The present and future of gravitational-wave astronomy

Photo credit: Mike White

Michael Coughlin

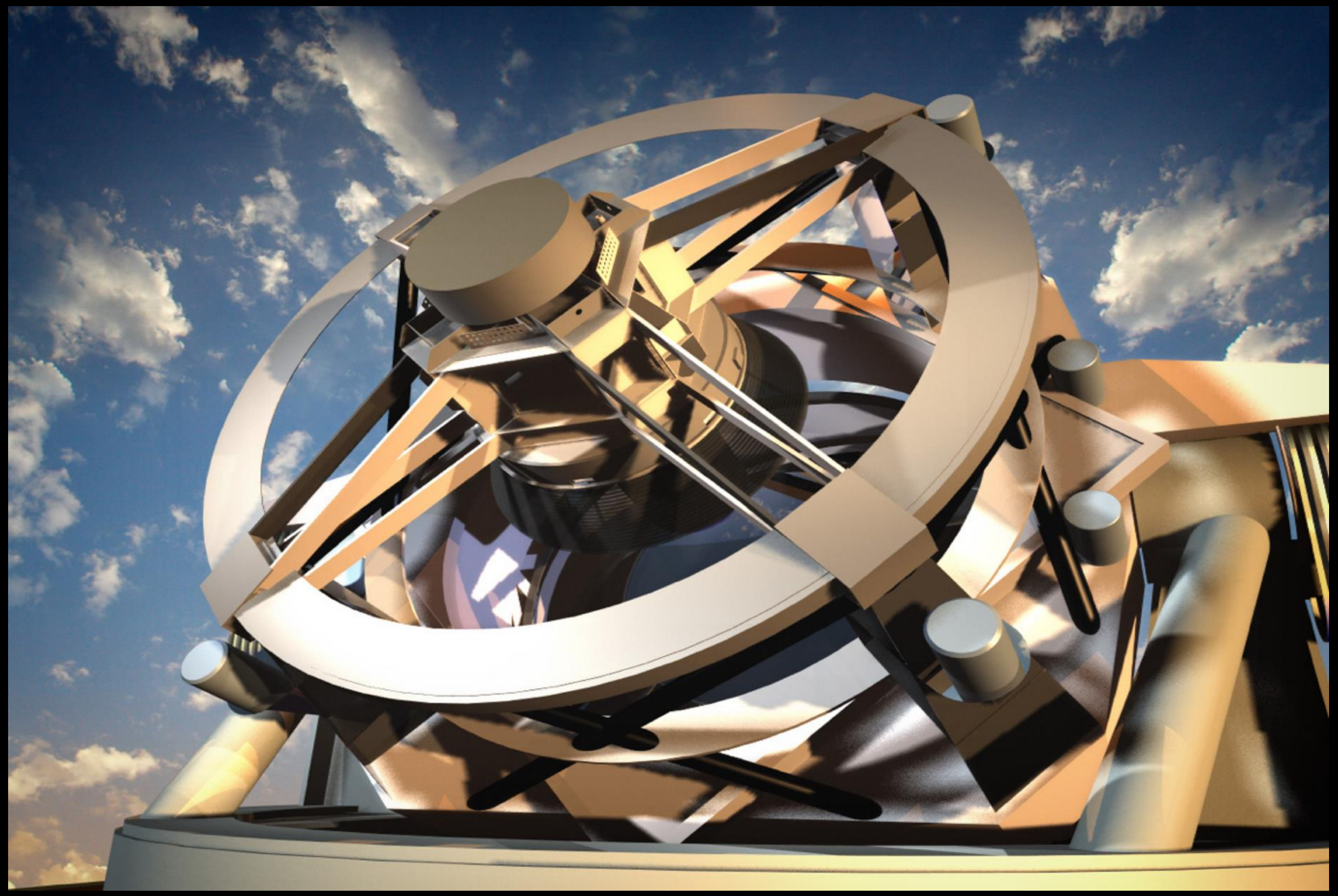
University of Minnesota - Twin Cities

HTC26

June 9, 2026

A Golden Era for Astrophysics

Vera Rubin Observatory / LSST



JWST

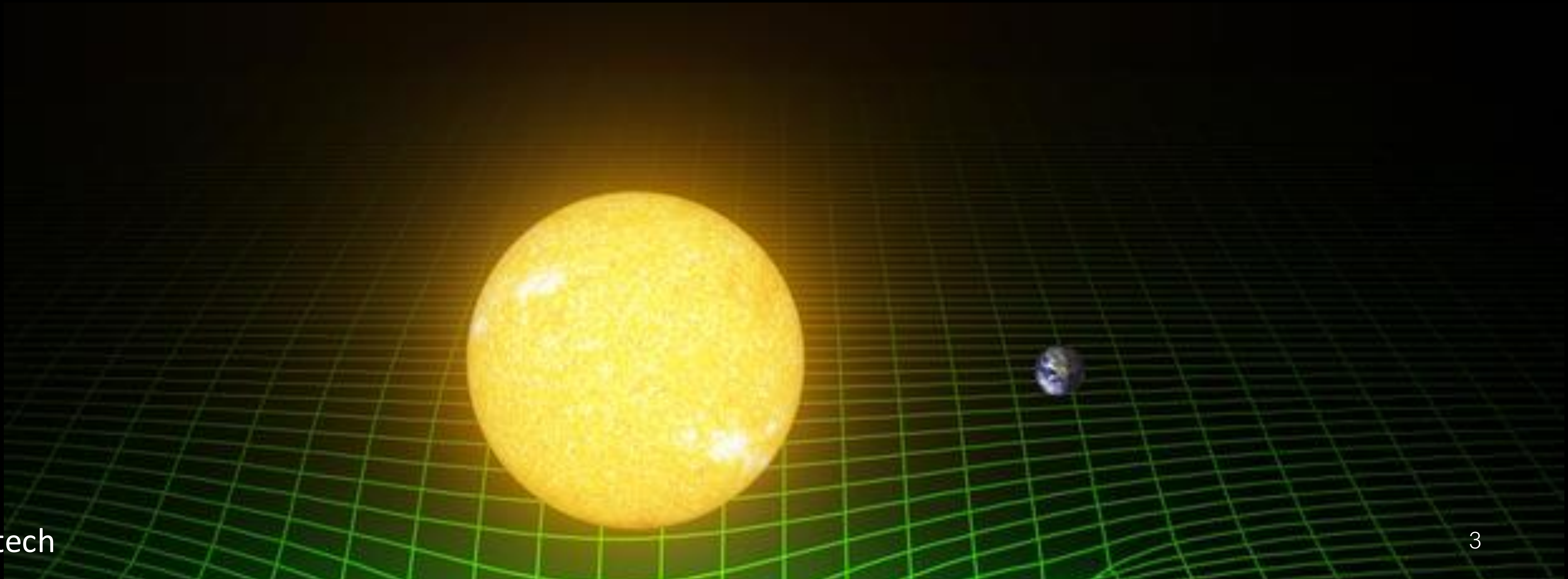
LIGO/Virgo/Kagra

IceCube

Einstein’s Gravity: General Relativity

$$G_{\mu\nu} = 8\pi \; T_{\mu\nu}$$

Matter tells spacetime how to curve
Spacetime tells matter how to move
- John Wheeler



LIGO/Caltech

Indirect evidence of gravitational waves

Hulse-Taylor Binary Pulsar

Won the Nobel Prize in Physics in 1993!

