

「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技術の開発」

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
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 AI・ロボット部
-----	---

更新履歴

更新日	更新内容
2024年11月28日	初版発行

目次

概要（事業全体）	1
プロジェクト用語集	1
1. 意義・アウトカム（社会実装）達成までの道筋	1-1
1.1. 事業の位置づけ・意義	1-1
1.1.1. 事業実施の背景	1-1
1.1.2. 政策的位置づけ	1-1
1.1.3. NEDO が関与する意義	1-2
1.1.4. 本事業の位置づけ	1-3
1.2. アウトカム達成までの道筋	1-4
1.3. 知的財産・標準化戦略	1-4
2. 目標及び達成状況.....	2-1
2.1. アウトカム目標及び達成見込み	2-1
2.2. アウトプット目標及び達成状況	2-1
3. マネジメント.....	3-1
3.1. 実施体制	3-1
3.2. 受益者負担の考え方	3-7
3.3. 研究開発計画	3-8
4. 目標及び達成状況の詳細	4-1
4.1. 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」	4-1
4.1.1. テーマ概要	4-1
4.1.2. 最終目標と根拠	4-1
4.1.3. 目標の達成度	4-13
4.1.4. 成果の普及	4-14
4.2. 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」	4-15
4.2.1. テーマ概要	4-15
4.2.2. 最終目標と根拠	4-15
4.2.3. 目標の達成度	4-16
4.2.4. 成果と意義	4-17
4.2.5. 成果の普及	4-19
4.3. 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」	4-23
4.3.1. テーマ概要	4-23
4.3.2. 最終目標と根拠	4-28
4.3.3. 成果と意義	4-31

4.4.	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」	4-36
4.4.1.	テーマ概要	4-36
4.4.2.	最終目標と根拠	4-37
4.4.3.	目標の達成度	4-37
4.4.4.	成果と意義	4-38
4.4.5.	成果の普及	4-39
4.5.	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」	4-41
4.5.1.	テーマ概要	4-41
4.5.2.	最終目標と根拠	4-41
4.5.3.	目標の達成度	4-43
4.5.4.	成果と意義	4-43
4.5.5.	成果の普及	4-45
4.6.	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」	4-46
4.6.1.	テーマ概要	4-46
4.6.2.	最終目標と根拠	4-47
4.6.3.	目標の達成度	4-48
4.6.4.	成果と意義	4-48
4.6.5.	成果の普及	4-53
4.7.	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」	4-54
4.7.1.	テーマ概要	4-54
4.7.2.	最終目標と根拠	4-54
4.7.3.	目標の達成度	4-57
4.7.4.	成果の意義	4-60
4.7.5.	成果の普及	4-60
4.8.	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」	4-61
4.8.1.	テーマ概要	4-61
4.8.2.	最終目標と根拠	4-62
4.8.3.	目標の達成度	4-62
4.8.4.	成果と意義	4-62
4.8.5.	成果の普及	4-64
4.9.	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」	4-65
4.9.1.	テーマ概要	4-65
4.9.2.	最終目標と根拠	4-65
4.9.3.	目標の達成度	4-65
4.9.4.	成果と意義	4-66
4.9.5.	成果の普及	4-76

4.10. 研究開発テーマ「セキュアオープンアーキテクチャ向けコンパイラバックエンドおよび対応ランタイム環境の設計・開発」	4-78
4.10.1. テーマ概要	4-78
4.10.2. 最終目標と根拠.....	4-79
4.10.3. 目標の達成度	4-79
4.10.4. 成果と意義.....	4-80
4.10.5. 成果と普及.....	4-81
●特許論文等リスト	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
◎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」	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」	7
◎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」	9
◎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」	13
◎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」	17
◎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」	24
◎研究開発テーマ「セキュアオープンアーキテクチャ向けコンパイラバックエンドおよび対応ランタイム環境の設計・開発」	27

概要（事業全体）

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 (旧：IoT 推進のための横断技術開発プロジェクト) METI 予算要求名称	プロジェクト番号	P16007
担当推進部/ プロジェクトマネージャーまたは担当者 及び METI 担当課	AI・ロボット部 PMgr 遠藤勇徳 (2024年11月現在) AI・ロボット部 PMgr 岩佐匡浩 (2024年11月現在) IoT 推進部 PM: 伊藤隆夫 (2020年7月～2022年3月) 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 社会の到来が近付くにつれ、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」の開発を実施している。		
1. 意義・アウトカム (社会実装) 達成までの道筋			
1.1 本事業の位置 付け・意義	<p>情報革命によるインターネットの普及から、すべてのモノがインターネットに繋がる IoT 社会の到来、AI の普及に生成 AI の登場など、世界では情報の蓄積と活用に係るハードウェア技術とソフトウェア技術の重要性が日に日に増している。政府も AI 戦略や半導体・デジタル産業戦略など各種戦略・政策を策定し、必要な支援も講じている。こうした中で、情報社会を支える技術として、先の時代のインフラ構造や想定される技術仕様など、情報処理を支える中核的な基盤技術を定め、早期に研究開発を開始することが、後の我が国の産業競争力の強化に資することは明白である。</p> <p>本事業では、増加する一方である社会全体が扱う情報量の増加に伴い、端末(デバイス)から得られる大量データ(情報)の利活用により、高度な制御や新たなサービスを実現する「IoT 社会」を支える IoT 情報処理基盤の開発。(研究開発項目③: 2016～2020) IoT 社会の到来により、急増するデータを即時的・効率的に処理するため、情報処理はサーバー(クラウド)集約型から、エッジ分散処理型へシフトする、エッジコンピューティングの実現にかかる技術の開発。(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: 2018～2022fy) ムーア則の終焉が叫ばれ、半導体の性能向上が限界を迎えると予想される未来社会において、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の処理に要する電力の劇的な低減に資する技術の開発。(研究開発項目②次世代コンピューティング技術の開発: 2018～2027fy) それぞれの時間軸において、IoT 社会の高度化に資する、革新的な技術の実現に貢献する。</p>		
1.2 アウトカム達成 までの道筋	<p>現代社会においてはクラウドコンピューティングが主流となっているが、一方でエッジ(端末)側におけるリアルタイム処理はもちろん、AI を用いた高度な情報処理、オフライン状態での処理などエッジコンピューティングの需要が高まっている。本事業では、研究開発項目①において、エッジコンピューティングに必要な各種基盤技術を 11 のテーマを通じて開発、展示会やセミナーでの成果発表、コンテストの実施、報道発表などを通じて成果の普及段階にあり、今後エッジ領域での情報処理サービスが世の中に普及していく中で、本事業の成果の普及、アウトカムの達成も見込まれている。</p> <p>一方で、こうしたコンピューティング技術の基礎ともいえるハードウェア、半導体の性能向上は限界が叫ばれている。製造、設計技術の向上や新たなトレンドの登場で年々高性能な半導体が誕生しているが、それでも限界は確実に訪れる。次の時代に向けて、既存の技術の延長がなく、また既存の技術以上に高性能かつ省エネルギーな次世代のコンピューティング技術の開発を、本事業では研究開発項目②において実施している。次世代のトレンドとなりえ</p>		

	<p>る技術の“種“の育成を先導研究から探索型研究枠として着手し、そのうえで、量子コンピュータ、光分散コンピュータ、脳型コンピュータというクラウドからエッジまでの中核的な位置づけを成すことが期待される技術の開発を実施している。</p> <p>高度な情報処理技術の保有は、ある種の国家競争、外交カードにもなりえる時代が到来している。2024年時点では本項目は残り3年の開発期間を残す形となるが、今後は成果の普及を見据え、広報活動以外に、ユーザとして想定される企業と連携した実証実験やアドバイザリー会議の開催、国際標準取得などを行い、日本独自の技術による、産業競争力強化、国内外市場の獲得に資する、新たなサービスの創出を目指すことで、アウトカムを達成する。</p>	
<p>1.3 知的財産・標準化戦略</p>	<p>情報産業は大きな変革期にあり、新たなトレンドがたびたび生まれるだけでなく、世界規模での競争が起こっている。成果最大化のためにはただ成果を知財化するだけでなく、世界的な技術動向を把握したうえで、戦略的な取り組みを進めることが重要と言える。本事業では、2022年度から知財戦略に係る調査を研究開発項目①、②においてそれぞれ実施。具体的にはAIエッジ、次世代（量子、光分散、脳型などの各種コンピューティング技術）において知財の出願動向や競合すると考えられる企業の知財保有状況、出願動向を調査するとともに、それらから考えられる各事業者今後の知財戦略に関して提言を行い、成果のオープン・クローズ戦略含めてNEDOのマネジメントのもとで実施した。</p> <p>また標準化施策についても、得られた研究開発成果について、必要に応じて標準化を図ることとし、評価手法の提案、データの提供、標準化活動等を事業内において実施した。</p> <p>なお、本プロジェクトは、「NEDOプロジェクトにおける知財マネジメント基本方針」を適用している。</p>	
<p>2. 目標及び達成状況</p>		
<p>2.1 アウトカム目標及び達成見込み</p>	<p>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。</p> <p>現在、開発状況、成果の達成状況は概ね順調であり、本アウトカムについては成果の普及とともに達成出来る見通し。</p>	
<p>2.2 アウトプット目標及び達成状況</p>	<p>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。（電力効率で10倍～100倍以上を達成。）</p> <p>研究開発項目① 事業期間：2018～2022年度 目標：電力効率10倍以上を達成。 達成見込み：すべてのテーマで当初の目標を達成。</p> <p>研究開発項目② 事業期間：2018～最長2027年度 目標：電力効率100倍以上を達成。 達成見込み：中間目標としては全てのテーマで達成見通し。</p>	
<p>3. マネジメント</p>		
<p>3.1 実施体制</p>	<p>プロジェクトマネージャー</p>	<p>AI・ロボット部 PMgr 遠藤勇徳（2024年11月現在） AI・ロボット部 PMgr 岩佐匡浩（2024年11月現在）</p>
	<p>プロジェクトリーダー</p>	<p>研究開発項目① 東京科学大学 教授 本村真人 研究開発項目② 産業技術総合研究所 特別顧問 金山敏彦</p>
	<p>委託先 ※2021～2024年度</p>	<p>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</p>

		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、アラクサラネットワークス(株)、等				
3.2 受益者負担の考え方	受益者負担の考え方： 本事業は委託事業であり、次世代の情報産業を担う基盤技術かつハイリスクハイリターンな技術の開発として国が支援するものであり、100%NEDO 負担で実施する。					
	主な実施事項	2021fy	2022fy	2023fy	2024fy	2025fy
	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研究 開発	委託	委託	-	-	-
	研究開発項目② 次世代コン ピューティング 技術の研究開発	委託	委託	委託	委託	委託
	その他調査など	委託	委託	委託	委託	-
3.3 研究開発計画						
事業費推移 [単位:百万円]	主な実施事項	2021fy	2022fy	2023fy	2024fy	総額
	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 研究開発	3,669	2,495	-	-	6,164
	研究開発項目② 次世代コン ピューティング 技術の研究開発	5,729	6,739	4,924	5,598	22,990
	その他 技術動向調査な ど	290	222	20	53	585
	事業費	2021fy	2022fy	2023fy	2024fy	総額
	会計(特別)	9,980	10,046	4,900	4,800	29,726
	追加予算・繰り 越しなど	1,648	1,916	725	798	5,087
	総 NEDO 負担額	11,628	11,962	5625	5,598	34,813
情勢変化への対応	<p><u>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</u> 昨今の AI を用いたコンピューティング技術に関する情勢は日々変化していると言っても過言ではない。中でも大きな情勢変化としては、半導体開発・生産に伴う世界的な動向の変化があげられる。政府は 2021 年 3 月から半導体・デジタル産業総合戦略に係る検討会を開催、当該戦略を発表するとともに、半導体の製造を始めとしたデジタルインフラの整備等に係る大規模な政策を実行している。これにより、将来的にはエッジコンピューティングを支える高性能な半導体の普及や、日本としての開発力、生産力の強化が期待できる状況となった。</p> <p>NEDO としても 2021 年度から「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に係る技術動向調査」及び「AI エッジコンピューティング技術領域に係る知財戦略調査」を実施。国内外の技術動向や今後の開発目標、事業化戦略にかかる情報収集を実施するなどし、政策の動向と合わせてマネジメントに活用することで情勢変化に対応している。</p>					
	<p><u>研究開発項目②次世代コンピューティング技術の開発</u> 10 年以上先の社会における、既存の技術の延長にないコンピューティング技術の開発を推進する本項目では、どのような技術に可能性、勝ち筋が生じるかの見極めが重要と考えられ</p>					

	<p>る。そこで当該項目においても開発項目①と同様に技術動向調査、知財動向調査を実施している。また、学術的には有効性が証明されているが、産業応用に向けてはまだ情報収集や検証が必要となる技術に関するフォローとして探索型研究枠（2018～2022年度）を実施。調査と探索研究の成果から今後必要性が高いと考えられる技術として</p> <ul style="list-style-type: none"> ・量子コンピューティング技術 ・光分散コンピューティング技術 ・脳型コンピューティング技術 <p>の3領域を定め、2023年度から集中的な開発を実施している。</p> <p>また、2023年度以降も、それぞれの領域においても継続して調査を実施し、刻々と変化する技術情勢の中で本事業の成果が勝ち筋を見いだせるよう、リソースの選択と集中、あるいは開発項目の新規追加などを実施している。</p>	
中間評価結果への対応	<p>2020年度に実施された第1回中間評価では、 事業の位置づけ・必要性：評価A（非常に重要） 研究開発マネジメント：評価B（良い） 研究開発成果：評価B（良い） 成果実用化に向けた取り組み：評価C（概ね妥当） とする評価を受けた。</p> <p>評価コメントとしては、事業の位置づけやNEDOのマネジメント、開発成果については評価を得たものの、全体を通してNEDOとして世界的な技術潮流に関する情報収集や、市場獲得に向けた方策を検討、拡充する必要があることが求められた。また、次世代のコンピューティング技術の開発に関しては、人材育成やユーザーとなりうる事業主体とのコミュニケーションの場を形成する必用についても求められた。</p> <p><u>評価結果への対応：</u> 世界的な技術潮流に関する情報収集としては、2021年度から技術動向調査、及び知財戦略調査を研究開発項目①、②ともに実施。得られた情報を事業者にフィードバックするとともに、NEDOとしてもマネジメントを実施した。加えて、研究開発項目①については、産業応用を推進する方策として、調査結果と中間評価の結果に基づき「省エネ AI 半導体及びシステムに関する技術開発」を新政策として立ち上げ、2022年度からの助成事業で研究成果の実用化・事業化を推進している。</p> <p>また、研究開発項目②については、人材育成としては開発した量子コンピュータ共通ソフトウェアをクラウド環境で無償公開するだけでなく、コンテストを開催するなどして先端技術に触れることでの人材の育成に貢献した。さらに、脳型コンピューティングでは開発した脳型チップ、アルゴリズムを活用することが想定される、まさにユーザーとなりうる事業者を集めたアドバイザリーボードを形成し、開発者から成果の共有、意見交換を行うことで事業化にかかる橋渡しを実施している。なお、光分散コンピューティングにおいては2025年度から100km圏内のデータセンタによる分散情報処理の実証実験を実施し、成果を公開することで早期の事業化に繋げる。</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	2020年度 中間評価実施
	終了時評価	2028年度 終了時評価実施予定
別添		
投稿論文	515件	
特許	349件 ※全事業期間	
その他の外部発表（プレス発表等）	<p>以下、NEDOとしてニュースリリースを実施。</p> <ul style="list-style-type: none"> ・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）） ニュースリリース実施日：2022/12/8 ・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）エヌエスアイテクス） ニュースリリース実施日：2023/6/8 ・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（KDDI（株）） ニュースリリース実施日：2023/3/22 	

	<p>・セキュアオープンアーキテクチャ向けコンパイラバックエンドおよび対応ランタイム環境の設計・開発（イーソル(株)） ニュースリリース実施日：2023/4/27</p> <p>・実社会の事象をリアルタイム処理可能な次世代データ処理基盤技術の研究開発（ノーチラステクノロジーズ(株)） ニュースリリース実施日：2023/7/11</p> <p>他、CEATEC 展等各種展示会に毎年度出展し、各事業者の成果を発表。CEATEC2024 では量子コンピューティング及び脳型コンピューティングの成果展示を実施。</p>	
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>(1) 2016 年 3 月、制定。</p> <p>(2) 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>(3) 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>(4) 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> <p>(5) 2020 年 12 月、研究開発項目②の分割、指名した PL・最新政策等情報の反映。ステージゲート審査に関する整理の追記。研究開発項目③の事業期間延期等による改訂。</p> <p>(6) 2021 年 12 月、研究開発項目④の追加による改訂。ステージゲート審査に関する整理の追記。</p> <p>(7) 2022 年 5 月、プロジェクトマネージャーの体制変更に伴う改訂。</p> <p>(8) 2023 年 2 月、研究開発項目④の移管、研究開発項目①の事業期間延期等による改訂。</p> <p>(9) 2024 年 7 月、NEDO 組織再編に伴う担当部署名変更に伴う改訂。</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。
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。
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。
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。
ECU	各種電子制御をするユニット。
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。
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。

用語	説明
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する、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。

用語	説明
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。

用語	説明
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イッチや情報を不揮発に保持するメモリに利用。

用語	説明
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。

用語	説明
複数深層学習	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。 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。
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. 意義・アウトカム（社会実装）達成までの道筋

1.1. 事業の位置づけ・意義

1.1.1. 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。

1.1.2. 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、コロナ禍、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。

1.1.3. 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。

1.1.4. 本事業の位置づけ

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。

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利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）リコンフィギャ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 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 ※1 沖 ※1 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 ※1		
			(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		

※1：2020年度終了
※2：2021年度終了

「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（2023年3月時点）

1.2. アウトカム達成までの道筋

現代社会においてはクラウドコンピューティングが主流となっているが、一方でエッジ（端末）側におけるリアルタイム処理はもちろん、AIを用いた高度な情報処理、オフライン状態での処理などエッジコンピューティングの需要が高まっている。本事業では、研究開発項目①において、エッジコンピューティングに必要となる各種基盤技術を13のテーマを通じて開発、展示会やセミナーでの成果発表、コンテストの実施、報道発表などを通じて成果の普及段階にあり、今後エッジ領域での情報処理サービスが世の中に普及していく中で、本事業の成果の普及、アウトカムの達成も見込まれている。

1.3. 知的財産・標準化戦略

情報産業は大きな変革期にあり、新たなトレンドがたびたび生まれるだけでなく、世界規模での競争が起こっている。成果最大化のためにはただ成果を知財化するだけでなく、世界的な技術動向を把握したうえで、戦略的な取り組みを進めることが重要と言える。本事業では、2022年度から知財戦略に係る調査を研究開発項目①、②においてそれぞれ実施。具体的にはAIエッジ、次世代（量子、光分散、脳型）において知財の出願動向や競合すると考えられる企業の知財保有状況、出願動向を調査するとともに、それらから考えられる各事業者に今後の知財戦略に関して提言を行い、成果のオープン・クローズ戦略を含めてNEDOのマネジメントのもとで実施した。

また標準化施策についても、得られた研究開発成果について、必要に応じて標準化を図ることとし、評価手法の提案、データの提供、標準化活動等を事業内において実施した。

なお、本プロジェクトは、「NEDOプロジェクトにおける知財マネジメント基本方針」を適用している。

2. 目標及び達成状況

2.1. アウトカム目標及び達成見込み

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。

現在、開発状況、成果の達成状況は概ね順調であり、本アウトカムについては成果の普及とともに達成出来る見通し。

2.2. アウトプット目標及び達成状況

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。

達成状況

以下に各テーマの目標達成状況を示す。

開発課題		研究開発テーマ 代表事業者	達成度 (中間評価)	達成度 (終了評価)
(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	○	○
(B)	リコンフィギャ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	△ 一部実施項目で遅延が発生	○ 一部実施項目で遅延が発生したが目標を達成
(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	ソシオネクスト	◎ 一部実施項目を前倒しで完了	◎ 一部実施項目を前倒しで完了
		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	○	○
(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	○	○
(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	TRASIO	○	○
		産総研	○	○
		イーソル	○	○

3. マネジメント

3.1. 実施体制

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。

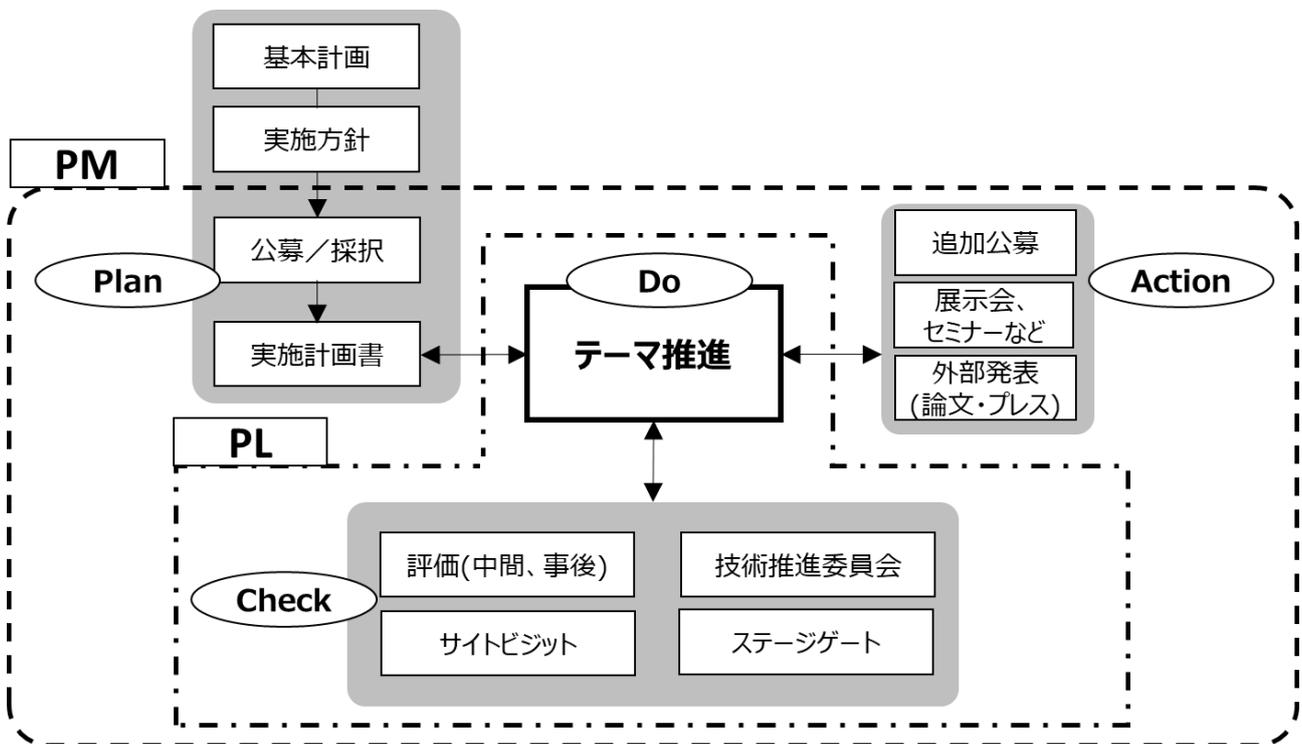


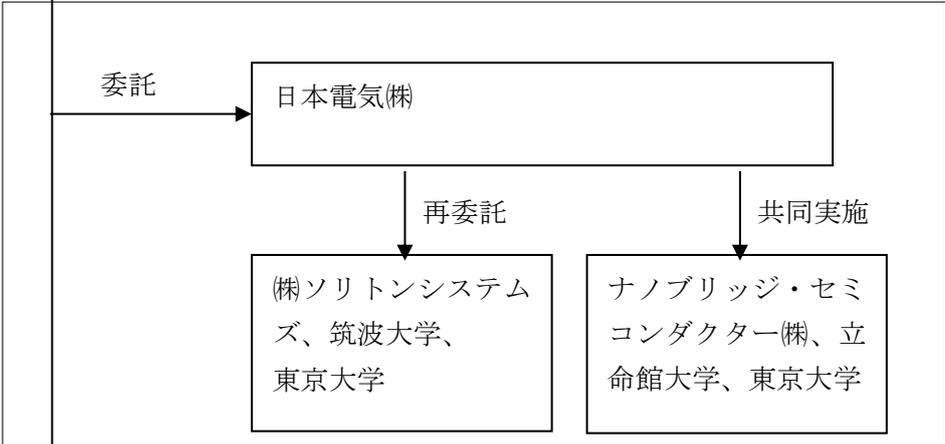
図 3.1.1 マネジメント体制

以下に各テーマの実施体制の詳細を示す。

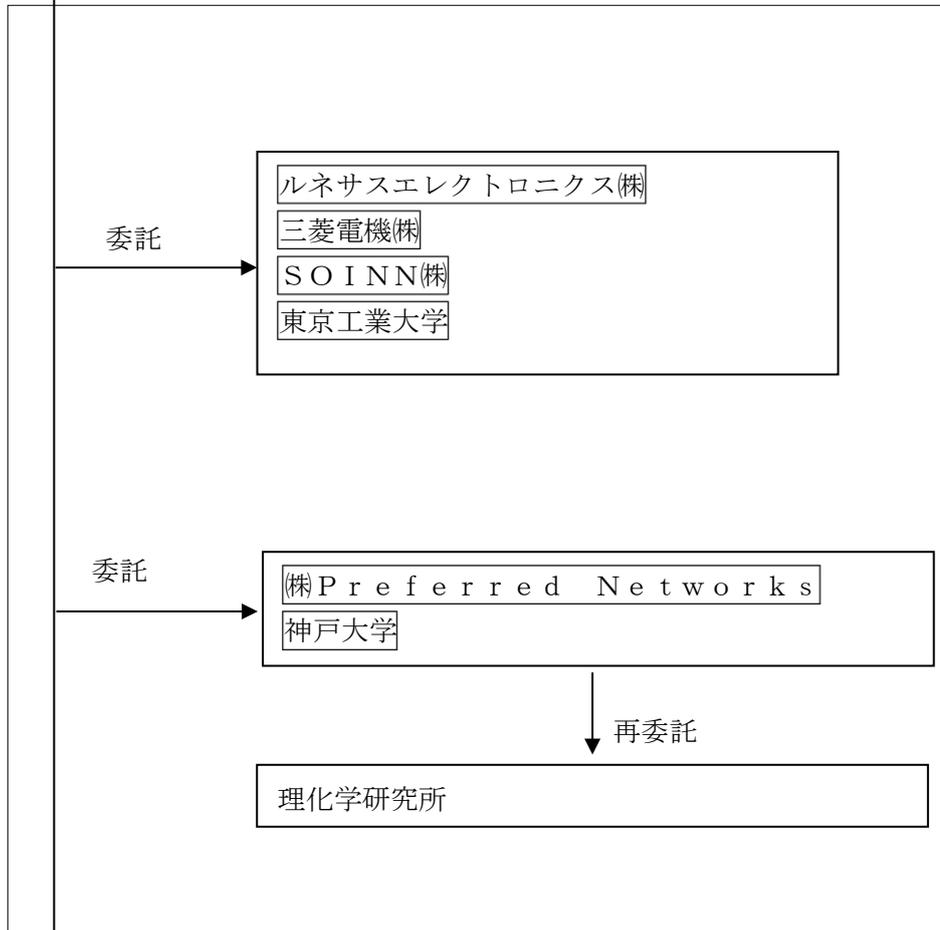
NEDO

プロジェクトリーダー
国立大学法人東京科学大学
本村 真人 教授

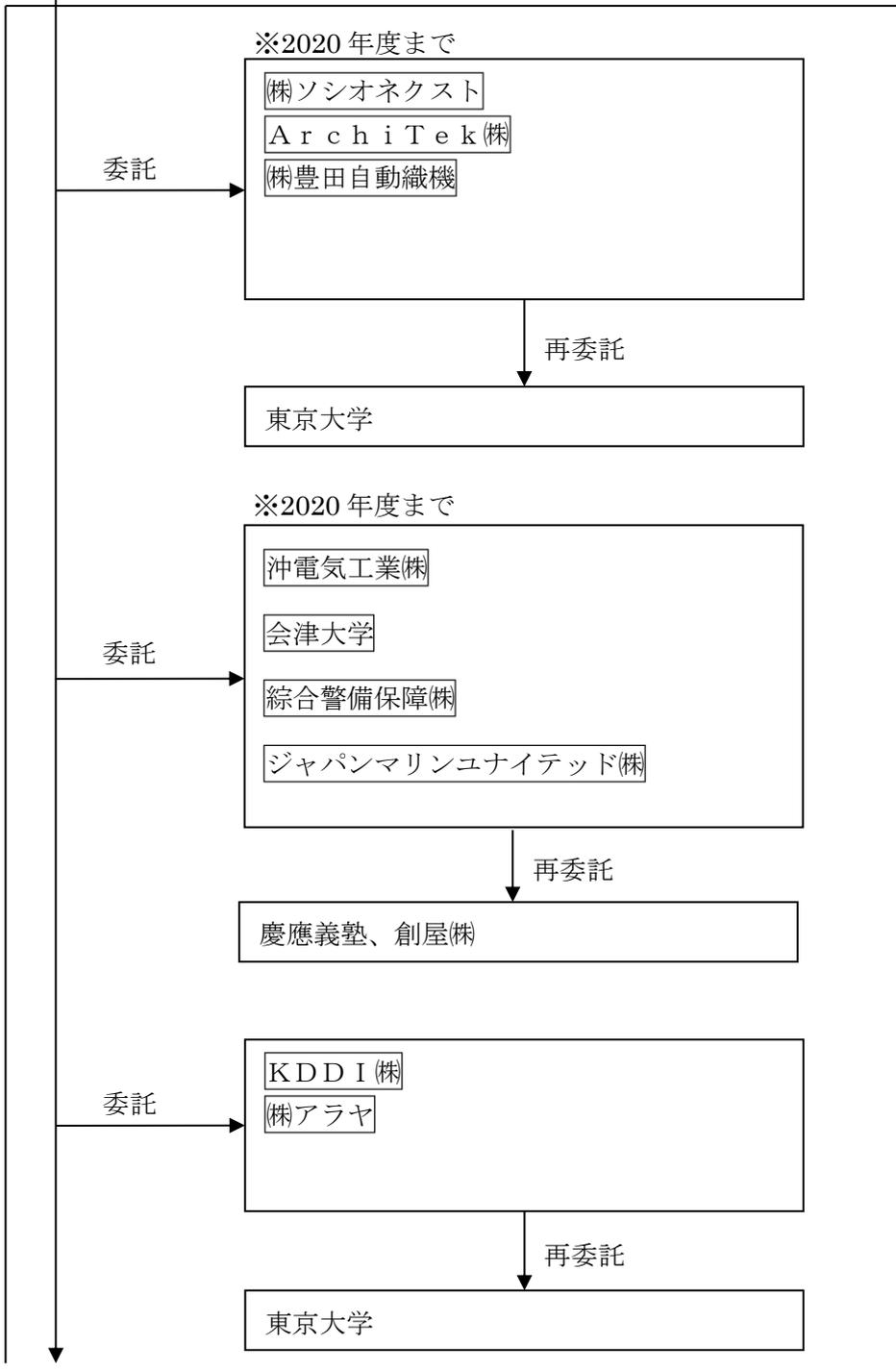
A) 不揮発性素子等のスイッチング機構を用いたコンピューティング技術 ※2021年度まで

