

Society5.0を支える 大規模研究施設連携による ビッグデータ収集・解析・利活用

課題代表者：佐藤 賢斗（高性能ビッグデータ研究チーム）



理化学研究所
計算科学研究センター
RIKEN Center for Computational Science

背景

Society5.0の実現に向けては、実験データと計算資源を統合的に活用するための高性能なデータ基盤が不可欠である。SPring-8やSACLAといった大型実験施設では、ペタバイト～テラバイト級の膨大なデータが生成されているが、それらを即時に処理・解析・活用するための計算基盤との連携は十分に整備されていなかった。特に、スーパーコンピュータ「富岳」との接続を通じて、データの圧縮・転送・解析・再利用を包括的に支援する基盤の構築が強く求められていた。

目的

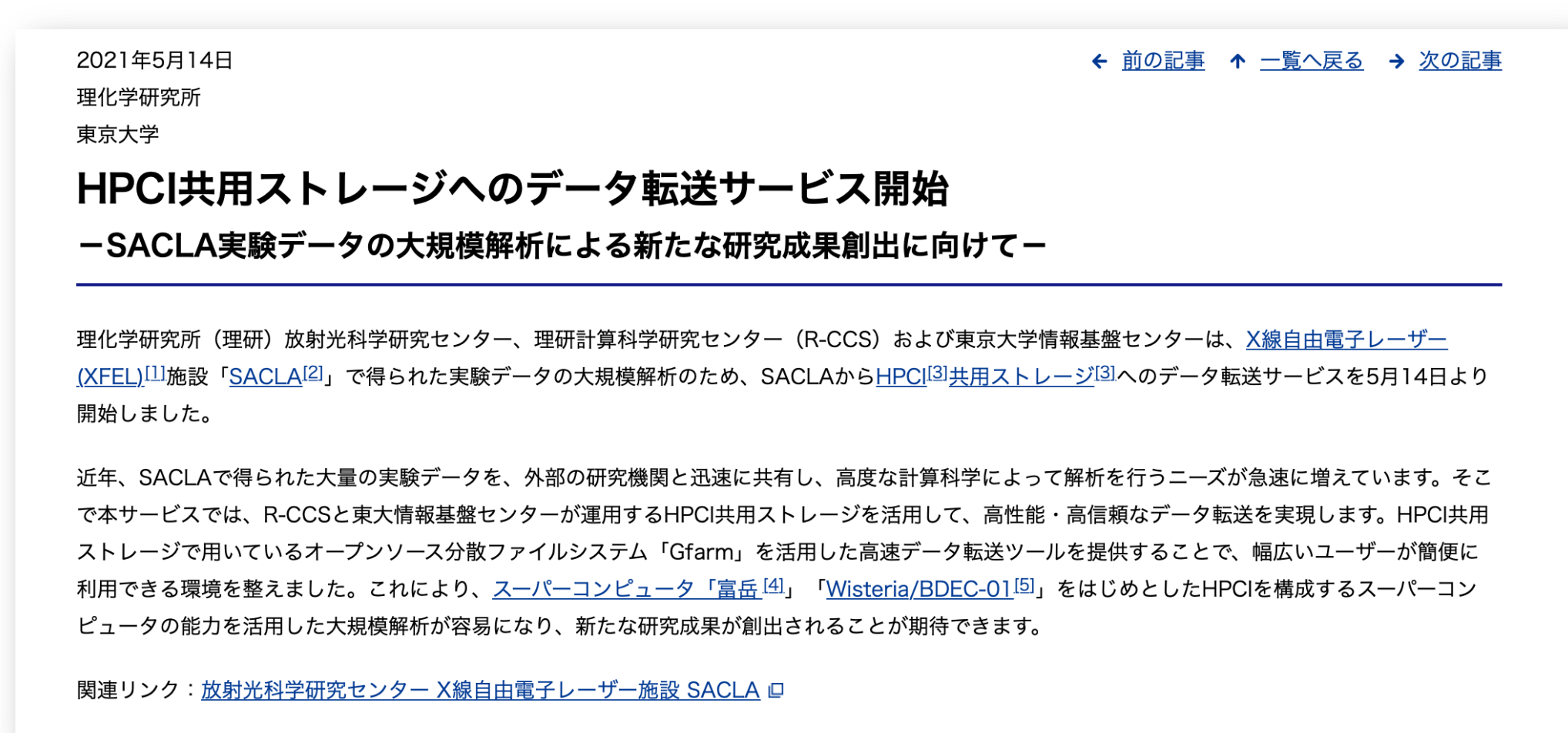
SPring-8やSACLAで生成される大規模かつ高頻度の実験データを、スーパーコンピュータ「富岳」や他のHPCIシステムにおいて迅速かつ高効率に解析・活用可能とするためのビッグデータ基盤を構築することが本研究の目的である。具体的には、SACLAとHPCI共有ストレージを結ぶデータ転送サービスの整備、並びにネットワーク帯域やストレージ効率の制約を克服するための高圧縮率かつ高スループットなデータ圧縮技術の開発を行う。また、これらの基盤を通じて、実験データのリアルタイム利活用や、解析結果の循環的な再活用を促進し、先端科学技術の加速と地域研究施設の価値向上に資することを目指す。

