Streams, AMPEL & ZTF

•••

Jakob Nordin, HU

Content

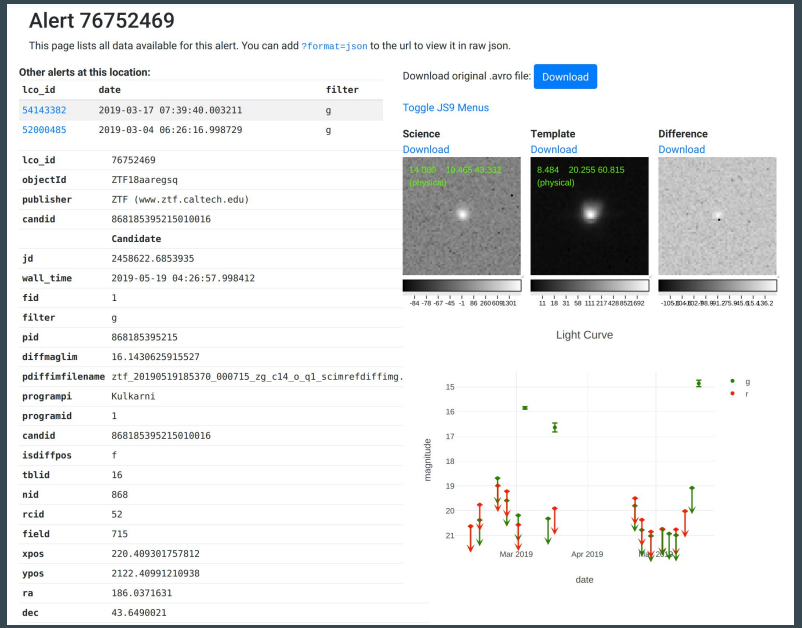
- What do we mean with alert streams?
- Why streams (data perspective)?
- Why streams (science perspective)?
- What is a broker?
- The Zwicky Transient Facility
- Alert Management, Photometry & Evaluation of Lightcurves (AMPEL)

What do we mean with alert stream?

- An *alert* is a collection of properties tied to a transient.
- A *stream* is a collection of alerts distributed one by one.

A *client* receives stream from a *server*. Can restart *pointer* from different timestamps and use parallel processing through multiple clients, but cannot arbitrarily jump to transient X.

Distinguished from “database” approach where a user actively queries for objects fulfilling certain criteria.



Alert 76752469

This page lists all data available for this alert. You can add ?format=json to the url to view it in raw json.

Other alerts at this location:

lco_id	date	filter
54143382	2019-03-17 07:39:40.003211	g
52000485	2019-03-04 06:26:16.998729	g

lco_id	76752469
objectId	ZTF18aaregsq
publisher	ZTF (www.ztf.caltech.edu)
candid	868185395215010016
	Candidate
jd	2458622.6853935
wall_time	2019-05-19 04:26:57.998412
fid	1
filter	g
pid	868185395215
diffmaglim	16.1430625915527
pdiffimfilename	ztf_20190519185370_000715_zg_c14_o_q1_scimrefdiffimg.
programpi	Kulkarni
programid	1
candid	868185395215010016
isdiffpos	f
tblid	16
nid	868
rcid	52
field	715
xpos	220.4093017578 12
ypos	2122.40991210938
ra	186.0371631
dec	43.6490021

Download original .avro file: Download

Toggle JS9 Menus

Science Download

Template Download

Difference Download

Light Curve

magnitude

date

Mar 2019

Apr 2019

What is a broker?

“a person who buys and sells goods or assets for others.”

It is something which receives the alert streams, so a logical consequence of the stream concept.

Not clear what they actually do. At its simplest a filter that narrows the information flow into something that can be saved into an accessible DB (refutes the purpose of streams).

Why streams (data perspective)?

- Raw data (detections) from instruments typically come as alerts.
- Processing these and inserting into a database takes (unknown) amount of time.
- Potentially large number of similar requests: “return the latest, brightest thing”.
- Easier for maintainers to outsource problem by providing non-interactive streams.
- Allows more creative ways to use data.

Why streams (science perspective)?