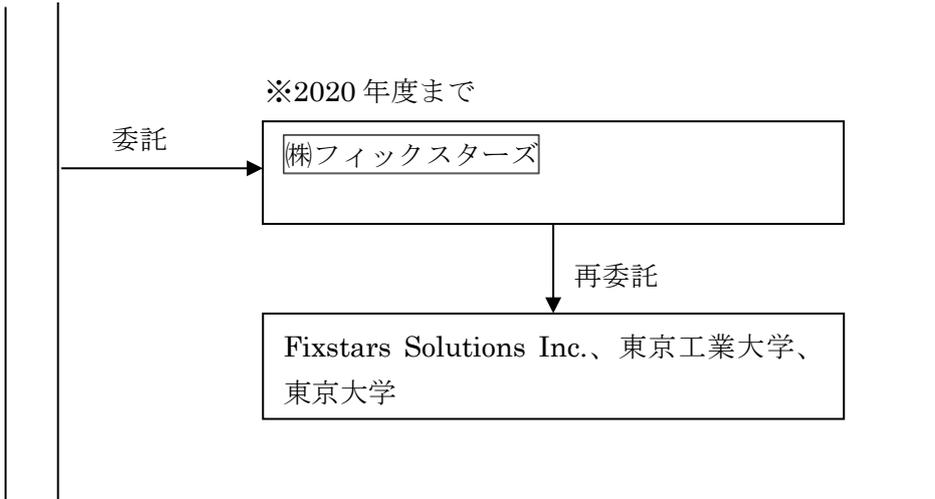


B) リコンフィギャラブルデバイスによるコンピューティング技術

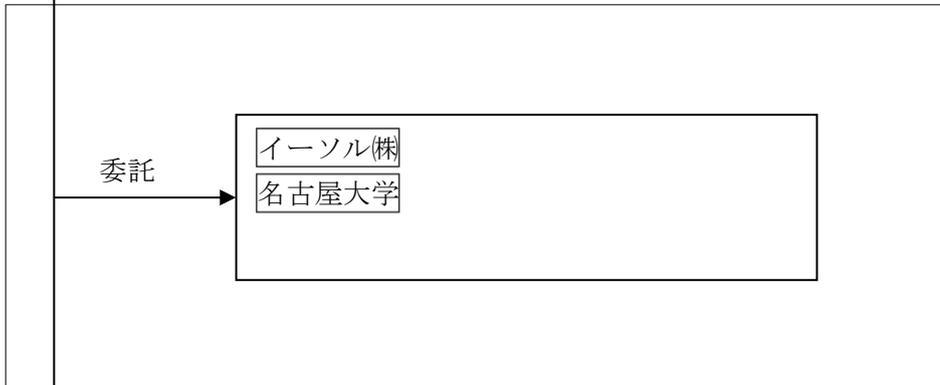


C) 演算処理量の軽量化を実現する AI 組み込みコンピューティング技術

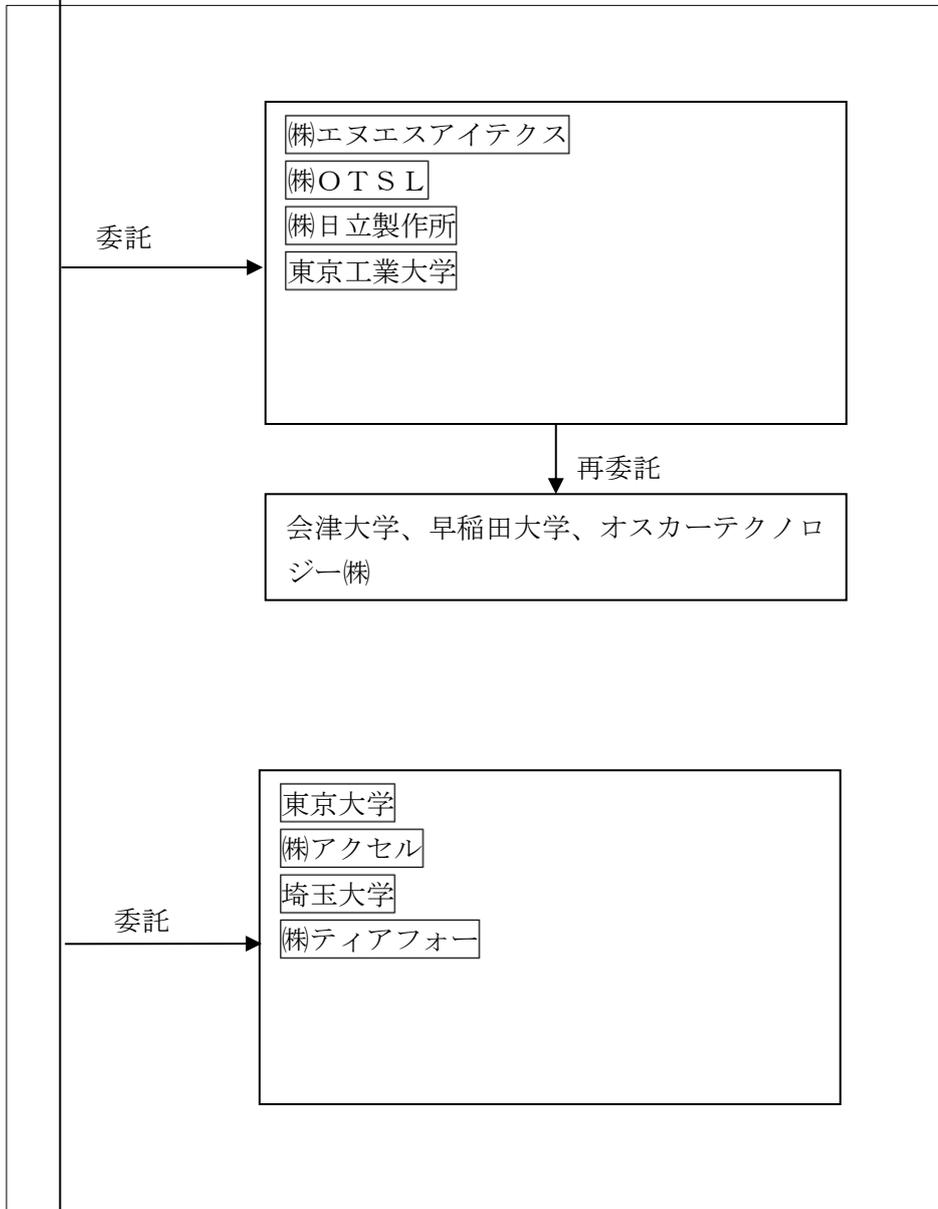




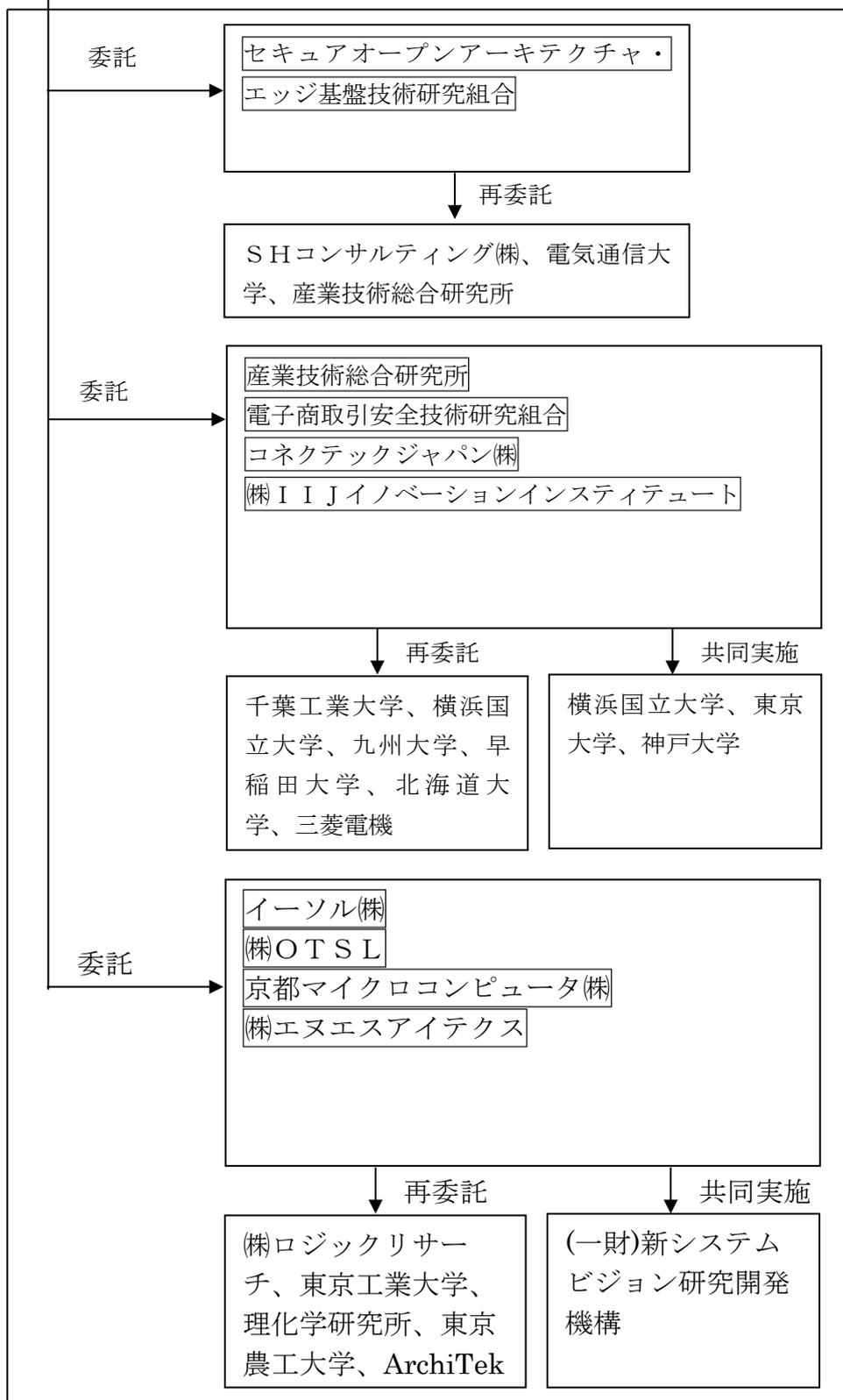
D) エッジコンピューティング向けリアルタイムソフトウェア制御技術 ※2020 年度まで



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



3.2. 受益者負担の考え方

本事業は、次世代の情報産業を担う基盤技術かつハイリスクハイリターンな技術の開発として国が支援するものであり、100%NEDO 負担の委託事業として実施する。

3.3. 研究開発計画

以下に研究開発スケジュールを記載する。



【2021年度】

以下の5課題に関するテーマについて、各テーマ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倍達成に向けて外部委員等による助言等を行った。

(課題)

- 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
- B) リコンフィギャラブルデバイスによるコンピューティング技術
- 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
- 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

【2022年度】

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

以下の 5 課題に関するテーマについて、最終年度となる研究開発を推進し、加速の検討、成果の取りまとめや情報発信等を実施した。

(課題)

- 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
- B) リコンフィギャラブルデバイスによるコンピューティング技術
- 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
- 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

4. 目標及び達成状況の詳細

4.1. 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」

4.1.1. テーマ概要

本研究開発では、学習はクラウド上で行い、推論をエッジで行う事を想定している。エッジで行う推論処理は、エッジサーバと IoT 端末の連携で行うものと、IoT 端末のみで行うものの 2 種類が存在する(図 4.1.1.1)。連携処理の場合は、IoT 端末に求められる AI 処理は前処理等であり、一般的に軽いものとなるが、一方で、収集すべきデータは端末に特化したものとなるため、端末へのプログラマビリティへの要求が高く、端末での AI 処理は FPGA でのみ行うことが想定される。このような連携処理の例としては映像や音声の配信端末への適用検討を行う。

本研究開発における ASSoC 開発は 2019 年度から 2021 年度の 3 カ年の間に 1 回の試作で完了する。2020 年度に ASSoC の設計を完了し、2021 年度にチップ製造・チップ評価を実施する。ASSoC の評価は日本電気株式会社およびナノブリッジ・セミコンダクター株式会社において実施する。筑波大学、東京大学稲葉研究室、ソリトンシステムズ株式会社においては、ASSoC チップに AI 処理を含むアプリケーションを実装し、各ユースケースにおいてその低電力性を評価する。ASSoC のパイロットユーザを開拓するための活動を 2021 年度に実施し、上市用チップの仕様検討や必要な CAD 製品の検討を行う。また、本プロジェクトにおける共同研究先として、東京大学塩谷研究室と連携して CPU コアとして RISC-V プロセッサの適用を検討する。同じく、立命館大学越智研究室と連携して、原子スイッチ FPGA に搭載する演算器の設計およびツールフローによる演算器利用について開発を行う。

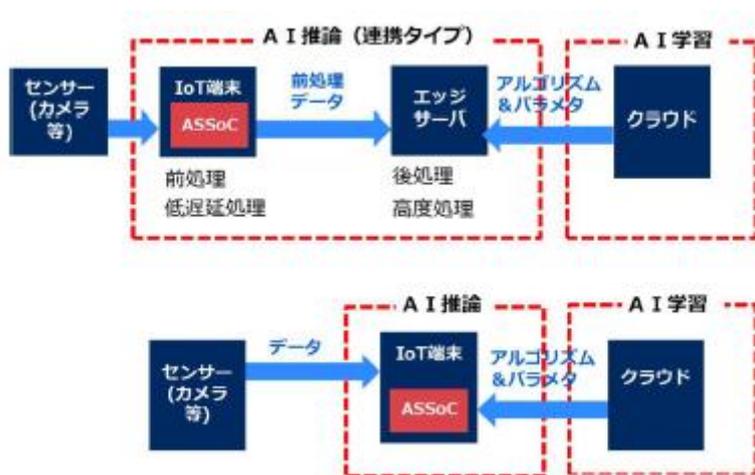


図 4.1.1.1 エッジ推論の 2 種類の形態、上、連携型、下、単独型

4.1.2. 最終目標と根拠

AI 処理用のコンピュータには、大量のデータの並列処理に向いている GPU が広く利用されている。CPU 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（ハードリアルタイム化）を満足することは専用設計の回路には及ばない。低遅延化の要求が強いアプリケー

ション領域（自動車、ロボット、介護介助、ドローンなど）への対応として、ハードウェア回路自体をプログラマブル化した FPGA に、AI 処理を行わせる試みが進んでいる。これは、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、低遅延で省電力な回路実装が可能だからである。

このような現状において、以下の課題がある。

- ・現状の最先端 FPGA は揮発性の 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 よりは低消費電力であるものの、数 W から数 10W 以上の消費電力を発生し、極めて消費電力制限の厳しい IoT 端末やエッジコンピュータには適用が難しい。
- ・比較的小規模・低電力な FPGA コア、メモリ、低電力 CPU、演算コア等を内蔵した SoC が近年製品化されており、処理中のデータは SoC 内で処理することにより、システムとしての省エネルギー化（一般的に 1/2 程度）が実現できているものの、電池駆動のデバイスでは十分ではない。
- ・半導体は世界情勢に依存した戦略的な部品でもあり、FPGA を含む SoC はキーデバイスとみなされているが、FPGA は米国の寡占状態であり、FPGA の利用技術は別として、チップの研究開発側における日本の存在感は薄かった。

FPGA の重要性は最近、認識されて始めており、2018 年度に打上げ予定の革新的衛星技術実証 1 号機では、日本発の革新的 FPGA として、原子スイッチ FPGA の軌道上評価が行われる。日本全体の競争力に資するためには、日本が競争力を持つ応用分野に適用するための FPGA 関連技術を開発することが必要である。この必要性に対して、本プロジェクトの研究開発では、原子スイッチ FPGA 技術を用いた SoC の開発・製造、チップ実証を実施するものである。

実施項目①においては、原子スイッチ素子を FPGA および CPU の動作に必要なコード ROM として混載可能な製造技術に関する開発を実施する。実施項目②-1 では、SoC 開発に必要な要素回路ライブラリの開発、SoC の統合設計・評価、ツール開発、開発されたチップの適用先拡大についての各項目を実施し、試作されたチップにおいて、電力効率 10 倍を達成する見通しを得る。

実施項目②-2 では、再委託先（株式会社ソリトンシステムズ）において、監視・見守り分野での利用を想定した映像音声識別 AI の実装評価を実施する。

実施項目②-3 では、再委託先（筑波大学）において、農業向けドローンを想定した実装評価を実施する。

実施項目②-4 では、再委託先（東京大学）において、ヒューマノイド型ロボットおよびロボット義足を想定した実装評価を実施する。

以下に、各項目の目標について説明する。

実施項目① 「AI 向け省エネ FPGA コア及び省エネメモリコア開発」

(a)省電力書換えが可能な原子スイッチの製造技術、(b)原子スイッチメモリコアの低電力回路技術、さらに(c)28nm 世代における FPGA コアおよびメモリコアの混載を可能とする製造技術の開発を行う。以下に最終目標および、根拠を説明する。

①-(a)「省エネルギー書換え原子スイッチ製造技術」(担当：日本電気株式会社)

電池駆動の IoT 端末上での書換えを考慮すると書込み電圧、および書込み時間を短縮して、書込みのエネルギーを低減する必要がある。現時点での FPGA 向けの原子スイッチの書き込みエネルギーは 14nJ であり、このエネルギーを 1/10 以下(0.5nJ 程度)に低減する素子技術を開発する。

原子スイッチの構成要素の中では、書込み特性への影響が最も大きい固体電解質の膜質の改良と加工プロセスを改良する。書込み特性値のばらつきの最悪値がチップの特性を決めてしまうため、ばらつきをいかに抑えるかも改良点である。2019 年に原子スイッチ単体の開発を完了、2020 年 9 月にアレイでの評価を完了し、書き込みエネルギー従来比 1/10 以下(0.5nJ 程度)を達成する。その結果製造技術について 2021 年 9 月に確定する。

①-(b)「28nm メモリコア回路技術」(担当：ナノブリッジ・セミコンダクター株式会社)

本実施項目では、ファウンドリで得られる標準的な CMOS を用いた原子スイッチメモリコアを開発する。動作時電力については、メモリコアでは読み出し時の電力が重要であり、ビット当たりの消費エネルギー 0.1pJ/bit を実現することを目標とする。28nm 世代で最も低電力な NOR-FLASH をメモリマクロ(Cypress 社製)と比較すると 1/10 以下の電力にあたる。

2020 年にメモリコア回路の設計を完了、2021 年 9 月に評価を完了し、従来のメモリコアと比較して 1 ビットあたりの読み出し電力を従来比 1/10 以下(0.1pJ/bit 程度)とする。

①-(c)「ON 抵抗混在安定化原子スイッチ製造技術」(担当：日本電気株式会社)

本実施項目では、28nm 世代においてメモリコアと FPGA コアの双方の混載を可能とする製造プロセスを開発する。メモリコアと FPGA コアでは要求される ON/OFF 比が異なるため、書込み条件も異なる。違った条件で書き込んだ素子でも、異なる ON/OFF 比が安定して確保できるような製造技術を開発する。

2021 年 9 月までに 28nm メモリコア向け原子スイッチの ON 抵抗安定性を確保し、下地 CMOS と多層配線の特性シフトがないことを確認する。2021 年度において、ASSoC のチップを製造し、原子スイッチの歩留まりが問題ないことを確認する(2022 年 2 月)。

実施項目②「端末-エッジ連携 AI 処理用 AI チップ(ASSoC)の開発」(担当：ナノブリッジ・セミコンダクター株式会社、東京大学、立命館大学)

本実施項目②において、2021 年 1 月以降はナノブリッジ・セミコンダクター株式会社、東京大学および立命館大学が ASSoC の回路およびツールの開発を行う。さらに、筑波大学、ソリトンシステムズ株式会社および東京大学が ASSoC の低電力性能の実証を進める。

ASSoC のチップの製造は日本電気株式会社が発行項目①を通じて行い、本実施項目②の実証先に ASSoC チップを提供する。日本電気株式会社はまた、実施項目②の活動が円滑に進められるよう情報共有に努め、進捗を管理・取りまとめを行う。

実施項目②-1「端末-エッジ連携 AI 処理用 AI チップ設計」

②-1(a)「原子スイッチ AISoC チップの混載要素技術開発」(担当：ナノブリッジ・セミコンダクター株式会社、日本電気株式会社)

IoT 端末用に AI 処理可能な低消費電力プログラマブル SoC を実現するため、原子スイッチメモリコア、CPU さらに原子スイッチ FPGA を 1 つの SoC チップに混載するための設計技術を開発する。市販されているプログラマブルな SoC は、主に CPU と SRAM

ベースの FPGA のロジックセルを混載したチップであり、動作時電力が 5W 程度かつ待機時の電力も 0.1W 程度と大きい。このようなチップは電池駆動を想定した IoT 端末に導入することは難しい。IoT 端末や医療用ロボットで利用するためには、入出力データの処理をプログラマブルロジック部分で高効率に実行することに加え、データの入出力が生じない期間は SoC チップもしくはチップ内のプログラマブルロジック部分の電源を遮断し、待機電力を数 mW クラスまで低減することが必要である。しかしながら、市販チップではプログラマブルロジック部の SRAM へのデータ再書き込みのために電源復帰後の起動時間が数 10 ミリ秒から 100 ミリ秒程度必要であり、実質的には頻繁に電源を遮断するような間欠動作を行うことが難しいという技術的課題がある。本開発項目の ASSoC では、プログラマブルロジック部に原子スイッチ FPGA を適用することで動作時の消費電力を従来チップに比べ 1/10 とすることを目標とする。さらに、原子スイッチメモリコアに CPU のプログラムを保存することでチップ全体の待機時電力を従来品に比べ 1/10 とし、間欠動作に適用するためにプログラマブルロジック部の起動時間として 1 ミリ秒以下を実現する ASSoC を開発することを目標とする。

ASSoC に必要な回路 IP の調達を 2020 年度末までに完了する。また、原子スイッチ FPGA コアの電源を ON/OFF 可能なパワーゲーティング対応 FPGA セル設計を完了する。ASSoC の待機時電力を従来品に比べ 1/10 となること、さらに、間欠動作に適用するためにプログラマブルロジック部の起動時間を 1 ミリ秒以下となることをシミュレーションにより確認する。また、原子スイッチを用いたプログラマブルバス回路の設計においては、2020 年度末までに完了する。2021 年度においては、CPU 等の要素回路の動作確認を行う。

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」(担当：ナノブリッジ・セミコンダクター株式会社、日本電気株式会社)

混載メモリにおいては、用途に応じた構成を得るためにビット長およびワード長、さらに ECC などのオプションが自由に選べ、自動的にレイアウトが生成される設計ツール(メモリコンパイラ)が一般的に用いられている。FPGA においても、AI チップの可変アクセラレータ用混載 IP コアとして広く使われるためには、用途やチップサイズに応じて構成を容易に変更できることが必要である。本開発項目では、原子スイッチ FPGA を構成する FPGA セル(ロジックセル、ブロック RAM および乗算器ユニット)と書き込みドライバー回路の IP 化を行う。IP 化されたそれぞれのブロックをライブラリとして、リソースのパラメータ指定により原子スイッチ FPGA コアを自動生成可能なファブリックコンパイラを開発する。

AI 推論処理の低電力には積和演算の効率が重要であり、立命館大学では、乗算ユニットの設計を行い、積和演算の計算効率化を高めたアーキテクチャを実現する(本乗算ユニットに関して、国内外で成果発表または情報収集を実施予定)。

また、原子スイッチ FPGA コアと開発項目②-1(a)により開発済みのプログラマブルバス回路、周辺回路 IP 等と組み合わせて ASSoC のチップ統合設計を完了する。チップ製造後に、半導体テストでの評価環境を構築し、低電力性評価を実施する。さらに、フィールド評価用のプロトタイピング基板を作成完了し、共同研究先とのアプリケーション実装評価を完了する。

2020 年 2 月までに原子スイッチ FPGA を構成する FPGA セル(ロジックセル、ブロック RAM および乗算器ユニット)と書き込みドライバー回路についてそれぞれ IP 化を行う。

また、2021年3月までに、原子スイッチ FPGA コアと実施項目②-1(a)により開発済みのプログラマブルバス回路、周辺回路 IP 等と組み合わせて ASSoC のチップ統合設計を完了する。

ASSoC を製造するためのマスクセット用データに関して、第4配線層までを2021年4月、原子スイッチ層から最上層までを2021年6月までに準備する。また、2021年12月までに、半導体テストでの ASSoC 評価環境を構築し低電力性評価を実施する。2021年9月を目処にプロトタイピング基板を作成完了し、2022年2月を目処に共同研究先とのアプリケーション

実装評価を完了する。

②-1(c)「原子スイッチ AISoC チップのツール開発」(担当：ナノブリッジ・セミコンダクター株式会社・立命館大学)

既存の CPU の SDK(Software Design Kit:ソフトウェア開発キット)およびシステムレベル設計ツール(上位ツール)と連携して、ASSoC をプログラムするための設計ツール群を開発する。ASSoC のプログラミングフローでは、まず、上位ツールにより、ソフトウェア設計とハードウェア設計の切り分けを実施し、ハードウェア設計情報である

RTL(Register Transfer Level)回路記述およびプログラマブルバス構成情報を生成する。バス構成情報に基づいてプログラマブルバス回路をコンフィグレーションする。本開発ではこのバスマッピング技術を開発する。また、プログラマブルバス構成結果に応じて原子スイッチプログラマブルロジック部に効率的に RTL 回路記述をマッピングするバス接続フロアプランナを開発し、既存の原子スイッチ FPGA のマッピングツールと連携可能とする。演算器部分のツールに関しては、立命館大学と共同で開発を行う。

2020年10月までに、バス接続フロアプランナおよび開発項目②-1(b)において設計完了した ASSoC のファブリック構成に対応したプログラマブルロジック部のマッピングツールを開発完了し共同研究先へ提供する。また、プログラマブルロジック部と CPU や外部メモリ等との柔軟な接続を可能にするバスマッピング技術、及び、演算器の効率的な割り当て技術を2021年10月までに開発する。2021年度においては、ASSoC チップの実証のために本ツールを利用するあたり、デバックの対応等のサポートを実施する。

②-1(d)「適用領域拡大のための共通化仕様の策定」(担当：ナノブリッジ・セミコンダクター株式会社)

ASSoC チップを低遅延処理が必要な領域に広く適用できるようにして、広く普及させるためには、例えば自動車や医療機器向けチップ等で最低限必要な周辺回路や通信方式をあらかじめ備えている必要がある。このために2019年度中に再委託していないが協力関係にあるデンソー等の自動車部品メーカーとも協議し、自動車適用に必要な周辺回路やプログラマブルロジックの規模について明確化し、実証を行うロボットのニーズと仕様の共通化を実施する。2021年度作製のチップは自動車向けで実証する際の要求仕様を満足するようにする。2019年度中にデンソー等の自動車部品メーカーとも協議し、自動車適用に必要な周辺回路やプログラマブルロジックの規模について明確化し、実証を行うロボットのニーズと仕様の共通化を実施する。試作したチップについては、その他の用途においても低電力性の評価や耐放射線性の評価などによるユーザー評価を実施し、その結果を踏まえて適用領域の拡大を図る。

実施項目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：再委託先・株式会社ソリトンシステムズ)

2019 年度に、ASSoC と既存の SRAM ベースの FPGA を用いた SoC 型 FPGA の双方において同等の映像および音声伝送の評価が可能な評価基板の仕様を確定し、ベースとなる親基板と既存 FPGA のための評価小基板を設計する。並行して、作成された ASSoC チップを評価するための基板の基本設計を完了する。2020 年度には確定した ASSoC チップの仕様に基づいた評価基板を作成し、高位言語を用いて FPGA 向けのファームウェアと CPU 向けのソフトウェアを実装する。2021 年度には、評価基板に ASSoC チップを実装し、映像・音声のリアルタイム処理において想定された低消費電力が達成されることを確認する。

実施項目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：再委託先・筑波大学)

人間中心社会をキーワードとした次世代 IoT プラットフォームとしてドローンというフレームワークに注目が集まっている。日本再興戦略 2016 にドローンというキーワードが登場し、経産省「平成 30 年度経済産業政策」、文科省「次世代火山研究・人材育成総合プロジェクト」、農水省「G 空間プロジェクト」、また官民研究開発投資拡大プログラム(PRISM)など、様々な場所でドローンは注目技術として取り上げられている。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。そこで本実施項目では、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(他アーキテクチャとの差別化)について比較および改善の検討を行う。本実施項目の成果は、ドローンと言う ASSoC のキラーアプリケーション発掘だけでなく、他の次世代 IoT システムに対して ASSoC を応用する可能性についても示唆すると考える。

本実施項目では、上記を踏まえ、本事業終了時のアウトカムを明確にするためにスマート農業向けの ASSoC ドローンについて研究開発を進める。スマート農業において計測は非常に重要な要素であり、例えば、畑作物の生育状況を確認するためには、高度 30m 程度から 5mm~1cm 程度を判別できる分解能を持つ画像処理が必要となる。高度 30m という数値は、ドローン規制を考慮しつつ、ドローンの滞空時に生ずる風が農作物に与える影響を無視できる高度として選択されている。検証する動画処理技術は、火山などの災害地域探査、野生鳥獣の生態調査、高速道路や橋梁などの検査、などにおけるシステムへの応用も期待できる。

2019 年度より、ASSoC の仕様検討に並行し、代替チップ(SRAM 型 FPGA)を用いたスマート農業用システムの実証実験を開始する。農作物の育成状況を正確かつ高速に空中より判断する演算手法は現在確立されていない。そこで、既に農業用として利用されているドローンを利用し、検出アルゴリズムについてまず検討する。この検討では、判別精度に加え、ハードウェア化(ASSoC への実装時)に最も適したアルゴリズムの選定を図る。そして、2019 年度後半から 2020 年度にかけて、代替チップを利用した実証実験を開始する。この時、使用回路量およびメモリ量などについて詳細なデータを得ることで、ASSoC アーキテクチャの仕様検討に対して適宜フィードバックをかける。これにより、多種多様な農作物や計測環境に対応する素地が整い、農業用ドローン利用において ASSoC を広く利用可能にするための議論が可能となる。2021 年度より、代替チップから

試作 ASSoC への移植を開始する。また、これと並行して、組込み向け GPU および代替チップとの電力性能比較実験を繰り返し行い、ASSoC の優位性(性能、機能、電力など)について明らかにしていく。農業用ドローンは育成および収穫時期である夏～秋の運用が考えられ、システムが非常に高温となることも容易に想像される。利用条件において、代替チップ(SRAM 型 FPGA)と耐環境性があり、高温でも消費電力が増えない ASSoC との電力効率の比較においてその違いは大きくとなると考えられ、ASSoC 導入の有意性がこの実験を通して明らかになると考えている。上記実施内容により、最終目標である ASSoC による演算部電力効率 10 倍の見通しを得る。

実施項目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：再委託先・東京大学) (担当：再委託先・稲葉研究室)

③ -4-(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
産業用ロボットでは位置制御がなされているが、人のそばで人と接触しても安全に動作するためにはトルク制御が不可欠となる。そのために、道具を利用して環境との接触力を制御する必要のある大出力等身大人型ロボットで実現してきた技術を産業用へも展開できるように原子スイッチ FPGA を利用して実装評価する(図 4.1.2.1)。トルク制御を行うためには関節の力学パラメータのキャリブレーションを関節毎に行う必要がある。経年変化に伴うパラメータの変化に対応するため、関節駆動時の角度センサデータと出力トルクの時系列データを元にパラメータの学習を行い、各モータドライバ上でオンラインパラメータ推定を行うことでキャリブレーションレスでのトルク制御を行う。

2019 年度から 2020 年度にはモータドライバ上の FPGA にオンラインパラメータ推定法の実装を行う。2020 年度から 2021 年度にはモータドライバへの原子スイッチ FPGA の適用評価を行う。



図 4.2.1.1 大出力等身大人型ロボットにおいてトルク制御用に用いられるモーターおよび FPGA を含むドライバー

②-4-(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
産業用ロボットでもロボットの体内に複数のセンサ・アクチュエータが分散配置されるシステムとなってきたが、それらの分散通信によって身体システムの状態を学習する機

能を研究してきており、その実装を原子スイッチ FPGA の AI 処理機能を利用して実装し評価する。

図 4.1.2.2 は本研究室にて研究開発を行っている筋骨格ヒューマノイド臆悟郎である。臆悟郎には図に示す筋肉制御ユニットが 100 台以上搭載されており、制御コンピュータと通信モジュールを介して全身制御されている。これまでに身体構造をオフラインで学習し制御する手法を研究してきている。筋肉制御ユニットと制御コンピュータ間の通信を行う通信モジュールの FPGA 上でパラメータの学習を複数の筋肉制御ユニットからのセンサデータを元にオンラインで行うことで、動作精度の向上を目指す。また、ロボット義足等の待機状態が長いロボットへの適用を想定し FPGA の不揮発性を活用した電源制御を検討する。

2019 年度は、筋肉制御ユニットと通信モジュールに搭載した FPGA の原子スイッチへの置き換え検討を行う。この中で、最終的にロボット義足への適用を狙い、将来的には介護ロボット等への展開を狙う上での、ロボット向けの ASSoC のあるべき仕様について、日本電気と共に検討し、日本電気の設計する医療・ロボット用途に向けた ASSoC の仕様策定に寄与する。また、例えばモーターの制御電力が少ない状態が継続した場合は制御周期を長くする事や、モーターの制御電力がゼロである場合はモーター駆動波形を演算する FPGA 部分の電源を遮断するなどして、義足の状態に合わせた低電力化適応制御のモードを検討する。その上で、各モードへの移行時間や制御信号、各モードでアクティブにしておく必要のある回路モジュールについて検討した結果を日本電気と議論し、ASSoC において義足ロボット向けの低電力化適応制御に必要な機能として仕様に反映させる。