成果

- 成果1**：SACLAから「富岳」を含むHPCIシステムへのデータ転送サービスを構築し、2021年度より本格的に実運用を開始
- 成果2**：時系列画像データを対象としたAIベースの高効率圧縮フレームワークTEZipを新たに開発し、高い圧縮率を達成
- 成果3**：TEZipを圧縮ライブラリ共通APIであるLibPressioと統合することで多様な言語環境からの利用を可能にし応用範囲を拡大
- 成果4**：X線CTデータに特化したデータ圧縮手法も開発し従来手法よりも高い圧縮率を達成しつつ科学的精度を維持を実現

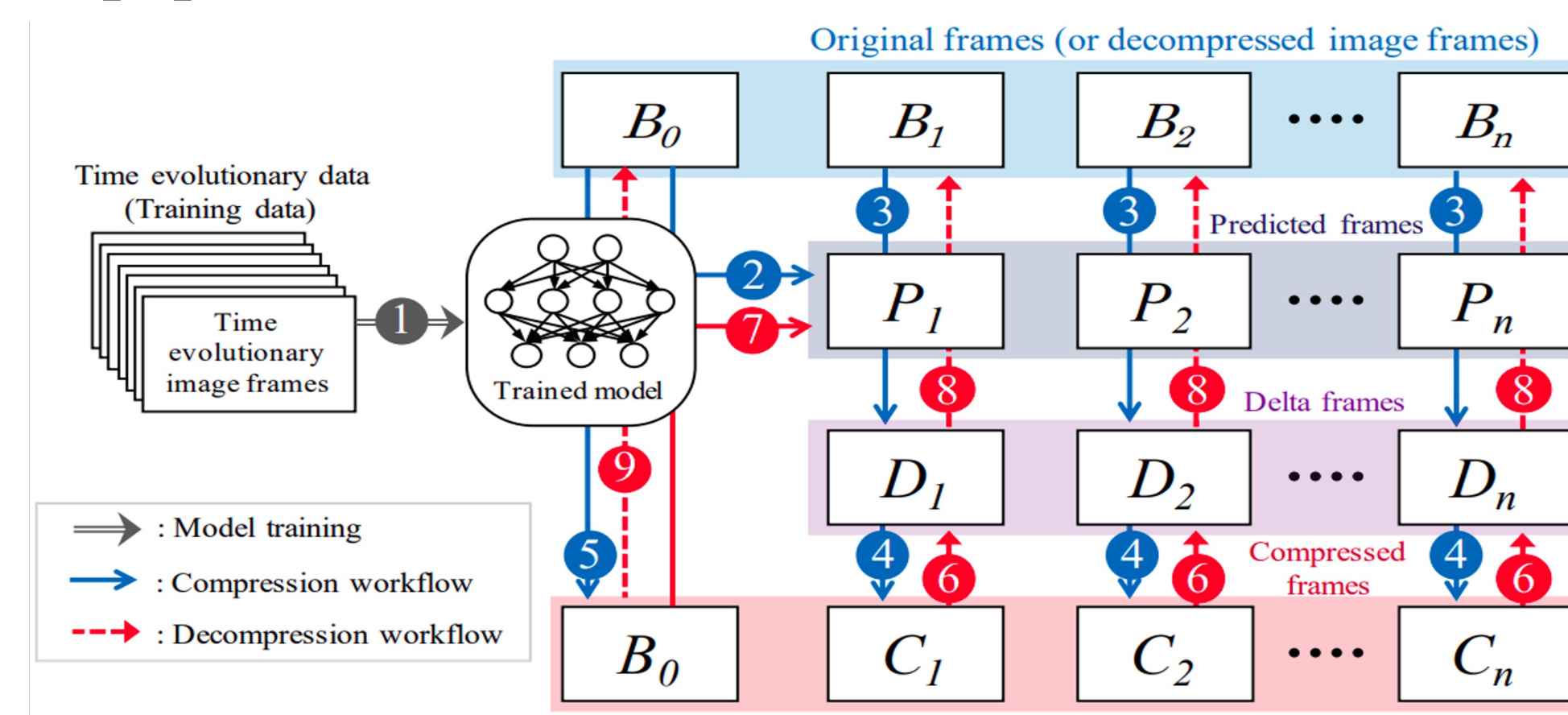
成果1 SACLAからHPCIへのデータ転送サービスの運用を開始

SACLAで得られる大規模な実験データを富岳で解析可能とするため、Gfarmを用いた高信頼かつ高速なデータ転送サービスを構築し、2021年より実運用を開始した[1]。これにより、放射光施設とHPCIシステム間のデータ連携が容易になり、データ解析の利便性が向上した。



