Anh-Tien Le, Ba-Anh Dao, Kuniyasu Suzaki, Cong-Kha Pham	電気通信大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Experiment on Replication of Side Channel Attack via Cache of RISC-V Berkeley Out-of-order Machine (BOOM) Implemented on FPGA	Fourth Workshop on Computer Architecture Research with RISC-V (CARRV 2020)	2020/5

番号	発表者	所属	タイトル	会議名	発表年月
17	Fumio Arakawa, Makoto Ikeda, Akira Tsukamoto, Kuniyasu Suzaki	東京大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Examination of applicability of RISC-V security specifications to low-end processors	IEEE first International Workshop on Secure RISC-V	2020/8
18	Ba-Anh Dao, Anh-Tien Le, Trong-Thuc Hoang, Akira Tsukamoto, Kuniyasu Suzaki, Cong-Kha Pham	電気通信大学, 産業技術総合研究所, セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	Dynamic Frequency Scaling as a countermeasure against simple power analysis attack in RISC-V processors	IEEE first International Workshop on Secure RISC-V	2020/8
19	Akira Tsukamoto, Kuniyasu Suzaki	産業技術総合研究所 セキュアオープンアーキテクチャ・エッジ基盤技術研究組合	TEEP (Trusted Execution Environment Provisioning) on RISC-V	RISC-V Global Forum	2020/9
20	Kenta Nakajima, Kuniyasu Suzaki	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所	Portable Implementation of GlobalPlatform API for TEE	RISC-V Global Forum	2020/9
21	Kuniyasu Suzaki, Kenta Nakajima, Tsukasa Oi, Akira Tsukamoto	セキュアオープンアーキテクチャ・エッジ基盤技術研究組合, 産業技術総合研究所	WiP: Performance and Behavior of Trusted Execution Environment and Rich Execution Environment on ARM, X86-64, and RISC-V	Hardware and Architectural Support for Security and Privacy (HASP) 2020	2020/10

番号	発表者	所属	タイトル	会議名	発表年月
22	須崎 有康, 中嶋 健太, 大居 司, 永田 貴彦, 菊池 正史, 塚本 明, 宮澤 慎一, 磯部 光平, 伊藤 大輔, 木村 貞弘, 安達 浩次, 高橋 睦史	産業技術総合研究所 セキュアオープンアーキテクチャ・エッジ基盤技術 研究組合, セコム株式 会社, 株式会社エヌエ スアイテクス 電気通信大学,	RISC-V TEEを強化するためのSecure CoProcessorとそれを活用 するソフトウェア	コンピュータセキュリティシンポジ ウム 2020	2020/10
23	須崎 有康	セキュアオープンアーキテクチャ・エッジ基盤技術 研究組合, 産業技術 総合研究所	招待講演: 「TEE (Trusted Execution Environment)は第二 の仮想化技術になるか?	情報処理学会 第 32 回コンピュータシステム・ シンポジウム (ComSys2020)	2020/12
24	Akira Tsukamoto, Kuniyasu Suzaki	産業技術総合研究所 セキュアオープンアーキテクチャ・エッジ基盤技術 研究組合	TEEP (Trusted Execution Environment Provisioning) Implementation on RISC-V Keystone and Arm TrustZone	Open Source Forum Japan	2020/12

(b) 新聞・雑誌等への掲載

番号	所属	発表内容	発表先	発表年月
1	セキュアオープンアーキテクチャ・エッジ 基盤技術研究組合	広報文タイトル: RISC-V の上に オープンなセキュリティ基盤を構築す るための技術研究組合 「TRASIO」が活動開始		2019/11/6

◎研究開発テーマ「AI エッジデバイスの横断的なセキュリティ評価に必要な基盤技術の研究開発」

【特許】

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
1	産業技術総合研究所	特願 2019-190852	国内	2019/10/8	出願済	識別補助データ生成技術及び識別情報抽出技術	古原
2	産業技術総合研究所, 早稲田大学	特願 2020-176747	国内	2020/10/21	出願済	個体識別子の付与方法および個体識別子を有する対象物	法元, 島本, 水野, 他

【論文】

番号	発表者	所属	タイトル	発表誌名、ページ番号	査読	発表年月
1	吉田, 松本, 他	横浜国立大学	白色干渉計を読取装置とした光学的ナノ人工物メトリクス	レーザー研究, 第 47 巻, 第 6 号, pp.295-299	有	2019/6
2	野高, 水野, 他	早稲田大学, 他	QFP/quartz adhesive bonding with surface treatment for physical security of edge artificial intelligence devices	Transactions of The Japan Institute of Electronics, Packaging Vol.13	有	2020/12 (採録決定)