さらに、通信モジュール上の FPGA で複数の筋肉制御ユニットからのセンサデータを元にパラメータのオンライン推定を行い、最後に、筋肉制御ユニット・通信モジュールの FPGA を原子スイッチ FPGA および ASSoC への置き換え評価を行う。

加えて 2019 年度から 2020 年度には、通信モジュール上の FPGA で複数の筋肉制御ユニットからのセンサデータを元にパラメータのオンライン推定を行う。2020 年度から 2021 年度には筋肉制御ユニット・通信モジュールの FPGA を原子スイッチ FPGA および ASSoC への置き換え評価を行う。

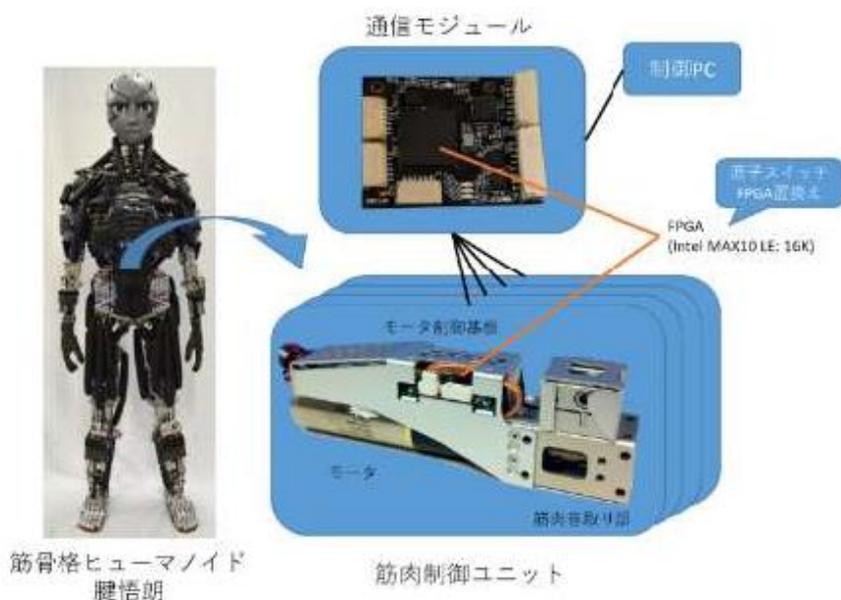


図 4.1.2.2 筋骨格ヒューマノイド臆悟郎と筋肉制御ユニット

②-4-(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

センサネットワークを実現する上で、柔軟物にも配置可能な分散センサシステムはリハビリ器具・屋内健康器具・多自由多関節ロボット柔軟外装等へ大いに期待されてきており、低消費電力・小型処理機能とセンサを組み合わせたユニットが不可欠となっており、その実装評価を行う。図 4.1.2.3 は本研究室にて研究開発を行ったヒューマノイドロボット小太郎に搭載されている 64 個の触覚センサを有するセンサ分布ユニット、図 4.1.2.4 はヒューマノイドロボット HRP-4R に搭載した 96 個の気圧センサを有する空気室型柔軟触覚センサである。近年、MEMS 技術の発達に伴い数ミリ角の IMU、距離センサ、光センサなどが開発され、これらを利用することで大量のセンサを分散配置することでより多くの情報取得が可能である。触覚センサ側のセンサデータと接触情報との関係性をあらかじめ学習し、センサ側に搭載した原子スイッチ FPGA によりセンサデータから接触情報への推論を行うことで、制御用コンピュータへの負荷低減と通信量の削減を実現する。2019 年度から 2020 年度は既存の FPGA を用いて多用途柔軟センサ分布ユニットの開発を行う。2020 年度から 2021 年度には FPGA を原子スイッチ FPGA へと置き換え評価を行う。



図 4.1.2.3 ヒューマノイドロボット小太郎と搭載されている 64 個の触覚センサ

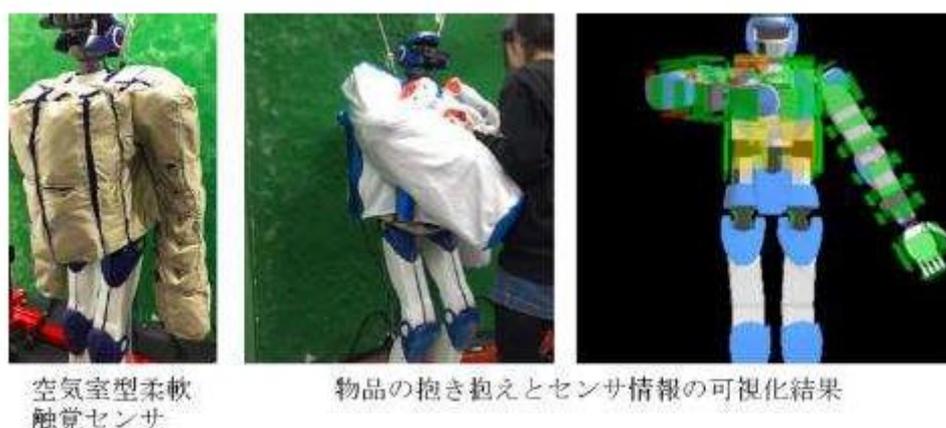


図 4.1.2.4 空気室型柔軟触覚センサを搭載したヒューマノイドロボット HRP-4R

②-4-(d) 介護介助ロボット向け AI チップ適用評価(AI による制御パラメータ調整・ユーザー行動推論開発)

ロボット向け AI 処理の社会実装として、ロボット技術を応用したロボット義足へと上記で開発された技術を適用する。

下肢切断者向けの義足として従来、パッシブ義足と呼ばれる機構的に膝関節の粘弾性特性を調節することで歩行の支援をしている。一方で障害者の生活の質の向上を目的に膝関節の粘弾性特性を能動的に可変とする電子制御パッシブ義足、あるいは、膝関節をモーターにより駆動する形式のロボット義足が従来研究開発されている。特に膝関節をモーター駆動するロボット義足は立ち上がり動作、階段昇降動作など従来の義足では実現不可能な動作を可能にする点で画期的な義足である。

このロボット義足における課題としては、一つ目は、ユーザー毎に足を踏み出すタイミングや足の踏み出し方などの歩行動作が異なるため、人間の平均的な歩行動作をロボット義足で実現するだけではユーザーは違和感を覚えてしまう。また、ロボット義足はユーザーが歩行中なのか、階段を登ろうとしているのか等、ユーザーの行動に合わせて制御プログラムを変える必要がある。現状の義足はユーザー毎にチューニングが必要であり、使いこなすために1か月のトレーニングが推奨されている。そこで、ユーザーの歩行時のデータを取得しクラウド上で学習を行い、AI チップにおいてユーザーの行動推論を行うことで手動でのチューニングなしに制御プログラムの切り替えを行う。

2019年度から2020年度は従来FPGAにおける推論の実装を行う。2021年度にはASSoCへの置き換え評価を行う。最終的にはユーザーのトレーニングなしに数日の調整で使用可能な仕組みを実現する。



図 4.1.2.5 現在市販されている義足

②-4-(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)

ロボット義足における二つ目の課題は、消費電力である。ロボット義足はバッテリー動作機器で1日中動作することが求められる。日常生活では長距離の歩行は想定されないため、モーターのパワー系で消費する電力は少なく、それ以外の制御回路部における電力消費量が無視できない。そこで、現在FPGAとMPUにより構成されている制御回路のFPGAをASSoCへと載せ替えることで低消費電力化を実現する。ロボット義足に使用する制御ユニットは筋肉制御ユニットと共通化しており、FPGAから原子スイッチFPGA

への置き換えによる省電力効果を評価する。ASSoC を用いることで MPU 含めた置き換え評価を行い、バッテリーで 1 日中動作可能とする見通しを得る。

2020 年度は FPGA から原子スイッチ FPGA への置き換えによる省電力効果を評価する。2021 年度は ASSoC を用いることで MPU 含めた置き換え評価を行い、バッテリーで 1 日中動作可能とする見通しを得る。

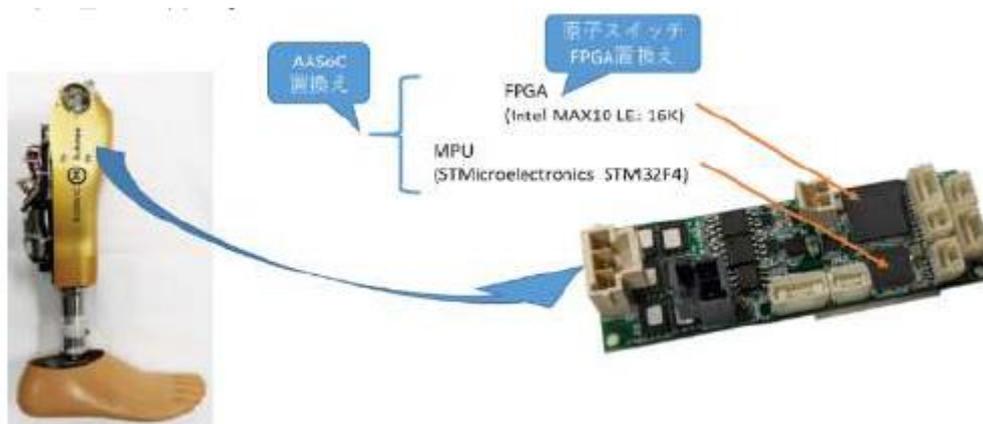


図 4.1.2.6 開発中のロボット義足と制御基板

以上に説明したプロジェクト目標を図 4.1.2.7 に纏める。

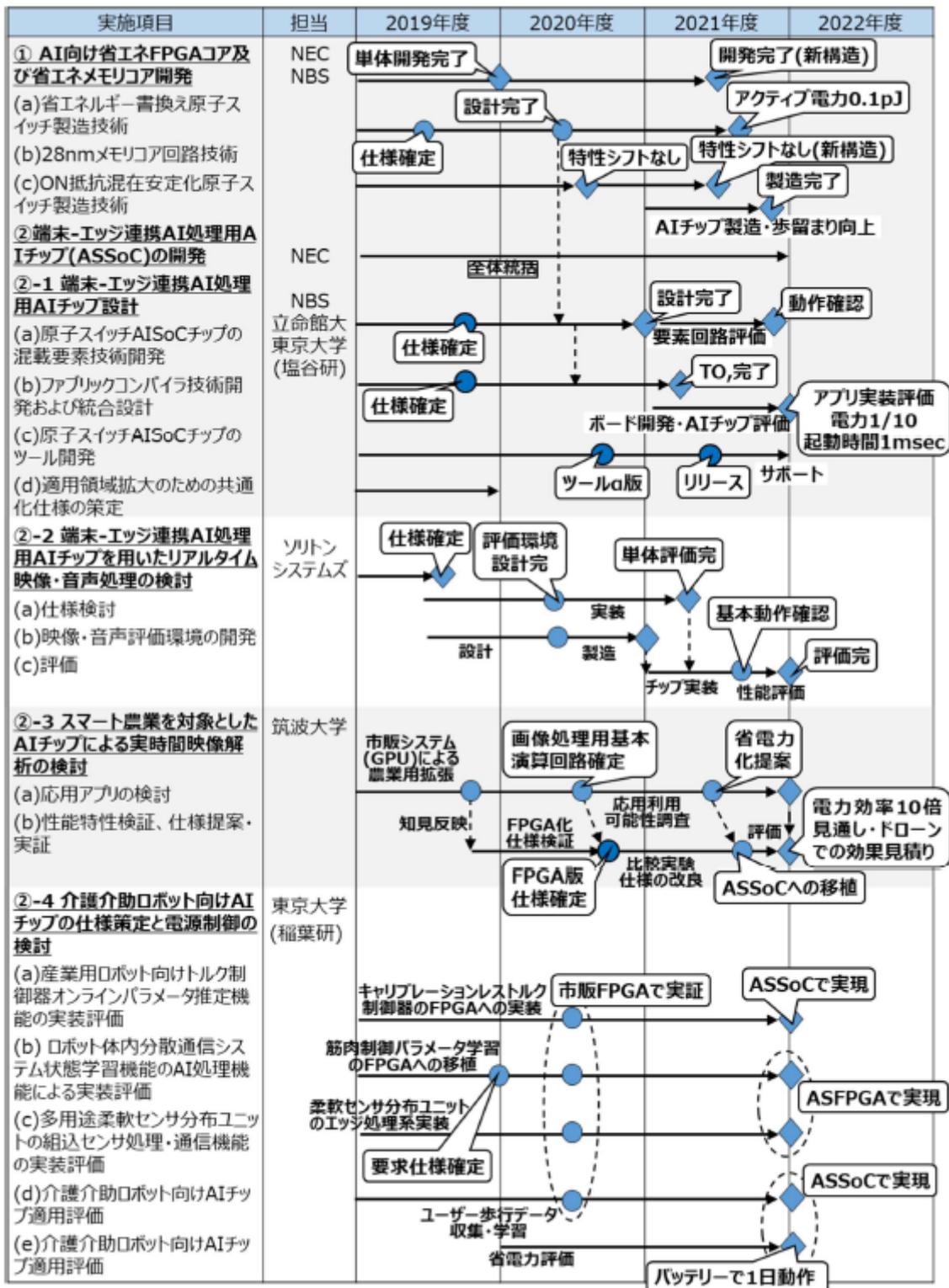


図 4.1.2.7 プロジェクト目標要約

4.1.3. 目標の達成度

各研究項目の達成状況を表 4.1.3.1 に示す。同表に記載の通り、最終目標はすべて達成した。

表 4.1.3.1 各実施項目の最終目標と達成状況

項目	最終目標	達成状況
実施項目① AI 向け省エネ FPGA コア及び省エネメモリコア開発		
(a)省エネルギー書換え原子スイッチ製造技術	<ul style="list-style-type: none"> 書き込みエネルギー従来比 1/10 以下(0.5nJ 程度)を達成 製造技術確定 	<ul style="list-style-type: none"> 書き込みエネルギー0.5nJ 以下を達成 製造バラつき安定化プロセスレシピ確定
(b)28nm メモリコア回路技術	<ul style="list-style-type: none"> 読み出し電力を、混載 Flash メモリと比較し 1/10 以下 	<ul style="list-style-type: none"> 読み出し電力 0.1pJ/bit (混載 Flash の 1/10 達成)
(c)ON 抵抗混在安定化原子スイッチ製造技術	<ul style="list-style-type: none"> メモリ・FPGA の両者で ON/OFF 比安定 	<ul style="list-style-type: none"> 安定性：125°C10 年相当不良なし確認 クロスファブ間の正常な電気コンタクト確認 ASSoC 製造完了
実施項目②-1 端末-エッジ連携 AI 処理用 AI チップ設計		
(a)原子スイッチ AISoC チップの混載要素技術開発	<ul style="list-style-type: none"> 回路 IP(バス、CPU)の調達 FPGA セル設計完了 FPGA の起動時間 1 ミリ秒以下をシミュレーションで確認 原子スイッチを用いたプログラマブルバス回路設計完了 	<ul style="list-style-type: none"> RISC-V プロセッサコア実装完了 FPGA 起動時間 1.6 マイクロ秒確認 プログラマブルバスを含めた ASSoC の論理設計を完了 RISC-V コアの動作確認完了
(b)ファブリックコンパイラ技術開発および統合設計	<ul style="list-style-type: none"> プロトタイピング基板を作成完了 低電力性評価 再委託先とのアプリケーション実装評価を完了 	<ul style="list-style-type: none"> テスト評価基板作成完了 ASSoC チップのテスト評価完了 再委託先へチップ出荷実施
(c)原子スイッチ AISoC チップのツール開発	フロアプランナおよび ASSoC のファブリック構成に対応したプログラマブルロジック部のマッピングツールを開発完了し共同研究先へ提供	<ul style="list-style-type: none"> ツールリリース完了 電力評価ライブラリ作成 演算器利用フロー開発完了
(d)適用領域拡大のための共通化仕様の策定	<ul style="list-style-type: none"> ASSoC 周辺仕様確定 原子スイッチ FPGA および ASSoC の耐放射線評価を実施 	<ul style="list-style-type: none"> ASSoC 周辺仕様確定 製造チップを Pj 外ユーザーに貸与し耐放射線評価を実施
実施項目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	<ul style="list-style-type: none"> ASSoC チップを実装し、映像・音声のリアルタイム処理電力効率 10 倍の見通し 	<ul style="list-style-type: none"> Intel 社 CycloneV、Microchip 社 PolarFire、および ASSoC の評価基板を製作完了

映像・音声処理の検討		<ul style="list-style-type: none"> 画像識別 AI アプリケーションの評価により間欠動作にて電力効率 10 倍の見通し確認
実施項目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	<ul style="list-style-type: none"> ASSoC チップに実装・評価 電力効率 10 倍の見通しおよびドローンでの効果の見積もり 	<ul style="list-style-type: none"> ステレオマッチング計測システムを搭載したドローンの製作・性能検証 飛行実験で得られた画像を対象としたステレオマッチングの実装 実装アルゴリズムが ASSoC の仕様を満たすことを確認
実施項目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		
(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	<ul style="list-style-type: none"> モータドライバ基板・筋肉制御ユニット基板・通信モジュール基板の設計 	<ul style="list-style-type: none"> トルク制御器オンラインパラメータ推定用の学習ネットワークの構築完了 推論機能を原子スイッチ FPGA に実装し、従来 FPGA に対して約 80% の消費電力削減を確認
(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	<ul style="list-style-type: none"> 市販 FPGA 利用 ASSoC への置換え実装・動作評価 	<ul style="list-style-type: none"> 状態学習機能を試すアームユニットの一部関節の設計改良 ASSoC 置き換え基板の設計・製造完了
(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	<ul style="list-style-type: none"> MCU を含め ASSoC および ASFPGA への置換え実装 バッテリーで 1 日中動作可能とする見通しを得る 	<ul style="list-style-type: none"> 多用途柔軟センサ基板の試作完了
(d) 介護介助ロボット向け AI チップ適用評価(推論開発)		<ul style="list-style-type: none"> 実機での路面に応じた歩行データ取得し推論アルゴリズムを開発
(e) 介護介助ロボット向け AI チップ適用評価(電力評価)		<ul style="list-style-type: none"> 置換え基板への ASFPGA 評価により、原子スイッチ FPGA を搭載したロボット義足を 1 日中動作可能である見積もりを得た

4.1.4. 成果の普及

これまで 16 件の研究会・学会での講演、1 件の学会誌への発表、さらに 1 件の展示会への出展を通じて、本事業で開発した成果の普及を行った。

4.2. 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」

4.2.1. テーマ概要

FA 機器や産業/サービスロボットなどの AI 化で求められるリアルタイム性・低電力化・知覚～制御までの複合的な処理への対応を実現するため、動的再構成機能を持つ Dynamically Reconfigurable Processor(DRP) (開発項目①)をベースに、AI アルゴリズム (開発項目③)とアーキテクチャ、回路技術(開発項目②)との連携を重視した開発により、高い柔軟性を有し、かつ従来技術比 10 倍の電力性能となる推論向けシステムを開発する。さらに、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実現性・実用性の検討(開発項目④～⑥)や、これらを統合するツール(開発項目⑦)の開発も行う。

4.2.2. 最終目標と根拠

【推論向け AI チップ (開発項目①②③⑦)】

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(DeepNeural 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 10TOPS(1 秒あたり 10 兆オペレーション)程度とすると、10TOPS/W の電力効率が求められる (この値は現在入手可能な AI チップのおよそ 10 倍)。本プロジェクトでは、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 (10TOPS/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。ターゲットとする電力性能比 (10TOPS/W) は、現状入手可能な AI チップや、JST のプロジェクト (ACCEL プログラム「近接場結合集積技術による革新的情報処理システムの実現と応用展開」)にて開発した動的リコンフィギュラブル AI チップと比較しても、10 倍の効率化が見込まれる。

最終目標の実現課題としては、アーキテクチャが AI 向けに最適化が進んでおり、今後大幅な電力性能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し、AI チップ性能向上の限界を打破する。具体的には、
・DRP アーキテクチャデザイン、コンパイラの技術開発・DRP 構成や軽量化を意識した先進的な単位演算 (積和演算や非線形演算など) 回路・アーキテクチャの開発・小フットプリント・軽量の計算を実現する DNN 変換機能の開発
・実製品への AI 組込みを容易化するツールの開発を行う。

【エンドポイント学習システム (開発項目④⑤⑥)】

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、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目標とする。また、

DNN(Deep Neural Network)以外の小型ニューラルネットワークを使う方法についても併せて研究開発を行っていくことで、エンドポイント学習の実現可能性および適用範囲を広げる。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～2022年度で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ことを最終目標とする。

4.2.3. 目標の達成度

全ての実施項目（①～⑦）に対して、研究開発目標を達成した。

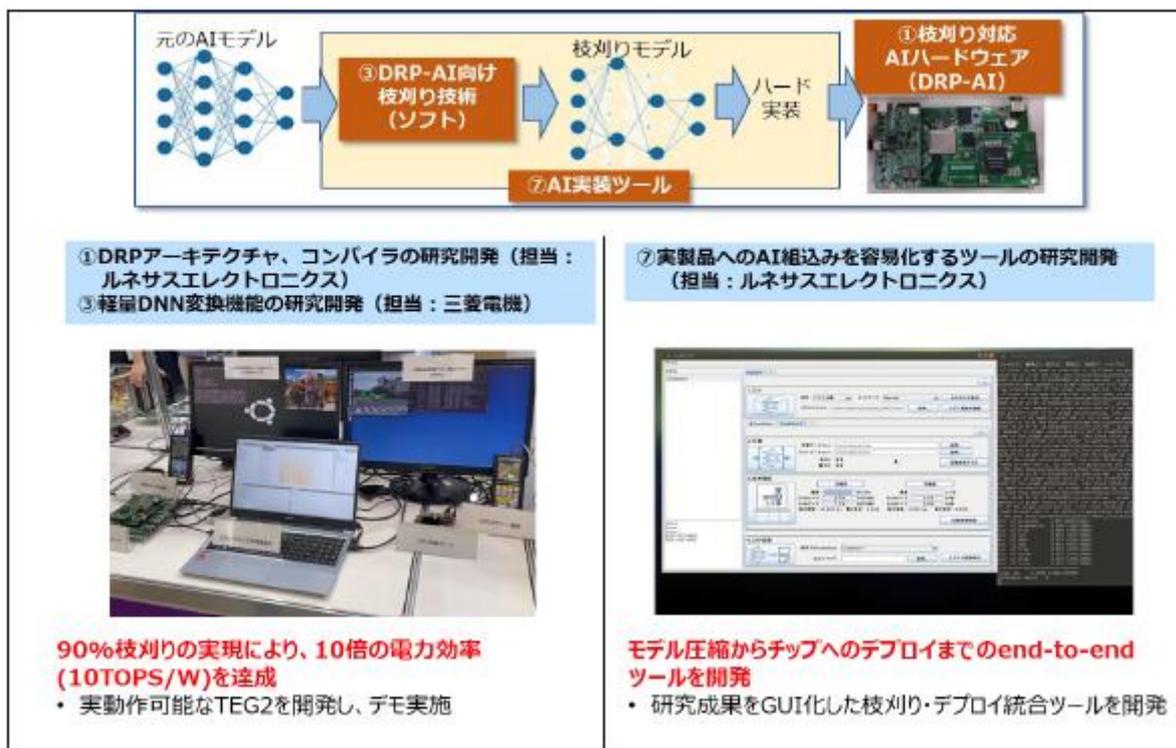
項目番号	個別実施項目(担当事業者)	当初計画の研究開発目標	達成状況と結果
①	DRP アーキテクチャ、コンパイラの研究開発（ルネサス）	ディープラーニング処理を行った際に、研究開発開始当時に入手可及な他社AIチップの10倍の電力性能比(10TOPS/W)を実測動作実証	予定通り達成。 2次TEG評価により、推論の10TOPS/W動作を実証した。
②	DRP構成や軽量化を意識した先進的な単位演算回路・アーキテクチャの開発（東京工業大学）	超軽量・超低電力エッジDNN推論エンジンの設計と試作を進め、実チップ評価を行う。また国際会議に投稿する。	予定通り達成。 計画通り設計・試作・評価を進め良好な結果(サブmW低電力、3-5TOPS/W)を得た。 VLSIシンポジウムに採択。
③	軽量DNN変換機能の研究開発（三菱電機）	DRP-AIチップに実装可能なDNNの軽量変換機能および試作ソフトウェアを開発し、DRP-AIチップと本ソフトウェアの連携により電力性能10TOPS/Wの動作実証を行う。	予定通り達成。 ソフトウェアによるINT8量子化対応と、枝化率90%以上の実現により目標電力性能が実現できることを確認。
④	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（三菱電機）	エンドポイント学習に有効なアプリケーションを抽出し、ユースケースのエンドポイント学習機能、およびシミュレータの開発により、エンドポイント学習の有効性を実証する。	予定通り達成。 ターゲットアプリとして基板の外観検査を抽出。ライン稼働後のメンテナンス等の環境変化を想定したシミュレーションでエンドポイント学習の有効性を確認。
⑤	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（ルネサス）	エンドポイント学習をTEG実証する。性能目標：約1日の学習更新(1枚当たり約3秒、約	予定通り達成。 DRPを活用したバックグラウンド学習構造を開発し、約3.8秒、約30,000枚で認識率向上

	ス)	30,000 枚で学習可能)	を TEG 実証。
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)	DRP で SOINN2.0 アルゴリズムを動作させ、6 軸ロボットを制御し制御周期 2ms 及び位置決め精度±3mm を確認する。	予定通り達成。 6 軸ロボットを DRP+SOINN を使用して制御し、制御周期 2msec を達成。
⑦	実製品への AI 組込みを容易化するツールの研究開発 (ルネサス)	上記①～⑥を含め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(2022 年度末)	予定通り達成。 2 次 TEG にて、ツールで生成したコードを使って動作を実証。

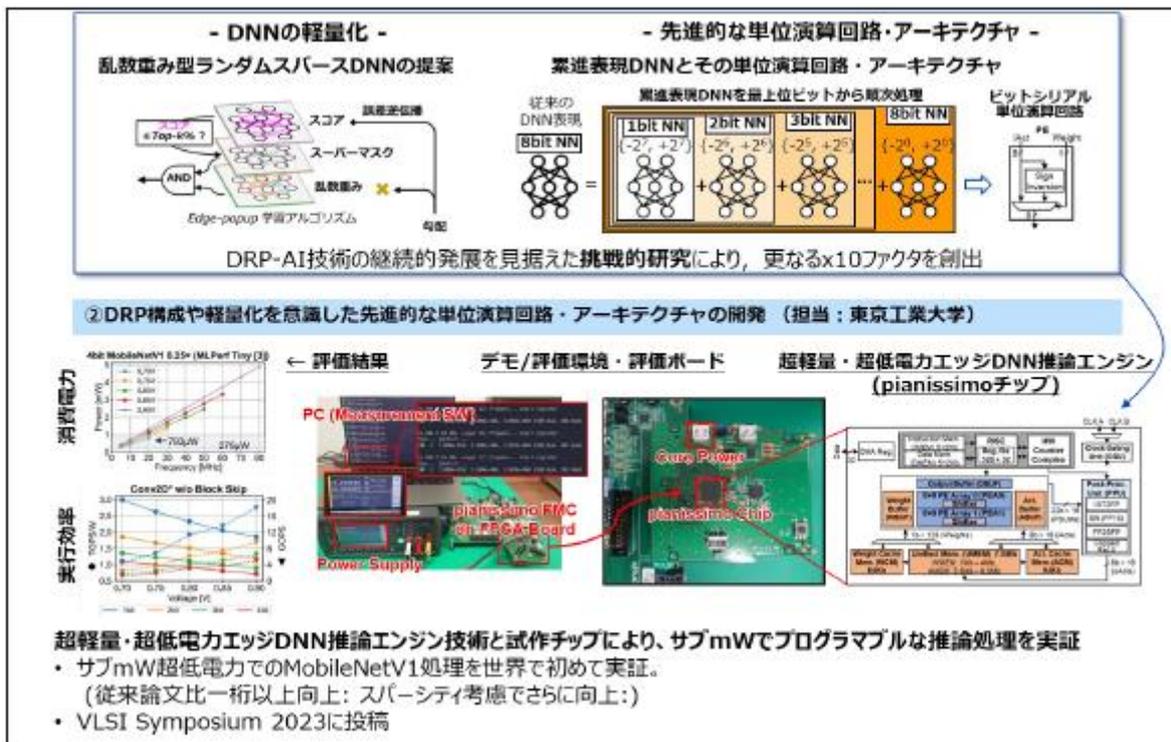
4.2.4. 成果と意義

各実施項目に関する主な成果を以下に示す。

【推論向け AI チップ (開発項目①②③⑦)】



【図 4.2.4.1】



【図 4.2.4.2】

(成果の意義) エンドポイントのデバイスから開発ツールまでの垂直連携の効果により、従来技術比10倍以上の電力性能が実現でき、開発ツールやライブラリも含めて容易に利用できる環境が整うことにより、

- 複数のセンサ信号からなる判断や画像認識を組み込み機器で実行し、ロボットなどのリアルタイム制御を可能にする

- 動的再構成機能により高い柔軟性を持つことで、多様な種類の入力情報、DNNのアルゴリズムを取り扱い可能とする
- 組み込み機器自身が判断することにより、生データをクラウドへ転送する場合に比べ大幅に通信量を削減、あるいは通信がない場合の自律を可能とする
- AIに関わる技術の進化を取り込めるプラットフォームは、AI技術の蓄積と利活用を促進する

これらの効果により、製品のイノベーションを促し、付加価値向上、生産性向上、品質向上、省エネを実現できる。(図 4.2.4.1~図 4.2.4.2)

【エンドポイント学習システム（開発項目④⑤⑥）】

リアルタイム推論 同時に実行 学習

リアルタイム推論を止めずに同時に実行可能なオンチップ学習環境を構築

④ DNNによるエンドポイント学習用ソフトウェアの開発 (担当: 三菱電機)
⑤ DNNによるエンドポイント学習用ハードウェアの開発 (担当: ルネサスエレクトロニクス)

学習データ削減技術と推論しながら学習する技術により約1日でupdateする学習システムの動作実証

- 外観検査の学習アルゴリズム、軽量学習機能の構築
- 外観検査のTEG動作実証 (実証実験)

簡単教示・エッジ学習・リアルタイム制御 (学習+推論)

模倣学習により、ティーチング作業を不要にするオンチップ学習環境を構築。ロボットタスクに必要なリアルタイム性能を同時に実現。

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 (担当: SOINN)

DRP+SOINNによるロボットのマルチモーダル模倣学習とリアルタイム制御

DRP+SOINNによるパテ塗作業 タスク選択機構GUIモデル比較表示

制御周期と位置決め精度を動作実証

- 学習時間 1 ~ 3 分
- 2msの制御周期
- ±3mm以下
- タスク選択可能な仕組みとGUIを開発

【図 4.2.4.3】 エンドポイント学習システムの概要

(成果の意義) 生産現場の動作環境やタスクの変化に自律的に適応していくエッジでの学習システムを構築することが可能となる。また、AIの実行を止めずにバックグラウンドで学習することや、少ないティーチング作業での教授が可能となるため、追加学習のための学習時間の確保やデータ収集の手間が大幅に削減され、現場での運用が容易になる。これらの成果の普及により、生産現場などにおいて、生産性の飛躍的な向上が実現できる。

(図 4.2.4.3)

4.2.5. 成果の普及

主な成果に関して、主要な国際学会、展示会、広報などを通じて積極的に情報発信を行い、普及に向けた活動を継続的に実施している。

AI チップ動展示・NEDO との共同ニュースリリースによる国内外への情報発信製品展開に向けたエコシステム拡大・国内外企業への promotion を活性化 (図 4.2.5.1)

本研究開発成果をCEATEC2022(NEDOブース)に展示

NEDOとの共同ニュースリリース(2022/12月)



従来技術に比べて最大10倍の電力効率を実現した人工知能(AI)チップを開発
一瞬時に自律的に対応し、リアルタイムで応答する組み込みAI機能の開発が可能に—

2022年12月15日
NEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）
ルネサス（有）とソイン（有）の共同プレスリリース

NEDOが主催する「高効率・高実用性を可能とするAIチップ・エッジAIコンピュテナブの技術開発」でルネサス エッジAIのニア
ズ（物）は、種別もクラスも異なる組み込みプロセッサ（DPP）を用いた人工知能（AI）チップを開発しました。電力効率は
1割（ワット）当たり10倍（W/Wh）で、従来技術に比べて最大で10倍の電力効率を実現しました。

⇒共同リリースの内容が多数記事に掲載(9件)

