

# ビッグデータ太陽系天文学の創成のための ALMA キャリブレーション観測データの悉皆的解析システムの検討と構築

飯野孝浩<sup>\*1</sup>, 並木美太郎<sup>\*2</sup>, 山田崇貴<sup>\*3</sup>

## A feasibility study of exhaustive analysis of ALMA calibration data for the creation of big-data driven solar system astronomy

Takahiro Iino<sup>\*1</sup>, Mitaro Namiki<sup>\*2</sup> and Takayoshi Yamada<sup>\*3</sup>

### 概要

2011 年から科学運用を開始した ALMA は、ミリ波・サブミリ波帯域においてこれまでにない高感度・高空間分解能観測を実現する測器であり、太陽系天体のリモートセンシングにおいてもブレイクスルーをもたらすと考えられる。観測毎に蓄積されるキャリブレーション観測データを科学観測用途に転用することで、周波数・時間・天体方向に対して巨大な太陽系天文学観測データの入手が可能となる。一方で、そのデータ容量は膨大であり、最終段階までデータ処理を行わなければデータ品質の見極めが困難であり、大量のデータ処理を可能にするソフトウェア・ハードウェアを必要とすることが分かった。我々は ALMA キャリブレーション観測データの科学観測への転用を目指し、データ処理の歩留まり向上とバッチ処理化、そして専用システムの構築に取り組んだ。キャリブレーション観測データを生データから fits ファイルおよびクイックルック用画像ファイルまで処理するシステムを構築し、海王星の Band6 以上のデータについては公開済み全ファイルの処理を完了した。大容量データに対応するスケーラビリティと高速性を持つ SAS ストレージを導入し、広く用いられている 1 Gbps のネットワーク接続ストレージと比べて、最も処理時間のかかるプロセスであったキャリブレーション処理時間を 19-36% に短縮させることに成功した。さらに、同プロセスの並列化により、処理時間を約 13% に短縮した。これにより、研究室で保有できる程度の計算機システムにより、現実的な時間で ALMA キャリブレーション観測データの悉皆的解析を行うための道筋をつけることができた。

### Abstract

The ALMA that started its operation in 2011 has significantly high sensitivity and spatial resolution in millimeter and sub-millimeter waveband and expected to introduce a breakthrough on solar system remote sensing. Use of calibration data accumulated during every observations enables us to obtain the vast amount of solar system astronomical data. In turn, the total amount of data is quite large and

---

doi: 10.20637/JAXA-RR-17-009/0003

\* 平成 30 年 1 月 18 日受付 (Received January 18, 2018)

<sup>\*1</sup> 東京農工大学科学博物館

(Nature and Science Museum, Tokyo University of Agriculture and Technology, E-mail: iino@nagoya-u.jp)

<sup>\*2</sup> 東京農工大学工学部情報工学科

(Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology)

<sup>\*3</sup> 東京工業大学大学院総合理工学研究科

(Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology)

all of the data need to be processed to see if the data is suitable for the scientific analysis. Therefore, we attempted to construct an automated ALMA data processing system and a new computer system specified for the conversion of ALMA calibration data into the scientific data. As a result, we succeeded to construct the automated data processing system that produces fits cube file and quick-look data of spectrum and 2-D brightness map, and to process the whole data of Neptune above Band 6. In addition, introducing a new SAS storage that enables both capacity, scalability and high-speed access succeeded to reduce the time required for calibration process to  $\sim 19\text{--}36\%$  comparing with a standard NAS storage installed on 1 Gbps LAN network. A parallel data calibration system also decreased the processing time up to  $\sim 13\%$ . In summary, we succeeded to draw a path for the comprehensive analysis of ALMA calibration data in a realistic machine-time.

Keywords: Radio astronomy, ALMA, Planetary atmosphere

## 1. はじめに

干渉計型望遠鏡は、複数の望遠鏡で同時に観測された天体からの信号に相関処理を施し、高い空間分解能を実現する測器である。中でも、2011年に運用が始まったアタカマ大型ミリ波サブミリ波干渉計 (Atacama Large Millimeter/submillimeter Array; ALMA) は、66台の12 m および 7 m 望遠鏡からなる電波干渉計であり、31 GHz から 950 GHz という広い周波数帯において連続波および分光観測を可能とする。実現される空間分解能は、2017年現在 (Cycle-5 シーズン) に実用化されている観測バンドにおいて、最高の空間分解能を達成できる最長基線モードで 0.042 秒 (Band 3, 100 GHz)–0.024 秒 (Band 10, 870 GHz)<sup>\*4</sup> であり、また、最大 66 台のアンテナの同時稼働により、アンテナ 1 台による単一鏡型望遠鏡に比べて高感度での観測も期待できる。

高感度・高空間分解能観測という ALMA の特性は太陽系天文学においても重要である。太陽系内天体のうち、単一鏡型望遠鏡で空間分解可能な天体は太陽・金星・火星・木星・土星にとどまる。しかし、ALMA は視直径 2.2–2.3 秒の海王星や、0.6–0.7 秒の土星衛星タイタンなど、多くの惑星・衛星の空間分解観測を実現する。探査機に匹敵する高空間分解能での分光・連続波観測は、大気組成や大気ダイナミクス、熱分布の導出を可能とし、太陽系天文学にブレイクスルーをもたらすと期待されている<sup>1)</sup>。しかし、ALMA を用いた科学観測の機会は限定されており、サイクル 0 からサイクル 3 までの観測プロポーザル採択倍率は 8.2, 5.8, 3.9, 3.9 倍で推移してきた。また、通常のプロポーザルにおいて申請できる 1 回あたりの観測時間は、12 m アレイを用いる場合で 50 時間までと制限されている<sup>\*4</sup>。現状では観測機会が限定されているのみならず、長大な観測時間を投下したラインサーベイ観測や、分子・熱分布の時空間変動取得を目指した長期の観測も困難な状況である。

