

戦略的国際共同研究プログラム(SICORP)
日本－米国共同研究
終了報告書 概要

1. 研究課題名：「国際共同研究で切り開く AI と高性能計算の学際共同研究推進」
2. 研究期間：令和 4 年 2 月 ～ 令和 6 年 3 月
3. 主な参加研究者名：
日本側チーム

	氏名	役職	所属	研究分担
研究代表者	片桐 孝洋	教授	名古屋大学 情報基盤センター	自動チューニング
研究参加者	森下 誠	博士課程学生	名古屋大学 大学院情報学研究科	渡航研究者、AI 適用、AT評価
研究期間中の全参加研究者数			2名	

相手側チーム

	氏名	役職	所属	研究分担
研究代表者	Osni Marques	Staff Scientist	Lawrence Berkeley National Laboratory (LBNL), Computational Research Division	数値アルゴリズム
研究参加者	Andrew Canning	Staff Scientist	LBNL, Computational Research Division	計算科学
研究参加者	Yang Liu	Research Scientist	LBNL, Computational Research Division	AI適用、自動チューニング
研究参加者	Doru Thom Popovici	Research Scientist	LBNL, Computational Research Division	AI適用、数値計算
研究期間中の全参加研究者数			4名	

4. 国際共同研究の概要

多数の科学技術計算アプリケーションをエクサスケールのスーパーコンピュータ環境で高性能に動作させるだけではなく、開発効率を高め保守管理コストを下げることは喫緊の課題である。本国際共同研究では、同分野の米国トップレベルの研究機関に研究者を派遣することで、近年発展が目覚ましい AI 分野と、高性能計算（HPC）分野との研究融合を目指す学際領域研究を進展させ、新たな知の循環の創成を目的にする。米国エネルギー省（DoE）の Exascale Computing Project (ECP)の研究開発物を活用する。

5. 国際共同研究の成果

5-1 国際共同研究の学術成果および実施内容

■実施内容

HPC 分野が牽引するスーパーコンピュータ技術は科学技術分野の基幹技術であり、「富岳」に代表されるスーパーコンピュータは、我が国の科学技術進展において大きな影響を与えている。米国においては、スーパーコンピュータ技術は最も力を入れている科学技術上の分野の1つであり、最もホットなトピックとして世界最高性能を達成する「エクサスケールコンピューティング」技術が盛んに研究されているだけではなく、複数のエクサスケールコンピュータの開発と運用がされている。

本共同研究で交流する米国 LBNL の Marques 博士らのフェンドである米国 DoE の ECP

は、当該分野で最も知られている研究プロジェクトの 1 つである。一方、AI 技術は新興技術であり、米国でも盛んに研究されている。そのため、HPC と AI が融合する研究領域となる本提案は、我が国の科学技術において広く適用できる技術分野に該当し、意義が大きい。

本課題では、以下の項目を実施した。そのことで、最新の HPC ソフトウェアに必要とされる原理と AI 適用性を明らかにした。

- **課題 A1**：米国 ECP 開発ソフトの機能・性能・生産性の調査
- **課題 A2**：A1 の対象における AI 適用可能性調査・性能評価
- **課題 A3**：A1 の対象における高性能実装の性能調査・性能評価

■成果

- **課題 A1**：米国 ECP 開発ソフトの調査を行った。具体的には、AI 手法の 1 つであるベイズ推定を基にした AT ソフトウェアである GPTune の機能・性能を調査した。なお、AT ソフトウェアはチューニングの自動化の観点から生産性向上に資するものである。加えて、米国 DoE の成果物として広く使われている、分散メモリ環境用の密行列用数値計算ライブラリ ScaLAPACK の機能・性能調査を行った。加えて、工学上重要な信号処理である FFT の DoE 成果物である FFTX の機能調査を実施した。
- **課題 A2**：ScaLAPACK と FFTX への AT 機能の拡張性の調査を行った。また、GPTune の性能の精査を行い、特に ScaLAPACK への AT 適用と AT 効果の調査を行った。
- **課題 A3**：GPTune のインストールと性能確認を行った。また、ScaLAPACK の LU 分解ルーチンに GPTune を組み込む実装方式を開発した。加えて、PC の環境において、GPTune による AT の検証を行った。

5-2 国際共同研究による相乗効果

HPC 分野におけるソフトウェア生産性の研究プロジェクトとして、米国 DoE の IDEAS-ECP が世界で最も知られている。加えて Marques 博士は、広く使われている数値計算ライブラリ LAPACK の開発者の 1 人である。代表者が 20 年来進めてきた、我が国独自の AT 技術をさらに進展させ AI 技術への展開を図ることで、HPC 分野における AI 技術による世界的な研究を推進し、若手研究者の人材交流を通して新たな知の循環を形成することが期待できる。