【図 4.2.5.1】 CEATEC 展示や共同ニュースリリース（ルネサス）

SOINN：展示会等で Promotion し、特にエンドポイント学習に関してニーズ把握や顧客獲得を進めている（図 4.2.5.2、図 4.2.5.3）。



【図 4.2.5.2】 展示会におけるプロモーション活動（SOINN）



【図 4.2.5.3】 エンドポイント学習システムの紹介 (SOINN)

東京工業大学：成果が国際トップ学会 (Hotchips)に採択され、記事も多数掲載 (図 4.2.5.4、図 4.2.5.5)

人工知能ニュース:

実効効率で世界トップクラスのエッジAIプロセッサアーキテクチャ、東工大が開発

<https://monoist.atmarkit.co.jp/mn/articles/2108/24/new067.html>

NEDOと東京工業大学は、エッジ機器で高効率なCNN(畳み込みニューラルネットワーク)による推論処理が可能なプロセッサアーキテクチャを開発したと発表した。同プロセッサアーキテクチャに基づく大規模集積回路(LSI)も試作し、「世界トップレベル」となる消費電力1W当たりの処理速度で最大26.5TOPSという実効効率を確認している。

2021年08月24日 06時30分 更新

[朴尚洙, MONOist]

NEDO(新エネルギー・産業技術総合開発機構)と東京工業大学は2021年8月23日、エッジ機器で高効率なCNN(畳み込みニューラルネットワーク)による推論処理が可能なプロセッサアーキテクチャを開発したと発表した。同プロセッサアーキテクチャに基づく大規模集積回路(LSI)も試作し、「世界トップレベル」(ニュースリリースより)となる消費電力1W当たりの処理速度で最大26.5TOPSという実効効率を確認している。今後は、ス



【図 4.1.5.4】 NEDO と共同のニュースリリース (東工大)

TECH Powered by マイナビニュース

企業IT テクノロジー Members+

エッジで高効率AI処理が可能なプロセッサ、東工大などが開発 - Hot Chips 33

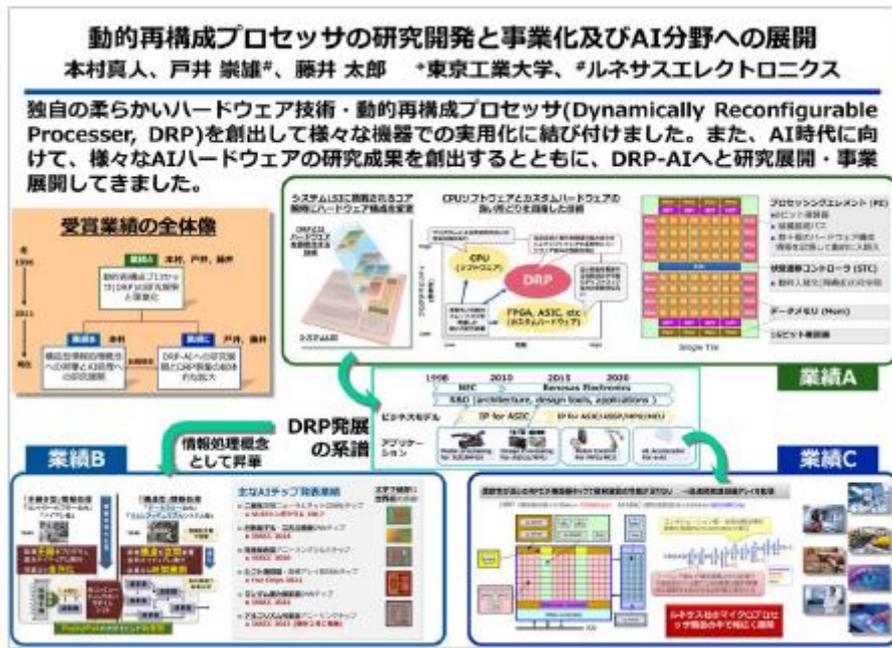
© 2021/08/24 20:32 2021/08/24 21:47 画像: 読者投稿

Twitter
Facebook
Blogger
URLをコピー

新エネルギー・産業技術総合開発機構(NEDO)と東京工業大学(東工大)は8月23日、スマートフォンやロボットなどのエッジ機器において、高効率な畳み込みニューラルネットワーク(CNN)推論処理を行うプロセッサアーキテクチャを開発し、それをを用いたLSIを試作。実測で最大26.5TOPS/Wの実効効率を達成したことを発表した。

【図 4.2.5.5】 Hot Chips 33 採択記事 (東工大)

DRP に関して、東工大とルネサスエレクトロニクス共同で、2022 年度山崎貞一賞を受賞（図 4.2.5.6）。



【図 4.2.5.6】 2022 年山崎貞一賞受賞（東工大とルネサス共同）

4.3. 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」

4.3.1. テーマ概要

(a) 研究開発の概要

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) 用の 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な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 テクノロジーを使った場合で 10 Tops/W、5nm テクノロジーでは 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、高性能・大規模な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。

(b) 背景

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、開発当初は Intel が買収した MobileEye は、2020 年頃に 7nm テクノロジーの EyeQ 5 で、2.4 Tops/W を実現するとしていた。

これらから、2018 年（課題採用時）での、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。言い換えると、本プロジェクトの目標である

<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には、最低 10Tops/W 程度の性能を実現する必要がある。エッジにおける AI プロセッサの社会実装には、絶対性能や電力性能の向上は不可欠である。

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) 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で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。

また、エッジでの多様な応用の開発を困難にする要因として大きなものは、半導体の開発コストの高騰である。この30年ほどの間、LSIの開発コストはデザインルールの縮小にほぼ反比例した指数関数的な上昇を続けてきており、7nmプロセスでは数十億円になる。このため、エッジ向けで大規模な大量生産が期待できる場合ですら、最先端のプロセスの利用は困難になってきている。

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さなければ汎用の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。

(c) 必要性等

現在AIエッジコンピューティングで柔軟な処理を実現することの障害になっているのは、ソフトウェアで柔軟な処理が可能なGPGPUベースの技術では必要な電力性能が実現できておらず、一方ハードウェア再構成によって柔軟性を実現する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は難しいということである。ソフトウェアでプログラム可能なシステムでは電力性能を画期的にあげるのは難しい。また、エッジで必要になるリアルタイム性を実現するのも困難である。このため、柔軟なハードウェアの再構成が可能なFPGAに基づいた技術で、学習処理等に必要な高い電力あたり性能を実現することが必要になっている。

このためには、高い電力性能を実現できる演算コアと、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。これは、FPGAが少数のチップ専業メーカーの独占技術であり、カスタムLSI向けのIPでは市販FPGA並みの回路密度・速度を実現できなかったためである。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、半導体技術としては既に最先端ではない TSMC の 40nm プロセスを利用して、同じ TSMC の現時点での最先端プロセスである 12FFC によって製造された NVIDIA Xavier とほぼ同等の電力性能を実現した ((推測する最高性能値として) Xavier が FP16 では 0.5Tops/W、それに対して我々も 0.5Tops/W)。これは、同等な半導体技術であれば 3 倍程度我々のアーキテクチャのほうが電力性能が高いことを意味する。これは、演算精度を犠牲にしたものではなく、演算方式やデータパスの深層学習アプリケーション向けの最適化によって実現されているものである。最適化の具体的なものは、行列ベクトル積に特化した回路による演算器自体の無駄の削減、レジスタ・メモリ等の消費電力の削減(特許出願中)、チップ内での放送・縮約を可能にするネットワークによるデータ移動に伴う消費電力の削減(特許取得済)等である。

また、我々は、人工知能ソフトウェア技術の開発、実社会・産業への応用を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、国内外で様々な応用に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に適用でき、大きな波及効果を実現できる。

(d) 全体計画

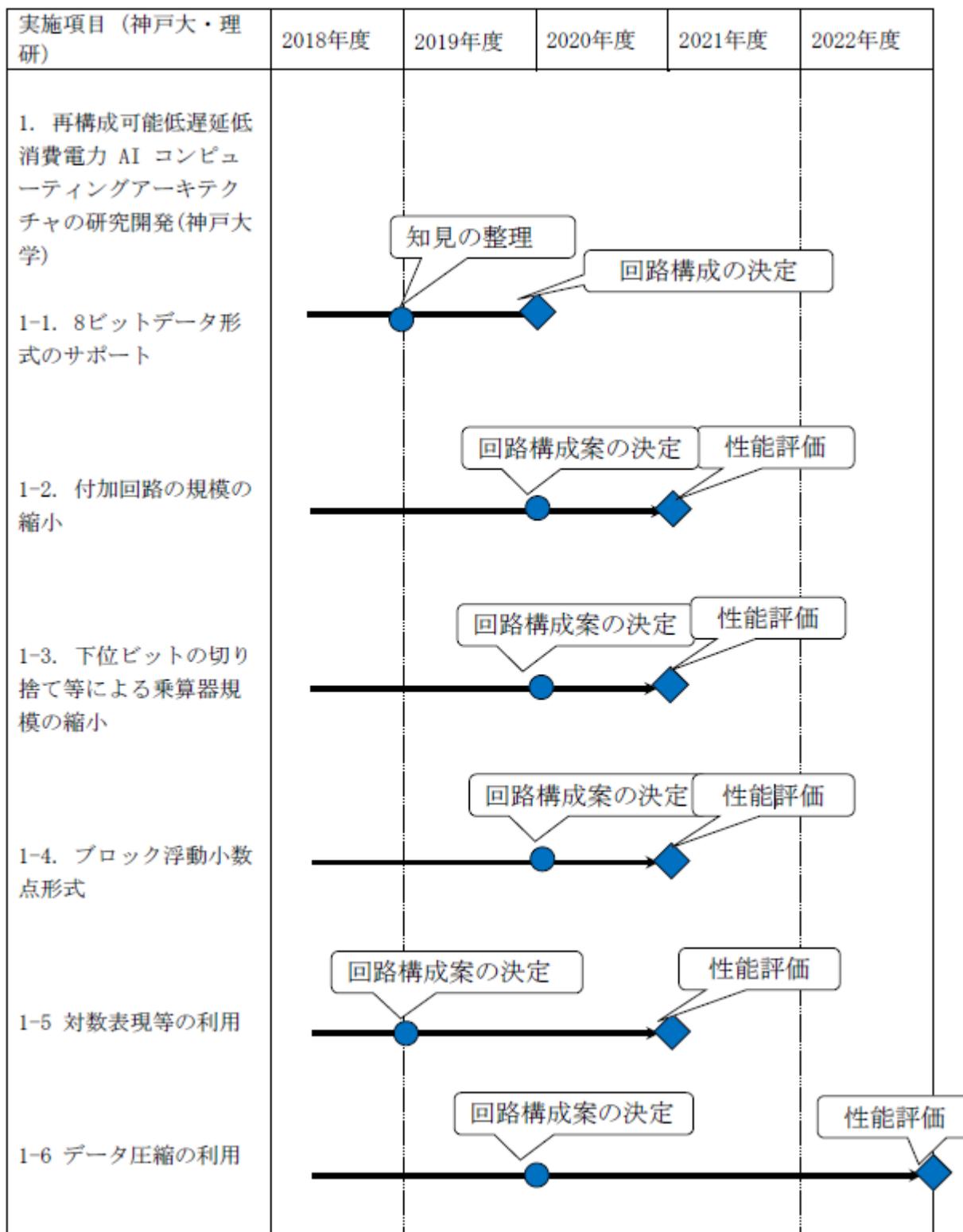
以下に 開発当初の計画を示す。

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思わ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標とし、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。

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する。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る。

2017 年度までの 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 5nm (N5) プロセスで 10 Tops/W を実現できるプロセッサを開発する。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。これに並行して、実際の産業・社会応用を進める。

表 4.3.1 実施項目毎の線表



4.3.2. 最終目標と根拠

以下に研究開発の目標、設定根拠について述べる。

(1) 研究開発目標

【アウトプット目標】

① 実施項目毎の目標の詳細

No	実施項目	担当	中間目標 (2020年度末)	最終目標 (2022年度末)	目標の設定根拠
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	神戸大・理研	8ビットデータ形式のサポート、付加回路の規模の縮小等の改良項目の評価を行い、コア部分のアーキテクチャを確定する。さらに、eFPGAでのチップ構成方法を確立する。	40nmテクノロジー相当で2Tops/Wの性能を実現できることを実証する。	この数値は2018年時点の自動運転用プロセッサ(7nmで実現)の2倍程度であり、40nmでの実現には技術的な飛躍が必要だが、我々が既に実現した数値からは4倍程度であり、演算方式等の改良で十分実現可能と考える。
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	PFN	実施項目1で開発したアーキテクチャを使い、7nmプロセスを想定した場合に15Tops/Wの性能を実現する。	中間目標までに開発したプロセッサをさらに改良し、プロセスの性能向上も合わせて20Tops/W以上の性能を実現する。	2018年の水準に比べると同等のテクノロジーで10倍以上の電力性能に相当するが、実施項目1の数値から、半導体の性能向上を考慮すれば実現可能となる。
3	ソフトウェアフレームワークの開発	PFN	ソフトウェア開発項目・ミドルウェア構成を明確化する。	フレームワークを実際に利用した応用ソフトウェアを動作させる。	ソフトウェアを動作させることが目標であり、数値目標は設定しない。
4	ロボティクス応用のためのソフトウェア開発	PFN	応用対象、ソフトウェアの構成を決定する。	開発したプロセッサ、フレームワークを実際に利用した応用ソフトウェアを動作させる。	ソフトウェアを動作させることが目標であり、数値目標は設定しない。

② 提案事業の最終目標（2022年度末）及び基本計画との合致性

提案全体計画における目標達成度評価	技術レベル	エネルギー消費効率/電力効率	技術・性能に係る目標
基本計画中の研究開発目標の該当項目 (1. (2) ①)	NVIDIA Xavier 相当（半導体は 12FFC)	1Tops/W	事業開始時点における同等の技術と比較し、10倍以上提案事業の最終目標
本テーマの最終目標	TSMC N5 を想定	20Tops/W	事業開始時点における同等の技術の 20 倍を実現

【アウトカム目標】

① 提案事業の 2037 年温室効果ガス排出削減効果への貢献

基本計画中の研究開発の目標の該当項目	研究開発を幅広く実施することにより生まれた成果が社会に適用されることで、2037 年時点における CO ₂ 削減効果が事業全体として年約 4,900 万 t を実現し得ることについて、要素技術レベルの成果を踏まえた見直しを確認する。
2037 年時点における CO ₂ （温室効果ガス）削減効果	年間 1220 万 t
CO ₂ 削減効果の達成見込みの確認方法・形態（予定）等	本提案によるプロセッサが目標通りの電力性能を実現できることの確認をする。

提案事業範囲：ロボティクス応用及びサービスクラウドのハードウェア

- 事業を実施しなかった場合のシナリオ：クラウドのハードウェアプラットフォームの市場規模が、2021 年度で 2000 億円、2030 年度で 5000 億円との予想があり 2037 年には 1 兆円に達するものと推測される。現状のプロセッサでは、ハードウェアコストと電力コストは同等になってきており、ほぼこの程度の電力コストになると想定される。これは電力量として年間 700 億 kWh である。
- 事業を実施しなかった場合の温室効果ガス排出量（万 t/年）：年間 700 億 kWh は温室効果ガス排出量として 3400 万 t/年にあたる。（排出係数として 0.000512tCO₂/kWh、電力単価として 15 円/kWh を仮定した）
- 事業を実施した場合のシナリオ：国内で 2037 年までに、本研究開発による技術、あるいはそれにならって電力性能を向上した技術が市場の 50%程度で使われるようになると想定する。本研究開発では、将来にわたって競合技術と同等の半導体技術を利用した場合に、現在の延長である競合技術に比べてほぼ 10 倍の電力性能の向上を実現する。言い換えると、温室効果ガス排出量は 1/10 になる。
- 事業を実施した場合の温室効果ガス排出量（万 t/年）：前項シナリオから、40%の部分が 4% になるので、全体では 64% になり、2180 万 t/年となる。
- 温室効果ガス排出削減効果量（万 t/年）：事業を実施しなかった場合の温室効果ガス排出量（万 t/年）－事業を実施した場合の温室効果ガス排出量（万 t/年）

- 事業を実施しなかった場合の温室効果ガス排出量（万 t/年）－事業を実施した場合の温室効果ガス排出量（万 t/年）であり、3400 万 t/年 - 2180 万 t/年 = 1220 万 t/年となる。

提案研究開発テーマに係わる 2027 年度、2032 年度、2037 年度時点における市場獲得規模は以下のとおりである。

	2027 年度	2032 年度	2037 年度
市場規模	AI ロボティクス及びサービスクラウドプラットフォーム市場 3 兆円	AI ロボティクス及びサービスクラウドプラットフォーム市場 5 兆円	AI ロボティクス及びサービスクラウドプラットフォーム市場 6 兆円
市場シェア	5%	10%	20%
提案事業成果による市場獲得規模	1,500 億円	5,000 億円	1 兆 2500 億円

③ 本研究開発によって想定されるその他のアウトカム

深層学習・人工知能応用だけでなく、ビッグデータ処理、データ同化、シミュレーション等のあらゆる高性能計算が要求される分野で、発展の妨げとなっている消費電力の増大の問題を解決し、シミュレーションと実験・観測が一体となった新しい科学の発展に寄与する。

(2) 研究開発目標の設定根拠

競合技術、既存技術との比較表

技術名称	技術保有者	年月	性能① (電力性能) Tops/W	性能② (チップ面積あたり (NVIDIA Volta を 1 とする))	エコシステム、ビジネスモデルの強み	全体市場規模 (兆円)	獲得市場規模と市場シェア(%)	総合評価 (LD、DH、RA)
提案技術	本技術 (現状)	2018/6	0.5	1	△	0.3	0	RA
	本技術(事業終了時)	2023/6	7	10	△	1	2	LD
	本技術(実用化時点)	2025/4	7	10	○	0.5	3	LD
	成果普及段階	2027/4	15	30	○	2	5	LD
NVIDIA 社 (GPGPU 技術)	NVIDIA (現状)	2018/6	0.1	1	○	0.3	95	RA
	NVIDIA (事業終了時)	2023/6	0.5	3	○	1	95	RA
	NVIDIA (実)	2025/4	1.0	5	○	1.5	90	RA

	用化時点)							
	成果普及段階	2027/4	2	10	○	2	90	RA
Intel 社 (LakeCr est 技 術)	Intel (現状)	2018/6	0.05	0.5	△	0.3	0	RA
	Intel (事業終 了時)	2023/6	0.25	1.5	△	1	1	RA
	Intel (実用化 時点)	2025/4	0.5	2.5	△	1.5	1	RA
	成果普及段階	2027/4	1	5	△	2	1	RA

・当該技術の世界における競争力について RA (Run After)、DH (Dead Heat)、LD (Leading) で記載

最終目標は、最終目標時点で利用可能な 5nm テクノロジー の次の世代テクノロジーで製造した場合に 10Tops/W 程度を実現できるアーキテクチャの AI プロセッサを設計、その性能を実証し、ロボティクス応用ソフトウェアを動作させることである。

4.3.3. 成果と意義

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
(国立大学法人 神戸大学)

電力性能向上のための回路方式の検討

本項目の目標は、

- 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である。このため、以下のような回路構成を検討した。
 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. 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
 3. 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
 4. 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
 5. 外部メモリとのデータ転送にデータ圧縮を利用する
 6. 2021 年度には第 1 世代 Chip の結果をフィードバックする
 7. 2022 年度には、深層学習以外の分野への適用可能性を考慮したアーキテクチャの改良を検討する

(1)については、固定小数点 8 ビットは乗算器の規模はある程度小さくなるが、入力のダイナミックレンジが小さくなるために利用困難であるとわかった。このため、通常の FP16 よりは仮数部を短くすることで、電力効率を向上させた。

(2) については、単精度、倍精度についてこの方法を適用して、有意に計算精度が落ちないことを確認した。これによって、単精度、倍精度の性能を大きく落とさずに回路規模を大きく縮小できた。

(3) ブロック浮動小数点形式は、特に半精度で大きく回路規模を削減できた。

(4) 対数変換については、半精度だけで確かに大きく回路規模を削減できることがわかった。しかし、人工知能応用だけに限っても、単精度が必要な部分もあるため、採用はしていない。

(5) データ圧縮については、特に 16 ビット語長の場合に大きく圧縮できる可能性があることを確認した。今後の採用を検討する。

(6) フィードバックとして、さらに電力性能を向上させるための検討をおこなった。このためには、基本的に演算論理回路の消費電力に比べてレジスタ・メモリアクセスの消費電力を減らす必要がある。また、実用になる範囲でより短い語長にすることで演算自体の消費電力を削減することも重要である。

前者の検討として、行列サイズを大きくすることを検討した。その方法として、物理的に大きな行列演算器を構成する方法と、小さな行列乗算器を時分割動作させる方法を検討した。特に動作クロックを高く維持することが必要な場合には後者が有利となる。半導体面積あたりの性能が重要となる今後は時分割動作の検討が必要とわかった。

(7) 深層学習以外の分野への適用可能性を考慮したアーキテクチャの改良については、まず適用可能性の検討のため、アプリケーション向けのコンパイラを開発した。今回開発したものは、OpenCL のような汎用環境ではなく、粒子間相互作用計算に特化した DSL (ドメイン特化言語) である。DSL とした理由は、言語仕様を小さく保つことと、高度な最適化を容易にすることである。

本プロセッサアーキテクチャでは、通常の汎用プロセッサでは必須となる多ポートレジスタの利用を避け、複数のメモリブロックを機械語レベルでオペランドとして選択する。この方法によって低消費電力と高いレジスタアクセスバンド幅を実現している一方、変数割り当ての自動化が困難で人手による変数割り当てが必要であった。

今回、変数割り当てに対してシミュレーテッドアニーリングによる最適化を行なうことで、短時間 (CPU で 1 秒程度) の試行で理論的境界の近くまで最適化されたコードを生成できることがわかった。このことは、本プロセッサアーキテクチャが深層学習以外の一般的な数値計算にも十分適用可能であり、最適化コンパイラの開発も短期間で行なうことができることを実証できたということである。

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
(株式会社 Preferred Networks)

第 1 世代 Chip の ES 製造及びその電力性能評価

本項目の目標は、実施項目①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い、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 7Tops/W という高い電力性能を実応用で実現することである。

第1世代 Chip の機能検証及び電力性能評価の結果および第2世代 Chip の設計製造については以下の通りである。

- 第1世代 Chip 機能検証及び電力性能評価結果サマリー

1. GDDR6 インターフェイス機能

2020年度から継続して評価を実施した。評価の過程にて GDDR6 コントローラの初期化ステップに一部冗長な部分があることがわかり、初期化プログラムの修正を行った。この結果 GDDR6 インターフェイスについては機能的に必要な帯域でエラーなく動作することが確認できた。

2. 演算器機能

評価ボードを用いた第1世代 Chip 内の行列演算器コアの電力性能の計測を実施した。演算器コアを 0.55v 610Mhz で動作させた場合に 8.55 TFlops/W の電力性能であることを確認した。

- 第2世代 Chip の設計製造サマリー

第1世代 Chip にて実装した行列演算器コア (MAU) を 5nm テクノロジー半導体での実装を行った。

実装を行うために、MAU を切り出し周辺回路の追加を実施した。

合わせて電力性能の向上や動作周波数向上を目的とした回路修正も実施した。

回路変更にもなって動作検証を行うための環境およびテストパタンの生成、パターンを使った機能動作検証を行い問題ないことを確認した。

その後、ASIC ベンダーへ回路 (RTL) をリリースして、物理実装を行った。物理実装の過程で発生したパフォーマンスの課題などについて ASIC ベンダーからフィードバック情報をもらい回路 (RTL) のブラッシュアップを行い、2022年10月に Tape Out を完了した。

Tape Out のときの設計データを用いたシミュレーションによる電力性能の値が 0.65v 1.1Ghz 動作で 8.44 TFlops/W であることを確認した。

TapeOut 完了後、第2世代 Chip の評価を行うための評価基板設計を ASIC ベンダーとともに進め 2023年1月に完了した。

これと並行して評価基板で用いるテストパターンについても準備を行った。

2023年3月に TSMC から ASIC ベンダーに第2世代 Chip の ES が納品され、評価基板を使った評価を開始した。

準備したテストパターンにて動作検証を行い機能には問題ないことを確認し、その後 MAU の電力性能が動作電圧 0.60V、動作周波数 950Mhz 時 11.26TFlops/W であることを確認した。

現状では、eFPGA は数値演算は実装可能であるが、高速動作は難しい。これは基本セル自体の特性によるもので、FPGA チップと同様の傾向である。

ステートマシン等の、制御回路の実装には適している。これは、我々の設計では制御回路は高速動作が必要ではない(命令生成は 4 クロックに 1 度のため)であることも理由である。

大バンド幅でのデータ移動の実装には適していない。これは配線リソースの不足のためである。

という特性をもつことが確認できた。また、我々の設計においては、命令生成等の制御部はチップ全体で 1 ユニットあればよいため、面積効率や電力効率よりも柔軟性が重要だが、演算部は分散して多数実装されるため、面積効率や電力効率も問題となる。このため、今回は特に制御部、具体的には命令生成ロジックの部分を eFPGA を使った実装にしている。

現在のところ、ホストから圧縮した命令を送り、これを展開して供給するロジックが実装され、必要バンド幅を半分程度に減らすことができている。また、命令生成自体を eFPGA 内部に実装したシーケンサで行なうことで、繰り返し構造を実行し、ホストからの実行時の命令供給をほぼ不要とすることも可能である。

ループ構造の実現は CPU を埋め込んでソフトウェアで命令列を生成することでも原理的には可能であるが、汎用性を高くすると非常に強力な CPU が必要となる。これは現実的とはいいがたく、大規模 SIMD プロセッサでは eFPGA によるシーケンサベースの実装が極めて有用であることが確認できた。

実施項目③ソフトウェアフレームワークの開発 (株式会社 Preferred Networks)

本項目の目標は、実施項目②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い、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・電力性能を考慮した上でプロジェクトの目標である 7Tops/W という高い電力性能を実応用で実現することができるミドルウェアの開発を行うことである。

深層学習ミドルウェアの検討においては、深層学習ワークロードをカスタムプロセッサで動作させるために、市中に存在する深層学習フレームワークの調査及び本研究開発におけるアーキテクチャの特性の両側面から特に必須となるランタイム要素について抽出し、アセンブラやソフトウェアエミュレータ、実機及び検証用に活用可能なランタイムソフトウェアなどの開発を行った。加えて、本研究開発におけるプロセッサの eFPGA 機能をソフトウェアから利用するために必須である命令切り替え機能の実装及びソフトウェアからの活用も行った。研究開発の結果、試作を行ったソフトウェアフレームワークを用いて実際

の第1世代 Chip 上でワークロードを動作させることができることを実証した。加えて eFPGA 機能についても Workload ごとのハードウェア高速化をソフトウェアから利用可能であることを実際に確認した。

また、eFPGA を活用した命令圧縮アルゴリズムを検討・実装し、eFPGA の評価を実施した。eFPGA を使用した場合に想定した周波数の 1/2 程度の速度でしか命令供給部が動作しないことがわかったものの、eFPGA 上で物体検出手法プログラムの推論コアが動作することを確認した。eFPGA を活用した命令圧縮アルゴリズムでは、ホスト上でサイクル方向に類似な命令を圧縮することで命令サイズを 426bit から 292bit に削減し、命令転送に必要なバンド幅を約 68%に削減できることを確認した。

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 (株式会社 Preferred Networks)

本項目の目標は、実施項目②で開発したプロセッサおよび実施項目③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し、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その性能を測定することである。

本実施項目においては、実際にエッジのロボティクスに本プロセッサを応用するための研究開発として、(1)本計画で開発するチップ上での具体的なニューラルネットワークの実装検討および実装と、(2)ソフトウェアの観点からのハードウェア仕様要求を行った。第1世代 Chip の試作完了後、実際に試作コンパイラを用いた物体検出アプリケーションの実装を行った。実装を行った結果、ロボティクスにおける実応用において必要なレイテンシ(10ms 以下)を実チップにおいて安定的に達成できることを確認できた。加えて、第1世代 Chip が持つ eFPGA 機能を用いること事で、実際のロボティクスワークロードにおいて実行性能の大きな劣化無く、命令サイズを 68%にコンパクト化できることを実際に確認した。

これら結果により、実施項目②で開発を行ったプロセッサにおいて実用的なロボティクスアプリケーションを実行可能であることを確認できた。

4.4. 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」

4.4.1. テーマ概要

図 4.4.1.1 に本研究開発全体像と研究開発実施項目との関連図を示す。

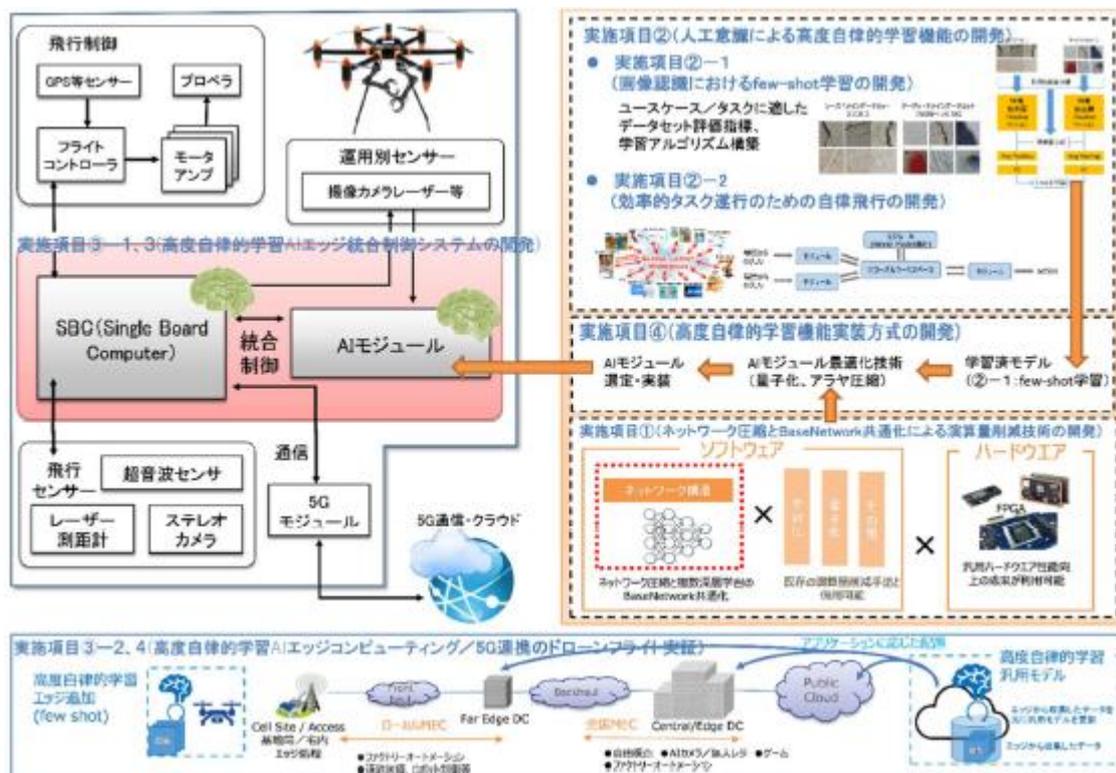


図 4.4.1.1 本研究開発全体像と実施項目との関連図

研究開発実施項目は以下より構成される。

- 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
- ② 人工意識による高度自律的学習機能の開発
- ③ AI エッジ統合制御システムの開発
- ④ 高度自律的学習機能実装方式の開発

①では、既存の深層学習・深層強化学習を対象に、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とともに、実機検証のために主に FPGA への実装を行う。また④では、①で開発したネットワーク圧縮技術及び実装方式を活用し②において高度自律的学習機能として開発するアルゴリズムの AI モジュールへの実装方式の開発を行う。【チップ開発】

②では、より高度な自律化のためのアルゴリズム開発として、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機能の開発を行う。そして、③において、①及び④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【システム開発】、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ドローンフライト実証を行う【総合実証】。

4.4.2. 最終目標と根拠

研究開発実施項目	最終目標
①：ネットワーク圧縮 BaseNetwork 共通化による演算量削減技術の開発	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)
②-1：画像認識における fewshot 学習の開発	学習に必要なデータ量 従来比 1/10
②-2：効率的タスク遂行のための自律飛行の開発	<ul style="list-style-type: none"> マルチタスク実行可能環境下におけるマルチタスクの効率的実行：2 種類以上のタスクを同時に実行するマルチタスク実行効率の 300%増加 ドローン飛行ミッションにおける複雑環境下での衝突回避：静的・動的障害物環境下での既存手法に対する衝突回避性能の 100%増加
③：高度自律的学習 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	<ul style="list-style-type: none"> 検証ユースケース設定とデータ整備、エッジ・クラウド機能配置及び 5G 連携設計 AI エッジ統合制御システムとのインテグレーションを含む、5G・AI エッジ・クラウドシステム検証環境の構築及び、高度自律的学習機能に係る、5G・AI エッジ・クラウドシステム検証環境における実証の実施。 <p>以上を総合し、学習量・計算量削減効果による、システムレベルでの消費電力 (対 GPU) 従来比 1/10 の達成</p>
④：高度自律的学習機能実装方式の開発	高度自律的学習機能アルゴリズムの AI モジュール実装