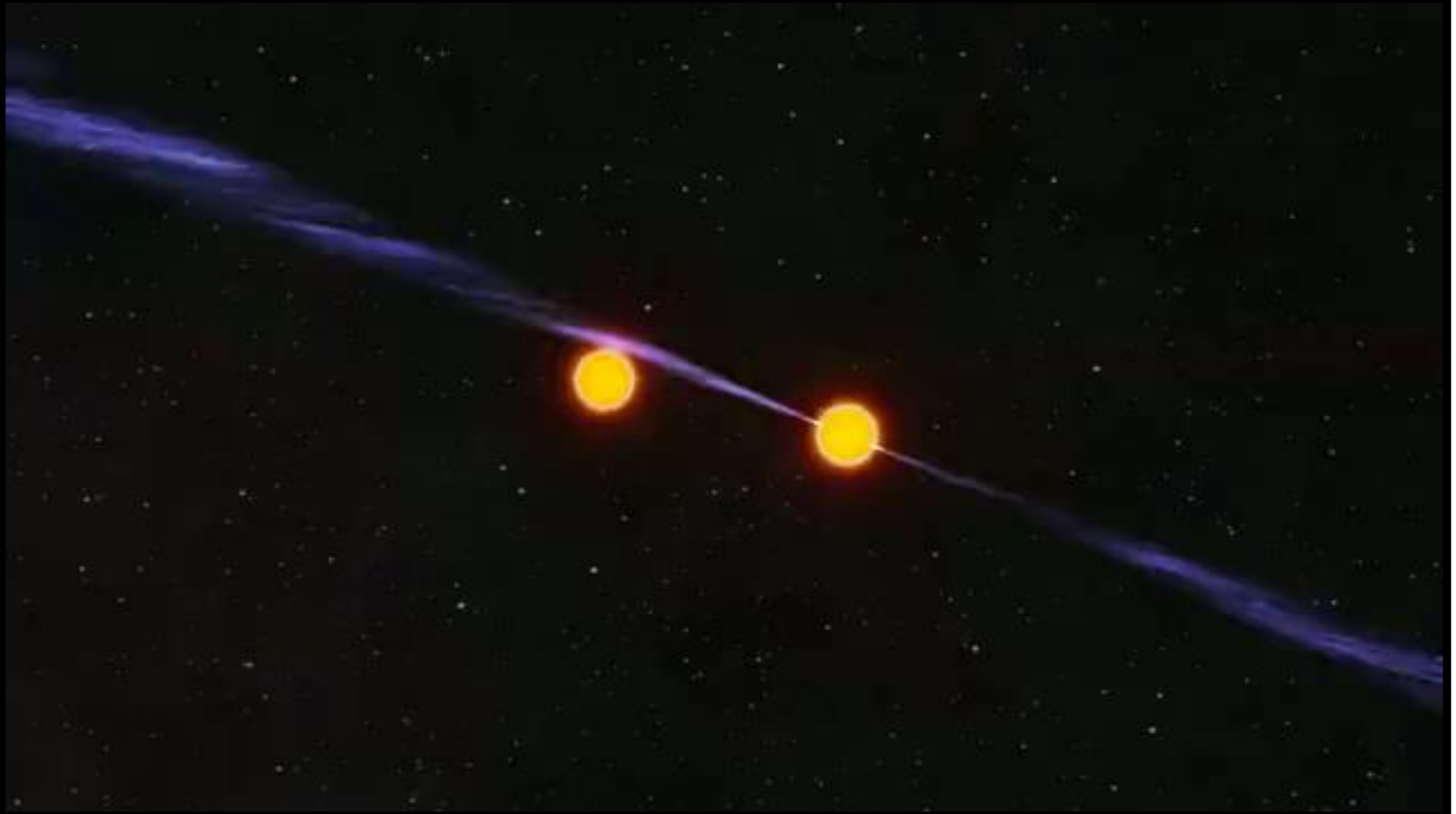