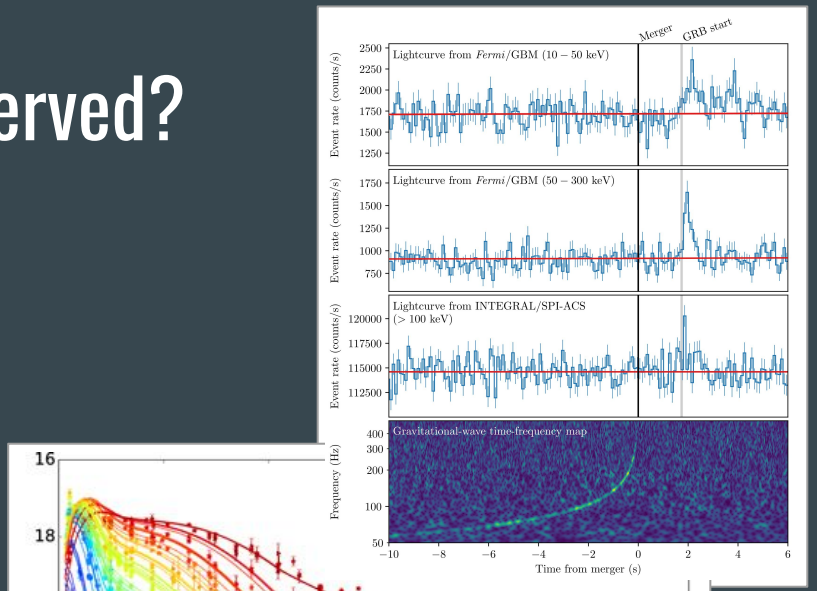
- Allows fully real-time reactions.
- Native multi-messenger studies through combining streams.
- Current (at least optical) analysis chains inflexible, discouraging provenance as well as proper software development. Consider an example...

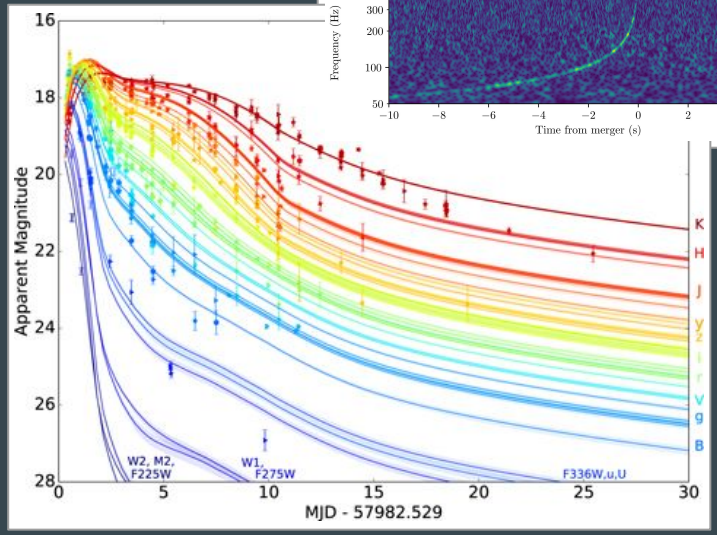
How many kilonovae have we observed?

GW170817 led to the discovery of a completely new kilonova transient class.

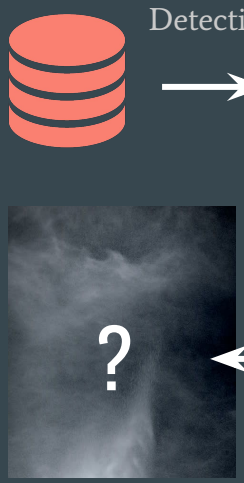
How many kilnovae have we observed so far?
We have no idea!

Why? Because prior survey can be described through a process of irreversible table transformations not aimed at GW170817.





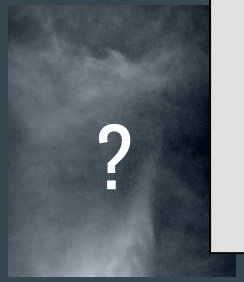
Linear process of irreversible table transformations



Detection → Calibrate → Filter → Inspect → Trigger → Observe → Inspect → Analyse → Inspect → Analyse → Observe → Calibrate → ? / Analyze → Inspect → Analyse → Inspect → Inspect → Analyse → Inspect → nature

Involves code running, human choices and input that depends on external circumstances (eg weather).

