

木曾超広視野高速CMOSカメラ Tomo-e Gozen による サーベイデータとデータ解析フロー

東京大学理学系研究科
天文学教育研究センター 研究員
大澤 亮

theme:

トモエゴゼンのデータが出てくるまで&データの使用方法

table of contents:

1. トモエゴゼン計画とデータの特殊性
2. トモエゴゼンのデータができるまで
3. トモエゴゼンの使い方
4. 京大3.8m望遠鏡とのコラボレーション

トモエゴゼン計画とデータ

トモエゴゼンの立ち位置について

プロポーザル型

木曾観測所(~2016), 京大 3.8m, すばる, など
KISS/KWFC, KISOGP/KWFC
HSC SSP など

SDSS, Pan-STARRS, CRTS, LSST, など

プロジェクト型

トモエゴゼンの立ち位置について

プロポーザル型

木曾観測所(~2016), 京大 3.8m, すばる, など
KISS/KWFC, KISOGP/KWFC
HSC SSP など

← トモエゴゼン

SDSS, Pan-STARRS, CRTS, LSST, など

プロジェクト型

トモエゴゼンの立ち位置について

プロポーザル型

木曾観測所(~2016), 京大 3.8m, すばる, など
KISS/KWFC, KISOGP/KWFC
HSC SSP など

← トモエゴゼン

プロジェクト型の特化性能
+
木曾観測所の柔軟さ

SDSS, Pan-STARRS, CRTS, LSST, など

プロジェクト型

トモエゴゼンの特殊性

1. 広い視野 ($\sim 20 \text{deg}^2$) と高い感度を両立
2. オーバヘッドなしの連続読み出しを実現

⇒ これらの特徴を活かした観測を実施したい

トモエゴゼンの特殊性

1. 広い視野 ($\sim 20 \text{deg}^2$) と高い感度を両立
2. オーバヘッドなしの連続読み出しを実現

⇒ これらの特徴を活かした観測を実施したい

最大レートで観測し続けた場合

$84 \text{ chip} \times 2.4 \text{ Mpix} \times 4 \text{ byte} \times 2 \text{ fps} \times 10 \text{ 時間} = \text{およそ } 50 \text{ TiB/day}$
($\sim 1600 \text{ MiB/s}$)

6TB HDD に保存した場合: 1日10台, 年間100晩でおよそ3,000万円必要

木曾 ⇄ 外部のネットワーク (フレッツ光) 速度: $\sim 125 \text{ MiB/s} \ll \sim 1600 \text{ MiB/s}$

運用可能な状態での保管は困難 & 木曾から出られない

オンサイト解析の必要性

基本方針:

解析済みのデータのみを残す

研究の主軸はタイムドメイン天文学

- ⇒ 時間変動していない信号については重要度が低い
- ⇒ 全動画データを残さなくてもサイエンスが可能

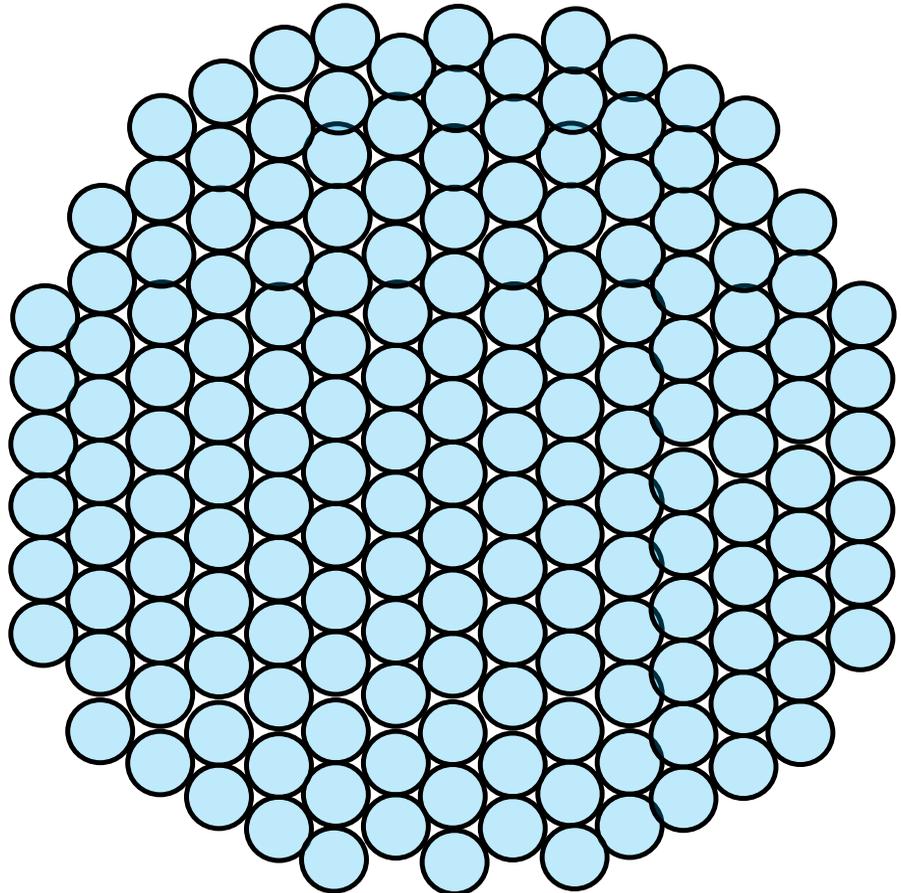
自動解析パイプラインの整備

- ⇒ 多様な観測提案に柔軟に対応したい
- ⇒ 基本となるデータセットを定義・提供する*

* 基本データセット以外を用いる観測提案を却下するわけではない(議論参照)