4.4.3. 目標の達成度

研究開発実施項目	目標達成内容	達成状況
①：ネットワーク圧縮 BaseNetwork 共通化による演算量削減技術の開発	ドローンを用いた人物追尾タスクで開発した、カメラ制御を行う深層強化学習モデルに対してネットワーク圧縮技術を適用した結果、モデル全体のパラメータサイズが圧縮前の 8.4% (10 分の 1 以下) となる事を確認し目標を達成した。	○
②-1：画像認識における fewshot 学習の開発	1/10 量のデータ量で評価値 3%劣化の結果となり目標を達成した。また、モデルは DeepLav v3+の軽量版を使用しドローン上に搭載しフィールド実証を実施した。	○
②-2：効率的タスク遂行のための自律飛行の開発	<ul style="list-style-type: none"> マルチタスク実行可能環境下におけるマルチタスクの効率的実行：既存手法に対して提案モデルは、平均報酬で2倍以上、タスク成功率では既存手法が0%に対して20%の結果が得られ、目標を達成した。 ドローン飛行ミッションにおける複雑環境下での衝突回避：成功率(障害物に衝突せず目標位置まで到達した割合)において、既存手法が約20%のところ、提案手法では約40%と目標の約100%増加を達成した。 	○

<p>③：高度自律的学習 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</p>	<p>・錆び点検のユースケース構築とエッジ環境にてリアルタイムにサビ検知を行い、クラウドへデータ連携を行うエッジ・クラウド機能配置及び 5G 連携設計を実施。</p> <p>・AI エッジ統合制御システムの開発が完了し、2022年12月にKDDIグループの通信鉄塔の撮影およびサビ点検を行う、ドローンフライト実証を行った。フライト中リアルタイムでエッジAI処理による鉄塔錆検出動作が可能であり、AI エッジ統合制御システムが実ユースケースにおいて有効に機能することを確認。また 5G 通信を利用した地上での映像伝送も行い、最大 60Mbps で伝送可能なことを確認。また電力量・作業量において最大約 90%の効率化・システム観点での効率の従来比 1/10 の実現性を確認した。</p>	○
<p>④：高度自律的学習機能実装方式の開発</p>	<p>プルーニング手法に基づくネットワーク圧縮技術と独立に適用可能な圧縮手法である量子化に関する技術開発として、整数 8 ビット (INT8) よりも演算精度が低い量子化ビットを用いた深層学習モデルの圧縮手法と、それに対応可能な FPGA 実装手法として整数 4 ビット (INT4) に対応する実装 IP を開発した。</p> <p>開発した量子化技術を用いて多クラスを扱うタスクにおいて量子化ビット INT8 以下でも精度劣化が抑えられることを確認した。ドローンフライトに量子化技術を適用したモデルを実装し、人物検知タスクにおいて実用可能な推論速度となることを確認した。</p>	○

4.4.4. 成果と意義

研究開発実施項目	成果と意義
<p>①：ネットワーク圧縮 BaseNetwork 共通化による演算量削減技術の開発</p>	<p>複数深層学習の共通化技術である Base Network 共通化技術、および、FPGA に実装可能なプルーニングに基づくネットワーク圧縮技術の開発し、デバイス実装において性能評価と効果確認ができた事は、エッジに搭載可能な学習・推論演算量低減・低消費電力のための技術として今後の FPGA を含むエッジ AI モジュールへの深層学習モデルの実装開発において重要な意義がある。</p>
<p>②-1：画像認識における fewshot 学習の開発</p>	<p>本技術開発により、これまでにデータを取得したことがない条件であっても、そのデータを数枚～数十枚だけ取得して追加で学習させることで、新しい条件でも高い精度の few-shot 学習が実現できた。</p>
<p>②-2：効率的タスク遂行のための自律飛行の開発</p>	<p>・マルチタスク実行可能環境下におけるマルチタスクの効率的実行：提案モデルのアーキテクチャとタスク内容の両方に新規性があり、ドローンのユースケースとの関連については、ドローンの配送分野において、巡回セールスマン問題を(深層)強化学習や他の最適化手法を用いて解く課題は最近でも関連研究があり、ドローンの産業応用という観点から重要な課題の一つであることが示唆されている。</p>

	<p>・ドローン飛行ミッションにおける複雑環境下での衝突回避： DNN を用いた動的障害物の軌道予測、軌道予測の対象・タイミングを限定・調整するための注意機構、回避動作を出力する MPC(モデル予測制御)、を組み合わせたモデル構成は類がなく新規性があり、既存手法に対して、動的障害物が静的障害物の死角に隠れ、ドローンから観測できない場合も含まれる状況において、衝突リスクが少ない障害物回避行動を実現できた。</p>
③：高度自律的学習 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	<p>錆び点検のユースケース構築とエッジ環境にてリアルタイムにサビ検知を行い、クラウドへデータ連携を行うエッジ・クラウド機能配置及び 5G 連携設計を実施し、AI エッジ統合制御システム、エッジクラウド連携システム、5G 通信を利用した地上での映像伝送を行い、最大 60Mbps で伝送可能なことを確認できた。また電力量・作業量において最大約 90%の効率化・システム観点での効率の従来比 1/10 の実現性を確認できた。</p>
④：高度自律的学習機能実装方式の開発	<p>プルーニング手法に基づくネットワーク圧縮技術と独立に適用可能な圧縮手法である量子化技術について、演算精度が整数 8 ビット以下でも深層学習モデルの学習からエッジ化実装までを行える技術要素とフローとを開発し、デバイス実装において性能評価と効果確認ができた事は、エッジに搭載可能な学習・推論演算量低減・低消費電力のための技術として今後の FPGA を含むエッジ AI モジュールへの深層学習モデルの実装開発において重要な意義がある。</p>

4.4.5. 成果の普及

研究開発実施項目	成果の普及
①：ネットワーク圧縮 BaseNetwork 共通化による演算量削減技術の開発	<p>FPGA/組み込み GPU/AI チップ NPU を含むエッジ AI モジュールをターゲットとして AI モデル・アルゴリズムをエッジ AI 実装する、エッジ AI コンサルティング・開発支援を通じたエッジ AI 社会実装支援を継続する。また、AI チップ開発におけるエコシステム形成として AI チップ開発拠点 (AIDC) との連携継続、及び、AI アクセラレータ実証チップ開発 (AI-One/Two) の次期開発計画への参加による、開発した AI-IP の実用検討を推進する。</p>
②-1：画像認識における fewshot 学習の開発	<p>本研究開発における人工意識アーキテクチャの開発成果を踏まえてアルゴリズムおよびシミュレータ開発を進め、KDDI プラットフォーム等を利用して収集した動画像・データを活用した AI モデル学習、及び、ドローンへのエッジ AI 実装を通じて社会実装・事業化を進めることを基本方針として (想定タスク：人物検知、錆点検、障害物回避、アーム付きドローンにおけるアーム動作活用、等)、研究成果の積極的な学会発表・対外発表を進め、意識的な情報処理の機能を備えた高度自律的 AI の実現に向けた研究開発、情報発信により成果の普及を進める。</p>
②-2：効率的タスク遂行のための自律飛行の開発	
③：高度自律的学習 AI エッジコンピューティ	<p>「KDDI スマートドローン」お客様運用メニューとして 2020 年 2Q に提供開始。AI サービスをオプション提供した。2020</p>

<p>ング／5G 連携のドローンフライト実証</p>	<p>年8月に KDDI Accelerate5.0 構想発表で本研究目的の NW・AI・PF・ドローン融合社会を訴求していく。</p>
<p>④：高度自律的学習機能実装方式の開発</p>	<p>(研究開発実施項目①と同様)</p>

4.5. 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」

4.5.1. テーマ概要

本研究開発テーマでは、エッジビジョン AI を超軽量化し短 TAT（短期間での処理）で実装する技術の研究開発に取り組む。

ビジョン AI を活用したエッジコンピューティングの必要性が高まる上、専用アーキテクチャを搭載したハードウェアが登場する一方、半導体プロセスの技術停滞と必要計算量・メモリ量の増加によってハードウェアの種類が多様化し、専用ハードウェア開発の需要も加速している。

このような状況での課題として、ビジョン AI システム開発における開発工数が増大しており、特にビジョン AI システムをエッジデバイス上で高効率に動作させるために、多大なコストが生じている。

このような背景を踏まえ、本研究開発テーマでは、ビジョン AI アルゴリズムからドメイン特化アーキテクチャを自動生成するコンパイラ基盤を開発し、ユーザの目的に真に最適化されたハードウェアアーキテクチャを即時に構築できる開発環境を実現させることを目指す。

4.5.2. 最終目標と根拠

本研究開発では、最終目標として、データ処理の軽量化の技術開発（実施項目 1）、ドメイン固有言語とそのコンパイラの技術開発（実施項目 2）、そしてベンダ非依存の高位合成技術（実施項目 3）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。

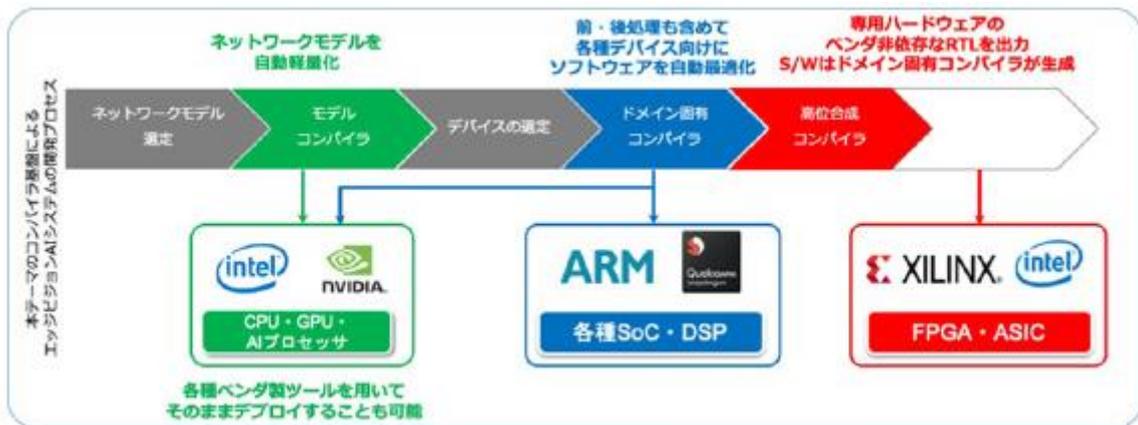


図 4.5.2.1 本テーマのコンパイラ基盤を用いた、エッジビジョン AI システムの開発プロセス

● 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（実施項目1）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（実施項目2および実施項目3）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てているケースは少なく、これが実現できれば、今後増え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、日本では株式会社フィクスター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	本技術(現状)	2018/6	△	×	自社AI	中国市場での	△	×	低	RA

AI 実装技術 (DNNC)					ボード実装	製造工場との連携				
	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
- X・・・競合よりも劣るレベル
- LD・・・Leading
- DH・・・Dead Heat
- RA・・・Run After

4.5.3. 目標の達成度

実施項目 1 では、軽量化実装する技術について、認識精度が 2.4 ポイント劣化したものの、モデルのメモリサイズを 256 分の 1 に圧縮し、組込み GPU (JetsonNano) と比較して推論は 11.9 倍、電力を 48%削減した。

実施項目 2 では、DSL コンパイラの Xilinx FPGA 向けバックエンドを開発し、手動最適化実装と比べて、同等の性能と使用リソース量を保ちつつ、LOC (lines of code : ソフトウェアの規模の指標) を 1/10 以下に削減した。更にこれらのコンパイラを利用可能なエッジビジョン AI アプリケーションの開発&ベンチマークのための Web サービス

「GENESIS DevEnv」を開発した。更に DNN のコンパイル&ベンチマークフレームワーク『Arachne』を開発し、複数種類デバイス向けのエッジ DNN モデル開発&評価の MLOps を確立し、開発工数削減事例を確認した。

実施項目 3 では、DNN コンパイラ TVM の中間表現 Relay および同等レベルのグラフ表現に基づく、より DNN/AI 処理に特化したベンダ（製造元供給者）非依存の高位合成コンパイラの開発を行い、データパス回路のサイズを削減する圧縮方式を開発した。

以上の通り各実施項目において成果をあげた結果、本研究開発テーマ全体として、既存の組み込み GPU に対し、電力効率 20 倍以上となるモデル軽量化手法及び回路のアーキテクチャ設計が完了した。更にコンパイラによる自動コード生成により、開発工数も削減できることが確認できた。

これらにより、エッジビジョン AI を超軽量化し短 TAT で実装する技術を確立するとともに、NEDO 基本計画における 2022 年度最終目標「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」を達成した。

4.5.4. 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った。

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

(reconfigurable : 設計変更が可能な) 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

熱源により機能劣化するデバイス (例 : CMOS イメージセンサ等のセンサ) 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 Windows/Intel、最近では NVIDIA 社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。

4.5.5. 成果の普及

株式会社フィクスターズでは、研究開発で構築したコンパイラ基盤や MLOps 基盤を利用して、エッジビジョン AI の開発や評価を高品質かつ低 TAT で行うソリューションサービスと、その成果によって開発されたエッジビジョン AI の評価を行うことができる Web サービスの提供を行う。

4.6. 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」

4.6.1. テーマ概要

全体として、以下3つの方策にて研究開発の取り組みを行う。

方策1) アーキテクチャー検討と定義

提案する並列プロセッサのアーキテクチャー検討と定義を行う。本並列プロセッサは、グラフ構造で現わされた処理内容の各枝部分をスレッド化して、そのスレッドを実行条件が整った順に動的に実行していく構造を与える。

上記の実現のために、演算器の基本構造、データの供給と格納系などの構造定義と共に、制御系の定義と検証を行うアーキテクチャー検討を最初に実施し、この期間に公立大学法人会津大学（以下、会津大学と記載）において、世の中にある他方式のアーキテクチャーの評価を行い、開発するアーキテクチャーがもたらす計算性能、電力性能の推定について評価並びに見積もりを並行して進める。

アーキテクチャーの検討完了に前後して、以下の順番で開発を進める。並列プロセッサ単独コアで完結する基本ハードウェア設計を開始する。並行して、大規模グラフ向けのスケジューリング制御を担うソフトウェアについて詳細の構成検討を進める(Phase1)。このスケジューラ・ソフトウェアの検討結果をもって、大規模グラフに対応した並列プロセッサを複数コア用いたクラスタ構成のハードウェア設計を行う(Phase2)。論理設計を行い FPGA での実装と性能評価、電力性能評価ができる状況にもっていく。下記の開発環境の整備と併せて、利用できる環境が揃った段階で、ベンチマーク、実証実験と進めていき、速やかに商用化の流れに向けた活動を加えていく。なお、Phase1 で開発した基本構成のハードウェアはエヌエスアイテクス単独事業により製品化を予定する。

方策2) ヘテロジニアス対応コンパイラとハイパーバイザーの開発

本並列プロセッサは、一般の CPU のような逐次実行する機構ではなく、GPU のような同種の処理を多量の並列演算装置を用いて実行するプロセッサでもない。処理内容をグラフ構造で表し、それをスレッドという単位に分解し、並列化して依存関係を踏まえて実行するというプロセスをとるため、ユーザプログラムから並列構造を捉え、これをスレッド単位に分解して、並列性を引き出す必要がある。

将来的には、クラスタ構造の中に特定機能向けの専用アクセラレーターなどの導入の可能性もあるので、それらを組み込んだ場合にもプログラムモデルを大きく変える必要のない仕組みを与える必要がある。これらはオスカーテクノロジー株式会社(以下、オスカーテクノロジーと記載)のもつヘテロジニアス対応のコンパイラ技術を本並列プロセッサ向けに発展させることで実現を図る。

また、本並列プロセッサは、汎用 CPU との連携でシステム動作させることを考えており、実際に適用される時代においては、先に述べた如く多様な処理を実現するためそれぞれの特徴をもつ複数の OS（オペレーティング・システム）を汎用 CPU において同時に実行し、その中で必要となる処理を本並列プロセッサと連携して進めていく必要がある。このソフトウェア機構を評価するために OS を仮想化して複数実行できるハイパーバイザーと言われるソフトウェアが必要となり、この部分を株式会社 OTSL(以下、OTSLと記載)と協力して進めていく。

方策 3) 性能評価指標ならびに実証の進め方

本並列プロセッサを事業化・実用化するには、ハードウェアとしてのプロセッサ、それを駆動するソフトウェアをつくるための開発環境が必要であることは述べたが、さらに、実応用面に即した利用を図って機能、性能を証明していくことにより、実用化に必要な課題を抽出して解決していく必要がある。その活動として、国立大学法人東京工業大学（以下、東京工業大学と記載）と協力して、本並列プロセッサ向けのベンチマークを開発し性能評価の指針として公表する活動をすすめる。また、産業の自動化をめざしている株式会社日立製作所(以下、日立製作所と記載)と協力して、特定の応用面、特にロボットのリアルタイム動作向けアルゴリズムを本並列プロセッサに実装をすることで機能、性能の実証、開発環境の実証を行い、必要に応じて改善を加え、世の中への PR を通じて実用化を加速する。

4.6.2. 最終目標と根拠

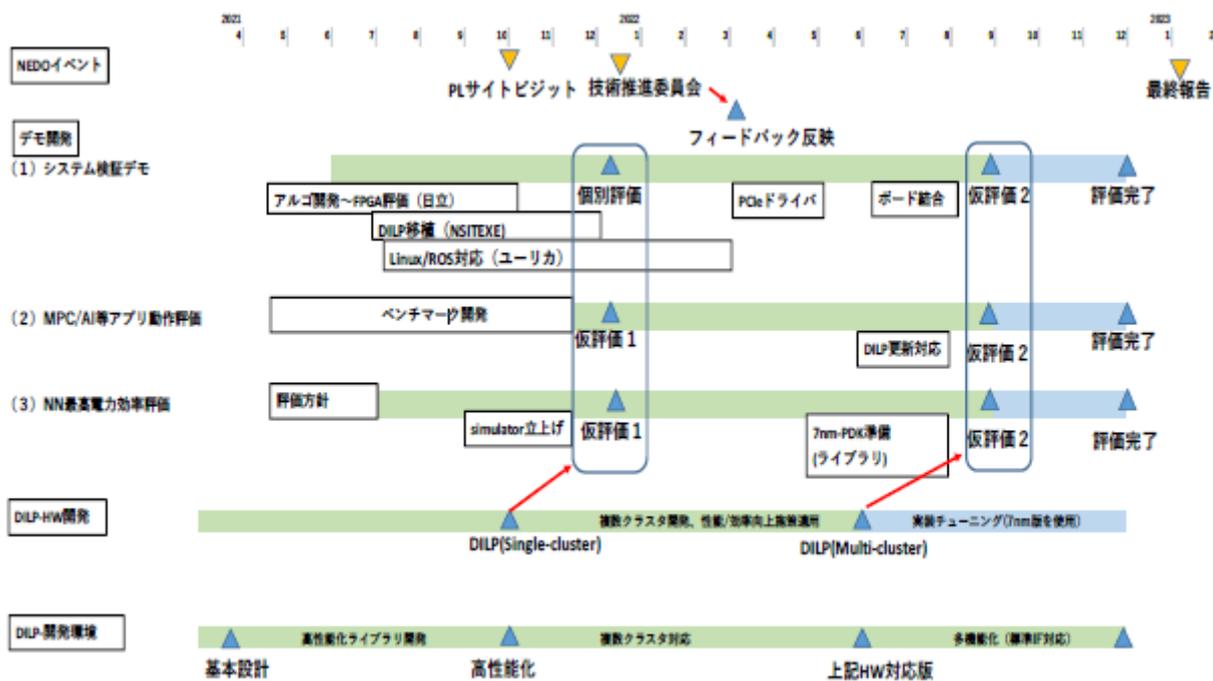


図 4.6.2.1 最終目標までのマイルストーン

テーマ全体としての最終目標は、3 種類のデモと評価を、2021 年 12 月に個別評価（仮評価 1）、2022 年 9 月に仮評価 2、2022 年 12 月に評価完了の 3 ステップで行う。

(1) システム検証デモは、4 社連携して構築したシステム検証環境に、地図生成、軌道計画アルゴリズムを DILP に実装してデモ動作を行う。DILP による高速化による効果を確認することで、DILP の早期事業化の目途をつけることを目標とする。

(2) MPC（Model Prediction Controls：モデル予測制御）/AI 等アプリ動作評価は、最高性能や最高電力効率など理論値での評価にとどまらず、DILP の適用範囲を広くとらえ、実際に DILP 適用アプリケーションとして評価を行うことを目的としている。これは技術推進委員会からのフィードバックを受け、実施した内容である。

(3) NN (Neural Network) 最高電力効率評価は、DILP で実際のニューラルネット処理を行い、15 TOPS/W の電力効率を実現することを確認する。本研究開発においては、方式開発を主眼としチップ開発はスコープ外とするが、電力効率を精度高く評価するために 7nm 物理ライブラリにて DILP を実装して評価を行う。15TOPS/W は従来技術から 10 倍以上となる値で、専用ハードウェアでないプロセッサとしては世界最高レベルの値となる。加えて同環境で実装した DILP の面積効率として、1TOPS/mm² を目標とする。これも 7nm 世代で世界最高レベルの面積効率となる数値目標である。

4.6.3. 目標の達成度

前節に記載した研究開発計画に沿って、3つの評価項目（デモによるシステム検証、MPC アプリ動作評価、電力効率）について仮評価 1（'22/1）、仮評価 2（'22/9）、最終評価（'22/12）月の 3ステップで実施し、目標を達成した。

デモによるシステム検証について、DILP を実装して 4社連携して構築したシステム検証環境に、地図生成、軌道計画アルゴリズムを搭載してデモ動作完了。DILP による高速化も確認まで完了。

MPC のアプリ動作評価では、制御用途でニーズの大きいモデル予測制御(MPC)アプリを開発し DILP に実装し、FPGA 環境にて高速高精度動作確認まで完了。

NN 最高電力効率評価では、仮評価での達成度合いを精査して、DILP 実装に関しては PPA (Performance, Power, Area) 改善を進め、後述する独自の電力効率向上施策も組み込んで最終目標とした電力効率と面積効率(15TOPS/W, 1TOPS/mm²)を 7nm プロセス環境にて達成した。

4.6.4. 成果と意義

(1) システム検証

システム検証環境のハードウェアについて下図に記すように NXP 社製の評価ボード（図 4.6.4.1 左）と DILP を搭載した PCIe 接続ボード（図 4.6.4.1 右）とした。このシステム検証ハードウェア上で、Host 側 CPU 上の Hypervisor で動作する各リソースパーティション内のアプリケーションから DILP を呼出し、空間地図作成やアームの軌道計画アルゴリズムを FPGA に実装した DILP で実行させロボットアームを制御した。DILP 適用により軌道計画時間を従来制御方法に比べて 1/10 以下にできており、作成した動作デモでも、ロボットアームが非常にスムーズに制御できることが実証できた。本システム検証により、DILP がロボット制御のような高精度かつ高速処理が必要なアプリケーション処理に適しており、早期事業化に向けて目途をつけることができたと考える。

■ NXP Semiconductors LX2160Aリファレンス設計ボード

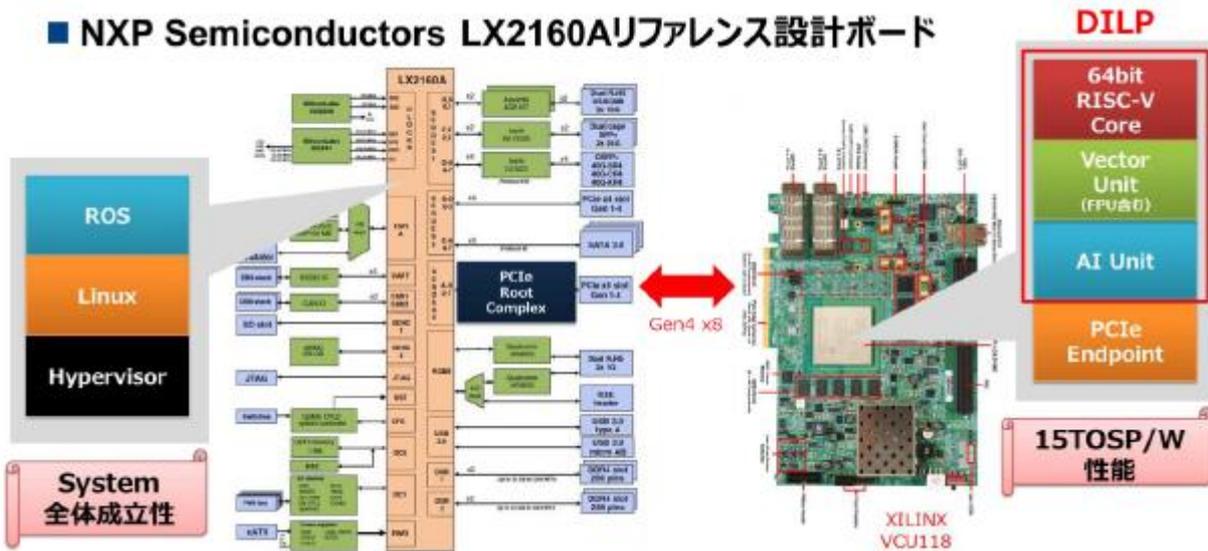


図 4.6.4.1 システム検証ハードウェア

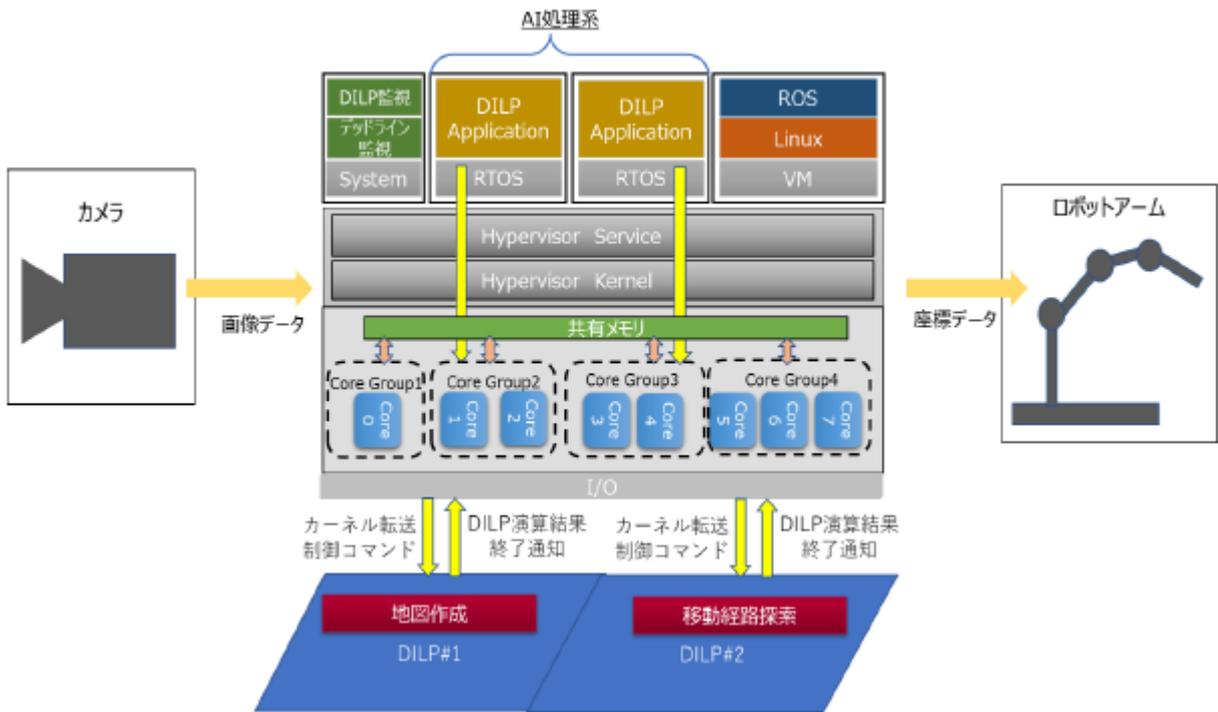


図 4.6.4.2 システム検証構成図

(2) MPC/AI 等アプリ動作評価

DILP は幅広く産業分野への適用が可能と想定しているが、その中でも実用ニーズの高いモデル予測制御(MPC)について処理フローを確立して評価を実施した。モデル予測制御とは、制御対象(プラント)の出力を予測し、予測された出力が制御目的をどの程度満たすかを評価関数 H で記述し、 H が最適となる操作量をリアルタイムで探索する制御手法であり、最適操作量探索手法(ソルバー)、プラント出力予測手法(モデリング手法)を組合せて活用する。

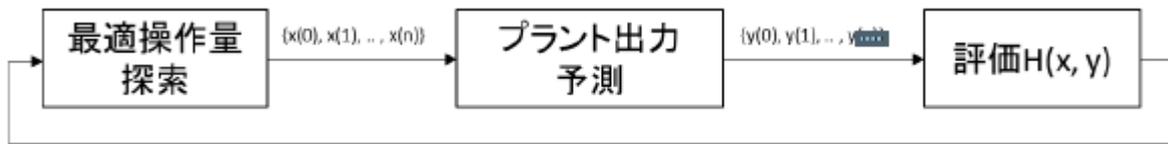


図 4.6.4.3MPC のフロー

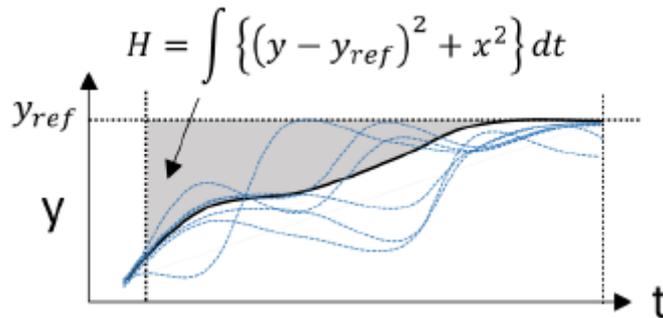


図 4.6.4.4MPC の評価関数

今回、MPC のユースケースとして、図 4.6.4.5 に示す、既存従来手法の PID 制御に、MPC による指令値生成をアドオンして制御性能向上を評価する環境を構築した。図 4.6.4.6 に評価結果を示す。従来 PID 制御では、出力値の誤差が最大で 0.291 秒、平均で 0.098 秒であったのに対して、MPC をアドオンした出力値では、誤差が最大で 0.120 秒、平均で 0.030 秒に短縮できた。これは最大誤差が 41%に、平均誤差が 31%に精度向上できていることを示している。この MPC 評価により、FA(Factory Automation) 分野をはじめ精緻な制御を要する分野で DILP の事業化につながる成果が得られた。