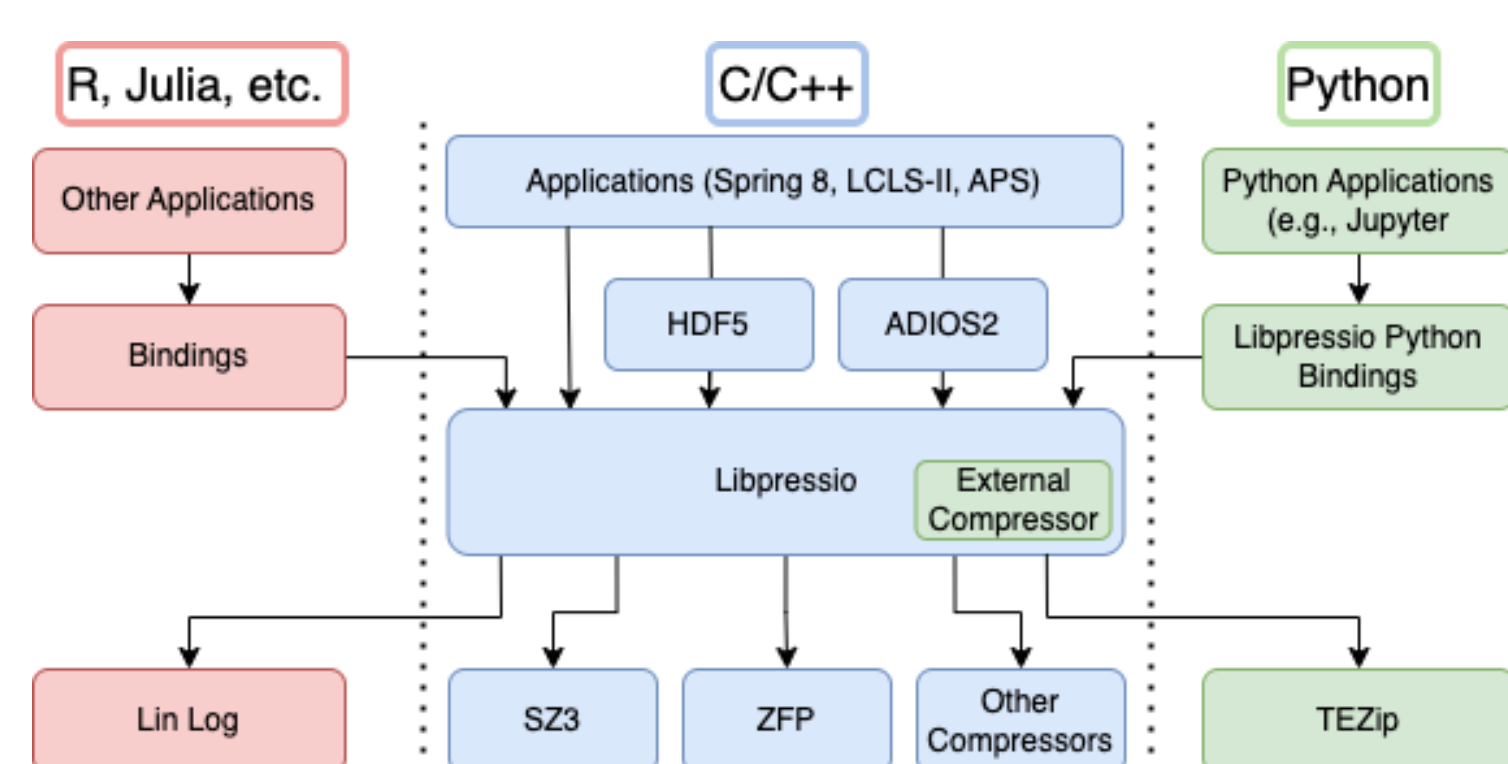
成果2 時系列画像向けAI圧縮フレームワークTEZipを開発

膨大な時系列実験データを効率的に保存・転送するため、深層学習に基づく画像予測と差分圧縮を組み合わせたAI圧縮フレームワークTEZipを開発した[2]。LibTorch版への移植やConvLSTMモデルの導入により、従来に比べ圧縮・復元の速度と柔軟性が改善され、実用性の高い圧縮ツールが開発された[3]。



