

放射光実験の大容量データの即時圧縮技術を開発

—SPring-8 のデータを 8,600 分の 1 に圧縮—

概要

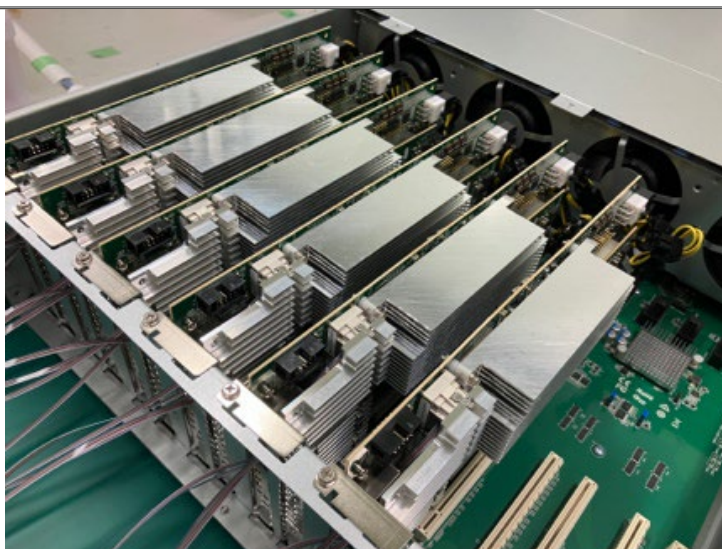
理化学研究所（理研）放射光科学研究センター制御情報・データ創出基盤グループの初井宇記グループディレクター、高輝度光科学研究センター研究 DX 推進室の西野玄記主幹研究員（理研放射光科学研究センター客員研究員）、城地保昌室長（理研放射光科学研究センタービームライン制御解析チームチームリーダー）、東北大学大学院理学研究科の齋藤真器名准教授（理研放射光科学研究センター客員研究員（研究当時））らの共同研究グループは、放射光実験における X 線画像検出器から出力される大容量データを即時圧縮するデータ処理基盤を開発し、大型放射光施設「SPring-8」^[1]において構築し、運用を始めました。

本研究で開発されたデータ処理基盤は、エクサバイト（EB、1EB は 100 京バイト、1 テラバイト（TB、1TB は 1 兆バイト）の記憶装置にすると 100 万台分）級のデータ生成が見込まれる、「SPring-8-II」^[2]などの次世代放射光施設を支える中核基盤技術です。また、本技術は、幅広い計測分野における大容量データ処理を支える技術としての展開も期待されます。

本研究で対象としたガンマ線準弾性散乱（QEGS）実験^[3]では、84 万画素の CITIUS 検出器^[4]を 17.4 キロヘルツ（kHz）で動作させます。このとき検出器からは毎秒 27 ギガバイト（GB、1GB は 10 億バイト）のデータが生成され、1 週間では約 19 ペタバイト（PB、1PB は 1,000 兆バイト、1TB の記憶装置にすると 1,000 台分）に達します。

共同研究グループは、FPGA データ処理基板上の書き換え可能な半導体回路、FPGA^[5]により画素ごとの前処理をデータ生成後に直ちに実行し、さらに可逆圧縮（元データを損なわず正確に復元できるように圧縮すること）を組み合わせることで、平均約 8,600 分の 1 という高い圧縮率を達成しました。圧縮後のデータは施設内データセンターへ自動転送され、高性能計算機を用いた並列解析により、ユーザーは実験中に解析結果を確認し、測定条件を最適化できます。

本研究は、科学雑誌『*Journal of Synchrotron Radiation*』オンライン版（3 月 25 日付）に掲載されました。



FPGA データ処理基板を収めた計算機

背景

大型放射光施設では、物質中の原子や分子の構造や運動を高精度に観測するため、検出器の高速化・高精細化が急速に進んでいます。時間変化を捉える測定では、1秒間に数万枚の画像を取得する必要があり、1回の実験で生成されるデータ量は飛躍的に増大しています。

本研究で対象とした QEGS 実験では、84万画素の検出器を 17.4kHz で動作させることで、ナノ秒（10億分の1秒）スケールの原子・分子の動きを観測します。その結果、1週間当たり 19PB もの大容量データが発生します。

従来のビームライン計算機環境では、このようなデータをそのまま保存・転送・解析することは現実的に難しく、解析には数日以上を要することもありました。そのため、実験中に結果を確認して測定条件を最適化することが困難でした。