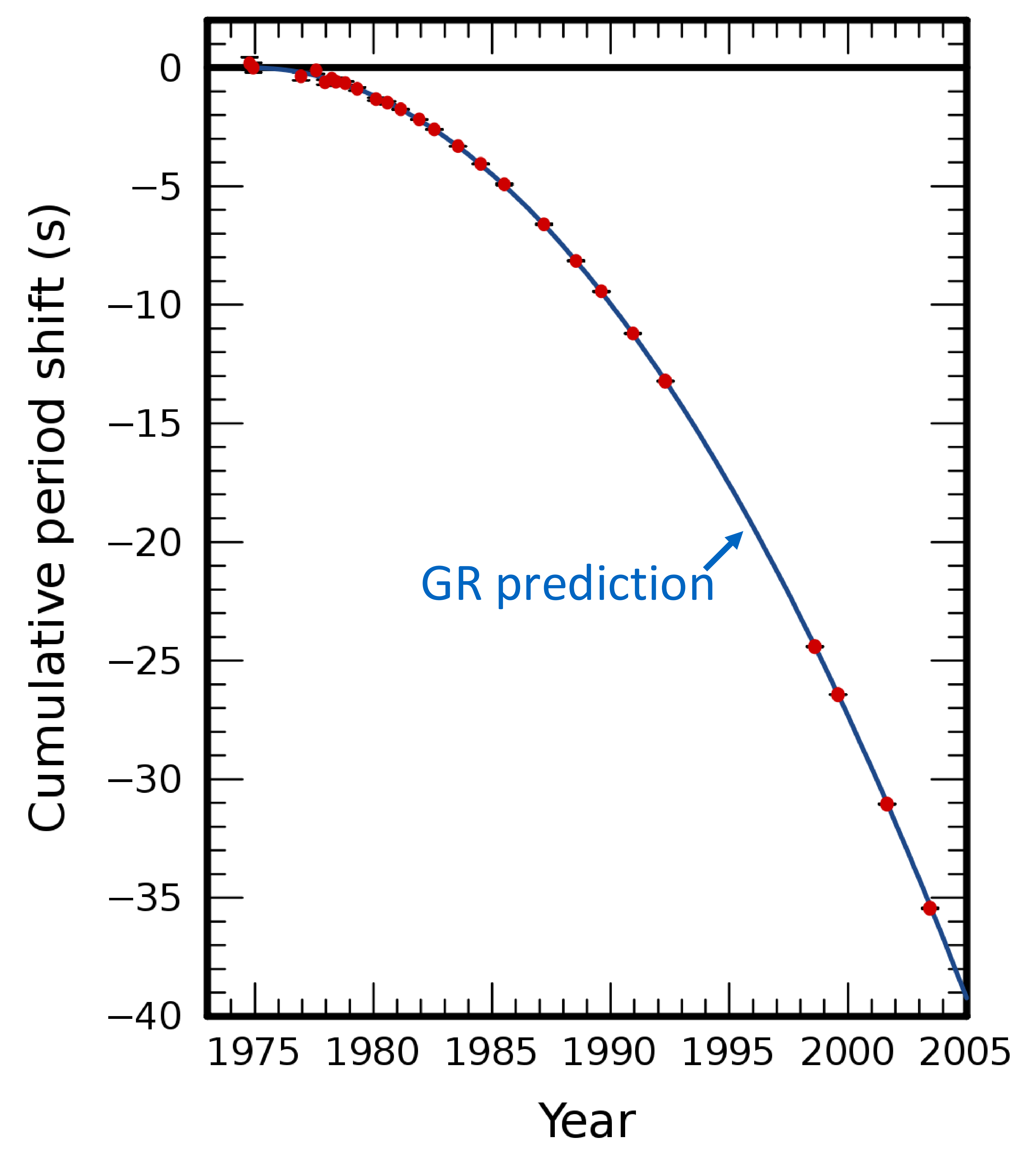