想定される観測提案の例

広視野高速サーベイ



数秒で視野を切り替え広域を複数回掃天する

2Hz 定点モニタリング



ひとつの視野を 2Hz で長時間モニタリング

トモエゴゼン計画で残したい情報

一般的な観測で期待される情報

天体の測光値, 画像(deep image), など

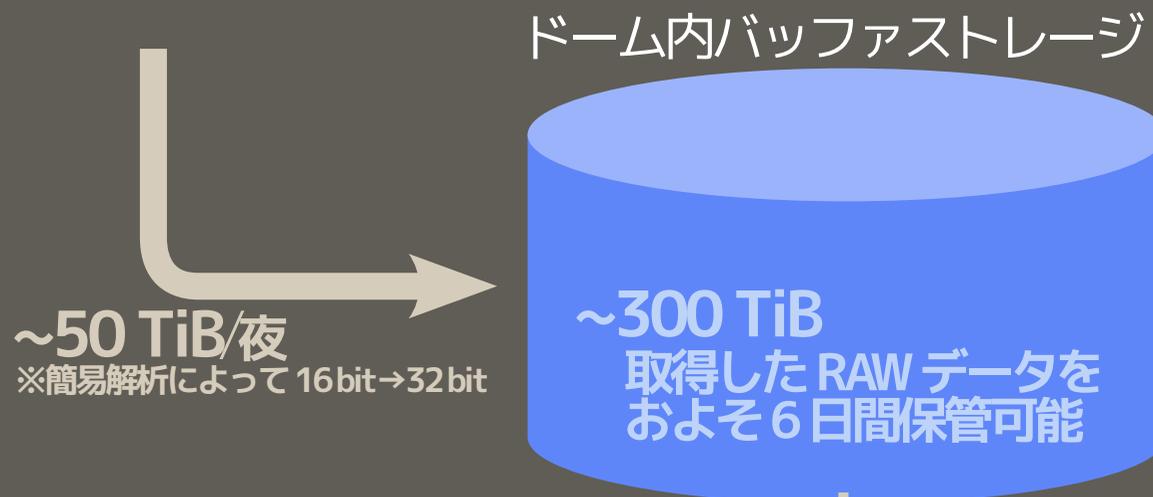
動画観測でしか得られない情報

変動・運動の時系列情報, 高速移動天体, など

トモエゴゼンのデータフロー

木曾シュミット望遠鏡ドーム

Tomo-e Gozen カメラ



ドーム内解析用計算機群

解析によってデータ総量を
およそ ~2 TiB/夜まで削減する
現在は 4 種類のパイプラインを検討中

※中間生成データは残さない



カメラから出力されたデータ



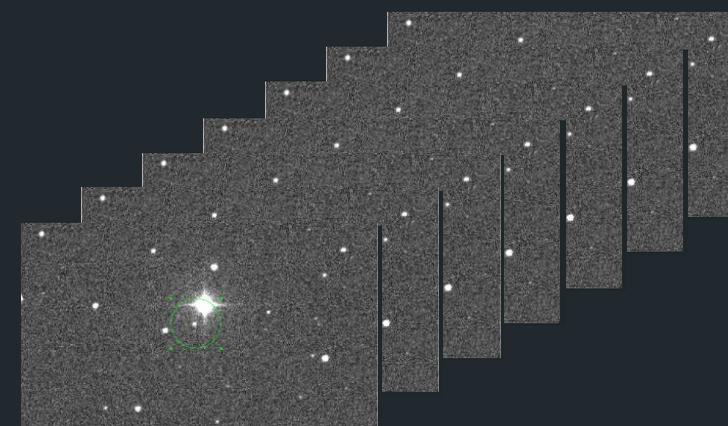
簡易解析(バイアス・フラット処理)

トリミングや 32bit 化を実施
データレートはおよそ ~50TiB/夜になる



ドーム内のストレージに一時保管

解析処理のバッファとして使用
およそ 6 日前のデータまで後から参照可能



3次元 FITS データ

木曾シュミット望遠鏡ドーム

Tomo-e Gozen カメラ

ドーム内バッファストレージ

~50 TiB/夜
※簡易解析によって 16bit→32bit

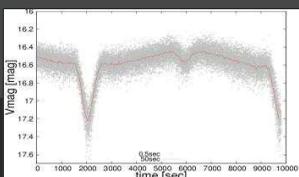
~300 TiB
取得した RAW データを
およそ 6 日間保管可能

ドーム内解析用計算機群

解析によってデータ総量を
およそ ~2 TiB/夜 まで削減する
現在は 4 種類のパイプラインを検討中

※中間生成データは残さない

1
全測光値
@2 Hz



2
長時間
積分画像



3
抜粋動画
@2 Hz



4
高速移動
天体



カメラから出力されたデータ



簡易解析(バイアス・フラット処理)

トリミングや32bit化を実施
データレートはおよそ ~50 TiB/夜になる



ドーム内のストレージに一時保管

解析処理のバッファとして使用
およそ 6 日前のデータまで後から参照可能



解析用計算機へ転送



自動解析パイプラインで処理

WCSの貼り付けなどはこの段階で実施予定
必要なデータだけ抽出して ~1-2 TiB/夜 まで圧縮
基本的に中間生成データは残さない
生成されるデータについては後述

本館のアーカイブに転送

解析済みのデータをおよそ1年分保持可能



木曾観測所本館

Tomo-e Gozen データアーカイブ

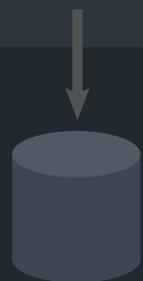
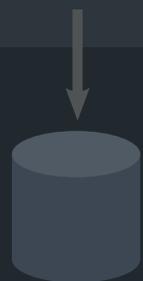
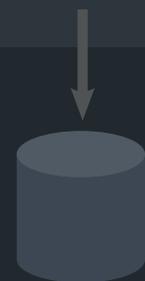


~300 TiB

解析済みのデータをおよそ1年分保管可能
※公開範囲は木曾観測所内部

※木曾観測所の計算機室にプロジェクト用のスペースを用意

各プロジェクトの計算機へ



ALERT!