Linear process of irreversible table transformations



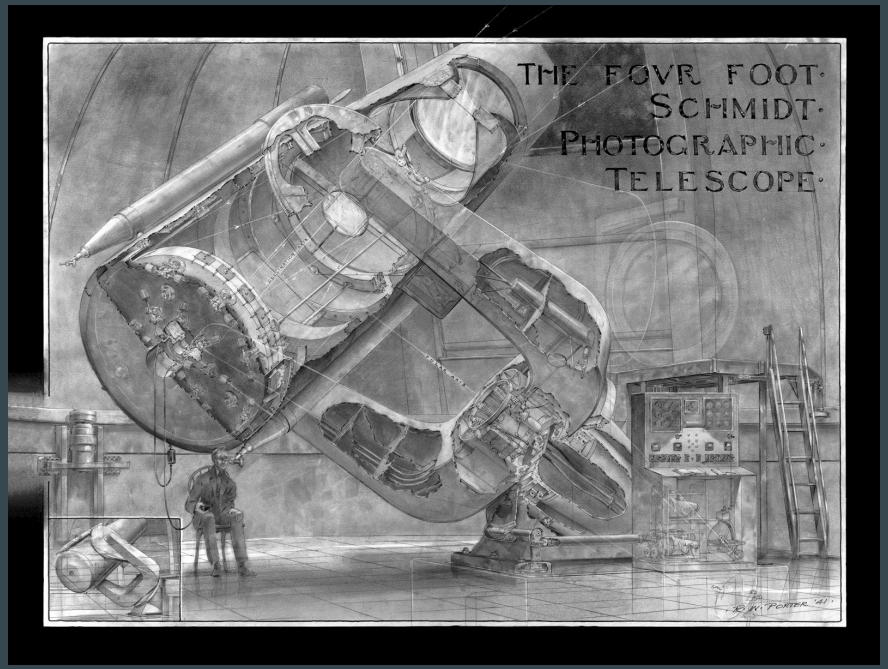
- While tables are stored, the real work occurs in transformations. This is often lost.
- No natural point to develop e.g. multi-messenger code.
- Not designed for large samples or automatic surveys.
- No easy way to answer what if questions.
  - What does it mean that we have no spectra of kilonovae? That they have never existed, never passed threshold, never selected?
  - Very real question for SNIa systematics.

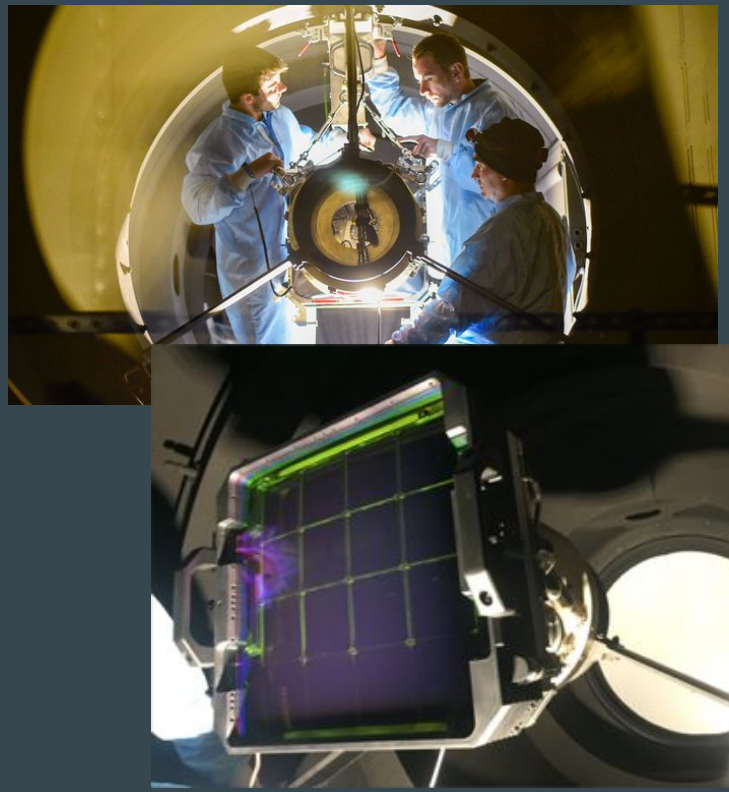
Involves code running, human choices and input that depends on external circumstances (eg weather).

Why streams? (summary)

- Data management choice has been made.
- Required to reach real-time.
- A chance to create an analysis framework that encourages provenance and good software solutions.