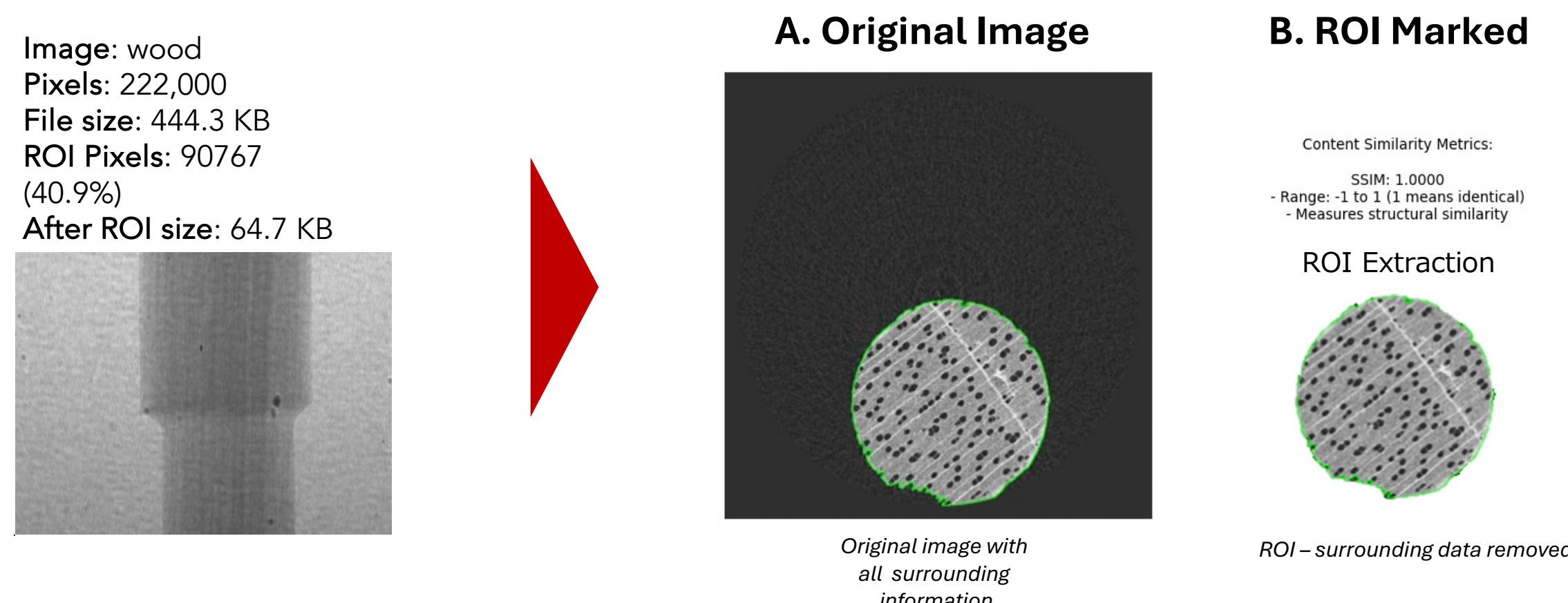
成果3 TEZipをLibPressioと統合し多言語環境での利用を拡大

米国ANLが開発する圧縮ツール共通API “LibPressio” とTEZipを統合することで、Python以外の言語環境（C/C++、R、Julia等）でもTEZipの圧縮機能が利用可能となった[4]。これにより、科学計算や可視化、データ解析など多様な用途での活用が容易になり、国内外の研究者による利用拡大が期待される。



成果4 X線CTデータに特化型圧縮手法を開発し高い圧縮率を実現

X線CT画像特有のノイズ特性と再構成要件に対応するため、背景除去およびSinogramへの可逆変換を組み合わせた新たな圧縮手法を開発した。この手法により、従来比17.2%の圧縮率向上を達成し、画像の本質を損なうことなく高効率な保存・転送を可能にした。



地域 貢献

- ❖ 貢献1：SPring-8/SACLAおよび富岳の連携強化により、県内研究施設の利便性と活用価値を向上
- ❖ 貢献2：兵庫県立大学との連携によるインターンシップ受け入れなど、地域人材育成に貢献
- ❖ 貢献3：大規模研究設備を核とした研究成果を地域イベント・ワークショップ等で成果発表

[1] HPCI共用ストレージへのデータ転送サービス開始 — SACLA実験データの大規模解析による新たな研究成果創出に向けて — (https://www.riken.jp/pr/news/2021/20210514_1/index.html)

[2] Rupak Roy, Kento Sato, Subhadeep Bhattacharya, Xingang Fang, Yasumasa Joti, Takaki Hatsui, Toshiyuki Hiraki, Jian Guo and Weikuan Yu, “Compression of Time Evolutionary Image Data through Predictive Deep Neural Networks”, CCGrid 2021, May, 2021

[3] Mina Atef Yousef, Amarjit Singh, Kento Sato, “Refactoring TEZip: Integrating Python-Based Predictive Compression into an HPC C++/LibTorch Environment”, HPDC '25, July, 2025 (Poster)

[4] Isita Talukdar, Amarjit Singh, Robert Underwood, Kento Sato, Weikuan Yu, “Integrating TEZIP into LibPressio: A Case Study of Integrating a Dynamic Application into a Static C Environment”, SC23, Research Poster, Denver, USA, Nov, 2023



本研究は公益財団法人計算科学振興財団 研究教育拠点（COE）形成推進事業の助成を受けたものです。

This work was supported by FOCUS Establishing Supercomputing Center of Excellence.