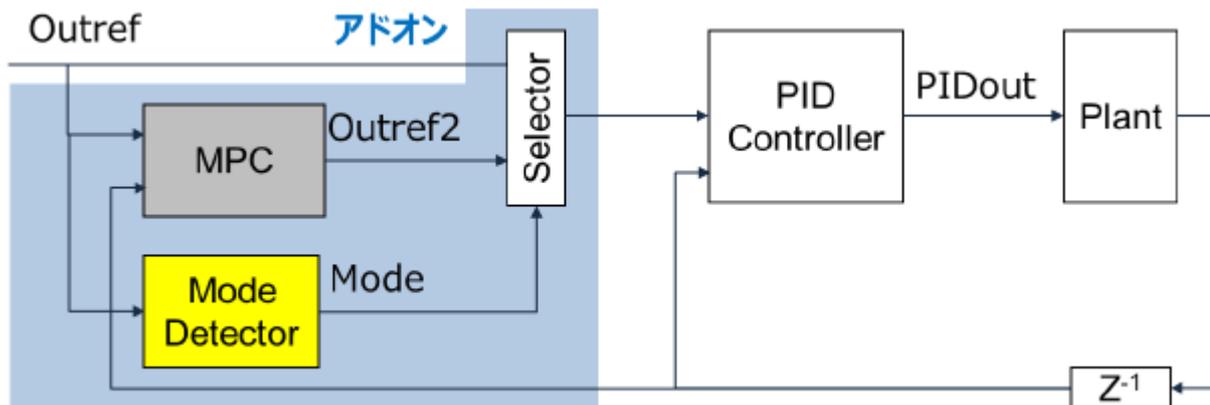


図 4.6.4.5MPC の評価環境

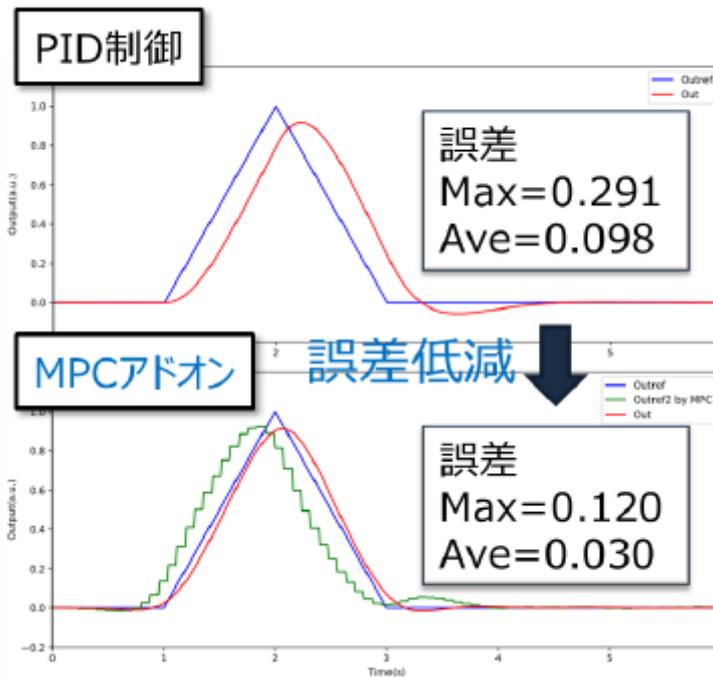


図 4.6.4.6MPC 評価結果

(3) NN 最高電力効率評価

下図に示すように DILP を 7nm プロセスライブラリで物理実装を行い、電力効率と面積効率評価を実施した。電力評価は VGG16 を評価プログラムとして、以下の低電力技術を DILP に対して独自に適用して電力効率を高めている。

①Tiling&Layer-Fusion

Tiling&Layer-Fusion 技術は、入力画像を分割 (Tiling) し、分割したデータごとに複数のレイヤー処理の入出力を連結 (Layer fusion) して処理することにより、中間データの外部メモリへの退避を削減し、電力効率を向上させる手法で、従来手法では単純な Layerfusion のため、重複した演算を実行する必要があったが、演算結果を再利用する独自方式を実装することにより重複演算を無くし、従来手法比で 50%の演算量を削減した。

②Winograd's minimal filter algorithm

Winograd's minimal filter は、畳み込み演算の乗算回数を削減する手法で、本開発では、Winograd's minimal filter のパラメーターを探索し、電力削減効果と回路面積のトレードオフ関係が最適となる 5×5 行列入力から直接 3×3 行列を出力する構成を適用した。これにより従来の畳み込み演算に対し演算量を 60%削減した。

③全結合層演算の低ランク化

全結合層の行列演算に特異値分解による低ランク化を適用した。本手法は一般的に適用されるプルーニング (Pruning) と異なり、ゼロ検出して演算をスキップする処理が不要であり、ベクタプロセッサで効率よく演算が可能となり高電力効率の実現に寄与した。全結合層の行列演算の低ランク化は AI 処理精度を評価しながら最適なパラメーターを探索し、全結合層の演算量を従来手法比で 17%削減した。

図 4.6.4.7 に DILP の内部構成を、図 4.6.4.7 に電力効率評価対象とフローを記す。

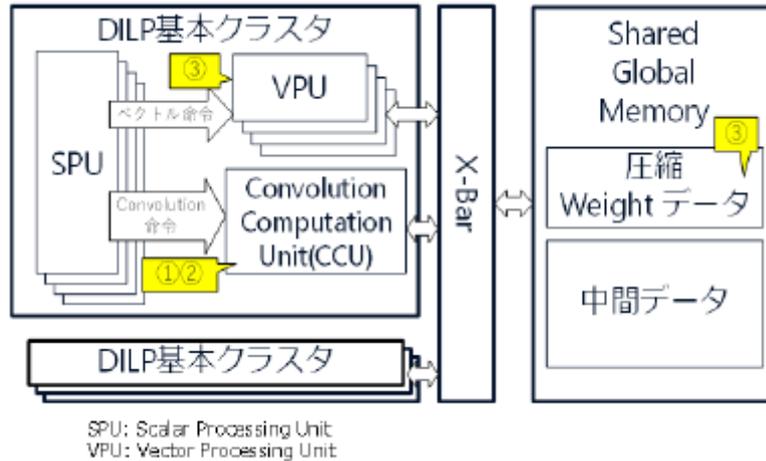


図 4.6.4.7. DILP 内部構成

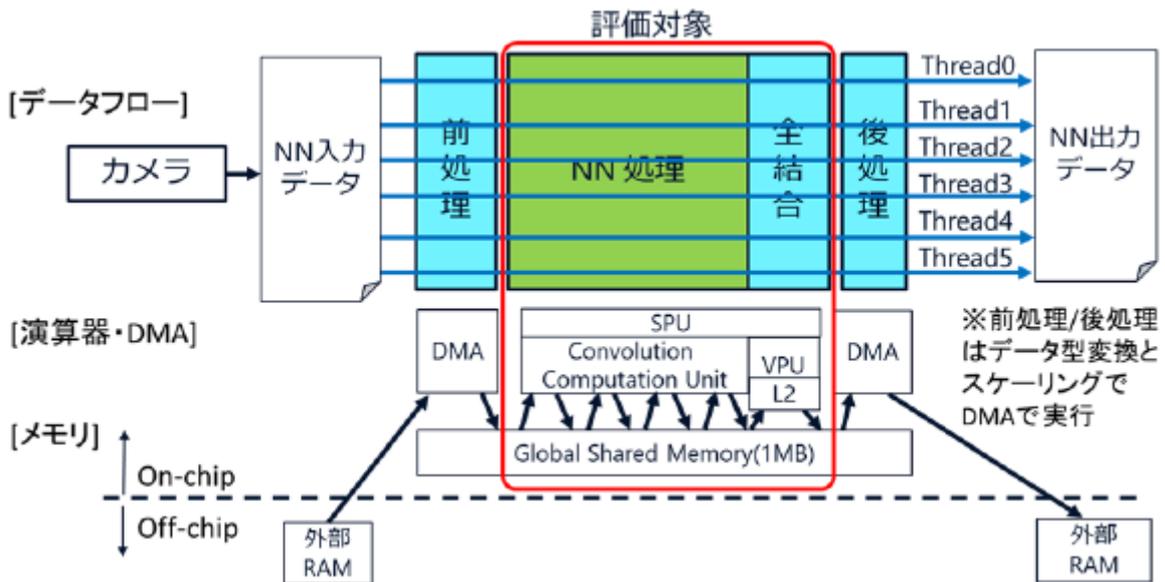


図 4.6.4.7. 電力効率評価対象とフロー

電力効率評価の結果を図 4.6.4.8 に示す。目標値 15TOPS/W に対して 15.1 TOPS/W、面積効率評価の結果、目標値 1TOPS/mm² に対して 1.36 TOPS/mm² となり、DILP のアーキテクチャーコンセプトおよびそのハードウェア実装が PJ 当初に掲げた目標を実現した。これらの値は 7 nm プロセスでは世界最高レベルの値であり、特に専用ハードウェアではないプロセッサで実現している点で非常に意義が高い結果である。

電力効率評価	評価値	面積効率評価	評価値
実行時間(msec)	14.5	Logic面積(mm ²)	0.95
電力(W)	0.141	SRAM面積(mm ²)	3.42
OP数(TOP)	0.0309	合計面積(mm ²)	4.37
TOPS	2.14	論理TOPS@500MHz	5.96
TOPS/W	15.1	TOPS/mm²	1.36

図 4.6.4.8 電力効率と面積効率の結果

4.6.5. 成果の普及

これまで述べたように本研究開発の成果により、早期の事業化が可能であると考えられる。DILP 事業化・実用化に向けて、まず、AI 処理を含めたアクセラレーターニーズが高い、制御系組込みでは FA 分野、自動車用途で車載制御分野を主要ターゲットとして、本 PJ でのシステム検証結果を活用して拡販活動を開始、システム検証デモも活用して幅広い分野での事業化を目指す。本プロジェクト内でのソフトウェアプラットフォームベンダー、ユーザーとの連携し、本プロジェクト外の自動車、FA、IoT などの各メーカーとの製品化に向けた協議を進めていく。また、アーキテクチャーに依存しない部分を中心に、本プロジェクト外のツールベンダーとも、コンパイラやデバッガの供給について協議を進める。さらに、広報活動としては、プレスリリースをはじめ出版社との市場活性化と普及に向けた活動も積極的に行っていく。

4.7. 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」

4.7.1. テーマ概要

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の達成を目標とし、それらの定量的な達成指数として、自動運転の電力効率を示す「消費電力あたりのユースケース数」を定義した。この指標を考慮しながら、Cargo Transport と Robo-Taxi において研究開発要素のある主要な処理に対してアクセラレータの研究開発を行い、各処理の消費電力を 1/10 以下（電力効率 10 倍以上）にした。リアルタイム性が必要な処理に対しては、RISC-V ベースのメニーコアプロセッサと OS とミドルウェアを開発して、各処理の処理時間のばらつきを 1/6.3～/10.3 に抑制した。

4.7.2. 最終目標と根拠

「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」として最終目標を設定した。

最終目標と設定根拠

最終目標（2022 年度末）	設定根拠
10 倍の電力効率（性能/消費電力）	自動運転において公道実験に取り組む多くは車両のトランクに汎用 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 200W 程度消費している。実用化に向けてこれを 20W(競合優位性を得るために設定した目標値、200W の 1/10)以下に引下げた上で、同等の性能が必要なため。

自動運転コンピューティングの性能/電力効率を表現する指標として「消費電力あたりのユースケース数」を提案した。(図 4.7.2.1)

本研究開発は、ハードウェアとソフトウェアの協調設計で、主にハードウェアによる省電力化を主目的に据えているため、自動運転ソフトウェアの能力を示すユースケース数の増加は研究開発対象外として、要求されるユースケースを達成するために必要な消費電力を 1/10 とすることで 10 倍の性能/電力効率を達成する。

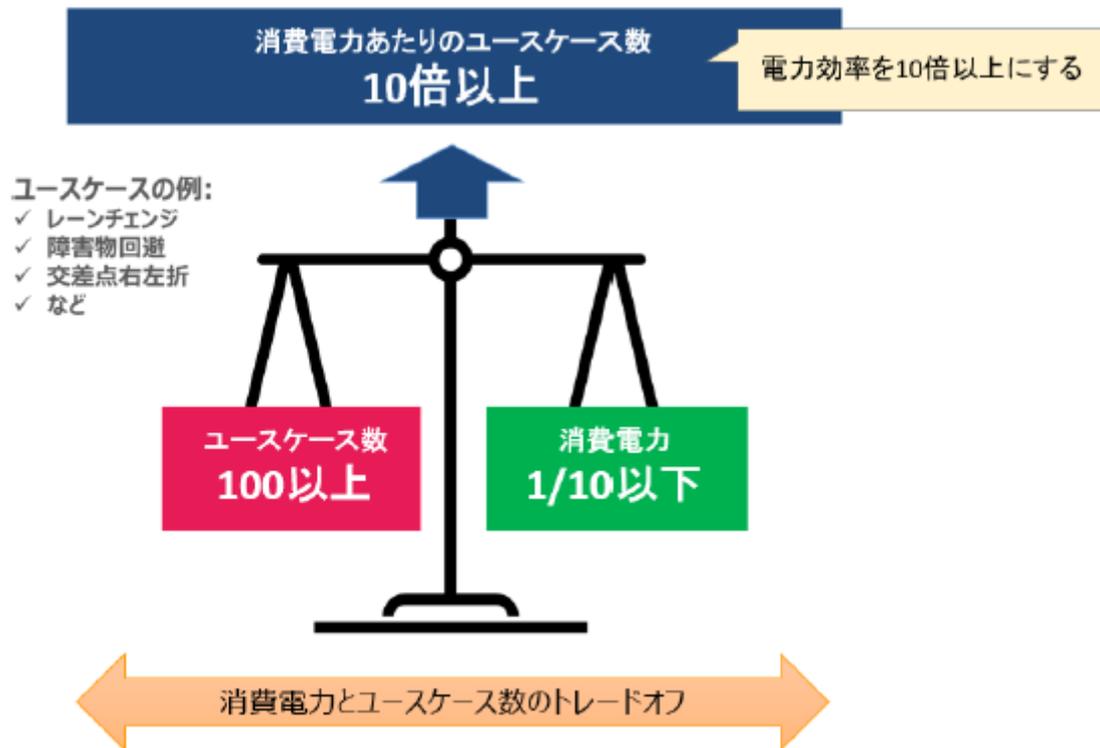


図 4.7.2.1 「消費電力あたりのユースケース数」の概念図

消費電力を削減するためには、消費電力の大きい処理から順に対策を打つことが効果的である。SoC とソフトウェアプラットフォームの研究開発であるため、自動運転コンピュータおよび SoC の消費電力を計測した。(図 4.7.2.2)

消費電力の内訳を大別すると、1. カメラ画像処理による物体認識および信号認識と、2. LiDAR 点群処理による物体認識 (GPU) と、3. その他の処理 (CPU) であった。

1. カメラ画像処理による物体認識および信号認識は、消費電力が大きく研究開発要素がある畳み込みニューラルネットワーク (CNN) を研究開発対象とした。2. LiDAR 点群処理による物体認識は、認識アルゴリズムの移り変わりが激しいため、本研究開発の対象外とした。

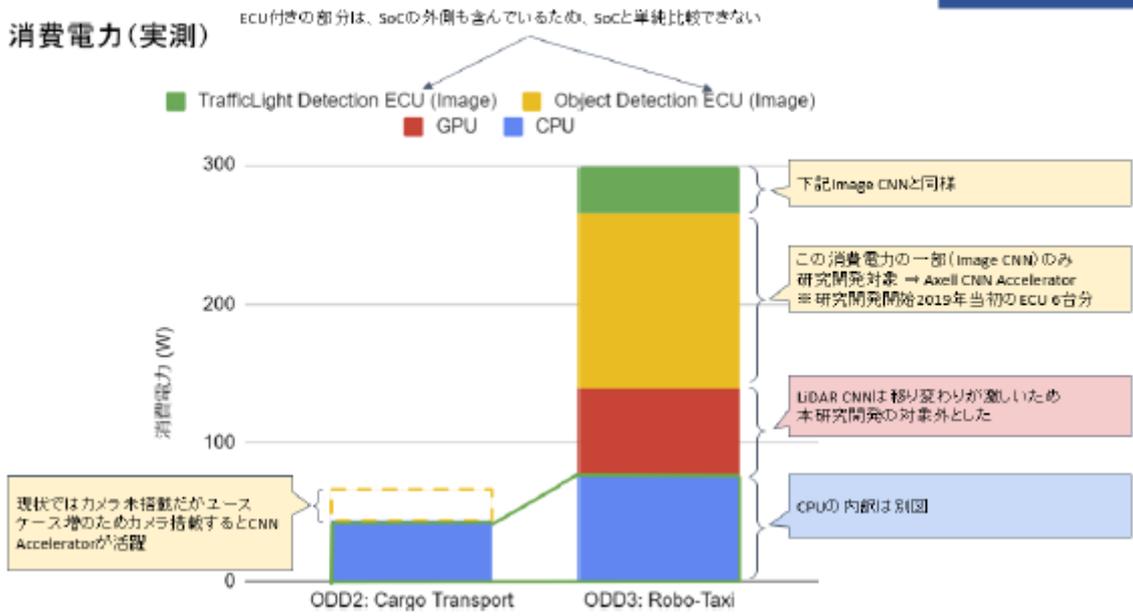


図 4.7.2.2 各 ODD における消費電力(実測値)

※ODD(Operational Design Domain)とは、運航設定領域を示す。ODD に応じて車両構成が異なり、ODD2 は工場等の敷地内を想定、ODD3 は公道走行を想定した車両を示している。

前述の 3. その他の処理 (CPU) は、別ハードウェアにオフロードしていないあらゆる自動運転処理の集合体である。「CPU 使用率」と「CPU の消費電力やメモリ量および帯域」は相関関係にあるので、CPU 使用率の高い LiDAR の Sensing (点群データの前処理) と、LiDAR の Localization (自己位置推定) を本研究開発対象とした。

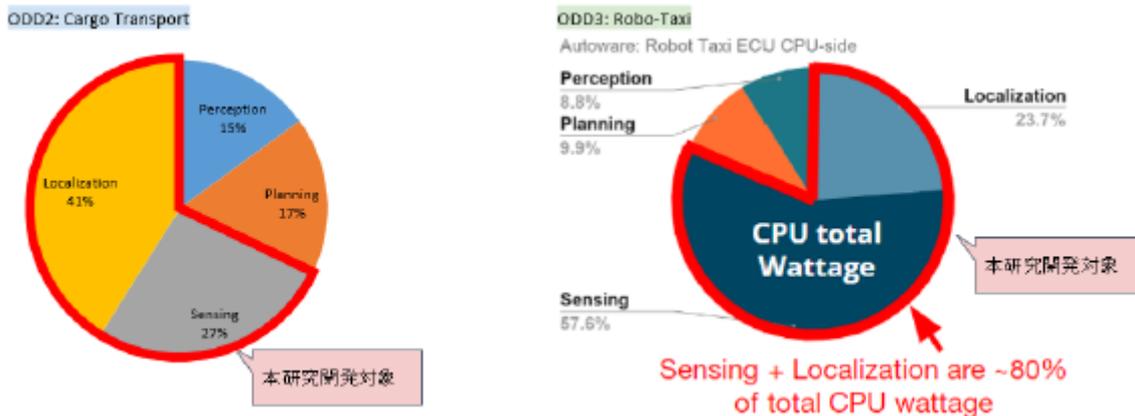


図 4.7.2.3 本研究の範囲 (ODD 毎の CPU 使用率)

本研究開発では、ターゲットとした自動運転処理を省電力化可能なハードウェアアーキテクチャを設計した。自動運転ソフトウェアは今後も成長することから、現状の自動運転ソフトウェアに対して SoC を最適化することは得策ではなく、ハードウェアアーキテクチャが今後の自動運転ソフトウェアに対してスケーラブルであることが重要である。そのため、研究開発対象の中でも主要な処理をピックアップして有用性を示す。

4.7.3. 目標の達成度

自動運転車両	指標	ベースライン	成果(電力:量産時)
Cargo Transport	消費電力 <small>(概算)</small>	118	(机上計算) 52 46%削減
	ユースケース数 <small>(概算)</small>	(2019年時点推定) 20	100
Robo-Taxi	消費電力 <small>(概算)</small>	300	(机上計算) 152 49%削減
	ユースケース数 <small>(概算)</small>	(2019年推定) 60	300

自動運転処理	指標	ベースライン	成果(電力:量産時)
カメラ物体認識	消費電力	30.0 [W]	2.30 [W] 92%削減
LiDAR自己位置推定(NDT)		21.7 [W]	1.90 [W] 91%削減
自己位置推定(EKF)		5.4 [W]	0.40 [W] 93%削減
点群除去(自車)		5.4 [W]	0.02 [W] 99%削減
点群除去(ミラー)		5.4 [W]	0.02 [W] 99%削減
縦横制御統合	変動係数	0.65	0.10
シフト制御		0.74	0.07
制御機能切替		0.60	0.07

消費電力
1/10以下

SoC : アクセラレータ

SoC : メニーコア

図 4.7.3.1 本研究開発成果とベースラインの比較

本研究開発成果とベースラインの比較を図 4.7.3.1 に示す。ベースラインは本研究開発成果の非適用時もしくは本研究開発開始時（2019 年度）の各指標値を示す。

図 6 の上表は、本研究の成果および将来量産時の想定プロセステクノロジーに適用した場合の電力削減効果の概要を示した表である。本研究開発成果であるアクセラレータを高性能 LiDAR に今後対応した場合を想定し消費電力概算を行っている。本研究開発成果である自動運転アクセラレータの対応 LiDAR 追加によって、Cargo Transport で消費電力 46%削減、Robo Taxi で消費電力 49%削減できる見込みである。

図 6 の下表は、電力削減効果の詳細（各ブロックごと、等）を示した表である。いずれの表も Gate-level simulation 等のシミュレーションで見積もった消費電力である。下表は、本研究開発でターゲットとした自動運転処理に関する消費電力を示している。

- カメラ物体認識 SoC：ベースラインは Autoware で使用している GPU ベースの既存 AI デバイスの TDP、研究開発成果はパワー解析の結果である。
- LiDAR 点群前処理および自己位置推定：ベースラインは現行装置（x86 3GHz CPU 搭載）の TDP に CPU 利用コア数を乗じた値。

本研究開発対象としてアクセラレータ化した主要な処理については性能/電力効率 10 倍を達成した。リアルタイム性が必要な処理に対しては、RISC-V ベースのメニーコアプロセッサと OS とミドルウェアを開発して、各処理の処理時間のばらつきを抑制することで、実用化に向けてアクセラレータとメニーコアを組み合わせたアーキテクチャの有用性を示した。

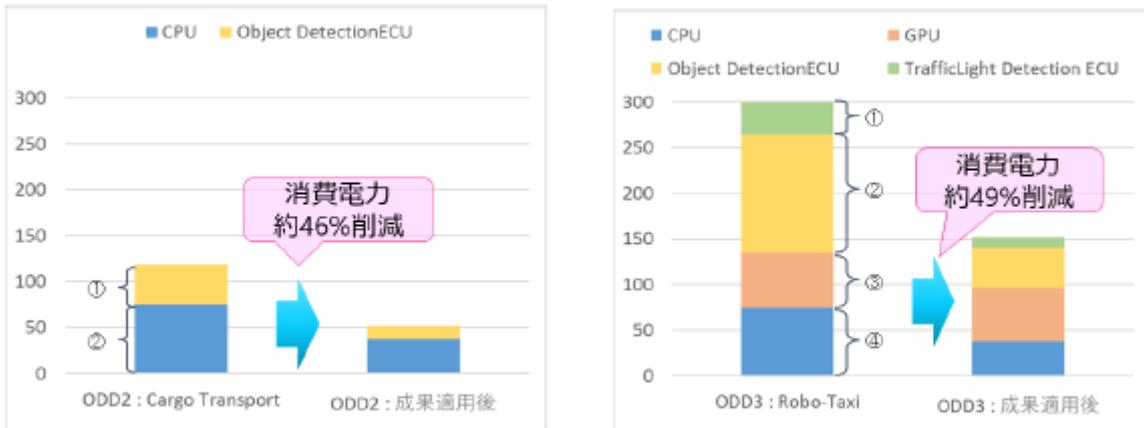


図 4.7.3.2 ODD2,3 ベースライン消費電力(縦軸 : W) (成果適用前 / 適用後)

自動運転車両コンピュータ消費電力のベースラインと研究開発成果の考え方を示す。ベースラインは本研究開発成果の非適用時の Autoware で使用されている x86 CPU および既存 AI デバイス GPU+SoC の消費電力実測値である。成果適用前と適用後の消費電力の変化を図 4.7.3.2 に示し、以下で Cargo Transport および Robo-Taxi それぞれの消費電力について説明する。

[ODD2 : Cargo Transport 消費電力]

下記の成果の組み合わせで約 46%の消費電力削減効果が得られる。

- ① 現状は物体認識用のカメラ使用していないが、工場内対応ユースケースが増加して、2台の前後カメラを搭載した場合は CNN アクセラレータの成果適用が可能
- ② LiDAR アクセラレータにより CPU 使用率を 1/2 程度低減可能で、Robo-Taxi とほぼ同様の効果が期待できる

[ODD3 : Robo-Taxi 消費電力]

下記の成果の組み合わせで約 49%の消費電力削減効果が得られる。

- ① CNN アクセラレータは、信号認識処理(Traffic Light Detection)に現時点では対応していないが、検出処理自体は類似したものであるため、将来応用可能であると見込んでおり、下記②とほぼ同様の効果が得られると見込んでいる
- ② 人や自動車を検出する自動運転のセンサー系主要処理 (Image CNN) に対し、新規開発の Axell CNN Accelerator を適用することで、1/3 程度まで消費電力を低下させることが可能

※(参考)研究開発開始 2019 年当初は、ECU 6 台分でありそれと比較

- ③ 既存 HostECU 内の GPU 処理 (=主に LiDAR CNN)については、アルゴリズムが定まっていないこともあり、今回は研究対象外とした
- ④ Sensing 点群処理アクセラレータで CPU 使用率を 1/2 程度に低減可能

[補足 : ODD2/3 CPU 全体消費電力低減について]

Localization アクセラレータと Sensing アクセラレータの効果はセンサー構成(=ODD)によって異なる。現行 CPU の消費電力低減についての詳細グラフを図 4.7.3.3 に示す。

Localization の CPU 使用率が 1/10 となる理由は、アクセラレート対象の処理である NDT(Normal Distribution Transform) が CPU 使用率の支配的な値であり、その処理をアクセラレータにオフロードした為である。研究成果チップ活用により 1/10 の使用率となることが確認できた。

Sensing アクセラレータの対象処理は車両サイドに設置してある近距離向けの LiDAR(VLP-16)であり、ここから得られる点群データに対する 5 つの処理 (Convert、Cropping、Transform、Distortion Correction、Outlier Filter) のアクセラレート機能を FPGA 実験で確認できた。メイン LiDAR(VLS-128)処理は、アルゴリズム自体は変わらないがデータ量が大いいため、成果物である近距離向け LiDAR(VLP-16)用のアクセラレータでは性能不足であるが、LiDAR アクセラレータの SIMT アーキテクチャのスケラビリティを活かして、今後のアーキテクチャ改善により処理可能とする計画である。以上の結果として、現時点では FPGA 実装段階ではあるものの、今後の SoC 化によって Sensing 処理全体を使用率 1/2 に低減させる効果が見込める。また対象処理を拡大するなどにより、さらなる改善の余地がある。

ODD2 では Localization の割合が大きいため、これを 1/10 にすることで全体 37%の低減が可能になる。Sensing も併せると 51%の消費電力低減が見込まれる。

ODD3 では Sensing アクセラレータ効果が高くなり、29%低減が可能となる。Localization に関しても 21%低減が可能となり合わせ 50%の消費電力低減を見込む。

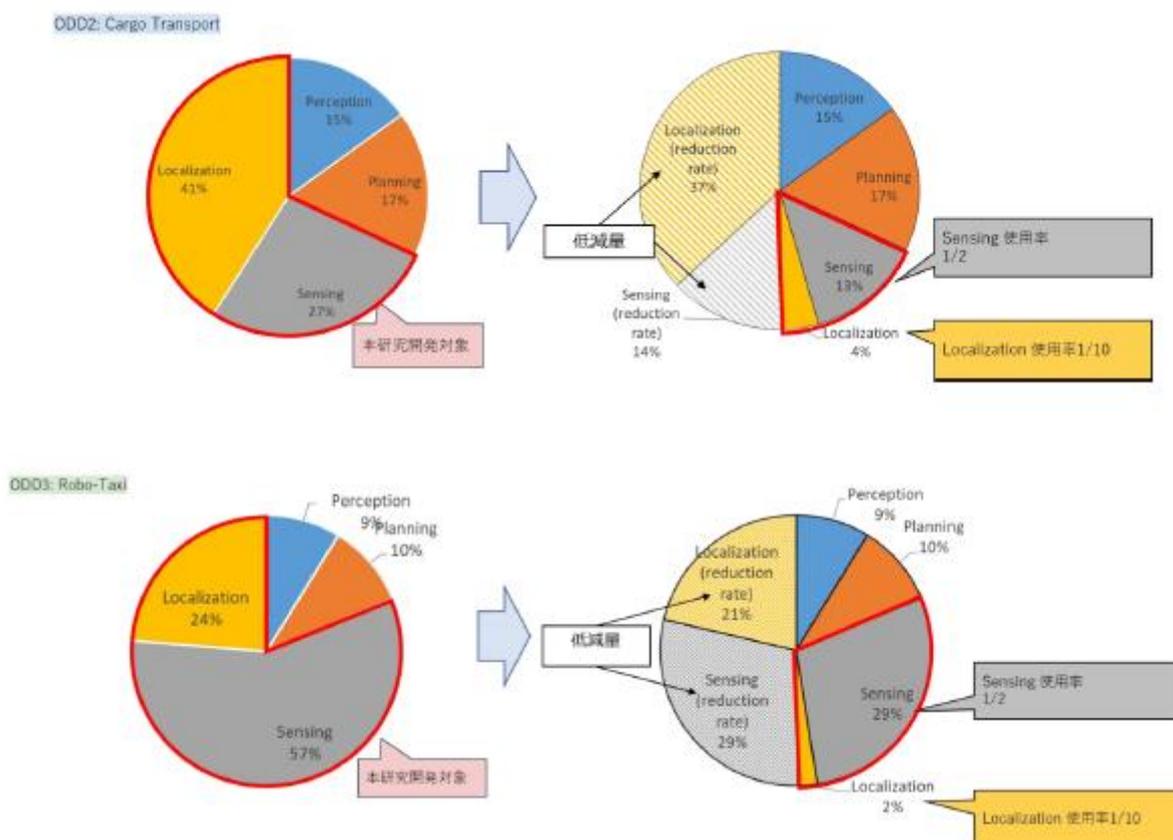


図 4.7.3.3 ODD2/3 CPU 効率化説明
(右側が成果適用後で、斜線箇所が消費電力低減可能箇所)

4.7.4. 成果の意義

本研究開発成果は、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ために要素技術を研究開発し、オープンソースソフトウェアである **Autoware** を利用した実証実験を通して、本研究開発で提唱した技術コンセプトの事業性があることを実証した。

4.7.5. 成果の普及

本研究開発成果は、事業性の高いコンポーネントからユーザに対して普及させていく。具体的には、ドキュメントの提供や品質の向上、開発キットの提供などを通して、ユーザが気軽に利用できるようにする。

4.8. 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」

4.8.1. テーマ概要

本テーマでは IoT エッジのセキュリティー基盤となるチップ統合環境を構築することを目的とし、実施項目を大きく以下の 4 つに分けて研究開発を進めた。

実施項目 1 : セキュア MCU アーキテクチャー

実施項目 2 : トラスト実行環境 (TEE)

実施項目 3 : アーキテクチャーの産業用途即応化

実施項目 4 : セキュリティー社会実装 PoC 設計

実施項目 1 : セキュア MCU アーキテクチャー (ハードウェア基盤技術開発)

AI チップでサイバーセキュリティー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ー主要概念を開発する。チップのセキュリティーでは安全な鍵注入や暗号技術の安全な管理がユーザーから求められている。これに応えるために RISC-V 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。セキュア MCU のための固有鍵サーバーと、セキュア AI エッジデバイスのための鍵管理サーバー (実施項目 4) の連携方法を開発する。セキュア MCU は RISC-V と同一 SoC 上に組み込まれることを想定し、低消費電力、物理的耐性を持ち、機能性、安全性を確保する。

実施項目 2 : トラスト実行環境 (TEE) (ソフトウェア基盤技術開発)

TEE の RISC-V 上での実装方式の開発と、RISC-V International 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 Global Platform において仕様が策定されており、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ため本プロジェクトでは技術開発と並行して、RISC-V International の WG にも参加して標準化の議論を進める。本プロジェクトの TEE ハードウェアで使われる RISC-V (64bit コア) の起動等の動作はセキュア MCU により検証され、更にその検証は TEE ソフトウェアによってネットワークを経由して第三者機関で検証出来るよう構築する。これにより組み込みユーザーがハードウェアラスト実行環境を提供する。

本プロジェクトで開発する TEE ソフトウェアの Trusted OS や Trusted Application は AI エッジ・産業用途に供する。特に Trusted Application としては、ネットワーク経由で OS の更新を行う OTA (Over The Air) やアプリケーションの実行をネットワーク経由で監視する Remote Attestation を作成することで、安全に OS/アプリケーションが更新・実行されることを確認できるようにする。OTA および Remote Attestation を TEE としてビジネスに用いた具体例はなく、本プロジェクトの開発成果を世界初の共通基盤とする。また、Trusted Application に関する管理プロトコル (TEEP: Trusted Execution Environment Provisioning) が IETF で提案されているため、これらとも連携を取り、標準化することで産業応用に供するものにする。Trusted Application の取得先および Trusted Application 自身の信頼を検証するための鍵管理手法を開発する。