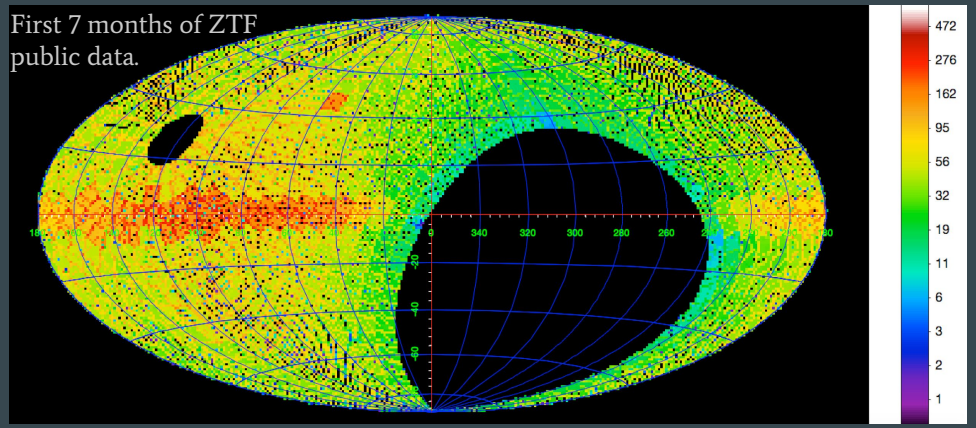
The Zwicky Transient Facility





Gigantic camera (47 sq deg) installed into old POSS P48.

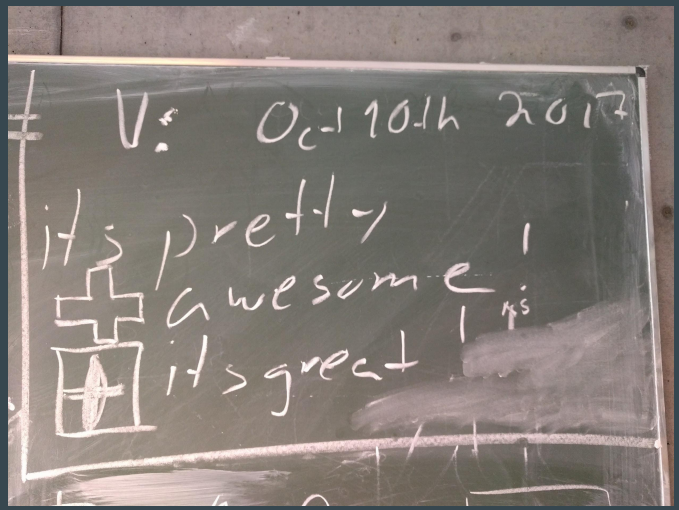
The Zwicky Transient Facility



- Can monitor the full sky to 20.5 mag every night. Three optical filters installed.
- Operated by a complex combination of collaboration, CalTech and NSF.
- We got involved to:
  - Scan sky for counterparts o IceCube extragalactic neutrinos.
  - Find thousands of Type Ia supernovae for SN Ia cosmology.
- Each image scanned for new transients at IPAC, and *alerts* issued for significant detections.

Alert Management, Photometry and Evaluation of Lightcurves

- Development started w. V. Brinnel to handle SN detection systematics.
- Prior to any alert distribution or the term “broker” introduced.
- Currently ~12k lines of Python over ~10 Git repositories.
- The acronym getting less and less valid each day.



Axiomatized, Modular, Provenance-ready Execution Layers?

Design goals

Scientist do what they are good at:
- Request input (eg optical lightcurve + IceCube sky map)
- Perform calculation in python (does the explosion time coincide with a neutrino detection?)
- Return answer (4.3 sigma disagreement)

An AMPEL *channel* requests a number of such units to be run on all transient *states**.

AMPEL takes care of the rest
- Logging, version control, replayability
- Ensure algorithm consistently applied to data

* A state is the core AMPEL entity, to which computations belong. New states are created by new observations, but also eg. updated calibration or data rights. A transient is connected to a set of states.