そこで、我々は ALMA がキャリブレーション目的で行っている惑星観測データの科学観測データへの転用に着目した。同様のアプローチにより、これまで土星の衛星タイタンについて、ALMA の空間分解能を活かしたタイタン大気中同位体比や窒素化合物検出が報告されている<sup>2),3)</sup>。図 6 に ALMA の典型的な観測プロシージャ (Project code: 2011.0.00628.S) を示す。R CrA IRS7B が科学観測天体であり、それ以外に海王

<sup>\*4</sup> Cycle-5 ALMA Proposer's guide による (<https://almascience.nao.ac.jp/proposing/documents-and-tools/cycle5/alma-proposers-guide>)

星およびクエーサー J1924-292 がそれぞれ Flux および Phase/Gain キャリブレーションとして観測されている。惑星は視直径や温度がある程度分かっているため、Flux キャリブレーションとして頻りに用いられており、各観測セットごとに1回ずつ観測される。科学観測と同じアンテナ台数が使用されており、観測時間は短いものの、高い観測性能を持つと期待できる。キャリブレーションとして用いられている太陽系内天体は、火星、木星、天王星、海王星といった惑星だけでなく、ガニメデ、カリスト、タイタンといった衛星、そして小惑星であるセレスと幅広い。

キャリブレーション観測データは、積分時間が短いためにプロポーザルをベースとした科学観測に比べて感度で劣るものの、1. 天体、2. 周波数、3. 時間軸、の3軸に対して巨大なデータ群になっていることが重要である。これらの特質により、それぞれ1. さまざまな対象天体に対し、2. 特定の瞬間における多様な分子観測および大気・地表温度構造の取得、さらにその3. 時間変動の取得、という観測手法の展開が可能になる。これにより、大気化学、大気動力学、惑星形成といった、多方面からのサイエンスの展開が期待される。一方で、太陽系内天体は絶対強度のキャリブレーションに用いられており、確からしい Spectral Energy Distribution モデルを再現するように、観測された天体のフラックスをスケールアップするのに用いられる。すなわち、各天体表面においてグローバルな連続波強度の変動を捉えることは原理上不可能であることに留意する必要がある。

本研究では、キャリブレーション観測データを太陽系天文学ビッグデータととらえ、その悉皆的な解析を通じて、上記のサイエンスにおけるブレイクスルーを目指す。その前段階として、本稿では、キャリブレーション観測データの個数や観測周波数・対象天体の特性等を明らかにし、期待されるサイエンスを整理する(2章)。また、悉皆的な解析において重要なキャリブレーション観測データのバッチ処理を可能にするとともに(3章)、専用の環境を構築し、将来の悉皆的な解析において必要とされるマシンタイムの見積もりを行った(4章)。

## 2. ALMA キャリブレーション観測データの概要と、期待されるサイエンス

### 2.1 ALMA キャリブレーション惑星データ概要

キャリブレーション観測データのフィジビリティ検討のため、主要な太陽系内天体(火星、木星、ガニメデ、カリスト、タイタン、天王星、海王星、セレス)について、観測日時、観測周波数およびデータ容量の調査(タイタンのみ)を行った。これら天体群は大気の分光観測が期待できる群(タイタン、海王星、天王星、火

Date	Timerange (UTC)	Scan	FldId	FieldName
04-May-2012	11:04:04.0 - 11:09:21.2	4	0	J1924-292 Phase, Gainキャリブレーション
	11:14:32.7 - 11:19:27.9	7	2	Neptune Flux キャリブレーション: 惑星
	11:23:38.5 - 11:24:08.8	10	0	J1924-292
	11:26:32.5 - 11:34:33.7	12	3	R CrA IRS7B 科学観測天体
	11:36:17.6 - 11:36:47.9	14	0	J1924-292
	11:38:34.2 - 11:46:35.4	16	3	R CrA IRS7B
	11:48:19.0 - 11:48:49.2	18	0	J1924-292
	11:50:35.9 - 11:58:37.1	20	3	R CrA IRS7B
	12:00:32.6 - 12:01:02.8	22	0	J1924-292
	12:03:01.1 - 12:11:02.2	24	3	R CrA IRS7B
	12:12:45.9 - 12:13:16.1	26	0	J1924-292
	12:15:01.9 - 12:23:03.1	28	3	R CrA IRS7B
	12:24:46.6 - 12:25:16.8	30	0	J1924-292
	12:27:02.6 - 12:33:41.8	32	3	R CrA IRS7B
	12:35:37.1 - 12:36:07.3	34	0	J1924-292

図1 海王星を含む観測プロシージャの例。観測天体として、1つの科学観測天体と2つのキャリブレーション天体が含まれる。

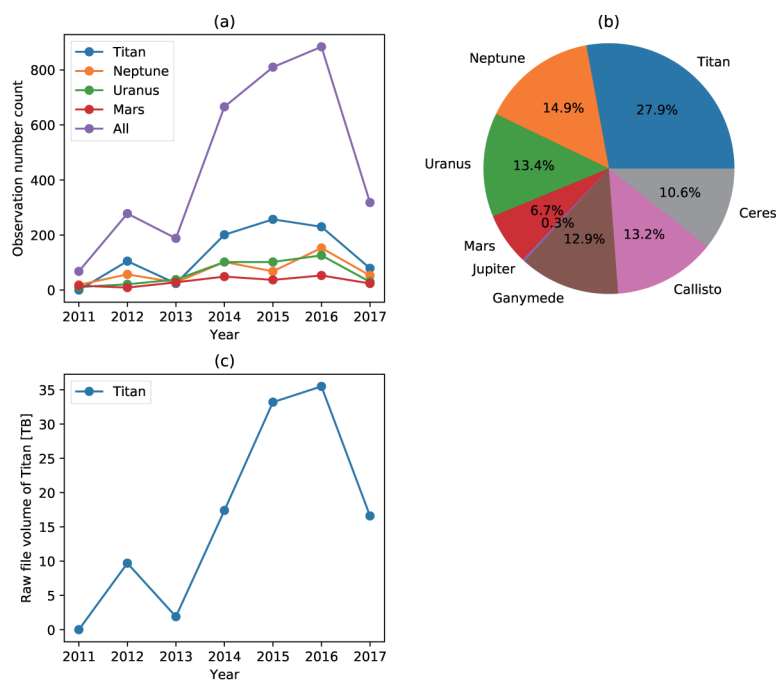


図2 (a) 年毎の太陽系天体全体及び主要天体の総観測数の推移．2017年は5月までの数値である．2014年以降は総数が600回を越えており、今後も同程度で推移すると期待される．(b) 天体ごとの観測数の割合．(c) タイタンの観測年ごとのデータ容量の推移