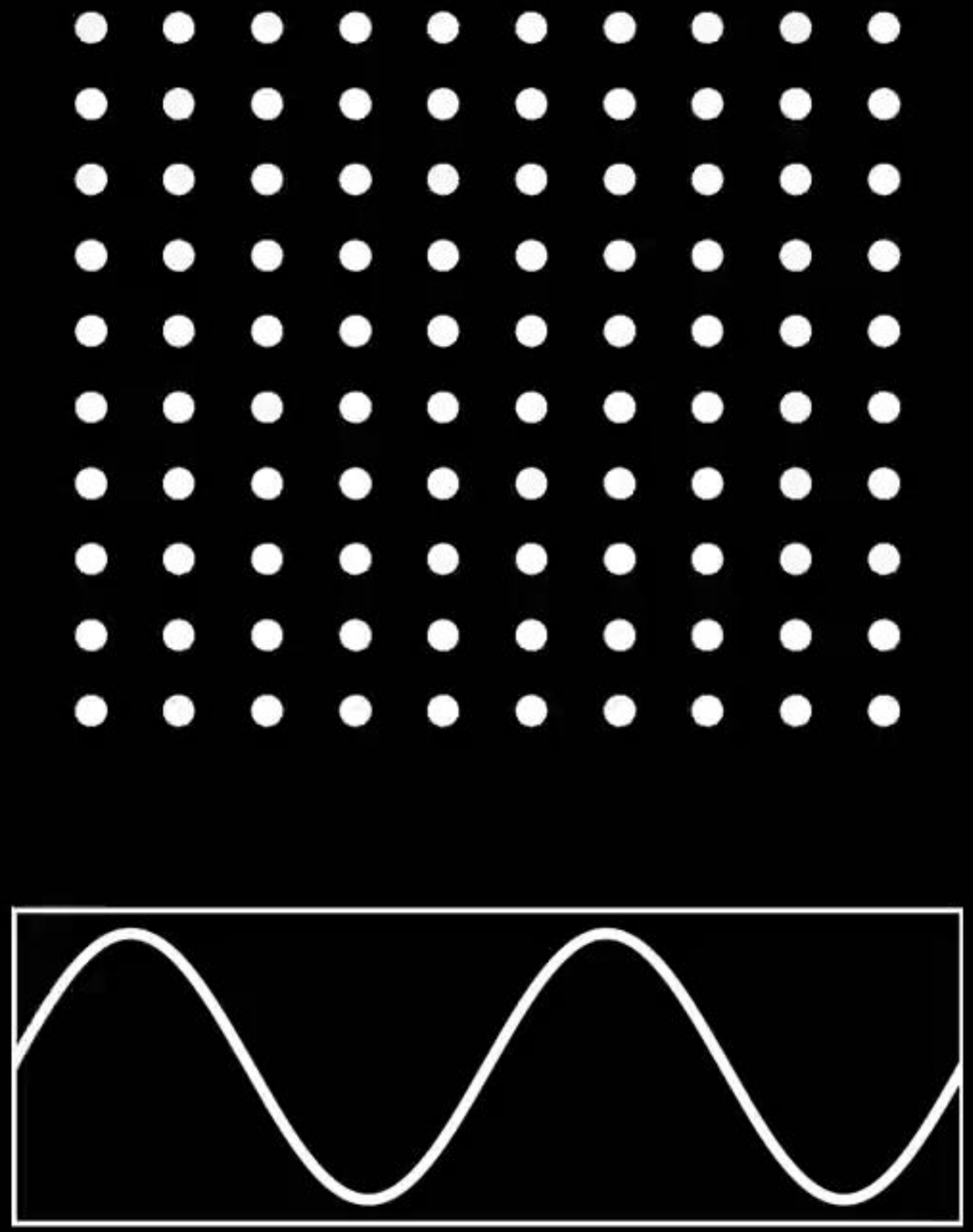