物理メディアを郵送して
データセンターにアーカイブ

※ 頻度は月に1度程度を想定

外部アーカイブシステム(予定)

※ 現在 SMOKA チームと協力して検討を進めている

Kyoto University 3.8m Telescope

本館のアーカイブに転送

解析済みのデータをおよそ1年分保持可能



木曾観測所本館

※木曾観測所の計算機室にプロジェクト用のスペースを用意

Tomo-e Gozen データアーカイブ



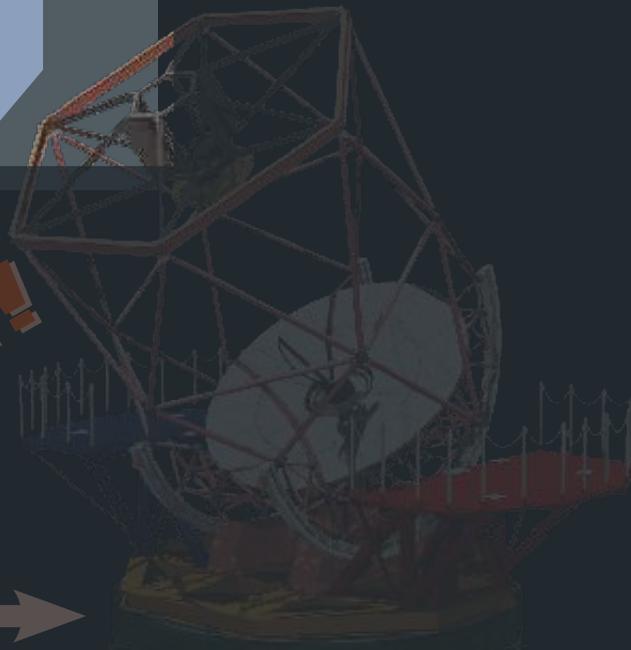
各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※ 頻度は月に1度程度を想定



ALERT!



外部アーカイブシステム(予定)

※ 現在 SMOKA チームと協力して検討を進めている

Kyoto University 3.8m Telescope

本館のアーカイブに転送

解析済みのデータをおよそ1年分保持可能

※木曾観測所の計算機室にプロジェクト用のスペースを用意

木曾観測所本館

Tomo-e Gozen データアーカイブ



各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※頻度は月に1度程度を想定



ALERT!



Kyoto University 3.8m Telescope

外部アーカイブシステム(予定)