星)と、固体天体群(ガニメデ、カリスト、セレス)に二分できる．ALMAアーカイブより上記全天体の公開データ(観測日、観測周波数、積分時間等)のクエリを行い、テキストデータとしてダウンロードし、調査を行っている．データの個数、天体ごとの内訳、データ容量(タイタンのみ)を図2に示す．2017年は5月までの値である．(a)に示した総観測数はほぼ右肩上がりであり、2016年には800件超となっている．一部の天体(タイタン、海王星、天王星、火星)の観測数推移も併せてプロットしているが、同様に増加を示している．(b)にはキャリアレーション天体のデータ数の内訳を示した．タイタンが最多で27.9%を占める．他の天体は10%台であるが、国内外に研究者の多い火星は6.7%と少ない．(c)はタイタンの総データ容量の推移を示す．2015–2016年は30TBを越える．後述するように、タイタンは複雑な大気組成を持つため、その大気組成や同位体比の導出のためには、可能な限りのデータを用いたラインサーベイ観測が重要である．トータルで100TB超という大容量のデータを高速で処理するために、本研究では大容量・高速ストレージが必要となること分かる．アーカイブでは他に空間分解能も公開されているが、楕円形となるビームの短辺・長辺の平均を示しており、ビーム形状の詳細についてはデータをダウンロードし、CLEAN処理を施すまで不明である．

## 2.2 ALMA キャリブレーション観測データ利活用の意義

### 2.2.1 周波数方向

惑星大気由来分子スペクトルの輝線幅は圧力に比例する「圧力広がり(Pressure broadening)」が支配的であり、数MHz/torr程度が一般的である．気圧が100mbar以下と低い成層圏に偏在する分子の線幅は典型的

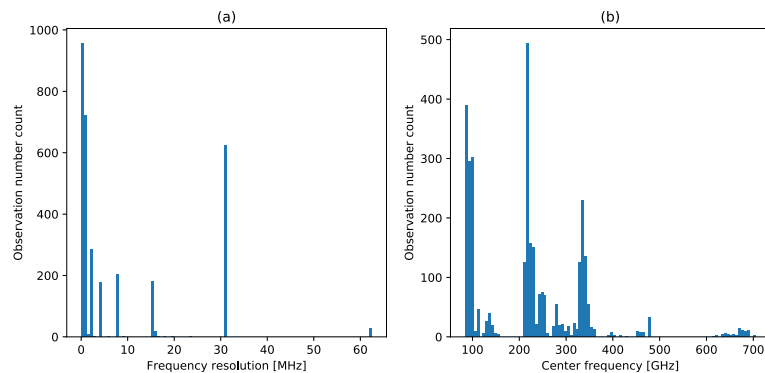


図3 (a) キャリブレーション観測データの周波数分解能のヒストグラム, (b) キャリブレーション観測データの中心周波数ヒストグラム. Band 8 以上の高周波数帯が乏しく, ミリ波・サブミリ波天文学で実績のある 100 – 300 GHz に集中している.

に数十 MHz 程度であり, スペクトル形状を十分に分解するには高い周波数分解能が必要となる. 図 3(a) からは, キャリブレーション用データにおいて多数を占めるのは, 1 MHz 以下という高い周波数分解能を持つデータであることが分かる.

周波数方向に広く (無バイアスに) 分光観測が行われていることは, 観測的天文学のメソッドとして重要である. 図 3(b) に示した中心周波数ヒストグラムからは, 観測データが Band 3, 6, 7 の 100~300 GHz 帯に集中していることが分かる. この波長域における観測数は少ないものの, 単一鏡型望遠鏡においても一般的ではなかった Band 8, 9 のデータが取得可能になっている意義は大きい. 図 4 にサイクル 0 におけるタイタンのキャリブレーション観測データのスペクトルを示す.  $\text{HC}_3\text{N}$  分子の回転および振動遷移, そして同位体がとらえられている. このような分光観測を周波数方向に多数結合することで, 大気中の微量分子や同位体比をもれなく観測することが可能になる. この手法は, 濃密な大気を持ち, 分光観測による効果が見込める火星, タイタン, 天王星, 海王星について重要である. 特に,  $\text{N}_2$  を主成分とし,  $\text{HCN}$ ,  $\text{HC}_3\text{N}$ ,  $\text{C}_2\text{H}_5\text{CN}$  等の窒化物を豊富に含むタイタン大気においては, 複雑な大気化学プロセスの解明が重要な課題であり, その観測的制約のために本アプローチが果たす役割は大きい. また, 分子の分布の導出も分光観測から可能になる. 特定の大気層から放射される分子輝線形状は, 主に自転によるドップラーシフトおよび, 圧力広がりによって決定される. 視線方向に分割された各大気層はそれぞれが異なる温度・圧力・分子存在量を持っており, 輝線形状を考慮して十分に周波数方向に分解した輻射輸送計算を行い, 得られた輝線形状と観測された形状を比較することで, 分子の鉛直方向の分布を決定することが可能である (=大気フォワードモデル). さらに高空間分解能データを用いれば, 分子 3 次元分布の導出も可能であり, 実際に一部の窒化物分子群の極域への局在が報告されている<sup>3)</sup>. 分子の存在量と併せて, 分布を観測的に制約することが大気化学の理解においては重要である. これはタイタン以外の天体においても重要であり, 例えば海王星においては成層圏に偏在している  $\text{HCN}$  分子の空間分布や, 小天体の降り込みによって生成される分子群の存在量の観測的導出から, 成層圏微量分子の起源の制約が見込める. これまで, 単一鏡型望遠鏡を用いた系統的な観測により, 海王星成層圏は硫化物に乏しいことが明らかになっている<sup>4)</sup>. キャリブレーション観測データを用いれば, より広範な微量分子の探索や, 検出済み分子の 3 次元構造の導出が可能になる.