5-3 国際共同研究成果の波及効果と今後の展望

米国と我が国とのスーパーコンピューティング、および、HPC 研究の交流が進展し、次期国策スーパーコンピュータ開発にも知見をフィードバックさせることができる。このことで、将来にわたる計算機環境としての研究基盤の強化につながる。一方、AI 技術が HPC 分野を中心とした十分に適用されなかった分野に浸透し、ソフトウェアの生産性向上、および従来では実現できなかった性能向上の方向性を示すことで、数値シミュレーション分野の研究促進にも寄与できる。

Strategic International Collaborative Research Program (SICORP)
Japan—U.S.A. Joint Research Program
Executive Summary of Final Report

1. Project title : 「Promotion of Interdisciplinary Research between AI and High Performance Computing by International Joint Research」
2. Research period : February 2022 ~ March 2024
3. Main participants :

Japan-side

	Name	Title	Affiliation	Role in the research project
PI	Takahiro Katagiri	Professor	Information Technology Center, Nagoya University	Auto-tuning
Collaborator	Makoto Morishita	Ph.D candidate	Graduate School of Informatics, Nagoya University	Oversea Traveling Researcher, AI adaptation, AT evaluation
Total number of participants throughout the research period:				2

Partner-side

	Name	Title	Affiliation	Role in the research project
PI	Osni Marques	Staff Scientist	Lawrence Berkeley National Laboratory (LBNL), Computational Research Division	Numerical Algorithm
Collaborator	Andrew Canning	Staff Scientist	LBNL, Computational Research Division	Computational Science
Collaborator	Yang Liu	Research Scientist	LBNL, Computational Research Division	AI Adaptation, Auto-tuning
Collaborator	Doru Thom Popovici	Research Scientist	LBNL, Computational Research Division	AI Adaptation, Numerical Computing
Total number of participants throughout the research period:				4

4. Summary of the international joint research

Addressing the urgent challenge extends beyond merely executing numerous scientific and technical computing tasks efficiently on an exascale supercomputer. It encompasses enhancing development productivity and curbing maintenance overheads. Through this international collaboration, experts will be deployed to premier research institutions in the United States, fostering interdisciplinary studies merging AI and high-performance computing (HPC), two rapidly evolving domains. The objective is to forge a fresh cycle of knowledge, utilizing research products from the U.S. Department of Energy's (DoE) Exascale Computing Project (ECP).

5. Outcomes of the international joint research

5-1 Scientific outputs and implemented activities of the joint research

Implementation Details: Supercomputing technology, driven by the HPC field, stands as a cornerstone in the realm of science and technology. In Japan, supercomputers like "Fugaku" wield significant influence over scientific and technological advancements.

Similarly, in the United States, supercomputing technology remains a focal point within the scientific and technological landscape. The forefront of discussion revolves around "exascale computing" technology, aiming to achieve unparalleled performance on a global scale, with multiple exascale computers currently under development and operation.

The Exascale Computing Project (ECP), sponsored by the US Department of Energy (DoE) and spearheaded by Dr. Marques of LBNL, represents one of the most renowned research endeavors in this domain. Concurrently, AI technology undergoes rapid development and active research in the United States. Hence, this proposal, which amalgamates HPC and AI research areas, holds significant importance, aligning with technical fields with broad applicability within Japan's scientific and technological landscape.

This assignment entailed the following tasks, elucidating the principles and AI integration essential for cutting-edge HPC software:

- **Assignment A1:** Investigation into the functionalities, performance, and productivity of US ECP development software.
- **Assignment A2:** Evaluation of AI applicability and performance for targets outlined in A1.
- **Assignment A3:** Performance assessment and evaluation of high-performance implementations for A1 targets.

Results:

- **Assignment A1:** Our investigation delved into the US ECP development software, focusing on GPTune, an Auto-Tuning (AT) software leveraging Bayesian estimation—a form of AI—to enhance productivity through automated tuning. Additionally, we scrutinized ScaLAPACK, a widely utilized numerical computation library for dense matrices in distributed memory environments, developed under the auspices of the US DoE. Furthermore, we conducted a functional analysis of FFTX, a DoE product facilitating Fast Fourier Transform (FFT), a pivotal signal processing technology in engineering.
- **Assignment A2:** We assessed the adaptability of AT functionalities to ScaLAPACK and FFTX, alongside scrutinizing the performance of GPTune. Specifically, we explored the application of AT to ScaLAPACK and its resultant impact.
- **Assignment A3:** We deployed GPTune and evaluated its performance, devising a methodology to integrate GPTune into ScaLAPACK's LU decomposition routine. Moreover, we validated AT utilization via GPTune within a PC environment.