実施項目3：アーキテクチャーの産業用途即応化（普及促進活動と効果検証（従来チップからの移植容易性検討））

本プロジェクト当初の目標である「産業用途で求められる RISC-V 企画に従った CPU の研究開発(マイクロアーキ検討等)」は、RISC-V International における追加アーキテクチャー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た。2020 年度からは「広範な市場ニーズを把握するためのユーザーコミュニティ組織の形成」を実施し、市場から探索したニーズを他の実施項目の開発仕様に追加反映した。ユーザーコミュニティ活動においては、別途定める研究成果のユーザーコミュニティポータルフォリオに基づいて、市場に対して開示可能な成果物から開示した。

実施項目4：セキュリティー社会実装 PoC 設計（具体的なユースケース検討）

上記開発技術を活用し、具体的な AI エッジの産業ユースケースの模擬環境での実証を行う。サービス運用などに使う業務管理サーバーとエッジ処理ユニットとのコネクションが成立するためにデバイス真贋証明、暗号鍵や証明書の発行、保管、失効等の暗号鍵管理機能の確認を通して、事業活用の有効性を確認する。

また、本プロジェクト開始から 2019 年度前半までの期間で初期検討を行い、エッジ AI のセキュリティーに関するエコシステムを構築するために必要な技術要素を検討する。初期検討の結果に基づき、2019 年度後半に開発技術の目標仕様を詳細に定義し、研究開発を推進する。

最新の業界動向を踏まえると、本プロジェクトで開発するオープンなセキュリティー技術を広く普及させるためには、開発用の上位環境をセットで提供することが重要である。初期検討の中で、準備すべき上位環境の形態について仕様検討する。そして、一案として、実施項目4「セキュリティー社会実装 PoC」の中で上位環境の開発に関する開発も強化する等の合理的な体制を議論し、予算の増額提案も視野に入れて検討する。

4.8.2. 最終目標と根拠

本テーマ全体として、IoT エッジのセキュリティー基盤技術となる、RISC-V をベースとしたチップ統合環境を構築し、第三者が検証可能なホワイトボックスなものとするため、ソースコードレベルで設計資産をオープンにして、業界全体で活用可能な環境を提供することを最終目標とした。

IoT は様々な業界の業務を高度化、効率化する可能性を持った概念であるが、様々なデバイスが自由に接続、切断を繰り返すその特徴から、セキュリティー管理がその実運用上の大きな課題となっている。本テーマが最終目標としたエッジセキュリティーの統合環境の構築と普及は、IoT システムを実運用する上で必要不可欠な要素である。

4.8.3. 目標の達成度

本テーマ全体として、エッジセキュリティーの統合環境である TrustedRV の開発と実装を完了し、普及促進活動として「オープンフォーラム」を開催し、本テーマの取組みの啓もうを行った。本テーマ全体として、目標は達成できたと考える。

4.8.4. 成果と意義

本テーマで開発した TrustedRV は、RISC-V ベースで構築された信頼の起点と TEE を統合した世界で唯一のチップアーキテクチャーである。これまで、チップセキュリティーの

実装のためには高価なライセンス料を伴う商用ツールしか存在しなかったが、オープンISAであるRISC-Vを活用したTrustedRVにより、各サービス事業者が、セキュアなチップ環境の構築を低コストで実現し、それぞれの競争領域に開発リソースを集中できる事業環境が提供できる。

本テーマの成果一覧

実施項目	研究開発の目標	達成度
① セキュア MCU アーキテク チャー	サーバー、クライアント PC 等の周辺システムを含めた全体仕様書。	SecureUnit 側のネットワーク対応を実施し、サーバー、クライアント PC とセキュア通信ができるハードウェアプラットフォームの仕様を策定し、FPGA 実装を完了。
②-1 RISC-V 版 TEE 仕様策定および ハードウェア開 発	RISC V International の TEE アーキテクチャー拡張規格準拠チップ動作完。 社会実装 PoC の統合フィールド実証システム向け、チップ提供完。全体結合テストにおいて、目標性能の達成完。	実施項目①と連動して動作する TEE アーキテクチャーのハードウェアプラットフォームをリリースし、PoC システムへ適用の完了と、全体結合テストを実施し、目標機能・性能の達成を確認。
②-2 トラスト実行環 境 (TEE)	統合環境上での TEE ソフトウェアの動作確認、TEE の既知の脆弱性に対する対処技術、セキュリティー、およびアーキテクチャーに関する一流国際会議での発表。	Remote Attestation を Trusted RV で実現。国際貢献として IETF で TEEP プロトコル仕様の作成。一流国際論文 IEEE Access 2 件を含む全 8 論文を発表。
③-1 産業用途 RISC- V	ユーザーコミュニティ活動の実施報告。	1. オープンフォーラム開催：合計 5 回 2. 賛助会員：8 法人 3. 研究成果物のリリース：20 件を達成 4. その他：10/18-21 CEATEC への NEDO 出展にて研究成果のデモ展示。
③-2 ソフトウェア移 植容易性	制御用 OS の移植容易性を実機で実証する。	FPGA 上で RISC-V に Bi-Endian とレジスタサイズ可変機能を追加し、制御 OS を動作させることで、Endian やレジスタサイズの異なる既存 CPU から RISC-V へのソフトウェア移植が容易化されることを実機実証完了。
④-1 セキュリティー 社会実装 PoC 設 計	PoC システムの模擬環境での実証およびシステム評価完。	車載 ECU 向けユースケースの模擬環境での実証を行い、コネクテッドカー乗っ取り等のハッキングをローカル検証局により防止するシステムの評価を完了。
④-2 セキュア AI エッジ応用ライ フサイクルの社 会実装シミュ	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完。	RISC-V TEE システムのライフサイクル仮説設定完検証情報の設定とヒアリング実施完了。

レーション		
-------	--	--

4.8.5. 成果の普及

事業期間中に実施したオープンフォーラムにて、チップセキュリティーの重要性の啓もう活動と TRV の要素技術及び応用技術の訴求活動は一定程度できたと考えるが、引き続き、各事業者が国際会議や標準化団体での活動を通じて、成果の普及を進めていく。

4.9. 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」

4.9.1. テーマ概要

エッジ AI 向けを含む世界の半導体市場は 2030 年に 100 兆円に拡大すると見込まれており [1]、それらの応用先も含めた CPS/IoT 関連市場は 2030 年に世界で 404 兆円規模に拡大すると見込まれている [2]。一方、セキュリティ上の脅威や攻撃が原因で市場の縮小・停滞・混乱事例も出始めている(例えば、米国で 140 万台の車のリコールが発生していること [3]、不正利用により 7pay 事業が停止となりセブン銀行の決算でセブン・ペイに約 30 億円の損失が計上されたこと [4]など)。本研究開発テーマでは、セキュリティが原因で本市場の拡大が阻害されることを防止するため、AI エッジデバイスに対する脅威や攻撃と対策候補の関係について分析し、また、最新の成果も取り込みながら俯瞰できるようにすることで横断的な評価を可能とした。また、これを実現するために、デバイスの入出力に係るサブテーマを実施項目 1、デバイスの内部の保護に係るサブテーマを実施項目 2、デバイスの個体管理に係るサブテーマを実施項目 3 で取り組み、テーマ全体のとりまとめを実施項目 0 で実施した。

4.9.2. 最終目標と根拠

本研究開発テーマの最終目標（アウトカム）は、AI エッジデバイスに関連する市場が拡大する中で、セキュリティが原因で本市場の拡大が阻害されることを防止することにある。セキュリティにより市場の拡大が阻害される要因としては、セキュリティリスクや対策効果が見えにくく判断が難しいため、実際に被害が出るまで対応が放置される傾向にあることが挙げられる。また、対応を取る際にも、全体のセキュリティのレベルは一番弱い箇所が決まり、攻撃技術も常に変化・進化するため、将来を見越した上で全体を俯瞰し、適用先の状況も考慮に入れながら適切な対策候補群を見極める必要がある。これを実現するために、本研究開発テーマでは、AI エッジデバイスにおいて今後問題となることが予想されるセキュリティ上の脅威を、科学的/学術的根拠に基づいて公正かつ正確に議論・評価するためのセキュリティ評価分析基盤を確立することを本研究開発テーマ期間の目標とする。ここでセキュリティ評価分析基盤とは、AI エッジのセキュリティを横断的に評価分析するための中核的な研究開発体制を指し、セキュリティ要件、評価分析手法、セキュリティ評価技術/機器/ツール/シミュレータなどで構成される。

4.9.3. 目標の達成度

本研究開発テーマ全体の成果及びエッジデバイス向け AI(機械学習)へのセキュリティ上の脅威や攻撃と対策候補の関係を評価分析しながら整理することで、AI エッジデバイスの適用分野に応じて適切な対策がとられているか否かの評価分析を可能にするためのセキュリティ評価分析基盤を確立し、これにより事業期間内での目標を達成した。

4.9.4. 成果と意義

上述のとおり、AI エッジデバイスに対する脅威や攻撃と対策候補の関係を俯瞰できるようにし、それらを対象に含む新たな評価分析カテゴリを確立させ、AI エッジデバイスに対する横断的な評価を可能とした意義は大きい。

以下においてその内容について報告する。本研究開発テーマでは、エッジデバイスを用いた AI システムを以下のように分類する。

- エッジ完結型
- エッジ推論/再学習型
- エッジ推論型
- エッジ・クラウド連携（連合学習）型
- クラウド型

表 4.9.4.1 の最右列に示すとおり、エッジデバイスを用いているため、推論対象はいずれもエッジ側にあることを想定するが、AI に求められる処理（初期学習、再学習、推論）をエッジとクラウドのどちらで行うかが異なるため、それに応じて上記及び表 4.9.4.1 のように分類する。

表 4.9.4.1 エッジデバイスを用いた AI システムの型と各処理の実行場所との関係

型 \ 処理	初期学習	再学習	推論	推論対象
エッジ完結型	エッジ	エッジ	エッジ	エッジ側
エッジ推論/再学習型	クラウド	エッジ	エッジ	エッジ側
エッジ推論型	クラウド	クラウド	エッジ	エッジ側
エッジ・クラウド連携型 (連合学習)	エッジ/ クラウド	エッジ/ クラウド	エッジ	エッジ側
クラウド型	クラウド	クラウド	クラウド*	エッジ側

* 処理の一部をエッジで行う場合を含む

また、これらの関係を図で表したのが図 4.9.4.1 である。本研究開発テーマでの AI エッジデバイスは、これらの分類のうちクラウド型以外、すなわち、学習済みモデルがエッジデバイスに格納されている場合を指す。

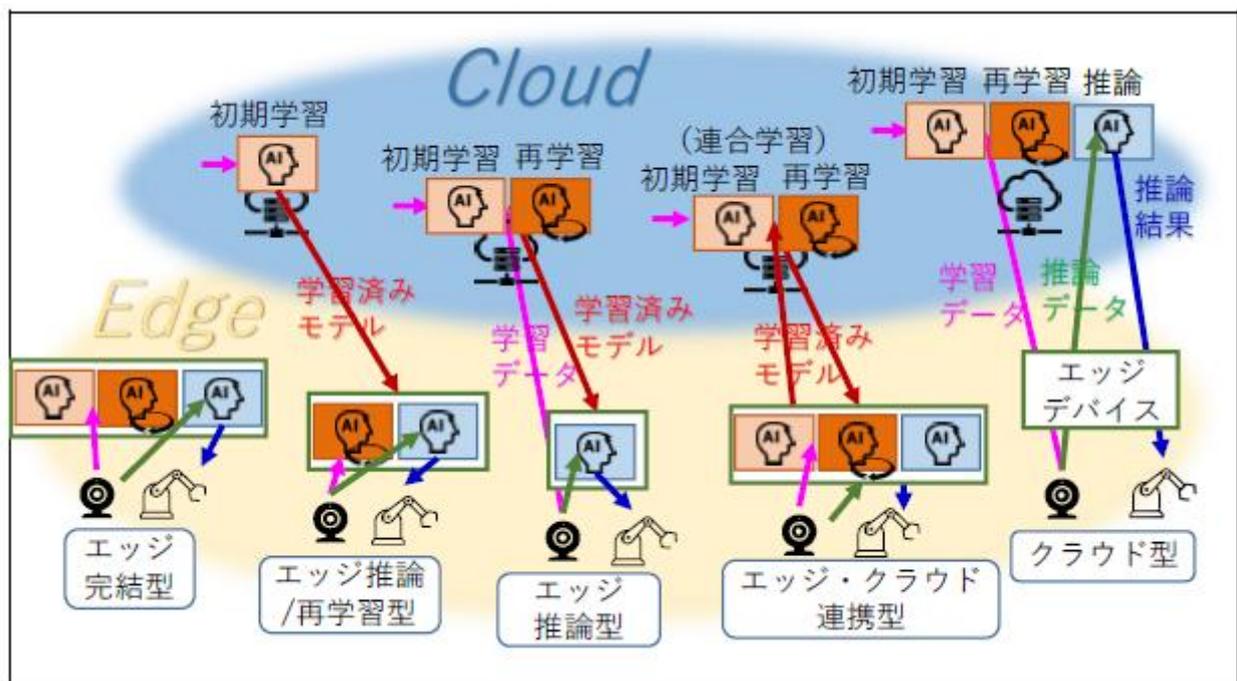


図 4.9.4.1 エッジデバイスを用いた AI システムの分類と構成図

AI エッジデバイスを用いる利点（エッジにおいて AI を用いる利点）としては、一つは、自動制御のようにリアルタイム性と通信障害耐性を確保する場合があります、もう一つは、少なくとも推論時において個人やプライバシーに関連する情報をエッジ側で処理することで、それらをクラウドに上げないようにする点などが挙げられる。前者の構成例を図 4.9.4.2、後者の構成例を図 4.9.4.3 に示す。

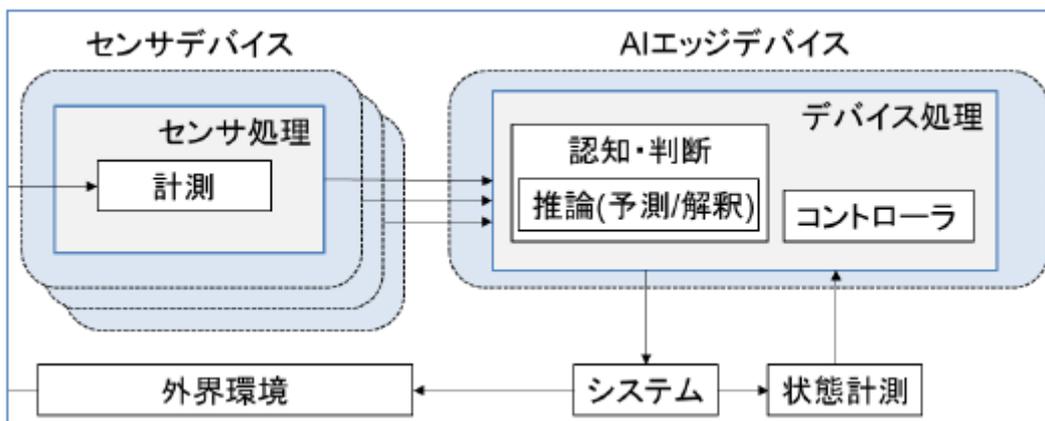


図 4.9.4.2 自動制御等で用いられる AI エッジデバイスの構成例

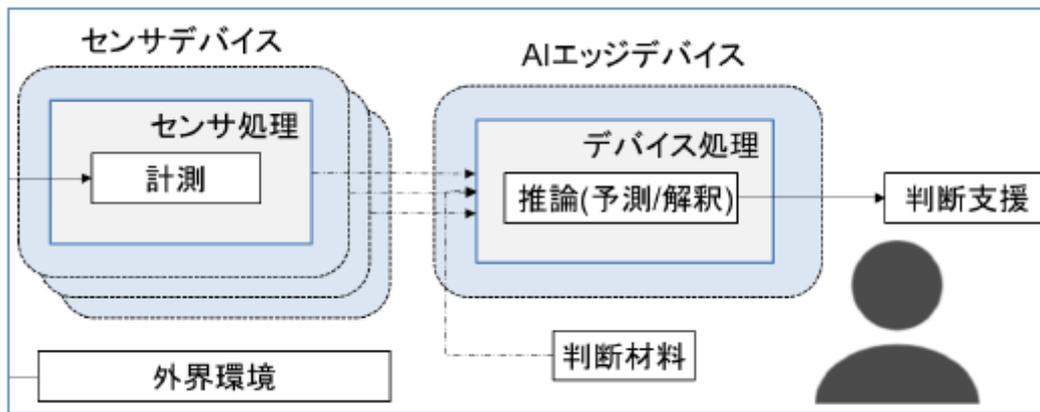


図 4.9.4.3 判断支援等で用いられる AI エッジデバイスの構成例

(1) 自動制御等で用いられる AI エッジデバイスの構成例

リアルタイム性と通信障害耐性が要求される自動制御等に用いられる AI エッジデバイスの構成例である。一つまたは複数のセンサデバイスが用いられ、外部インタフェースで AI エッジデバイスに接続する場合と AI エッジデバイスに物理的に据え付けられる場合がある。センサで得られた計測結果をセンサ処理で加工・デジタル化し、デバイス処理に受け渡す。デバイス処理において、センサデバイスから受け取ったデータの認知・判断を行う過程で予測・解釈のための AI による推論を実行し、認知・判断した結果をコントローラ経由で外部のシステムに受け渡す。

(2) 判断支援等で用いられる AI エッジデバイスの構成例

プライバシー保護が要求される判断支援等に用いられる AI エッジデバイスの構成例である。入力された判断材料やセンサデバイスから受け取ったデータをもとに、デバイス処理において予測・解釈のための AI による推論を実行し、判断支援のための情報を出力する。センサとの接続は、自動制御等で用いられる AI エッジデバイスと同様に、外部インタフェースで接続される場合と物理的に据え付けられる場合がある。

(クラウド型も含む) AI(機械学習)そのものに対する脅威や攻撃方法とそれらへの対策候補については最後にまとめるが、AI エッジデバイスに対しては、さらに、①入出力への攻撃、②内部への攻撃、③管理への攻撃も想定する必要がある。

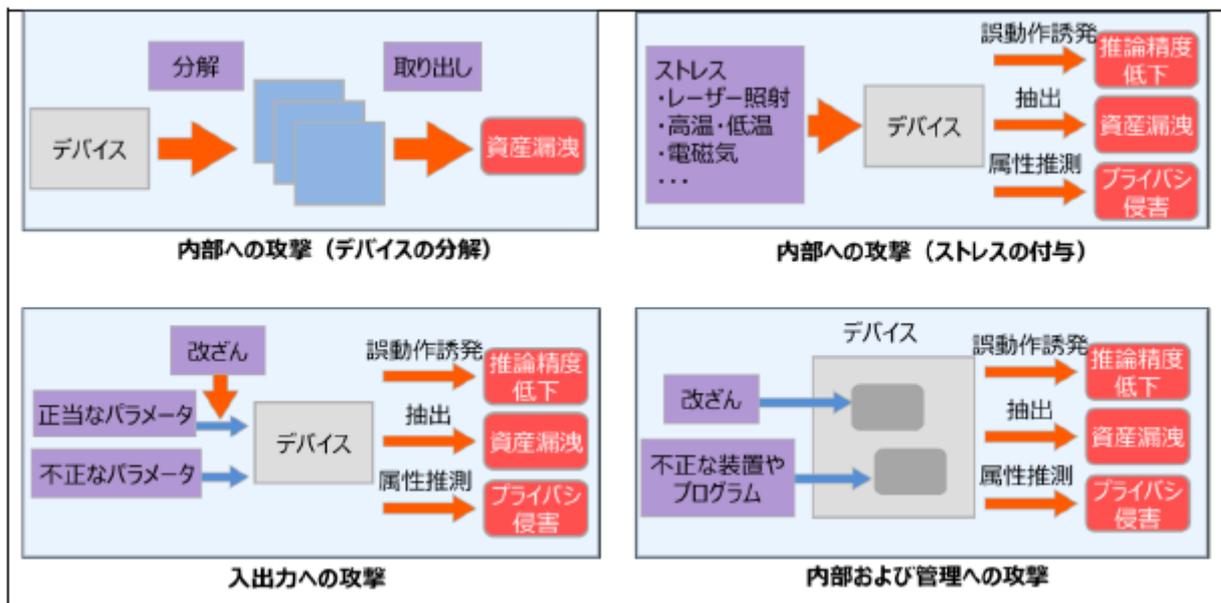


図 4.9.4.4 AI が組み込まれたデバイスに対する攻撃モデル

- 入出力への攻撃

センサ入力やデバイスの入出力インタフェース経由で行われる攻撃で、それらの入出力データを傍受し改ざんを加えるか、不正なデータを送り付けることで、誤動作を誘発するほか、内部データの抽出や属性を推測する攻撃である。不正データの入力に誤動作を誘発し、AIによる推論の精度低下や資産漏洩の発生につながる可能性がある。また、出力データの解析や漏洩したデータの属性を推測することによりプライバシー侵害につながる可能性がある。

- 内部への攻撃（デバイスの分解）

デバイスを物理的に分解し、ツール等を接続して、内部のデータを取り出す攻撃である。これにより、資産漏洩につながる可能性がある。

- 内部への攻撃（ストレスの付与）

デバイスに対して物理的なストレスを与え、負荷をかけて通常の保証動作環境を逸脱させ、動作に影響する状態を作り出すことで、内部データの抽出や属性を推測する攻撃である。外部からストレスを加えて誤動作を誘発し、AIによる推論の精度低下や資産漏洩の発生につながる可能性がある。また、漏洩したデータの属性を推測することによりプライバシー侵害につながる可能性がある。

- 内部及び管理への攻撃

デバイスの各ライフサイクルにおける管理が不十分な箇所を突き、デバイスの内部実装を直接改ざんするほか、不正な装置やプログラムを混入させることで、誤動作の誘発、内部データの抽出、及び属性を推測する攻撃である。管理不備を突いた物理的な操作が誤動作を誘発し、AIによる推論の精度低下や資産漏洩の発生につながる可能性がある。また、管理不備を突いた出力データの解析や漏洩したデータの属性を推測することによりプライバシー侵害につながる可能性がある。

これら AI エッジデバイス特有の攻撃方法に対しては、実施項目 1 において入出力、実施項目 2 において内部実装保護、実施項目 3 において個体管理のための識別子の詳細について

て評価分析を行った。実施項目0では、AI(機械学習)に対するセキュリティ上の脅威と攻撃方法について、最新の成果も取り込みながら以下のようにまとめた。

表 4.9.4.2AI(機械学習)への秘匿性に対する脅威・攻撃方法の分類

機械学習への脅威・攻撃方法	脅威・攻撃方法の説明
訓練データの漏洩/ (訓練データへの) 秘匿性攻撃	訓練データに関する情報や、そこからモデルの特性の証拠などを明らかにすることを目的とした敵対的な攻撃
モデル反転攻撃 モデルインバージョン攻撃	学習済みモデルや補助データから訓練データや機密データを発見する攻撃
メンバシップ推測攻撃 メンバシップ攻撃 メンバシップ推定	特定のデータが、訓練データセットに含まれていたか、訓練データセットと同じ分布の一部であったかという情報を判別する攻撃
属性推測攻撃	訓練用データに関わるセンシティブ情報を推測する攻撃
プロパティ推測攻撃	訓練用データセットについての大域的な性質を推測する攻撃
モデル情報の漏洩	訓練済みモデルの情報が漏洩する攻撃
モデル抽出攻撃	運用時に、入力データに対する出力の振る舞いを観察することで、訓練済みモデルと同様の動作をするモデルを抽出する攻撃
ホワイト/グレーボックス 攻撃	モデルの内部情報知識を悪用した攻撃、もしくは、モデルの部分的な知識を悪用した攻撃

表 4.9.4.3 AI(機械学習)への完全性・可用性に対する脅威・攻撃方法の分類

機械学習への脅威・攻撃方法	脅威・攻撃方法の説明
データポイズニング攻撃	機械学習に用いるデータセットに意図的な改変を加える攻撃
バックドア型データポイズニング攻撃	特定の入力に対してのみ機能するようなデータポイズニング攻撃
強化学習時の特定状態への利得の誤謬	強化学習時の特定の状態の利得に対して改変を加える攻撃
モデルポイズニング攻撃	訓練済みモデルに対して不正な動作などを埋め込む攻撃
回避攻撃 (敵対的データ)	悪意ある入力データ (敵対的データ) によって想定直感と異なる推論結果を出力させ訓練済みモデルを誤動作させる攻撃
解釈機能を誤動作させる攻撃	解釈機能による説明内容の価値を下げたり、間違った説明を生成したりする攻撃

次に、それらへの対策候補をその特性に応じて以下のように分類した。

- 機械学習モデル自体に高い汎化能力と安定性を持たせる対策候補
- 学習時に使用するデータセットを加工する対策候補

- 機械学習モデルへの入出力に対してあらかじめ処理を行うことで堅牢性を向上させる対策候補
- 管理/監視による対策候補
- プライバシ確保モデル（訓練データに含まれる可能性のある個人に関する情報を明らかにしない方法で学習されたモデル）を生成するための対策候補

表 4.9.4.4 機械学習モデル自体に高い汎化能力と安定性を持たせる対策候補

対策候補	概要
アーキテクチャ	アーキテクチャの面から、学習済みモデルが、訓練データセットに含まれない入力データに対して期待する反応を示すようにする
アンサンブル学習	複数の分類器を用いて、回避攻撃などに対する堅牢性を高めた分類法のこと
モデル軽量/最適化	モデルの軽量化及び最適化を通じて、学習済みモデルが、訓練データセットに含まれない入力データに対して期待する反応を示すようにする
ファインチューニング	用途に合わせて追加で学習させたりパラメータを微調整したりすること
量子化/計算精度の変更	パラメータの精度を落とした型への変換
プルーニング	必要性の低いノードの除去
重みクラスタリング/重み共有	重みをクラス分けし、各クラス内の重みをその代表値/重心値に設定する方法のこと
データ拡張	訓練用のデータ数の確保や過学習の防止など、または、テスト用のデータ数の確保のため、データの変種を生成してデータセットに追加すること
敵対的生成ネットワーク	敵対者（実際のデータと生成されたデータを区別しようとする識別器）の存在下でモデル生成器を学習させることで、モデル生成器の有効性を高める機械学習手法のこと
敵対的訓練/ロバスト訓練	誤推論しそうな近傍データを探索しながら訓練を行う手法のこと
ロバスト最適化法	敵対的学習は mini-max 問題として定式化される。内部の問題がワーストケースの有界摂動の下でトレーニングポイントを操作してトレーニングロスを最大化し、外部問題としてワーストケースの訓練損失を最小化するように学習アルゴリズムを訓練すること
誤差逆伝搬/損失関数への制約	誤差逆伝搬や損失関数の勾配を利用するなどの学習方法に制約を設ける方法のこと
正則化	訓練時に学習するパラメータが過剰に大きくなることを抑えること
勾配マスキング	勾配を最小化することで、「敵対的データ」に対するモデルの感度を下げる
短縮ネットワーク	「敵対的データ」に対する影響を軽減するために、防御の目的で「 smoothness penalty （滑らかさのペナルティ）」を適用する機械学習の手法の一つのこと。入力の変量に対する出力の変量にペナルティを課し、「敵対的データ」を生成するのに必要なばらつきを大きくすること
ドロップアウト	訓練中にランダムにニューロンを訓練対象から除外すること

表 4.9.4.5 学習時に使用するデータセットを加工する対策候補

対策候補	概要
データクリーニング /クレンジング データ無害化	データクリーニング/クレンジングは、データセットに対して、本来の特徴を訓練できるように、ノイズ除去、データ整形、欠損値補完、外れ値除去などを行うこと データ無害化は、操作されたサンプルデータを訓練データの外れ値として識別して扱い、検出・除去する防御方法のこと
プライバシー情報の削除/加工	データセット中のプライバシー情報を削除する手法。特に、匿名加工は、データセットに対して特定の個人を識別することができないように個人情報を加工することで、特定性を低減すること

表 4.9.4.6 機械学習モデルへの入出力に対してあらかじめ処理を行うことで堅牢性を向上させる対策候補

対策候補	概要
区分ベース防御	画像の断片にニューラルネットワークを適用し、画像中の敵対的なパターンに対する堅牢性を向上させる機械学習手法のこと
防御的蒸留	機械学習モデルに入力する前に、不正な判定・予測用データを検知・排除する、または、それらによる機械学習モデルの出力への影響を軽減する学習モデルを利用する手法のこと 不正な判定・予測用データによる誤った判定・予測の誘発への対策
オートエンコーダー	入力層の値を教師データとし、中間層で潜在変数に次元を圧縮するニューラルネットワークのこと
ロバスト学習	外れ学習サンプルに対する感度が本質的に低いロバスト統計（ Robust Statistics ）に基づく学習アルゴリズムのこと
出力低情報化	学習済みモデルの情報を攻撃者が得ることを防止するため、必要以上の推論結果を出さないようにすること
出力ランダム化	分類器の出力をランダム化し、攻撃者に不完全なフィードバックを与える防御方法。 Randomization defense （訓練データ、モデルの学習損失関数（ training cost function ）、学習パラメータ、モデル出力にランダムなノイズを付加し、プライバシーを保護する防御メカニズムのこと）のサブセットのこと
ランダムスムージング	ランダムノイズを付加したデータをニューラルネットに入力したときの出力の期待値を推論結果とする手法であり、ランダムスムージングによって安定性が向上することが報告されている

表 4.9.4.7 運用/監視による対策候補

対策候補	概要
運用対策	
データセットとその入手経路の管理	データポイズニング攻撃など機械学習に用いるデータセットに意図的な改変を加えること、メンバシップ推測攻撃などデータセットから情報を抽出する攻撃に対して、データセットやその入手経路を厳格に管理することで、攻撃の実現を抑制する
学習済モデルとその入手経路の管理	学習済モデル自体を厳格に管理することで、モデルポイズニング攻撃など、訓練済みモデルに対して不正な動作などが埋め込まれるような攻撃を抑制する
悪意のある利用者による大量計測の防止	学習済モデルへの(大量)問い合わせを通じたモデル抽出攻撃により、訓練済みモデルと同様の動作をするモデルを抽出するもの。これには、問い合わせ回数の制限等を組み込み攻撃の実現を抑制する
監視対策	
攻撃検知	異常と正常な動作、または「敵対的データ」と良性のデータを区別する動作のこと
精度モニタリング	訓練済み機械学習モデルの精度を直接測定すること
KPI モニタリング	モデルの単なる精度だけではなく、コンバージョン率などユーザ利益に即した KPI の観点での監視を行うこと
入出力モニタリング	訓練済み機械学習モデルによる推論結果またはその入力データの監視のこと。その監視方式は、人力監視（全数、サンプリング）、自動監視（アラート条件が既知）、フィルタリング（アラートの可能性が高い条件が既知）に分類される

表 4.9.4.8 プライバシ確保モデルによる対策候補

対策候補	概要
プライバシー保護/維持学習	訓練学習の中核をなす学習アルゴリズム中に差分プライバシーの方法を組み込むことで、導出結果である訓練済み学習モデルからの訓練データ保護を目的とする
連合学習	クライアント・サーバ方式による分散コンピューティングの考え方を訓練学習に応用する方法。大規模な訓練データセットを複数に区分けし、区分けしたデータセットごとに、これを入力とする訓練学習をクライアントで実行する このクライアント側の計算処理で得た中間的な訓練結果をサーバで集約する。クライアント・サーバ連携処理を繰り返して最終的な訓練済み学習モデルを得る

続いて AI エッジデバイスに対する攻撃の影響を横断的に評価分析できるよう、各攻撃の進展経路、最終的な影響、各対策がどの経路に効果を発揮するかについて評価分析を行い、右図のような凡例を用いて、図 4.9.4.5 のように整理した。凡例では、矢印始点が「前段階の脅威」を示し、これに対して、緑色の矩形が示す「対策候補」により、その脅威が軽減され、矢印終点が示す「次段階の脅威」が残存リスクとなることを示す。