### 2.2.2 天体方向

ミリ波・サブミリ波を用いた観測的研究は, キャリブレーション観測対象天体の中では, 海王星及びタイタ



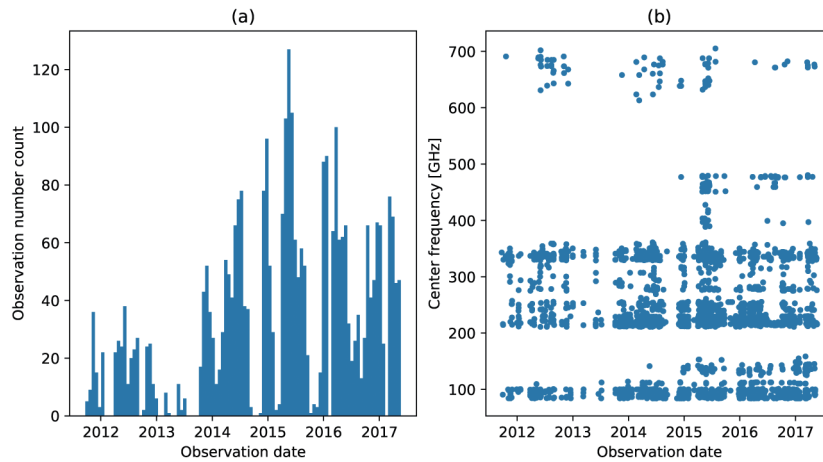


図5 (a) 調査天体に対する観測回数の推移. 2016年以降は絶え間なく観測が行われている. (b) 観測周波数の推移. これを周波数ヒストグラムにしたものが図3(b)である. 受信機が整備される毎に観測周波数が拡大している. 2017年からは, 200 GHz 周辺の Band 5 が共同利用に供される.

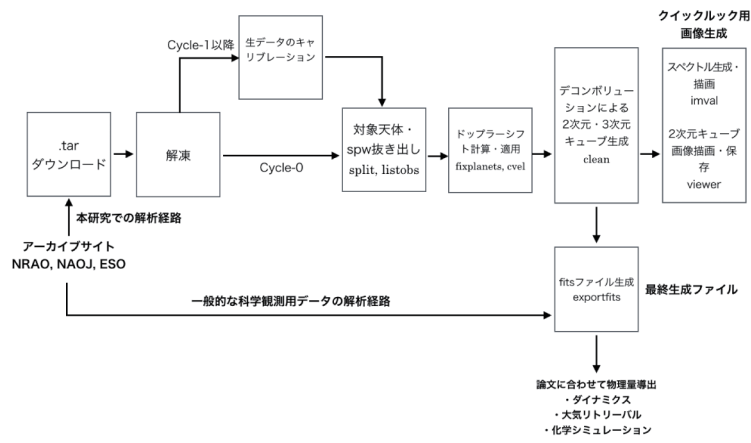


図6 本研究におけるALMAデータ処理のプロセス

### 3. データ処理プロセスとバッチ処理化

本研究では, 膨大なキャリブレーション観測データの悉皆的解析のために, データ処理プロセスのバッチ処理化と処理歩留まりの向上を検討した. ALMAにおけるデータ処理のプロセスを図6に示した. 科学観測の場合には, ALMAスタッフにより解析, 生成されたfitsキューブファイルが提供される. いっぽう, キャリブレーションデータの場合には生データ(ASDM)ファイルからfitsファイルの生成までを全てユーザーが行う必要がある. ASDMファイルは最終プロダクトであるfitsファイルと比べて容量も大きい. 例として, 火星をキャリブレータとして用いているデータのうち, IDの先頭が2013となるデータからASDMとfitsの含まれるtarファイルの容量・個数を調べると, それぞれ2.1TB(439個)および222GB(116個)であり, 容量でほぼ10:1であった. 解析用スクリプトはfitsファイルに同梱されているため, 本研究ではASDM, fits双方をダウンロードする必要がある. 本研究は対象のデータも多いため, 通常の科学観測データを用いたアプ

ローチに比べてはるかに大容量のストレージを必要とすると言える。データ処理には電波天文学データの汎用処理ソフトウェア CASA(Common Astronomy Software Application) を用いている。CASA は Linux もしくは Mac 上でのみ動作し、Windows はサポートされない。本研究ではデータ処理に Linux 計算機システムを用いている(表 1)。

全プロセス中で、最もマシンタイムを要するのはキャリブレーションプロセスである。キャリブレーションには観測データ毎に専用の Python スクリプト群が用意されており、CASA 上で呼び出す (`execfile` タスク) ことで動作する。処理すべきデータのスクリプトを順に実行することでバッチ処理が可能である。前処理として、事前に関連するキャリブレーションスクリプト群に対し、キャリブレーション後に生成されるファイル容量を抑制するオプションの記述や、ASDM とスクリプトのファイル数の不整合や予期せぬライブラリ読み込みによるエラーを回避するための処理を行っている。キャリブレーションスクリプトの実行歩留まりは低く、さまざまなエラーで CASA 自体が停止するため、失敗しても処理全体が停止しないよう、外部スクリプトからキャリブレーションスクリプトを呼び出している。