※現在 SMOKA チームと協力して検討を進めている

本館のアーカイブに転送

解析済みのデータをおよそ1年分保持可能

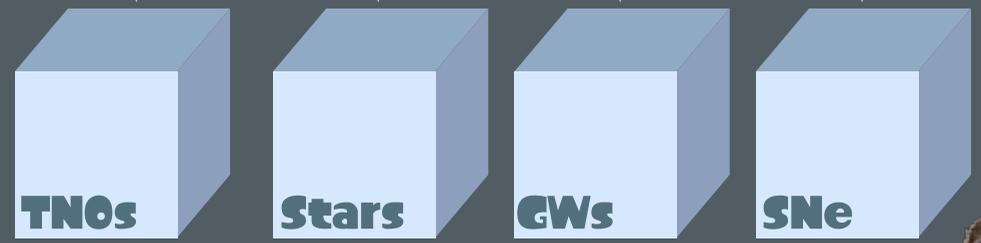
※木曾観測所の計算機室にプロジェクト用のスペースを用意

木曾観測所本館

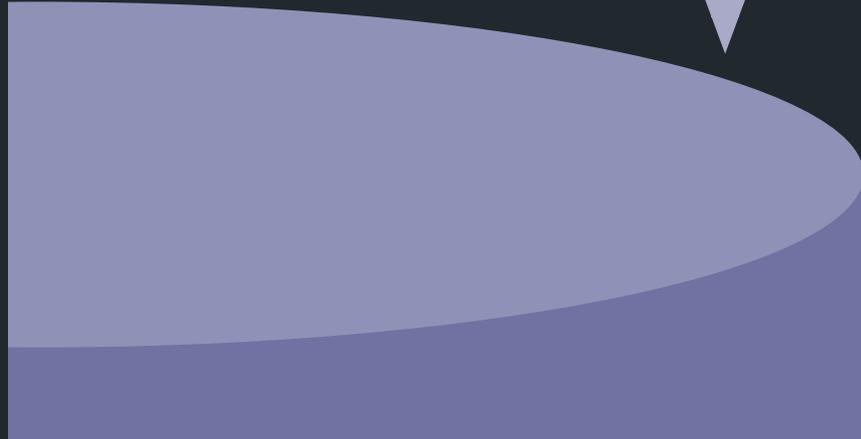
Tomo-e Gozen データアーカイブ



各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※頻度は月に1度程度を想定



ALERT!



Kyoto University 3.8m Telescope

外部アーカイブシステム(予定)

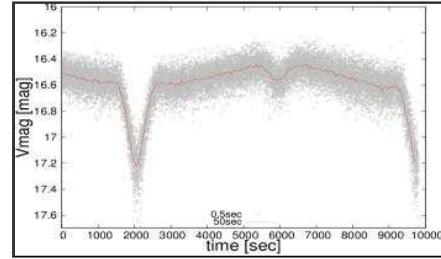
※現在 SMOKA チームと協力して検討を進めている

トモエゴゼンの使い方

提供されるデータ(案)

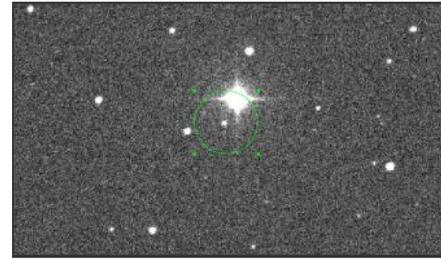
1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~18mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号

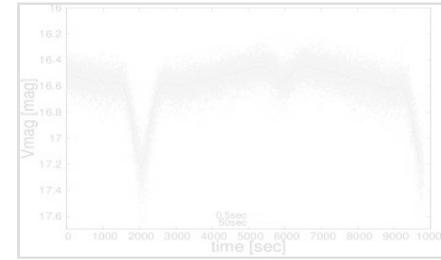


※ 現在 一木・大澤・酒向で基本パイプラインの開発を進めている

提供されるデータ(案)

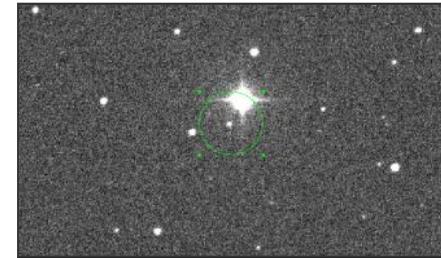
1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~18mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号

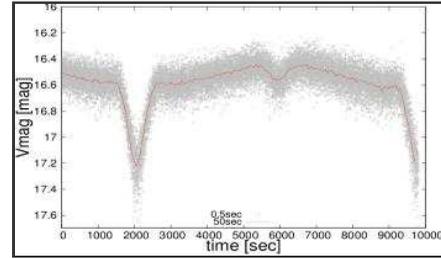


※ 現在 一木・大澤・酒向で基本パイプラインの開発を進めている

提供されるデータ(案)

1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~18mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号

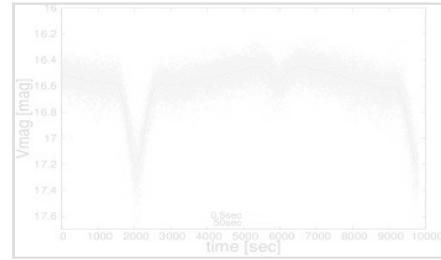


※ 現在 一木・大澤・酒向で基本パイプラインの開発を進めている

提供されるデータ(案)