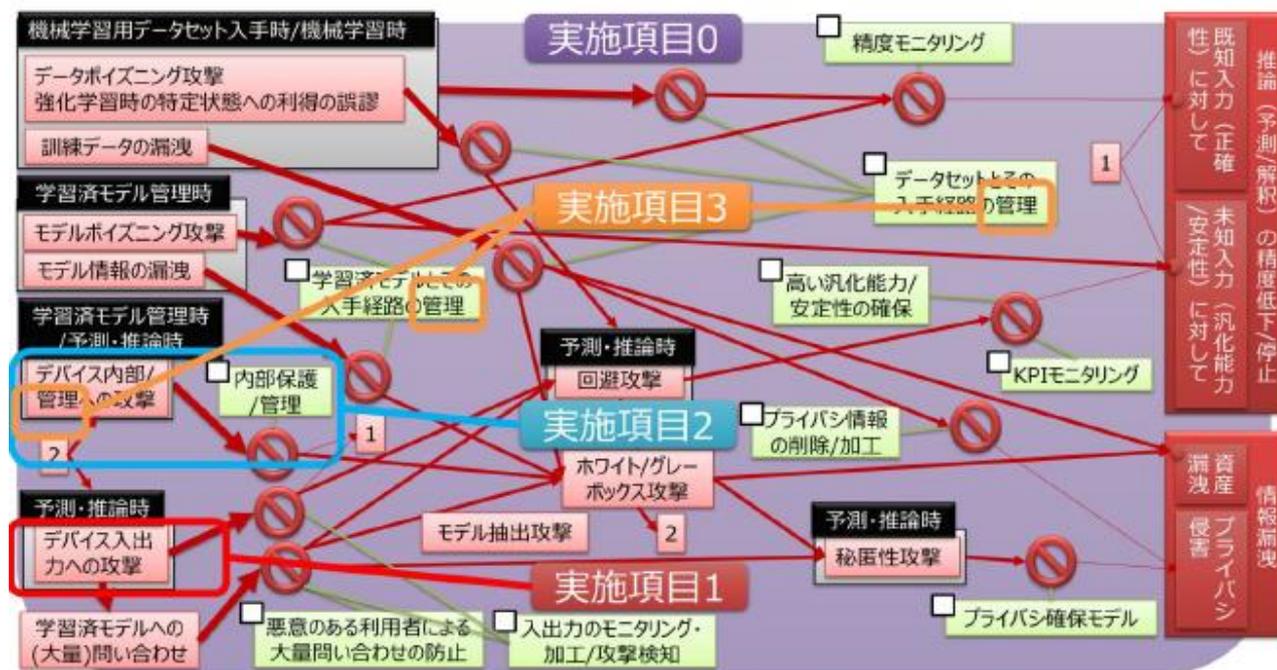
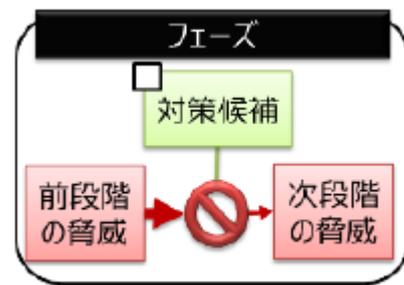


図 4.9.4.5 AI エッジデバイスへの脅威・攻撃とそれらへの対策候補の関係

ステークホルダや応用先によっては、AI のライフサイクル、サプライチェーン、バリューチェーンの一部の段階（フェーズ）のみしか担っており、全ての脅威や攻撃について考慮する必要が無い場合もあるため、各脅威や攻撃が問題となるフェーズについても以下のように分類した。

- 機械学習用データセット入手時/機械学習時
 - データポイズニング攻撃
 - 強化学習時の特定状態への利得の誤謬
 - 訓練データの漏洩
- 学習済みモデル管理時
 - モデルポイズニング攻撃
 - モデル情報の漏洩
- 学習済みモデル管理時/予測・推論時
 - デバイス内部/管理への攻撃
- 予測・推論時

- デバイス入出力への攻撃
- 学習済みモデルへの（大量）問い合わせ
- ホワイト/グレーボックス攻撃、モデル抽出攻撃
- 秘匿性攻撃
- 回避攻撃

各脅威や攻撃が適用先に与える影響は以下のとおりとなる。

(1) 推論(予測/解釈)の精度低下

推論精度の低下については、AI エッジデバイスへの入力既知か未知かによって対応の難しさが異なる。既知の入力に対しては正しい推論結果も分かっており、それを用いることでAI 以外による対処も可能であるのに対して、未知入力に対しては正しい推論結果は分からないためAI の推論精度の低下の影響を大きく受ける。例えば、自動制御/生産現場に関するユースケースでは、自動運転での予測精度の低下による事故や、重要インフラ保全での想定されない解釈結果等が導出されることによるサービス停止、判断支援サービスのユースケースにおいては、許認可等のように権利/資産等に係る判断やその基となる解釈を誤る、などにより社会の混乱などに繋がる場合がある。

(2) 情報漏洩

情報漏洩については漏洩する内容により、資産漏洩とプライバシー侵害に分類できる。AI モデルや訓練データに競争領域の業務知識やノウハウが含まれている場合は前者、プライバシー関連の情報が含まれている場合は後者の影響を受ける可能性がある。自動制御/生産現場に関するユースケースではプライバシー侵害に該当するケースは少ないと考えられるが、例えば、訓練データの背景等にプライバシーに関連する情報が含まれている場合や、入出力から特定の個人との紐づけが可能となる場合などが該当する。一方、判断支援サービスに関するユースケースにおいては、さらに、人種、病歴などの個人プロファイルや、在宅情報など犯罪に悪用され得る情報が漏洩した場合にプライバシー侵害が懸念され、法規制に抵触する場合、さらに大きな影響を受けることになる。

これらの成果を活用することで、適用先/応用先に応じてAI エッジデバイスの脅威や攻撃とそれらへの対策候補の関係を俯瞰できるようになり、横断的な評価を可能とした。

4.9.5. 成果の普及

前節の成果は、AI エッジデバイスを既に用いている、または、導入を検討している全ての分野に適用可能であるが、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（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。AI エッジデバイスなどの高い付加価値を持つ製品やそれらの部品などを物の物理的特徴の計測または測定により管理するための識別子（人工物メトリクス）に対しては、企業・大学などの有識者委員とともに「人工物メトリクスを用いた個体管理技術ガイダンス」を策定し2022年1月に公開し、本ガイダンスを解説するためのシンポジウムを2022年3月に開催した。また、関連する企業群、横国大、産総研、政府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、日本から提案した人工物メトリクスの国際標準化は2022年12月にISO 22387:2022 (Validation procedures

for the application of artefact metrics:人工物メトリクスを利用する際の妥当性確認手順)として成立した。セキュリティ評価時に信頼区間を厳密に計算するためのプログラム「EBCIC: Exact Binomial Confidence Interval Calculator」はオープンソースで公開すると共に、PyPI (Python Package Index)への登録を行った。

4.10. 研究開発テーマ「セキュアオープンアーキテクチャ向けコンパイラバックエンドおよび対応ランタイム環境の設計・開発」

4.10.1. テーマ概要

本プロジェクトの事業内容は、下記①～④であるハードウェアとアプリケーション・サービスの界面となる4つの開発と⑤、⑥の2つの効果検証によって構成される。ソフトウェアスタックで最上位層にあたる実施項目④の評価は他の項目に依存せず、④内にて完結した評価を実施する。今回の研究開発が対象とするエッジシステムは下記左のようなシステム全体像で示す。これに対して、研究開発の全体像は下図右の構成となっており、既存の世界標準を利用しつつ、本プロジェクトの目的である組込み（リアルタイム）システムに必要な差分を開発し、再び標準に戻す形となる。既存の世界標準を利用しつつ、本プロジェクトの目的である組込み（リアルタイム）システムに必要な差分を開発し、再び標準化活動へフィードバックする予定である。

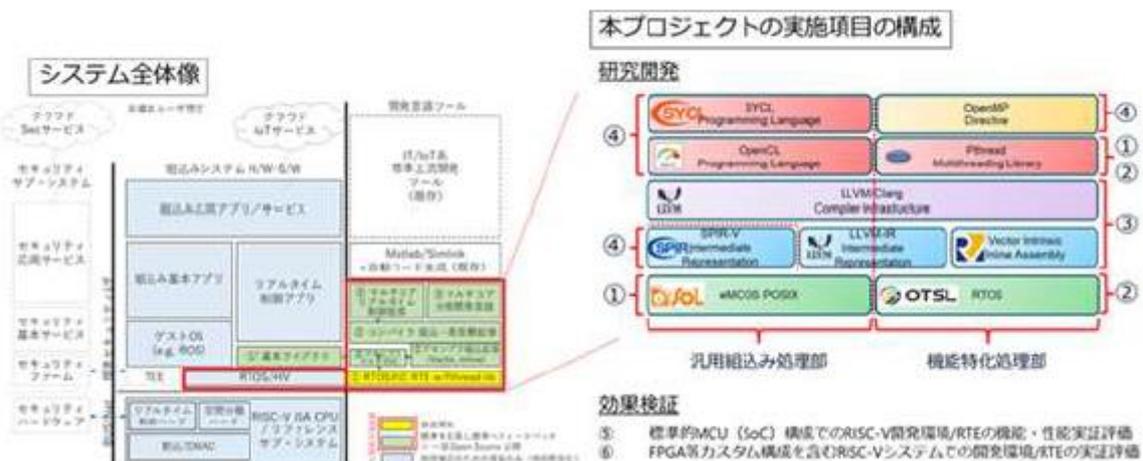


図 4.10.1.1 本プロジェクトの実施内容と想定するシステム構成の関係図

効果検証は開発環境が持つ「特定のシステム・構成に依存した仕様・実装・提案は標準として受け入れられない」という特性・事情から、以下の⑤、⑥の異なる2つのケースで実施する。

表 4.10.1.2 本プロジェクトの事業内容

種別	事業内容
研究開発	① RISC-V に最適化されたマルチコア対応の高性能 RTE の設計・開発
	② セキュアオープンアーキテクチャ対応 RTE のオープンソース化実証および対応評価環境の設計・開発
	③ セキュアオープンアーキテクチャ向けベクトル化 C コンパイラの設計・開発
	④ セキュアオープンアーキテクチャ向け並列化対応開発環境の設計・開発
効果検証	⑤ 標準的 MCU (SoC) 構成での RISC-V 開発環境/RTE の機能・性能実証評価
	⑥ FPGA 等カスタム構成を含む RISC-V システムでの開発環境/RTE の実証評価

4.10.2. 最終目標と根拠

本テーマでは、AI をエッジシステムで RISC-V アーキテクチャを利用する際に必要となる、ソフトウェア実行環境 (Runtime Environment, RTE) 、コンパイラ基盤、並列プログラム処理系といった高速かつ高効率・低消費電力なランタイム+開発ツールで構成される RISC-V 開発環境エコシステムを構築することが最終目標である。

4.10.3. 目標の達成度

本テーマの計画目標、達成状況は下表の通りである。

実施項目	計画目標	達成状況
①RISC-V に最適化されたマルチコア対応の高性能 RTE の設計・開発	RISC-V に最適化されたマルチコア対応の高性能 RTE の pthread 開発：RISC-V に最適化されたマルチコア対応の高性能 RTE により実行制御のオーバーヘッドおよび精度レンジが現状よりも 20%以上短縮 RTE と連携する並列化支援ツールの RISC-V 対応開発：高性能 RTE 実現のための開発支援環境の完成 デバッガのターゲット依存部開発：高性能 RTE 実現のための開発支援環境の完成	研究計画通り達成済み。eMCOS POSIX の RV64 移植を行い、また設計の改善を実施したことにより、実行制御のオーバーヘッドを約 71 % 削減させた高性能 RTE を実現した。 また、同 OS 上でデバッグおよび eMBP が正常に動作することを確認できた。
②セキュアオープンアーキテクチャ対応 RTE のオープンソース化実証および対応評価環境の設計・開発	RTE のオープンソース化実証および対応評価環境の設計・開発：オープンソースベースの RTE と LLVM およびセキュリティモジュールとの接続部分での無駄を抽出し、適切な I/F 仕様を策定し、標準的システム構成を検討後、FPGA ベースの PoC 環境を構築 機能特化処理部分においての RTE 環境の構築：構築された PoC を使用し、実行制御のオーバーヘッドおよび精度レンジの 20%以上の短縮	研究計画通り達成済み。PoC システム上にて RTE の動作を確認し、動作確認結果を RTE/RISC-V コアヘフィードバック・性能強化を行い、性能測定を完了した。
③セキュアオープンアーキテクチャ向けベクトル化 C コンパイラ的设计・開発	ベクトル化対応 LLVM の開発と、C 言語へのベクトル化対応専用数学ライブラリの導入：インラインアセンブラ、イントリンシック関数としてベクトル命令を処理できる LLVM を実現す	研究計画通り達成済み。RVV 対応のツールチェーンを用意し、C ソースコードで記述したプログラムから RVV 命令が生成されることを確認した。連携デバッグにおいて

	<p>る。また、ベクトル演算機能を用いる専用の数学ライブラリを用意する。これにより、従来のイントリンシックベースのベクトル化に対し、80%のベクトル化効率の向上を実現する</p> <p>LLVM デバッガインターフェースの開発：肥大化するソフトウェアのデバッグを効率化するために、コンパイラの技術を使うデバッグ手法を導入し、10倍のソフトウェア開発 TAT の短縮を実現する。</p>	<p>は、実行時に自動的にエラーを検出することを確認した。</p>
④セキュアオープンアーキテクチャ向け並列化対応開発環境の設計・開発	<p>組込み用途 OpenMP ディレクティブプログラミング環境：HPC 向け高スループット重視スケジューラと比較し、20%以上の性能改善</p> <p>RVV に対応する OpenCL/SYCL 処理系：イントリンシック関数により手作業で最適化するベクトル化コードと比較して 80%以上の実行効率</p>	<p>研究計画通り達成済み。RVV 対応の並列化プログラミングツールの開発を達成した。</p> <p>OpenMP 処理系で静的スケジューリング機能を実現し、スケジューラコストが 27%改善する事を確認した。</p> <p>OpenCL 処理系が RVV イントリンシックによる C プログラムと比較して 37%高い性能を発揮する事を確認した。</p>
⑤ 標準的 MCU (SoC) 構成での RISC-V 開発環境/RTE の機能・性能実証評価	<p>MCU (SOC)構成のシステム上でのプロジェクト成果物の動作検証</p>	<p>研究計画通り達成済み。Ai-One ボード環境上でのベンチマークを実施し、ベクトル化版の性能あたりの電力化効率がスカラ版と比較して、前者が約 68 倍優れている事を確認した。</p>

当初の目標である「RISC-V 開発環境エコシステムを構築すること」は 100%達成できたとと言える。

4.10.4. 成果と意義

本テーマでは RISC-V 向けに、図 4.10.1.1 にて示すシステム階層を包括的に構築し、ハードウェアとアプリケーション・サービスの界面となる基本ソフトウェア群を整備した。今後普及する RISC-V システムで主流となると予測されるヘテロジェニアシステムを想定し、マルチコア対応 RTE ソフトウェア長ベクトル命令に対応する LLVM/clang コンパイラ、OpenMP ディレクティブプログラミング環境、OpenCL/SYCL プログラミング処理系を実現し、既存の商用システムとも比較され得る 2023 年時点で最高水準の RISC-V 向けソフトウェアエコシステムを開発した事は、今後の国内組込み業界において RISC-V システムを普及・浸透させるためにも非常にアドバンテージが高い事であると言える。

4.10.5. 成果と普及

OpenCL/SYCL に関しては Khronos Group と連携しワールドワイドで共同でのマーケティング活動を実施している。

イーソル株式会社の提案者がチェアを勤める IEEE Std. 2804 で標準化している SHIMXML の RISC-V 版を RISC-V のオープンソースコミュニティに公開することで、RISCV における SHIM XML の標準化の普及を促進する。

株式会社 OTSL の他オープンソース対応ソフトウェアと連携させ、イベント等での紹介を行うとともに、ベータテスト企業などを選定し、早期に利用を開始してもらうことを予定している。

京都マイクロコンピュータ株式会社、株式会社エヌエスアイテックスの、オープンソース対応の取り組みとしては、2022/11/16-18 Edge Tech+（展示会）および 2022/11/25DSForum(展示会、講演)において、RVV 対応 RISC-V 用 LLVM/Clang コンパイラや連携デバッグ機能への取り組み内容を紹介した。

本プロジェクトの研究開発成果について、共同ニュースリリースを 2023/4/27 に実施した。

●特許論文等リスト

◎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
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	J P : 日本 国	2019.10.16	出願 継続 中	半導体装置およびその制御方法	
2	ルネサスエレクトロニクス株式会社	17/065169 (US)	U S : ア メ リ カ 合 衆 国	2020.10.7	出願 継続 中	SEMICONDUCTOR DEVICE AND METHOD OF CONTROLLING THE SAME	
3	ルネサスエレクトロニクス株式会社	2020111029 04.7 (CN)	C N : 中 国	2020.10.15	出願 継続 中	半導体装置およびその制御方法	
4	ルネサスエレクトロニクス株式会社	17/358579 (US)	U S : ア メ リ カ 合 衆 国	2021.6.25	出願 継続 中	SEMICONDUCTOR DEVICE, DATA GENERATION METHODS USED FOR THE SAME, AND METHOD OF CONTROLLING THE SAME	
5	ルネサスエレクトロニクス株式会社	2021107438 51.5 (CN)	C N : 中 国	2021.7.1	出願 継続 中	半導体装置、それに用いるデータ生成方法およびその制御方法	
6	ルネサスエレクトロニクス株式会社	2020-114592	J P : 日 本 国	2020.7.2	出願 継続 中	半導体装置、それに用いるデータ生成方法およびその制御方法	

7	SOINN株式会社	PCT/J P2022 /9043	PC T (全 指 定)	2022.3.3	出願 継続 中	情報処理装置、情報処理 方法及びプログラム	
8	ルネサスエレクト ロニクス株式会社	2023- 00645 2	J P: 日本 国	2023.1.19	出願 継続 中	半導体装置	
9	ルネサスエレクト ロニクス株式会社	2022- 12707 2	J P: 日本 国	2022.8.9	出願 継続 中	プロセッサ、およびコン パイラ	
10	SOINN株式会 社	PCT/J P2022 /1407 0	PC T (全 指 定)	2022.3.24	出願 継続 中	情報処理装置、情報処理 方法及びプログラムが格 納された非一時的なコン ピュータ可読媒体	
11	SOINN株式会 社	PCT/J P2021 /318 80	PC T (全 指 定)	2021.8.31	出願 継続 中	情報処理装置、情報処理 方法及び非一時的なコン ピュータ可読媒体	
12	SOINN株式会 社	PCT/J P2021 /318 81	PC T (全 指 定)	2021.8.31	出願 継続 中	情報処理装置、情報処理 方法及び非一時的なコン ピュータ可読媒体	
13	SOINN株式会 社	PCT/J P2021 /318 82	PC T (全 指 定)	2021.8.31	出願 継続 中	情報処理装置、情報処理 方法及び非一時的なコン ピュータ可読媒体	
14	SOINN株式会 社	PCT/J P2021 /317 89	PC T (全 指 定)	2021.8.31	出願 継続 中	情報処理装置、情報処理 方法及び非一時的なコン ピュータ可読媒体	

15	SOINN株式会社	PCT/J P2021 / 317 95	PC T (全 指 定)	2021.8.31	出願 継続 中	クラスタリング処理装置、クラスタリング処理方法、非一時的なコンピュータ可読媒体及び情報処理装置	
	SOINN株式会社	PCT/J P2023 / 278 8	PC T (全 指 定)	2023.1.30	出願 継続 中	情報処理装置、学習方法及びプログラム	
	SOINN株式会社	PCT/J P2023 / 304 2	PC T (全 指 定)	2023.1.30	出願 継続 中	情報処理装置、情報処理方法及びプログラム	
	SOINN株式会社	PCT/J P2023 / 2801	PC T (全 指 定)	2023.1.30	出願 継続 中	情報処理装置、情報処理方法及びプログラム	
	SOINN株式会社	PCT/J P2023 / 2793	PC T (全 指 定)	2023.1.30	出願 継続 中	情報処理装置、情報処理方法及びプログラム	

(Patent Cooperation Treaty: 特許協力条約)

【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

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
3	Junnosuke Suzuki, Tomohiro Kaneko, Kota Ando, Kazutoshi Hirose, Kazushi Kawamura, Thiem Van Chu, Masato Motomura, and Jaehoon Yu	東京工業大学	ProgressiveN: Achieving Computational Scalability with Dynamic Bit-Precision Adjustment by MSB-first Accumulative Computation	IJNC		2021.2
4	Junnosuke Suzuki, Tomohiro Kaneko, Kota Ando, Kazutoshi Hirose, Kazushi Kawamura, Thiem Van Chu, Masato Motomura, Jaehoon Yu	Tokyo Institute of Technology	ProgressiveN: Achieving Computational Scalability with Dynamic Bit-Precision Adjustment by MSB-first Accumulative Computation	International Journal of Networking and Computing 338-353		2021.5

【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
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
8	本村真人	東京工業大学	ポストノイマン・ポストムーア時代の情報処理アーキテクチャ	応用物理学会春季学術講演会	2021. 3
9	Masato Motomura	東京工業大学	"Hot" Techs for "Cool" AI Computing: Do We have Enough Tricks?	Cool Chips 24	2021. 4
10	中原啓貴	東京工業大学	二値・三値・量子化ニューラルネットワークの推論LSIと学習アルゴリズム	多値論理研究会 多値論理フォーラム	2021. 12
11	Masato Motomura	東京工業大学	Hiddenite: 4K-PE Hidden Network Inference 4D-Tensor Engine Exploiting On-Chip Model Construction Achieving 34.8-to-16.0TOPS/W for CIFAR-100 and ImageNet	International Solid-State Circuits Conference (ISSCC)	2022. 2
12	本村真人	東京工業大学	LSI技術が開く構造型情報処理の新展開	LSIとシステムとワークショップ	2021. 5
13	本村真人	東京工業大学	ビットスケラブルCNNにおける計算量・精度トレードオフ制御手法の検討	LSIとシステムとワークショップ	2021. 5

(b) 新聞・雑誌等への掲載

番号	所属	タイトル	掲載誌名	発表年月
1	SOINN株式会社	製造現場の最先端機械など披露 名古屋で展示会	産経新聞	2021/10

(c) 展示会への出典

番号	所属	展示会名	発表年月
1	SOINN株式会社	AI・人工知能 EXPO プース展示	2019/4

◎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」

【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 Preferred Networks	2023-048895	国内	2023.3.24	出願 継続 中	情報処理装置、命令挿入装置及び情報処理装置の制御方法	
2	株式会社 Preferred Networks	2023-048896	国内	2023.3.24	出願 継続 中	データ処理装置	

(Patent Cooperation Treaty: 特許協力条約)

【外部発表】

プレス発表等

番号	発表者	タイトル	雑誌名・学会名・ イベント名等	発表年月
1	株式会社 Preferred Networks	MN-Core Deep Learning Accelerator	SC2022 (Exhibit Booth: 2220)	2022.11

◎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	2021/3/31	公開	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 2020-071101	国内	2021/4/10	公開	情報処理装置及び情報処理方法	森達也、大鳥羽暢彦

(Patent Cooperation Treaty: 特許協力条約)

【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	Keiko Fujii, Hoshinori Kanazawa and 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) 137 - 142	有	2019年 8月
2	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 2D4-OS-18a-02	有	2020年 6月

【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	杉田博司	KDDI 株式会社	最優秀賞をハード・ソフトの2部門受賞	JAPAN DRONE 2021 主催 『Best of Japan Drone Award 2021』	2019年10月
2		株式会社アラヤ	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』 5G時代のAIエッジコンピューティングが切り開くリアルタイムな世界	2019年10月
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年11月
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年 6月

番号	発表者	所属	タイトル	会議名	発表年月
5	蓮井樹生	株式会社アラヤ	エッジ AI 実装の現状と、実装時に考慮すべき点	ITmedia Virtual EXPO 2021 (オンライン配信) 「秋組み込み開発&エレクトロニクス・AI EXPO」 基調講演	2021年9月
6	蓮井樹生	株式会社アラヤ	エッジ AI 実装の現状と、実装時に考慮すべき点	AI チップ設計拠点フォーラム (第 27 回) 技術講演	2021年9月
7	蓮井樹生	株式会社アラヤ	スケーラブルなモデル学習-デプロイ手法で広がるエッジ AI 技術の可能性	『Design Solution Forum 2022』 スポンサー講演	2022年11月
8	蓮井樹生 大島佑真井 上昌和	株式会社アラヤ	エッジ AI 技術の基礎と AI システム高速化への応用	『日本テクノセンター』 技術セミナー	2023年2月

(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「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
	株式会社日立製作所	特願 2020-204842	国内			経路計画システム、経路計画方法、および関節を備えたロボットアームの経路計画システム	
	株式会社日立製作所	特願 2021-003454	国内			情報処理装置、および、学習認識システム	
	株式会社日立製作所	US17/513968	US			INFORMATION PROCESSING DEVICE AND LEARNING RECOGNITION SYSTEM	

(Patent Cooperation Treaty: 特許協力条約)

【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
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

番号	発表者	所属	タイトル	会議名	発表年月
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
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

番号	発表者	所属	タイトル	会議名	発表年月
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 敦丈, 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, Tsuyoshi Isshiki, 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
14	Hironori Kasahara	University of Waseda	OSCAR Automatically Parallelizing Compiler and API for Heterogeneous Multicores	2021 IEEE/ACM Programming Environments for Heterogeneous Computing (PEHC)	2021/11
15	Fumiya Kono, Naohito Nakasato and Naru Hirata	University of Aizu	IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc-2021)	Acceleration of Gravitation Field Analysis for Asteroids by GPU Computation	2021/12

番号	発表者	所属	タイトル	会議名	発表年月
16	山本一貴, ヲ部拓吾, 小池穂乃花, 川角冬馬, 藤田一輝, 北村俊明, 川島慧大, 納富昭, 木村貞弘, 木村啓二, 笠原博徳	早稲田大学, オスカーテクノロジー, エヌエスアイテクス	ORB-SLAM3 におけるローカルマッピング処理の並列化	情報処理学会・電子情報通信学会 合同研究会 ETNET	2022/3
17	Masayuki Ito	NSITEXE Inc.	General-purpose scalar/vector AI processor for accelerating wide range of tasks including automotive and industrial applications	A-SSCC 2022(Asian Solid-State Circuits Conference)	2022/11
18	Masayuki Ito	NSITEXE Inc.	General-Purpose Scalar/Vector Processor for Accelerating Wide Range of Tasks Including Automotive and Industrial Applications	ASP-DAC 2023	2023/1

◎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

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
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 (ICESS2019)	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 (ICESS)	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

番号	発表者	所属	タイトル	会議名	発表年月
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 (ICESS2019)	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)	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
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
8	Tatsuya Miki, Takuya Azumi	埼玉大学	Porting Evaluation Framework for Self-Driving Software Based on ROS 2	Asia Pacific Conference on Robot IoT System Development and Platform (APRIS2021)	2021/11

番号	発表者	所属	タイトル	会議名	発表年月
9	Bo Peng, Yuqing Yang, Yoshikazu Okumura, Atsushi Hasegawa , Takuya Azumi	埼玉大学	ThreadedCallback: Improving Real-time Performance of ROS 2	Asia Pacific Conference on Robot IoT System Development and Platform (APRIS2021)	2021/11
10	Zihang Li, Atsushi Hasegawa , Takuya Azumi	埼玉大学	Autoware_Perf: Tracing and Performance Analysis Framework for ROS 2 Applications	IEEE International Conference on Embedded Software and Systems (ICESS2021)	2021/12
11	Takahisa Kuboichi, Atsushi Hasegawa , Bo Peng, Keita Miura, Kenji Funaoka, Shinpei Kato, Takuya Azumi	埼玉大学	CARET: Chain-Aware ROS 2 Evaluation Tool	IEEE international conference on Embedded and Ubiquitous Computing (EUC2022)	2022/12
12	Bo Beng, Atsushi Hasegawa , Takuya Azumi	埼玉大学	Scheduling Performance Evaluation Framework for ROS 2 Applications	IEEE International Conference on Embedded Software and Systems (ICESS2022)	2022/12

(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		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

番号	所属	発表内容	発表先	発表年月
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	国内	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
	SH コンサルティング株式会社, 国立大学法人電気通信大学	16/910,103(US)	US	2020/6/24	出願継続中	Root of Trust device that improves the safety of IoT device and its operating method	
	株式会社日立製作所	2021-153025	JP: 日本 国	2021.9.21	出願継続中	プロセッサおよびエンディアン変換方法	

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
	株式会社日立製作所	17/355251(US)	US: アメリカ合衆国	2021.6.23	登録済み	データ処理方法、エッジ装置、および、データ処理システム	
	株式会社日立製作所	10 2021 116 892.9(DE)	DE: ドイツ	2021.6.30	出願継続中	データ処理方法、エッジ装置、および、データ処理システム	
	国立研究開発法人 産業技術総合研究所	2022PRO-2850	JP: 日本国	2022.7.5	プログラム著作権登録済み	Open Trust Protocol (OTrP) proto	
	株式会社日立製作所	2022-091124	JP: 日本国	2022.6.3	出願継続中	装置システム、正当性検証システム及び正当性検証方法	
	国立研究開発法人 産業技術総合研究所	2023PRO-2897	JP: 日本国	2023.1.18	プログラム著作権登録済み	Trusted Execution Environment (TEE) 共通 API 提供部ソフトウェア (名称:TA-Ref) と、Trusted Execution Environment Provisioning (TEEP)プロトコル実行部のソフトウェア(名称:TEEP-Device)	
	国立研究開発法人 産業技術総合研究所	SN0465-2022	JP: 日本国	2022.4.27	プログラム著作権登録済み	Open Trust Protocol (OTrP) proto プログラム(ソフトウェア)	
	株式会社日立製作所	PCT/JP2022/026785	JP: 日本国	2022.7.6	出願継続中	プロセッサおよびエンディアン変換方法	

(Patent Cooperation Treaty: 特許協力条約)

【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・境界サイエンス誌 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&ロボットの委員会講演	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

番号	出願者	出願番号	国内 外国 PCT	出願日	状態	名 称	発明者
3	国立研究開発法人 産業技術総合研究 所	2022-13088 7	J P : 日本 国	2022.8.19	出願継 続中	電流波形分析装置及 び電流波形分析方法	
4	国立研究開発法人 産業技術総合研究 所	PCT/JP2022 /045655	P C T (全 指 定)	2022.12.12	出願継 続中	電流波形分析装置及 び電流波形分析方法	

(Patent Cooperation Treaty: 特許協力条約)

【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	白色干渉計を讀取装置とした光 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

番号	発表者	所属	タイトル	会議名	発表年月
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
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 イノ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

◎研究開発テーマ「セキュアオープンアーキテクチャ向けコンパイラバックエンドおよび対応ランタイム環境の設計・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	Jingyong Cai, Masashi Takemoto, Yuming Qiu, Hironori Nakajo	東京農工大学	Trigonometric Inference Providing Learning in Deep Neural	MDPI Applied Sciences (ISSN 2076-3417)		2021/8
2	Tomoaki Tanaka, Ryosuke Higashi, Hidetaro Tanaka, Takefumi Miyoshi, Yasunori Osana, Jubee Tada, Kiyofumi Tanaka and	東京農工大学, 北陸先端科学技術大学院大学, 琉球大学, わさば合同会社, 山形大学	Shared Vector Register of RISC-V for the Future Hardware Acceleration	Sixth Workshop on Computer Architecture Research with RISC-V (CARRV 2022)		2022/6

	Hironori Nakajo					
3	Tomoaki Tanaka, Ryosuke Higashi, Hidetaro Tanaka, Takefumi Miyoshi, Yasunori Osana, Jubee Tada, Kiyofumi Tanaka and Hironori Nakajo	東京農工大学	Shared Vector Register of RISC-V for the Future Hardware Acceleration	Sixth Workshop on Computer Architecture Research with RISC-V (CARRV 2022)		2022/6
	田中友章, 東良輔, 田中清史, 長名保範, 三好健文, 多田十兵衛, 中條拓伯	東京農工大学, 北陸先端科学技術大学院大学, 琉球大学, わさらぼ合同会社, 山形大学	ハードウェアアクセラレーションのためのベクトルレジスタ共有機構	電子情報通信学会 リコンフィギャラブルシステム研究会 (RECONF) 技術研究報告		2022/6

【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	Tomoaki Tanaka	東京農工大学	Shared Vector Register of RISC-V for the Future Hardware Acceleration	Sixth Workshop on Computer Architecture Research with RISC-V (CARRV 2022)	2022/6
2	田中友章	東京農工大学	ハードウェアアクセラレーションのためのベクトルレジスタ共有機構	電子情報通信学会 リコンフィギャラブルシステム研究会	2022/6

3	佐藤三久、辻美和子	理研	OpenACC Execution Models for Manycore Processor with ARM SVE	International Conference on High Performance Computing in Asia-Pacific Region Workshops (HPCASIAWORKSHOP 2023)	2023/2
---	-----------	----	--	--	--------

研究開発項目② 次世代コンピューティング技術の開発

※事業原簿 第2部に掲載