キャリブレーションスクリプト実行時間や生成前後のディレクトリの容量を表示させることで、実行効率の検証を行っている。実行時間とファイル容量の関係については後述する。

生成されたキャリブレーション済データ (`.ms.split.cal` 形式) は天体および、異なる観測周波数を持つ分光計のチャンネル群である Spectral window(`spw`) に分割する (`split` タスク)。強い分子輝線はマスクされている場合があるため、マスクを外す処理 (`unflag`, `plotms` タスク) を行う。観測周波数は科学観測天体のドップラーシフトに合わせて設定されているため、キャリブレーション天体のドップラーシフトの補正を行う (`cvel` タスク)。特に天体の分子輝線の同定において重要なプロセスであるが、本プロセスは本稿執筆時点では正常動作の歩留まりが低く、今後の重要な課題である。CLEAN プロセスにより、2 次元・3 次元輝度分布キューブを生成 (イメージング) する。同プロセスは輝度が集中すると予想される任意の領域を指定する必要があるが、太陽系内天体の場合には既に視直径が分かっているため、これもバッチ処理が可能である。イメージングにより生成されたキューブファイルの可視性を高めるため、クイックルック用に連続波マップ及びスペクトルを生成する。png 形式の連続波マップは、CASA のデータ可視化タスク (`viewer` タスク) を用いてバッチ生成する。スペクトル画像の生成時には、いったん天体の任意領域を指定して周波数-強度 (Jy/beam) の 2 カラムのテキストデータを生成したのち (`imval` タスク)、Python の可視化ライブラリ `matplotlib` で png ファイルを生成している。スペクトル図の例を図 4 に示す。同時に主要な分子輝線をオーバーレイすることで輝線同定に役立てている。

キューブファイルは CASA のみで読み込み可能であるため、`ds9` や `aplpy` といった汎用の天文学データ処理ソフトウェア・ライブラリでの利用のために、キューブファイルを `fits` 形式に変換する (`exportfits` タスク)。fits ファイルは最終アウトプットとして別ストレージにも格納し、冗長性を高める。

ALMA よりダウンロード・解凍されたデータの基本構造を以下に示す。Project code の下に、Science goal, Group, Member の各 ID 毎に階層が形成される。ひとつの Project code 中に複数の science goal が設定される際には、それぞれの science goal 以下に独自の ID を持つ階層が形成される。



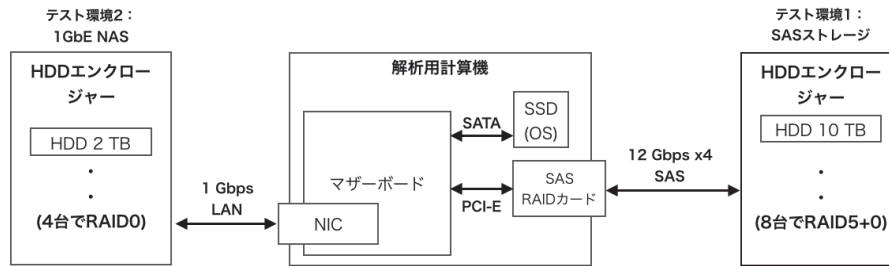
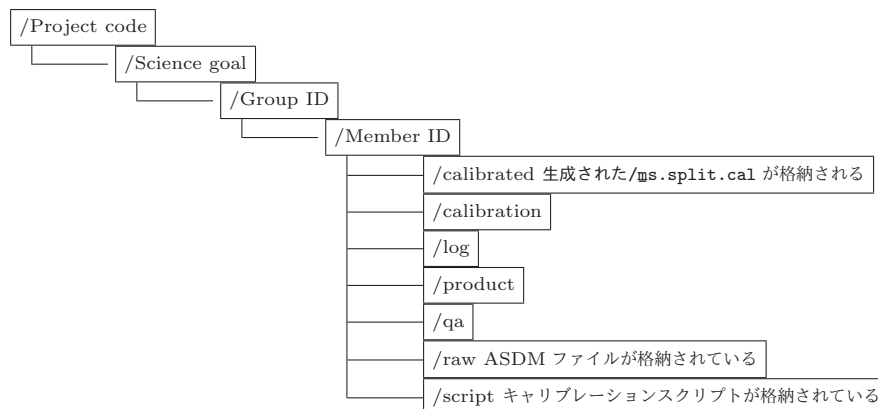


図7 本研究で用いているシステムのブロックダイアグラム.



本研究では、'Project code.Member ID.天体名.spw. 処理段階' という命名規則のもとに、処理後のファイル、ディレクトリを生成している。処理段階は図6の split 処理以降に対応する。

## 4. 生成データの概要と、解析処理高速化の取り組み

### 4.1 悉皆的解析に必要な新解析システムの要件と構築

表1 本研究に用いている計算機およびストレージの仕様

機能	仕様
型名	Precision T7500 (Dell 社製)
CPU	Xeon 5690 (6 コア 12 スレッド, 最大動作周波数 3.73 GHz) 2 基搭載
チップセット	Intel 5520
RAM	DDR3 1333MHz 駆動 ECC Registered 96 GB (4 GB × 8 + 16 GB × 4)
システムディスク	250 GB ソリッドステートドライブ, SATA 6 Gbps 接続
OS	Ubuntu Linux 16.04 LTS
SAS RAID カード	ASR8885Q(Adaptec 社製), 内外 8 ポート, PCI-Exp. 3.0 ×8 接続
HDD ケース	ARC-4038(areca 社製), SATA 8 ベイ, 48 Gbps SAS 接続, SAS エキスパンダー搭載

本研究では、ガスを持つ天体のラインサーベイ観測および、固体天体・ガス惑星輝度マップ観測、の2テーマが大きな観測ゴールである。そのためには、最大の観測データ数を持つタイタンを含む、大容量のアーカイ

ブデータの悉皆的な解析が必要である。そのため、大容量のストレージがシステムに求められる要件のトップに挙げられる。さらに、データの増加に対応できるスケーラビリティ、耐障害性、計算機の入れ替えへの対応、そして高速性が併せて求められる。

