

O-13

全天 X 線監視装置 MAXI のデータを用いた突発天体発見のための再解析準備

Preparation for the reanalysis of MAXI data to find undiscovered transient objects

○田中一輝¹, 根来均²*Kazuki Tanaka¹, Hitoshi Negoro²

Abstract: MAXI, Monitor of All-sky X-ray Image, is mounted on the International Space Station. MAXI is working more than 7 years. We try to find undiscovered X-ray transients by re-analyzing MAXI data since the current analysis program to discover transients has been enhanced compared with that used in the initial phase of the mission. We modified a program to retrieves data from a database containing past data, and successfully shortened the time to acquire data.

1. 導入

1.1 全天 X 線監視装置 MAXI

MAXI (Monitor of All-sky X-ray Image) は, 2009 年 7 月に国際宇宙ステーション (ISS) の日本実験棟「きぼう」船外実験プラットフォームに搭載された全天 X 線監視装置である。MAXI は, ISS が約 92 分で地球を一周するのにともない全天をスキャン観測し, 発見した突発天体³を世界に速報している。しかし, 運用当初はバックグラウンドの処理や, 突発天体の閾値などが最適化されておらず, 明るい天体しか発見できていなかった可能性がある。よって最新のシステムを用いた過去のデータを再解析が望まれる。本研究では, その効率的な再解析を行う。

1.2 MAXI-DB

再解析に必要な MAXI のデータは MAXI-DB と呼ばれるデータベースに格納されている。一般的な定点 X 線観測衛星では一定の観測時間で同じ天体を観測し, その観測データを 1 つのファイルなどで管理するが, MAXI では ISS の軌道運動により常時スキャン観測を行っているので 1 つの天体だけを見ているわけではない。

このように大量のデータが 1 つのデータベースに格納されているため, 再解析を行うには時間がかかる。これを迅速に行うためにデータベースからデータを取り出すためのプログラムを改良した。

2. システム改良

MAXI-DB に格納されたデータを取り出すとき, その後の天体解析をするための FITS⁴ というフォーマットで出力する。このとき高橋知義が 2007 年に作成した FITSdump というプログラムを使用する [1]。その後浅田真がこのプログラムによるデータ取得の高速化ができることを示した [2]。今回, これを実際使用しているプログラムに実装し, さらに改良を加えたことによってデータ取得にかかる時間が約半分程度から最大 5 分の 1 程度まで短縮することに成功した。

プログラムがデータを取り出す方法は次の 3 つの手順で行われている。1. 取り出すデータが全部で何行あるかを確認し, 保存に必要なメモリを確保する。2. 確認したデータを取り出し, 1. で確保したメモリ領域に保存する。3. FITS ファイルとして書き出す。浅田 [2] は, 2 の手順でデータを `ecpg5` により 1 行 1 行取り出す方法ではデータベースとの通信に時間がかかってしまうので, `libpq6` を使うことで 1 度に全ての行を取り出すことによって高速化ができることを示した。今回の改良ではリスト形式のデータ保存方式を用いることにより 1 の手順でデータを確認すると同時にデータを取得するよう変更した。

これらの改良によりデータ取得速度が約 1.3 倍から最大 5 倍程度まで上がった。今回の計測はどのような環境で再解析用のデータベースを構築するのが効率的かを複数の環境で調べた。そのために 3 つの CPU の異なる計算機, 3 種類の記憶媒体を用いて同じ計測を行った。計測したデータは再解析に使う時間スケールの 1 scan (ISS が地球を 1 周する内の 40~100 秒間), 4 orb (同 4 周回分) 1 day, 4 day と, 長期間の場合として 30 day, 60 day のデータを取り出すときにかかる時間を計測した。使用したデータは Crab という星の周囲 10° のデータで, 検索対象の

¹日大理工・院(前)・物理²日大理工・教員・物理³X 線新星やガンマ線バーストのような突然輝き出す天体。⁴<https://heasarc.gsfc.nasa.gov/docs/heasarc/fits.html>。⁵Oracle 等で用いられる C 言語上でデータベースを操作する一般的な埋め込み SQL プリプロセッサ。⁶PostgreSQL 特有の C 言語でデータベースを操作する埋め込み SQL プリプロセッサ。