【外部発表】

(a) 学会発表・講演

番号	発表者	所属	タイトル	会議名	発表年月
1	呂, 葛西, 他	北海道大学, 他	ナノ人工物メトリクスのためのレジスト倒壊ランダムパターン形成と評価	第 6 6 回応用物理学会春季学術講演会	2019/3
2	葛西, 他	北海道大学, 他	ナノ人工物メトリクスのための 2 次元ランダム構造形成プロセス最適化と電氣的読出しの実験的検討	電子情報通信学会 SDM ED CPM 合同研究会	2019/5
3	吉田, 松本, 他	横浜国立大学	複数の受光素子を用いたパルス方式測距 LIDAR の計測セキュリティ	電子情報通信学会技術研究報告, vol.119, no.143, HWS2019-36	2019/7
4	葛西, 他	北海道大学, 他	Formation and Characterization of 2D Random Si Nano-Pattern Using Resist Collapse for Nano-Artifact Metrics	2019 Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices (AWAD 2019)	2019/7
5	豎, 他	九州大学	ナノ人工物メトリクスのための白色干渉画像群の取得と解析	第 20 回情報フォトニクス研究グループ研究会	2019/9

番号	発表者	所属	タイトル	会議名	発表年月
6	葛西, 他	北海道大学, 他	[招待講演]ナノ人工物メトリクスを実現するランダムナノ構造形成と電気的読み出し技術	電子情報通信学会シリコン材料デバイス研究	2019/10
7	豎, 他	九州大学	Evaluation of original nano-artifact images obtained by a simplified optical equipment based on experimentally calculated FMR/FNMR	The 4th Asian Applied Physics Conference	2019/11
8	豎, 他	九州大学	Security performance of nano-artifact metrics based on customized white light interferometric system	The 9th Korea-Japan Workshop on Digital Holography and Information Photonics (DHIP2019)	2019/12
9	鈴木, 他	三菱電機株式会社	自律システムを想定したセンサ攻撃シミュレータの開発	2020 年暗号と情報セキュリティシンポジウム(SCIS2020)	2020/1
10	梨本, 他	三菱電機株式会社	ファスト FMCW 方式の MIMO レーダの攻撃評価シミュレーション	2020 年暗号と情報セキュリティシンポジウム(SCIS2020)	2020/1
11	松本, 他	横浜国立大学	超音波センサに対する攻撃を分析する音響 シミュレータ	2020 年暗号と情報セキュリティシンポジウム(SCIS2020)	2020/1
12	吉田, 松本, 他	横浜国立大学	局所特徴量に基づく 1 対 N 認証人工物メトリックシステム	2020 年暗号と情報セキュリティシンポジウム(SCIS2020)	2020/1
13	宮園, 吉田, 松本, 他	横浜国立大学	車線検出機能に対する色調改変攻撃	電子情報通信学会技術研究報告, vol.119, no.444, HWS2019-113	2020/3
14	泉田	株式会社 IJ インベーションインスティテュート	解析対象への前提知識を必要としないバイナリプログラム解析技術	Internet Infrastructure Review (IIR), vol.46, pp.24-31	2020/3
15	豎, 他	九州大学	高度物理セキュリティのための白色干渉像取得系が示す 個体認証性	第 81 回応用物理学会秋季学術講演会	2020/9
16	葛西, 他	北海道大学	ナノ人工物メトリクスのためのナノ構造埋込 Si MOSFET の試作と評価	第 81 回応用物理学会秋季学術講演会	2020/9
17	永田, 他	神戸大学	オンチップ擾乱検知に向けた SAR ADC 搭載 IC チップの評価	2020 年電子情報通信学会ソサイエティ大会	2020/9
18	永田, 他	神戸大学	A Dual-mode SAR ADC Enabling On-chip Detection of Offchip Power Noise Measurements by Attackers	International Conference on Solid State Devices and Materials (SSDM 2020)	2020/9
19	松本, 他	横浜国立大学	超音波センサに対するレーザー利用攻撃	電子情報通信学会技術研究報告, vol. 120, no. 211 (HWS2020-28)	2020/10

番号	発表者	所属	タイトル	会議名	発表年月
20	水野, 他	早稲田大学	5G 通信への応用可能な特性の実現を目指した LT/Quartz 複合 SAW 基板の作製	第 7 回電子デバイスフォーラム 京都 (招待講演)	2020/10
21	葛西, 他	北海道大学	Fabrication and characterization of nano-convexembedded Si MOSFET toward electrical nanostructure discrimination	33rd International Microprocesses and Nanotechnology Conference (MNC2020)	2020/11
22	水野, 他	早稲田大学	Short Note QFP/quartz adhesive bonding with surface treatment for physical security of edge artificial intelligence	2020 年 12 月 JIEP 英文誌 (Trans. JIEP)	2020/12

【受賞実績】

番号	発表者	所属	受賞種類	受賞内容	受賞年月
1	Renpeng Lu	北海道大学	2019 Asia-Pacific Workshop on Fundamentals and Applications of Advanced Semiconductor Devices (AWAD 2019)	Formation and Characterization of 2D Random Si Nano-Pattern Using Resist Collapse for Nano-Artifact Metrics	2019/3
2	梨本 翔永	三菱電機株式会社	2020 年暗号と情報セキュリティシンポジウム (SCIS2020) SCIS2020 論文賞	ファスト FMCW 方式の MIMO レーダの攻撃評価シミュレーション	2021/1 (予定)

以上