大容量ストレージとして 2017 年現在で一般的であるのは Network Attached Storage(NAS) である。2-8 台の内部ディスクに加え、USB や eSATA を用いた拡張にも対応可能である。また、ネットワークを介して複数の計算機からの同時アクセスが可能である。しかし一方で転送速度が難点であり、一般的には 1 Gbps に留まる。10 Gbps Ethernet を用いる NAS は普及途上であり、ネットワーク全体を再構成するコストが生じる。内蔵 SATA ポートを用いてソフトウェア RAID を組む場合には、各 HDD と 6 Gbps で接続できるため速度では優位性があるが、SATA ポートと HDD ベイの数が容量の上限を決めてしまうこと、計算機リプレースに対応しないことが問題となる。ハードウェア RAID 機能を搭載したハードディスクケースは計算機リプレースに対応できるが、接続速度とスケーラビリティにおいて劣る。Thunderbolt 3 規格に対応したハードディスクケースは速度面で優位であり、デジチェーンによる拡張性も併せ持つが、Linux 環境で十分にテストされた製品は見つからなかった。これら競合する規格との比較検討のうえ、本研究では Serial Attached SCSI (SAS) 接続によるストレージの構築を計画した。SAS はケーブル 1 本あたり 12 Gbps の高速接続速度（さらに接続本数で通信できる）と、デジチェーン・パラレル接続による拡張性を併せ持つ規格である。ハードウェア RAID 機能を持った拡張カードを用いることで、計算機のリプレースにも対応する。いっぽう、複数の計算機からの高速接続には対応しない点がデメリットである。

本研究で用いているシステムのブロックダイアグラムを図 7 に示す。本研究では、前段階として 1 Gbps 接続の NAS 類似ネットワークストレージ（図 7 中のテスト環境 1）を複数構築し、データ処理プロセスの構築やアーカイブデータのポテンシャルの確認を行い、図 4 のような高感度データが多く含まれることを確認した。その中で、キャリブレーション観測データの容量が大きく、複数のストレージにデータを分散させる不便が大きいこと、そしてデータ処理中は常に NAS と計算機間のデータ転送容量が 1 Gbps の上限値に張り付いているため、外部記憶との転送速度が処理のボトルネックになっていると見られたことから、SAS を用いた高速・大容量ストレージの構築（図 7 中の新ストレージ）を検討・実施した。本稿執筆時点では、10 TB のエンタープライズ向けハードディスク 8 台で RAID5+0(RAID5 のストライピング) を構築、48 Gbps で計算機と接続している。実効容量は 58.4 TB である。本ストレージは SAS ストレージの性能確認用として導入したものであり、近い将来に 24 ベイのエンクロージャを導入予定である。

SAS ストレージの転送速度の試験として、`hdparm` コマンドを用い、3 秒間のシーケンシャルリードで転送した容量を 10 回計測し、平均値と標準偏差を計測した。結果は  $1309.6 \pm 7.7$  MB/s であり、1 Gbps LAN の理論上の転送速度の 10 倍程度を実現できた。転送帯域を使い切っていない状況であり、今後は RAID5 を構成するディスク数の増加による速度の向上も期待できる。

#### 4.2 実行効率とビーム形状

2017 年 5 月～6 月に、海王星の Band 6 以上の公開済み全データのキャリブレーション処理を完了させることができた。図 9 に、海王星を対象としたキャリブレーションの所要時間 (a) とキャリブレーションの実行効率 (ASDM データ容量/実行時間) (b) を示す。所要時間は大まかに ASDM 容量と比例していることが分かる。(b) の実行効率はデータによる差異が大きく、複数のピークを持つが、おおまかに 2-9 MB/s の範囲に収まる。ここから、将来のより巨大なデータの処理に必要な時間を見積もる事が可能である。タイタンは提供されているデータ全体の処理によりラインサーベイ解析が可能であり、提供済みの総データ容量は 114.3 TB である。うち 90% を ASDM ファイルと、実行効率を海王星と同等の  $\sim 5$  MB/s と仮定すると、キャリブレーション

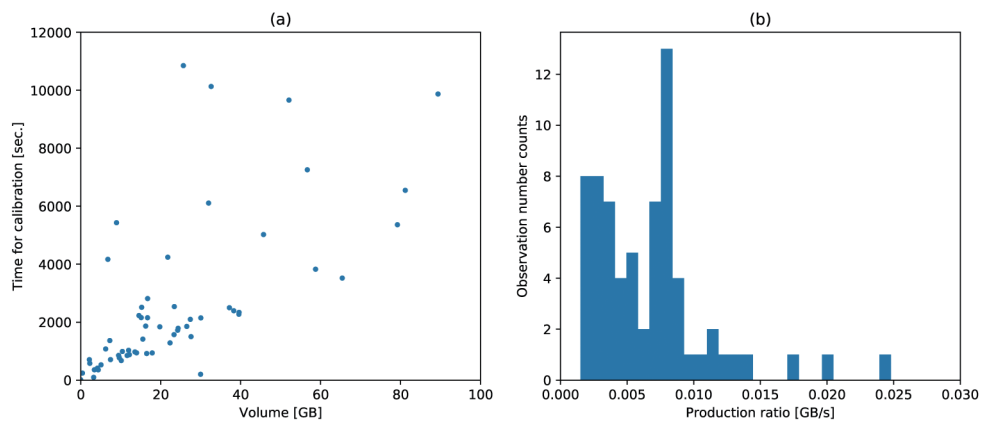


図8 (a) キャリブレーション所要時間 (縦軸) と ASDM ファイル容量 (横軸) の関係 (b) データ処理効率のヒストグラム. 2-9 MB/s に値が集中している.