さらに、SPring-8-II では光源の高輝度化に伴ってデータ量はさらに増加し、EB 級に達すると見積もられています。このため、取得後にまとめて処理する従来の運用から、取得と同時に処理して結果を返すデータ処理基盤への転換が不可欠です。

研究手法と成果

本研究では、検出器から出力される毎秒 27GB のデータをそのまま保存するのではなく、データ生成直後に前処理と可逆圧縮を行う新しいアプローチを採用しました。具体的には、FPGA データ処理基板に搭載した FPGA（書き換え可能な半導体回路）上で、画素ごとの背景信号を補正し、ノイズを除去するしきい値処理、短時間に連続して得られる画像を統合する処理を実行するようにしました。これらの処理により、科学的に意味のある信号を保ちながら、重複した情報や信号以外の成分を減らしました。その結果、データの情報量（エントロピー）

を低減でき、後段の可逆圧縮の効率が大きく向上しました。

FPGA で前処理されたデータは、ビームラインの計算機において可逆圧縮を施しました。その結果、平均約 8,600 分の 1 という高い圧縮率を達成し、実際の 1 週間の実験では約 19PB の元データを約 2.2TB まで削減しました。圧縮後のデータは SPring-8 内のデータセンターへ自動転送され、取得から 2~3 分以内に保存されます。この可逆圧縮処理により、転送に必要なネットワーク帯域も大幅に削減されました。

データセンターでは、高性能計算機（HPC）を用いた並列解析を行います。SPring-8 データセンターでは HPC 利用基盤として Open OnDemand^[6]が導入されており、本研究ではその既存基盤の上に、QEGS 実験専用の解析アプリケーション群を構築しました。これにより、専門的なコマンド操作を必要としないウェブブラウザベースの解析環境を実現しました。

この統合基盤により、初期的な解析結果は取得開始から約 6 分以内に得られ、さらに詳細な解析結果もその約 7 分後に得られるようになりました。これにより、ユーザーは実験中に解析結果を確認し、その場で測定条件を最適化できるようになりました。従来は実験終了後に数日以上かかる場合もあった解析を、実験中に完了できる体制を実現しました。

本研究の成果は、高効率な圧縮技術の実現にとどまらず、検出器によるデータ生成から、FPGA による前処理、可逆圧縮、転送、高性能計算機による並列解析、さらにウェブブラウザから利用可能な解析環境までを一体化した統合データ処理基盤（図 1）を、実際のユーザー実験で運用し、その有効性を実証した点にあります。

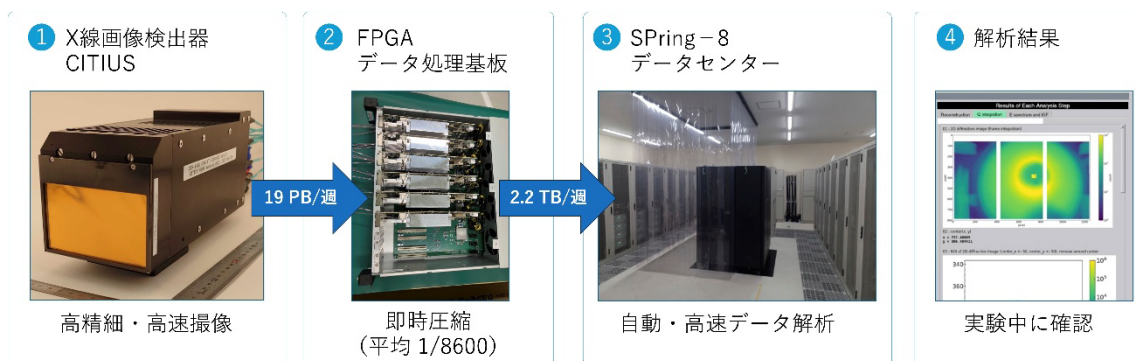


図 1 本研究で構築した検出器データ処理基盤の概念図

測定データの前処理、圧縮、転送、解析に至るまでのデータの流れを左から右へ示している。左から順に、① 毎秒 27GB のデータを生成する X 線画像検出器、② FPGA データ処理基板を収めた計算機、③ SPring-8 施設内のデータセンター、④ 高性能計算機上で動作する解析アプリケーションの結果表示画面を示す。