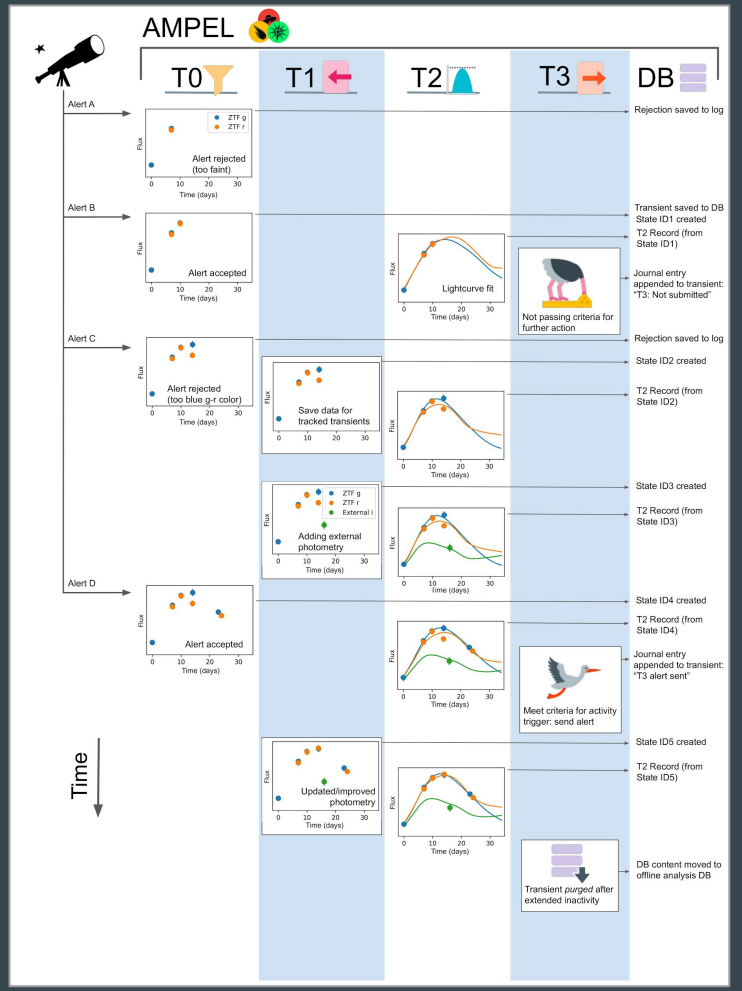
Design principles

A user creates a *channel* for a specific purpose. This calls on *units* to perform calculations and actions on transients stored in a central (Mongo) DB.

– Provenance and reproducibility are guaranteed by the combination of information stored in a permanent database, containerized software and alert archive in a system designed to allow autonomous analysis chains.

– A modular system provides analysis flexibility, and introduces a method for developers to allow software distribution and referencing.

– Designed to manage the alert rates expected from surveys such as LSST.



Transient states

Transient described through parallel *states*, rather than entries in sequential tables. A state is a collection of measurements.

States vary based on time-range, sources, data rights etc.

Code/algorithms do one of the following:

- Create new states (T1).
- Perform a calculation based on a state (T2).
- Rank/react to a collection of transients (T3).

Developers contribute units that operate in a tier on states with a required set of properties. AMPEL does all the rest.

Outstanding issues

- Currently at v 0.6.x.
  - Version 0.8 (stable) expected this fall.
  - Updates require manual work (read JvS).
- Users provide units and channels through github reps. that are incorporated into either a live processing instance or a data replay.
  - Need to make this easy!
  - Mirror DB accessible to users?
- Rerun facility (NERSC?)
- Processing two “master” streams.
- More details:
  - https://arxiv.org/abs/1904.05922
  - https://github.com/AmpelProject/Ampel-contrib-sample

AMPEL team & projects

Core development by V. Brinnel, J. van Santen, M. Giomi & JN.

Units by L. Rauch, M. Rigault, R. Stein, N. Miranda, S. Van Velzen, C. Ward, S. Schulze, ...

Active research programmes for cosmology, neutrino, AGN/TDE, supernova physics, GW, lensing (see next talk).


Valéry Brinnel


Matteo Giomi


Jakob Nordin


Ludwig Rauch


Mickael Rigault


Robert Stein


Nicolas Miranda

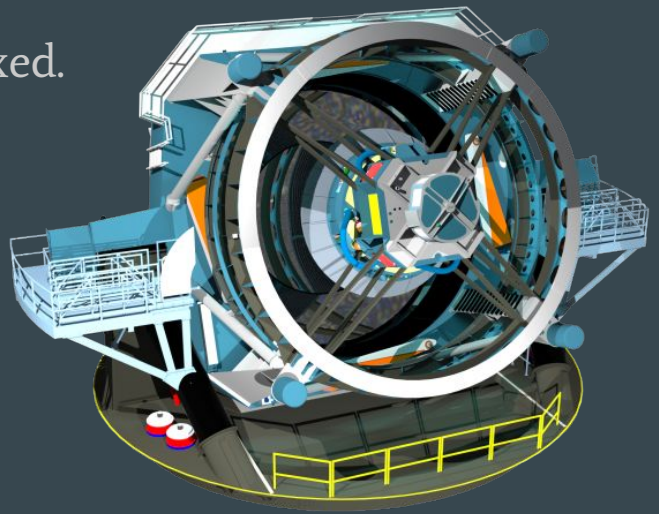

JvS

LSST & Ultrasat

...