ション処理に要するマシンタイムは、現状のシステムで ~265 日となる。

海王星のサイクル 1 以降のキャリブレーション後のデータとサイクル 0 のデータのうち、spw=0 のみに split, CLEAN 処理を行い、ビーム形状を調べた結果を図 9 に示す。(a) はビーム長辺の長さを示す。海王星の視直径は 2.2-2.3 秒程度であるが、データの多くは視直径以下のビーム径を持つことから、空間分解観測が可能なポテンシャルを持つデータが多いことがわかる。(b) は扁平率 (1-短辺/長辺) であり、値が大きいほどビームが扁平であることを表す。マッピング観測時にはビームが円形に近いことが求められるが、短辺:長辺=1:2 となる 0.5 を越える扁平率を持つデータも多い。本稿で扱ったデータ処理終了後には、輻射輸送計算により分子の存在量や気温分布といった物理量を導出することになるが、輻射輸送コードには楕円ビーム形状の処理の実装が重要である。これは円形ビームを仮定できる単一鏡型望遠鏡にはないデメリットである。(c)(d) はそれぞれ海王星視直径とビーム長辺比、視直径とビーム長辺の時間推移をプロットした。(c) は 1 を越えると海王星の空間分解がなされているデータとなる。最高で 4 を越えるデータがあるが、これはサイクル 0 シーズンに行われたものであり、現在論文化に取り組んでいる。干渉計に特有の問題として、一定以上に広がった輝度分布を取得できない Resolving out という現象が挙げられる。特に惑星は見かけの形状が円形であり、温度分布の観測時にこの影響を受けることが予想される。ALMA の場合には、空間分解能と Resolving out する角度の比は各バンドにおいて典型的に 1:10 程度であり<sup>\*4</sup>、海王星においては 0.2 秒程度が最大の空間分解能となるが、現状では該当する高空間分解能のデータは含まれていない。(d) からは、定常的に高空間分解能データが提供され続けていることが分かり、温度分布の時間変動が観測の可能性を示している。火星や氷天体群は季節変化のタイムスケールが短いため、同様のデータが提供されていけば温度や組成分布の季節変動をとらえられる可能性がある。

#### 4.3 高速ストレージ導入によるキャリブレーション処理高速化の検証

SAS ストレージの導入による高速化の検証のため、旧ストレージ (図 7 中「テスト環境 2」) と SAS ストレージでのキャリブレーション処理の実実行時間の比較を行った。キャリブレーション観測データとして、火星のデータを 5 個用い、両環境で 1 回ずつキャリブレーション処理を実行している。実実行時間の積み上げグラフを図 10 に示す。実実行時間は各データにおいてテスト環境 2 と比べて 19-36% に、全体として 32% に短縮した。キャリブレーション観測データは一つ一つが数十 GB と巨大であり、またキャリブレーション後の出

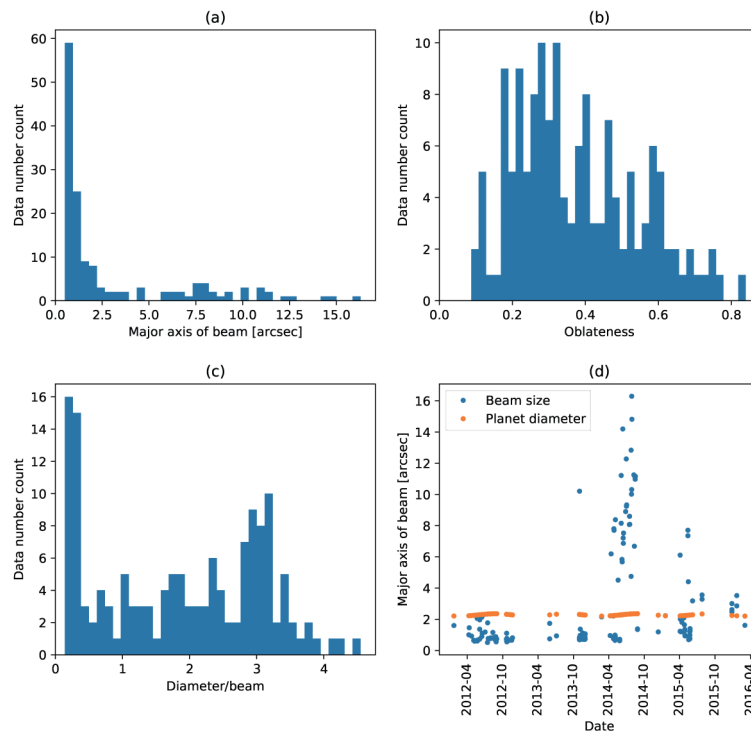


図9 イメージング後の海王星データ群から取得した、ビーム形状と海王星視直径の関係。(a) ビームの長辺の FWHM (b) ビームの扁平率 (c) 海王星視直径とビーム長辺の比 (d) ビーム長辺の大きさと海王星視直径の時間推移

カデータは元データより大きいため、外部記憶の転送速度がボトルネックになっていたことが示された。一方、処理速度の短縮はシーケンシャルリード速度の比ほど劇的ではないため、現状では CPU やメモリ速度がボトルネックと考えられる。

#### 4.4 並列実行によるキャリブレーション処理高速化

前節のストレージ高速化に引き続き、複数のキャリブレーション処理のプロセス並列による高速化に取り組んだ。並列処理には Python の `multiprocessing` モジュールを使い、キャリブレーションスクリプトをサブプロセスとして実行することにより実装した\*<sup>5</sup>。並列実行するプロセス数は、12 個の物理 CPU コア数に対して 10 プロセスとした。実行速度は単純に 10 倍にはならず、メインメモリ・キャッシュメモリの帯域や容量、外部記憶のオーバーヘッド等により実行効率が低下すると予想される。そこで、キャリブレーションスクリプトの実行時間から並列化の効果測定を実施した。シングルプロセス実行時は、前節と同様に複数のキャリブレーションスクリプトを連続実行し、12 のスクリプトの実行時間を計測した。マルチプロセス実行時は、常に 10 のプロセスが実行されるようにしながら、それぞれの実行時間を計測した。シングル、マルチプロセスそれぞれの総実行時間は 45146 および 61559 秒であり、マルチプロセス時には全てのスクリプトで実行時間が長くなった。双方での各スクリプト実行時間の積み上げグラフを図 10 に示す。各スクリプトでの実行時間の比 (マルチプロセス/シングルプロセス) は 1.34 であり、標準偏差は 0.16 であった。10 のプロセスを同時実行する今回の計測設定では、キャリブレーションの実行効率を 7.4 倍 (13.4%) に高速化することができた。