今後の期待

本研究で構築したデータ処理基盤は、放射光科学における「測定後に解析する」運用から「測定しながら解析する」運用への転換を可能にしました。これにより、実験中に結果を確認しながら測定条件を最適化できるようになり、研究効率の

向上が期待されます。

SPring-8-II では EB 級データが見込まれており、本研究の統合データ処理基盤はその中核基盤技術となります。

さらに共同研究グループは、1 日当たり約 30PB のデータ生成が見込まれる半導体向け X 線タイコグラフィ顕微鏡^[7]への応用にも取り組んでいます。タイコグラフィでは、ナノメートルレベル (nm、1nm は 10 億分の 1 メートル) の分解能で 3 次元再構成を行うため、QEGS を上回る大容量データが発生します。本研究の即時圧縮・転送・並列解析の技術は、こうした次世代 X 線計測の実現に不可欠な基盤となります。

本研究のデータ処理技術は、放射光施設や X 線自由電子レーザー (XFEL)^[8]に特有のものではなく、高速かつ高解像度な検出器から大量の画像データが出力される計測装置に広く応用できる可能性があります。大容量データの処理は多くの先端計測分野に共通する課題であり、本技術はその解決を支える基盤技術として期待されます。

論文情報

<タイトル>

FPGA-accelerated streaming data reduction achieving an average compression ratio over 8000 in a 17.4 kHz, 840 kpixel CITIUS detector for quasi-elastic gamma-ray scattering

<著者名>

Haruki Nishino, Masashi Kobayashi, Toshiyuki Nishiyama Hiraki, Yoshiaki Honjo, Kyosuke Ozaki, Mitsuhiro Yamaga, Nobumoto Nagasawa, Yoshitaka Yoda, Yasumasa Joti, Makina Saito and Takaki Hatsui

<雑誌>

Journal of Synchrotron Radiation

<DOI>

[10.1107/S1600577526000883](https://doi.org/10.1107/S1600577526000883)

補足説明

[1] 大型放射光施設「SPring-8」

理化学研究所が所有する兵庫県の播磨科学公園都市にある世界最高性能の放射光を生み出す大型放射光施設で、利用者支援などは高輝度光科学研究センター (JASRI) が行っている。SPring-8 (スプリングエイト) の名前は Super Photon ring-8 GeV に由来する。SPring-8 では、放射光を用いてナノテクノロジー、バイオテクノロジーや産業利用まで幅広い研究が行われている。

[2] SPring-8-II

SPring-8 の大幅な性能向上を目指した次期計画の名称。詳しくは、下記プレスリリースを参照。

2024 年 10 月 24 日プレスリリース「SPring-8 光源大改修の設計指針を公表」

https://www.riken.jp/press/2024/20241024_1/

[3] ガンマ線準弾性散乱 (QEGS) 実験

放射光を用いて、物質中の原子・分子の動きを 100 ピコ秒(1 ピコ秒は 1 兆分の 1 秒) ~100 ナノ秒 (1 ナノ秒は 10 億分の 1 秒) の時間領域で調べる測定手法。散乱された X 線のわずかなエネルギー変化を解析することで、材料や生体関連試料などの微視的な動的挙動を評価できる。詳しくは、下記プレスリリースを参照。

2024 年 6 月 18 日プレスリリース「10 億分の 1 秒の原子運動を見る放射光技術を開発」
https://www.riken.jp/press/2024/20240618_2/

[4] CITIUS 検出器

理化学研究所が開発した、画素サイズ 72.6 マイクロメートル (μm 、1 μm は 100 万分の 1 メートル)、フレームレート (1 秒間に撮像できる静止画の枚数) 17.4kHz で動作する X 線画像検出器。本研究で用いたシステムでは、3 枚のセンサーを組み合わせて合計 84 万画素の検出器として運用されている。

[5] FPGA

製造後でも回路構成を変更できる集積回路。並列処理や入出力制御を柔軟に実装できるため、計測装置などでデータ処理や制御を組み込む用途に用いられる。FPGA は Field Programmable Gate Array の略。

[6] Open OnDemand

高性能計算機(HPC)を、ウェブブラウザ上の画面から利用できるようにする仕組み。コマンド操作に頼らずに、計算の実行、ファイルの管理、利用状況の確認などを行える。

[7] X 線タイコグラフィ顕微鏡

試料に X 線を当てる位置を少しずつ変えながら多数のデータを取得し、そのデータを計算機で再構成することで、試料内部の構造を高い分解能で可視化する顕微鏡。厚みのある試料にも適用でき、半導体デバイスなどの内部構造を非破壊で観察する技術として期待されている。