Large Synoptic Survey Telescope

- Dark energy, dark matter, transients, solar system, MW formation.
- 8m telescope w. 3 billion pixel camera
- 800 images & 30 TB/ night
- 15s exposure in six bands, reach depth ~24.5 ®
- Fast and wide survey modes, but cadence not yet fixed.
- Survey start 2023-2024.



LSST data access and alert stream

LSST members (eg A.F.) will be able to access a DB containing both image data and transient detections. Products available there ~1 day after observations.

An immediate alert stream immediately public, but stream limited to 5-10 brokers (bandwidth limited).

Can expect up to 10 million alerts each night, with up to 1TB / night size.

The broker landscape

DESY submitted a letter of intent to host an LSST broker (PI Kowalski).

Full proposal guidelines expected Fall 2019 with due date Spring 2010.

A number of brokers exist today: ANTARES, MARS, LASAIR, ALERCE, ... Focused either on convenience (web interface) or machine learning (photometric classification).

Potentially unique focus on multi-messenger science and analysis flexibility.

Ultrasat

NUV-imager

15 deg FoV

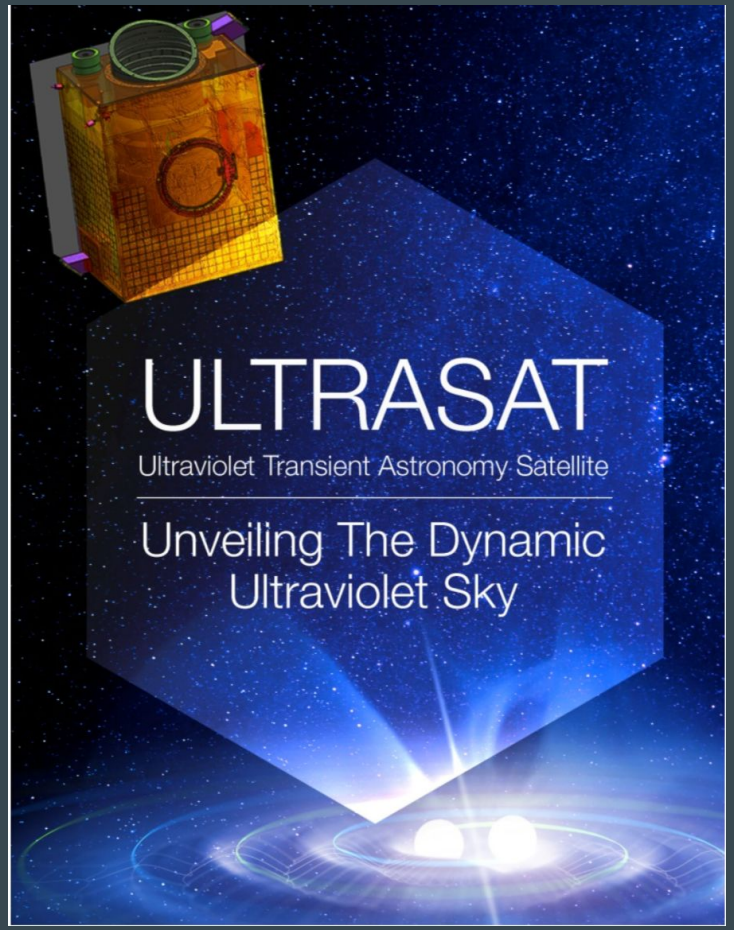
5 min cadence

Plan to be in orbit by 2023/4 and take data for 3 years.

Many research areas require real-time response.

GW, CC SNe, TDE.

Instrument output likely as stream.



Ultrasat

NUV-imager

15 deg FoV

5 min cadence

Plan to be in orbit by 2023/4 and take data for 3 years.

Many research areas require real-time response.

GW, CC SNe, TDE.

Instrument output likely as stream.