Illustration - John Rowe

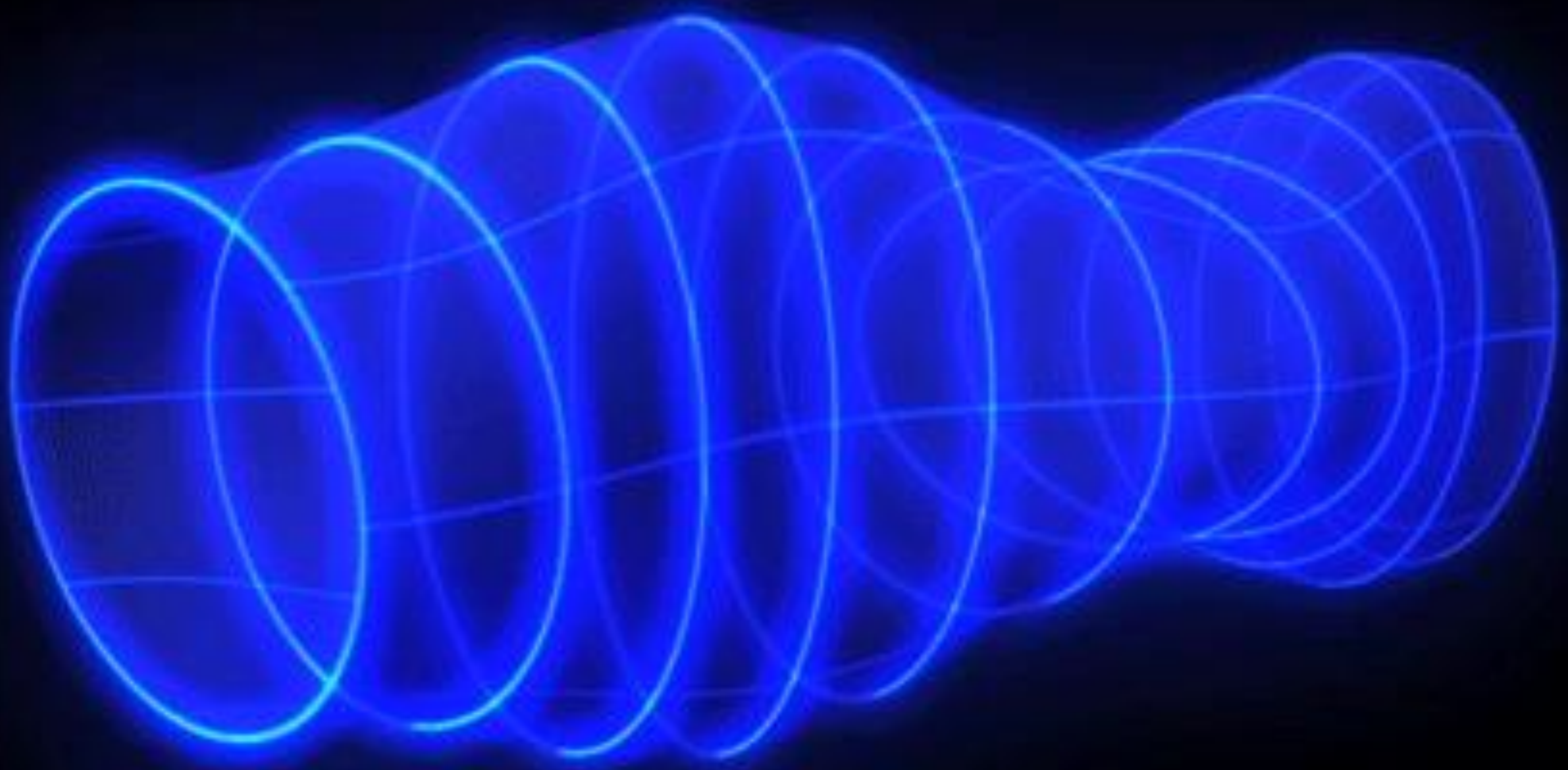


Weisburg, Nice & Taylor, 2010