[8] X 線自由電子レーザー (XFEL)

X 線領域におけるレーザーのこと。従来の半導体や気体を発振媒体とするレーザーとは異なり、真空中を高速で移動する電子ビーム (自由電子) を媒体とするため、原理的な波長の制限はない。また、数フェムト秒 (1 フェムト秒は 1,000 兆分の 1 秒) の超短パルスを出力する。XFEL は X-ray Free Electron Laser の略。

共同研究グループ

理化学研究所 放射光科学研究センター

制御情報・データ創出基盤グループ

グループディレクター 初井宇記 (ハツイ・タカキ)

特別研究員 (研究当時) 平木俊幸 (ヒラキ・トシユキ)

エンジニアリング部門 エンジニアリングチーム

テクニカルスタッフ I 尾崎恭介 (オザキ・キョウスケ)

テクニカルスタッフ I 本城嘉章 (ホンジョウ・ヨシアキ)

高輝度光科学研究センター

研究 DX 推進室

主幹研究員 西野玄記 (ニシノ・ハルキ)

(理研 放射光科学研究センター 客員研究員)

室長 城地保昌 (ジョウチ・ヤスマサ)

(理研 放射光科学研究センター ビームライン制御解析チーム
チームリーダー)

加速器部門 加速器制御グループ 機器制御チーム

チームリーダー 山鹿光裕 (ヤマガ・ミツヒロ)

(理研 放射光科学研究センター 客員研究員)

分光・イメージング推進室 精密分光チーム

特任研究員 依田芳卓 (ヨダ・ヨシタカ)

(理研 放射光科学研究センター 客員研究員)

研究員 永澤延元 (ナガサワ・ノブモト)

(理研 放射光科学研究センター 客員研究員)

東北大学 大学院理学研究科 物理学専攻

准教授 齋藤真器名 (サイトウ・マキナ)

(理研 放射光科学研究センター 客員研究員 (研究当時))

大学院生 (研究当時) 小林 正 (コバヤシ・マサシ)

(理研 放射光科学研究センター 研修生 (研究当時))

研究支援

本研究は、理化学研究所放射光科学研究センターSACLA/SPring-8 基盤開発プログラム 2022-2023 によるサポートを受けて実施され、日本学術振興会 (JSPS) 科学研究費助成事業基盤研究 (B)「ガラスの力学特性を支配する Johari-Goldstein 緩和過程の普遍的機構の解明 (研究代表者: 齋藤真器名、JP24K00592)」、科学技術振興機構 (JST) 戦略的創造研究推進事業 CREST「階層的時空構造と動的不均一性から紡ぐナノ力学機構の理解と制御 (研究代表者: 山本潤、JPMJCR2095)」の助成を受けて行われました。

本研究で用いた FPGA データ処理基板は、理化学研究所と東京エレクトロン デバイス株式会社が共同で開発しました。

また、本研究におけるデータ処理基盤の構築は、グローリーテクニカルソリューションズ株式会社および日東コンピューターサービス株式会社の協力を得て実施しました。

発表者・機関窓口

<発表者> ※研究内容については発表者にお問い合わせください。

理化学研究所 放射光科学研究センター 制御情報・データ創出基盤グループ

グループディレクター 初井宇記 (ハツイ・タカキ)

Email: hatsui@spring8.or.jp

高輝度光科学研究センター 研究 DX 推進室

主幹研究員 西野玄記 (ニシノ・ハルキ)

(理研 放射光科学研究センター 客員研究員)

室長 城地保昌 (ジョウチ・ヤスマサ)



(理研 放射光科学研究センター ビームライン制御解析チーム チームリーダー)
Email: haruki.nishino@spring8.or.jp (西野)

東北大学 大学院理学研究科
准教授 齋藤真器名 (サイトウ・マキナ)
(理研 放射光科学研究センター 客員研究員 (研究当時))

<機関窓口>

理化学研究所 広報部 報道担当
Tel: 050-3495-0247
Email: ex-press@ml.riken.jp

高輝度光科学研究センター (JASRI) 利用推進部 普及情報課
Tel: 050-3502-3763
Email: kouhou@spring8.or.jp

東北大学 大学院理学研究科 広報・アウトリーチ支援室
Tel: 022-795-6708
Email: sci-pr@mail.sci.tohoku.ac.jp