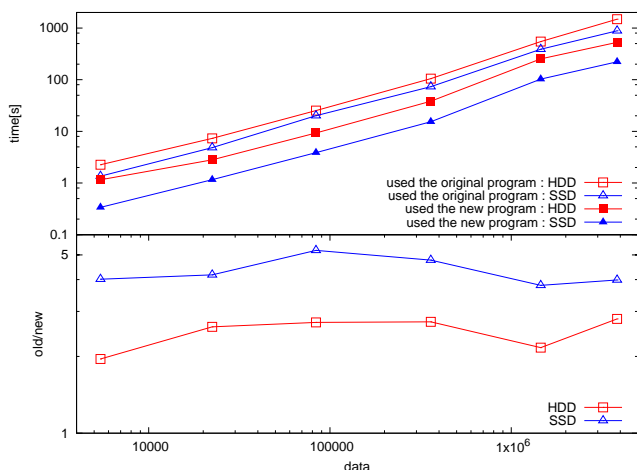


Figure 1: Retrieving data time from HDD (red) and SSD (blue) in Computer 1 using the original and modified programs (upper panel) and improved ratios (lower panel)

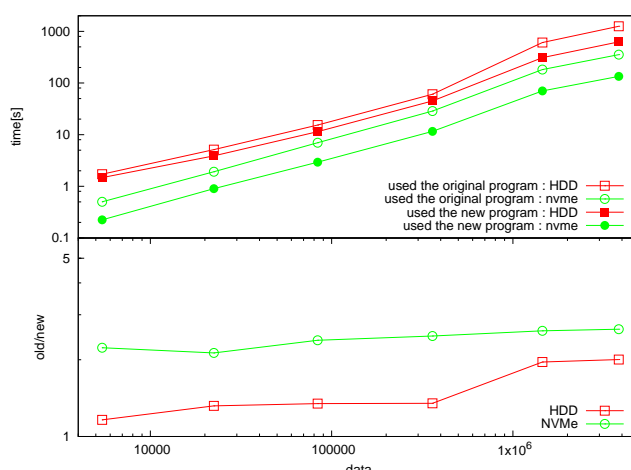


Figure 2: Same as Figure 1 but for HDD (red) and NVMe (green) in Computer 2.

時間は 2010 年 10 月 31 日 24 時を終点に取得したい時間分の始点を検索対象としている。

まず最初に計算機 1 に使われている HDD と SSD の 2 つの記憶媒体で計測を行い、Figure 1 の結果を得た。また計算機 2 でさらに高速な記憶媒体である NVMe⁶ でも速度に差が出るのかを検証した。計算機 2 の NVMe でも検索の時間が約半分になるという Figure 2 の結果が得られた。測定の結果、データ数が多くなっても時間の短縮率は大きく変化しなかった。また、計算機 2 の HDD の 30 day と 60 day で速い検索結果が出た理由は、改良前のプログラムだったためと考えられる。検索期間が長くなるとデータがキャッシュの容量を超えてしまい、検索に時間がかかった。また、本来 SSD の読み込み速度は 426.8 MB/s⁷、NVMe の読み込み速度は 1005 MB/s⁷ と差があるのに、今回の結果は 60 day のデータでも SSD で 149.6 秒、NVMe で 134.3 秒と 11%程度の差しかなかった。これはボトルネックになっている部分が補助記憶装置以外にあることを示している。

また計算機 1 と計算機 2 の測定結果で通信 1 回にかかる時間に差があることがわかった。これは先行研究の浅田 [2] が、データベースとの通信 1 回にかかる時間を計測結果より 100 μ 秒と計算し、計算機 1 でも近い値が出たが、の計算機 2 で計測した結果これが 10 μ ~20 μ 秒程度という結果になった。計算機 1,2 ではそれぞれ異なる OS の Cent OS 6,7 を用いていた。各計算機よりバス速度がより低速な計算機 3 で Cent OS 6 と Cent OS 7 の双方を同じ計算機で動かし、通信にかかる時間を計測したところ通信速度の差は得られなかった。

計測の結果からデータベースを構築するのは補助記憶装置として NVMe ではなく安価で大容量 (~20 TB) 化が容易な SSD を使うことを決定した。

3. 今後の課題

現在は再解析用のデータベースに筑波宇宙センターで FITS データとして保存されている過去のデータを挿入するため、FITS データをデータベースに再度読み込める形に直す作業をしている。また現在使われている解析用のユーザーインターフェイスを高速化し、過去のデータの再解析を進めていく。

参考文献

- [1] 高橋知義, 日本大学大学院 理工学研究科 物理学専攻 修士論文, 2007
- [2] 浅田真, 日本大学大学院 理工学研究科 物理学専攻 修士論文, 2012

⁶NVMe は計算機に SSD を接続する次世代規格である。

⁷fi-crysdiskmark を使って測定