\*<sup>5</sup> <https://docs.python.jp/3/library/multiprocessing.html>

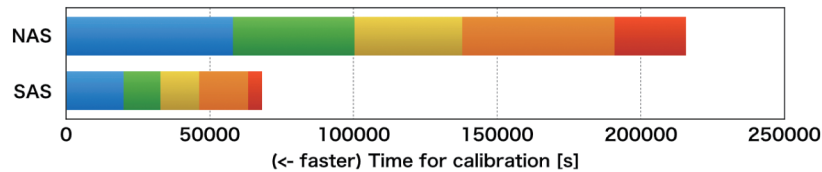


図 10 (a) 1 GbE-NAS・SAS 両システムでの実キャリブレーション時間の比較. SAS システムでは NAS システムに比べて処理時間が 19–36% に短縮している. (b) キャリブレーションスクリプトのシングルプロセス実行時とマルチプロセス (10 プロセス) 実行時の処理時間の比較. 各種オーバーヘッドによりマルチプロセスでは各スクリプトの処理時間がシングルプロセス時に対して長くなっているが、その比は  $1.34 \pm 0.16$  倍に留まっており、トータルでは 7.4 倍の高速化となっている.

4.2 節で推算したタイタンの総計算時間は 265 日であったが、これが 35 日まで短縮されることになる。

#### 4.5 今後の展開

本研究で得られる多様なデータの論文化のため、複数の研究グループの立ち上げ、クイックルックデータの共有が重要である。国内外では火星、タイタン、固体天体、ガス惑星と天体ごとに研究コミュニティが存在するが、観測データを通じてこれらコミュニティとの有機的な結合をはかることで、ALMA とキャリブレーション観測データの特徴を活かしたビッグデータ太陽系天文学時代の創成を目指していく。

#### 謝辞

本研究は、2017 年度電気通信普及財団研究助成、日本学術振興会科研費 (17K14420) および大学共同利用機関法人自然科学研究機構アストロバイオロジーセンター (AB291014) のサポートを受けて行われています。本テーマで発表を行った平成 28 年度 JAXA 宇宙情報解析シンポジウムでは、フロアから多くの有用なコメントをいただきました。また、解析手順等について慶応大学の田中邦彦助教および京都産業大学の佐川英夫准教授から、そして ALMA データ処理全般について国立天文台チリ観測所の西合一矢助教、中西康一郎准教授から多くのコメント、ご意見、サポートをいただきました。名古屋大学環境学研究科卒業生の仲本悟さんには、在学時に ALMA データ処理について多くの有益なトライアンドエラーの知見を提供いただきました。匿名の査読者の方には、専門的な視点からの確かなコメントを頂戴いたしました。ここに感謝申し上げます。 This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00405.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

#### 参考文献

- [1] Hideo Sagawa. アルマ望遠鏡による太陽系研究. 天文月報, 110(5):317–324, 2017.
- [2] Joseph Serigano, Conor A. Nixon, Martin A. Cordiner, Patrick G. J. Irwin, Nicholas A. Teanby, Steven B. Charnley, and Johan E. Lindberg. ISOTOPIC RATIOS OF CARBON AND OXYGEN IN TITAN'S CO USING ALMA. *The Astrophysical Journal*, 821(1):L8, apr 2016.
- [3] M. a. Cordiner, M. Y. Palmer, C. a. Nixon, P. G. J. Irwin, N. a. Teanby, S. B. Charnley, M. J. Mumma, Z. Kisiel, J. Serigano, Y.-J. Kuan, Y.-L. Chuang, and K.-S. Wang. Ethyl Cyanide on Titan: Spectroscopic Detection and Mapping Using Alma. *The Astrophysical Journal*, 800(1):L14, 2015.
- [4] Takahiro Iino, Akira Mizuno, Taku Nakajima, Takehiro Hidemori, Takashi Tsukagoshi, and Chihaya Kato.

- Search for sulfur-bearing species as remnant of cometary impact on Neptune. *Planetary and Space Science*, 104:211–215, dec 2014.
- [5] E Lellouch, C De Bergh, and B Sicardy. Exploring the spatial, temporal, and vertical distribution of methane in Pluto’s atmosphere. *Icarus*, 2014.
- [6] E. Lellouch, C. de Bergh, B. Sicardy, S. Ferron, and H.-U. Käufel. Detection of CO in Triton’s atmosphere and the nature of surface-atmosphere interactions. *Astronomy and Astrophysics*, 512:L8, apr 2010.
- [7] Takahiro Iino, Yasuhiro Hirahara, Takehiro Hidemori, Takashi Tsukagoshi, Taku Nakajima, Satoru Nakamoto, and Chihaya Kato. Observational constraint on Pluto’s atmospheric CO with ASTE. *Publications of the Astronomical Society of Japan*, 68(1):L1, feb 2016.
- [8] E. Lellouch, M. Gurwell, B. Butler, T. Fouchet, P. Lavvas, D.F. Strobel, B. Sicardy, A. Moullet, R. Moreno, D. Bockelee-Morvan, N. Biver, L. Young, D. Lis, J. Stansberry, A. Stern, H. Weaver, E. Young, X. Zhu, and J. Boissier. Detection of CO and HCN in Pluto’s atmosphere with ALMA. *Icarus*, 286:289–307, apr 2017.