1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~18mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号



※ 現在 一木・大澤・酒向で基本パイプラインの開発を進めている

データへのアクセス方法

基本的に観測所内部のアーカイブは外部に非公開

データ転送で観測所への回線を食い潰す恐れがあるため

共同研究プロジェクトにデータのアクセス権

⇒ 本館の計算機室にプロジェクトの計算機を置く

⇒ プロジェクト計算機から本館のアーカイブにアクセス

ドーム内の計算機には直接アクセスさせない方針

※ 外部のアーカイブシステムでのデータ公開も予定 (現在検討中)

トモエゴゼン計画 — データ編

トモエゴゼン計画:

特化型かつ柔軟な観測に対応することを目指す
プロジェクト単位での共同研究を想定

データ運用計画:

全データ保存は難しい ⇒ 重要な情報だけ抽出する
基本データセットを定義・パイプラインで生成
データへのアクセス権は共同研究チームに付与
各チームは木曾に計算機を置いて解析
最終的には外部にデータのホスティングを依頼予定

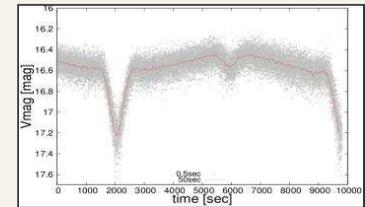
議論・検討事項

議論 — 提供されるデータセット

残すべき情報について

1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~19mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



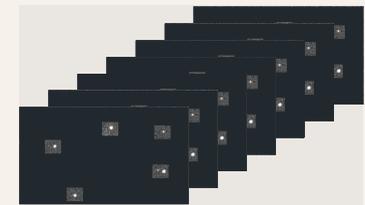
2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号