5-2 Synergistic effects of the joint research

The IDEAS-ECP project under the US DoE stands out as the foremost research initiative worldwide focusing on software productivity in the HPC domain. Additionally, Dr. Marques, a key figure in this endeavor, has contributed significantly to the development of LAPACK, a widely utilized numerical computation library. By propelling forward Japan's distinctive Auto-Tuning (AT) technology, championed by the representative for over two decades, and integrating it with AI advancements, we aim to foster global research at the intersection of AI and HPC. Through fostering exchanges among young researchers, we anticipate generating novel ideas and nurturing a cycle of knowledge dissemination.

5-3 Scientific, industrial or societal impacts/effects of the outputs

Collaborative exchanges in supercomputing and HPC research between the United States and Japan will advance, facilitating the reciprocal flow of knowledge into the development of the next generation of national supercomputers. This synergy will fortify research infrastructure, shaping a robust computing environment for future endeavors. Concurrently, AI technology has permeated into previously unexplored realms, notably the HPC domain. By elucidating pathways for enhancing software productivity and achieving performance milestones previously deemed unattainable, AI-driven advancements are propelling forward research in numerical simulation fields. This integration holds promise for furthering advancements in the realm of computational research.

国際共同研究における主要な研究成果リスト

1. 論文発表等

* 原著論文（相手側研究チームとの共著論文）発表件数：計 1 件

・ 査読有り：発表件数：計 0 件

・ 査読無し：発表件数：計 1 件

1. Morishita Makoto, Katagiri Takahiro, Marques Osni, Liu Yang, Hoshino Tetsuya, Nagai Toru, Kawai Masatoshi, "Performance Evaluation of GPTune for Parameter Auto-Tuning", 2023, 情報処理学会研究報告, 2024-HPC-194, 1-6

* 原著論文（相手側研究チームを含まない日本側研究チームの論文）

・ 査読有り 発表件数：計 0 件

・ 査読無し 発表件数：計 0 件

* その他の著作物（相手側研究チームとの共著総説、書籍など） 発表件数：計 0 件

* その他の著作物（相手側研究チームを含まない日本側研究チームの総説、書籍など） 発表件数：計 0 件

2. 学会発表

* 口頭発表（相手側研究チームとの連名発表）

発表件数：計 4 件（うち招待講演：3 件）

* 口頭発表（相手側研究チームを含まない日本側研究チームの発表）

発表件数：計 1 件（うち招待講演：1 件）

* ポスター発表（相手側研究チームとの連名発表）

発表件数：計 1 件

* ポスター発表（相手側研究チームを含まない日本側研究チームの発表）

発表件数：計 0 件

3. 主催したワークショップ・セミナー・シンポジウム等の開催

1. ミニ・シンポジウム「Challenges for Attaining High-performance in Numerical Software」、主催者：Takahiro Katagiri (Nagoya University, Professor), Toshiyuki Imamura (RIKEN, PI), Osni A. Marques (Lawrence Berkeley National Laboratory, Staff Scientist)、国際会議 ICIAM2023 の公募ミニ・シンポジウム)、東京都、早稲田大学、2023 年 8 月 23 日、参加者：15 名程度

2. 国際ワークショップ「Workshop on Numerical Computing and Software Productivity in Nagoya (WNCSP2023)」、主催者：片桐孝洋（名大・教授）、名古屋大学情報基盤センター、2023 年 8 月 30 日、参加人数：15 名程度

3. ミニ・シンポジウム「MS53 Aspects of Software Engineering and Extreme Scale Computing」、主催者：Takahiro Katagiri (Nagoya University, Professor), Toshiyuki Imamura (RIKEN, PI), Osni A. Marques (Lawrence Berkeley National Laboratory, Staff Scientist)（学会の主催者：米国応用数学会 (SIAM)、国際会議 PP24 の公募ミニ・シンポジウム）、Baltimore、Maryland、Load Baltimore Hotel (米国)、2024 年 3 月 7 日、参加者：10 名程度

4. 研究交流の実績（主要な実績）

1. 共同研究実施、2023 年 6 月 12 日～8 月 13 日、米国ローレンスバークレー研究所、63 日間滞在

5. 特許出願

研究期間累積出願件数：0 件

6. 受賞・新聞報道等

該当なし

7. その他

該当なし