議論 — 京都大学 3.8m 望遠鏡との連携

アラートを送信

基本構想: 木曾観測所 ⇒ 京大 3.8m

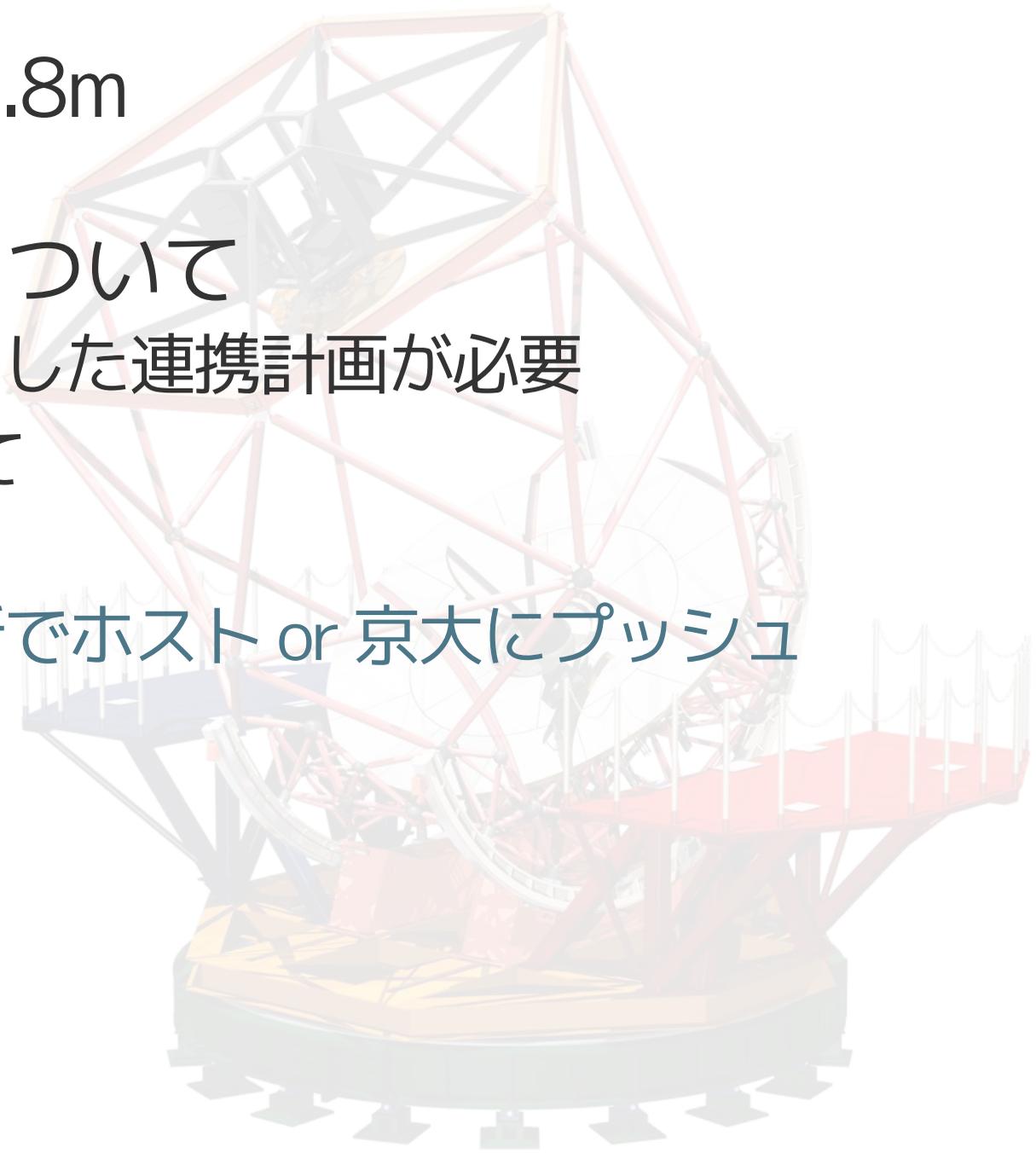
追観測依頼の望ましいかたちについて

アラート数 > 観測可能数を想定した連携計画が必要

アラートのフォーマットについて

アラートのかけ方について

候補天体リストを木曾観測所でホスト or 京大にプッシュ



議論 — より柔軟なプロジェクトに

特殊な観測計画の実施について

高速部分読み出し (e.g., TNO survey, pulsar survey)

non-sidereal objects (e.g., debris, meteors)

対物グリズム分光観測

⇒ 解析パイプラインの開発も含めた共同研究として考えたい

動画をより深く解析する計画について

動画解析技術の応用, 科学用途指向動画圧縮 (e.g., GoDec 圧縮)

ハードウェアパイプラインによる高速動画処理 (e.g., 重ねあわせ法)

⇒ 共同研究計算機をドーム内に増設することを検討

まとめ

トモエゴゼン計画:

特化型かつ柔軟な観測に対応することを目指す
プロジェクト単位での共同研究を想定

データ運用計画:

全データ保存は難しい ⇒ 重要な情報だけ抽出する
データは木曽でアクセスして木曽で解析する

検討事項:

最終的なアーカイブとしてデータのホスティングを依頼
基本パイプラインで残す情報セット
京都大学 3.8m 望遠鏡との連携観測・情報共有方法
より複雑な観測提案に対する共同研究体勢

提供されるデータについて

1. 全天体測光値テーブル at 2 Hz **monitoring: ~200GB**
survey: ~30GB

2Hz で映る全天体 (~18mag.) およそ 10^5 天体の測光結果を保存
測光結果リストを text テーブル or データベースとして提供

2. stacked deep images **monitoring: ~75GB**
survey: ~1.5TB

サーベイなら 3s, モニタリングなら 180s 積分画像に相当

3. cropped movies of fast transients **monitoring: ~250GB**
survey: ~40GB

10×10 の領域を 8,000 天体分確保, 3次元 FITS 画像として提供

4. 高速移動天体スナップショット **monitoring: ~500GB**
survey: ~150GB

流星であれば 10^3 – 10^4 イベントが期待, 3次元 FITS 画像として提供