Gravitational wave strain

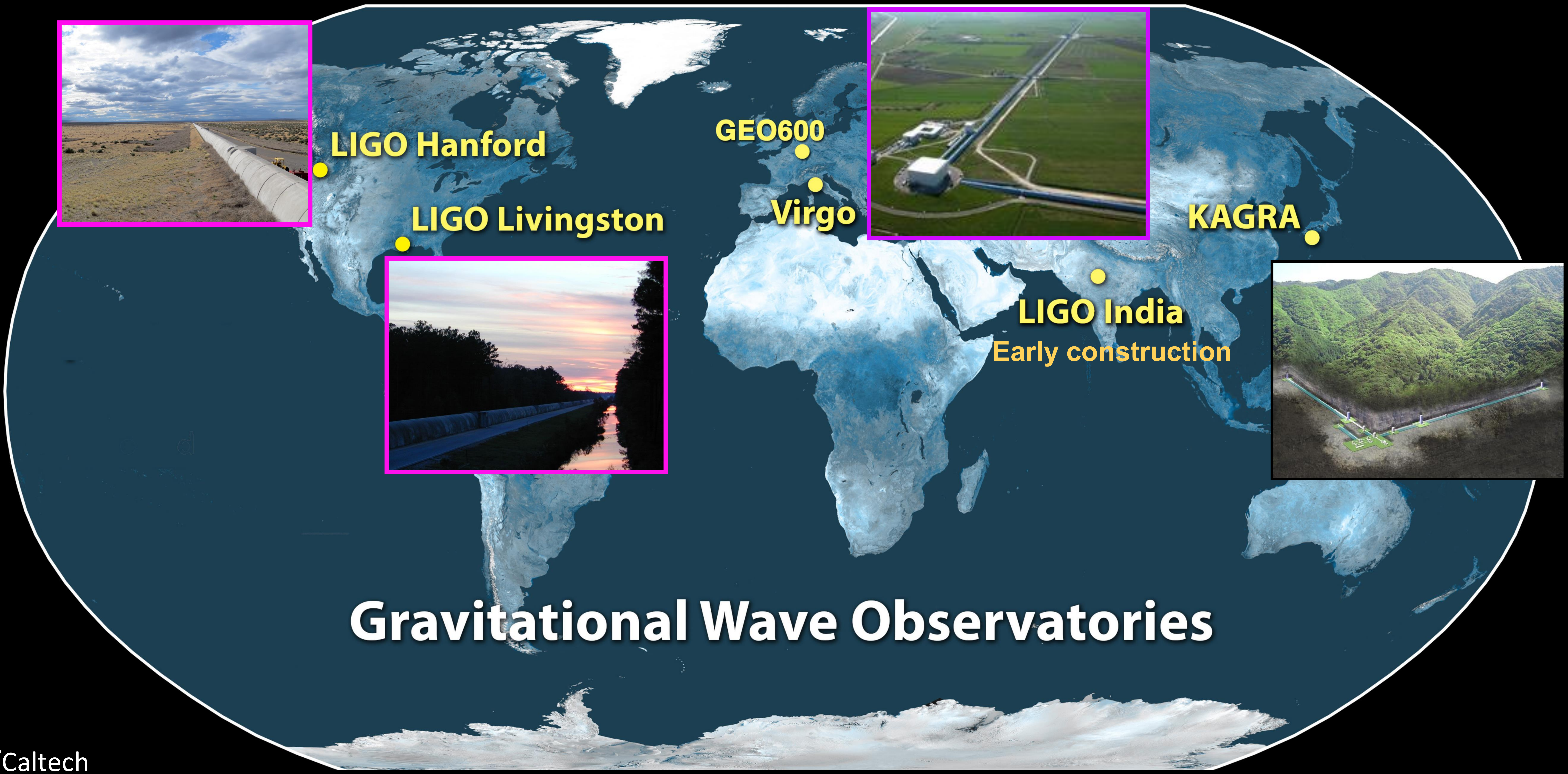


Movie: Carl Rodriguez



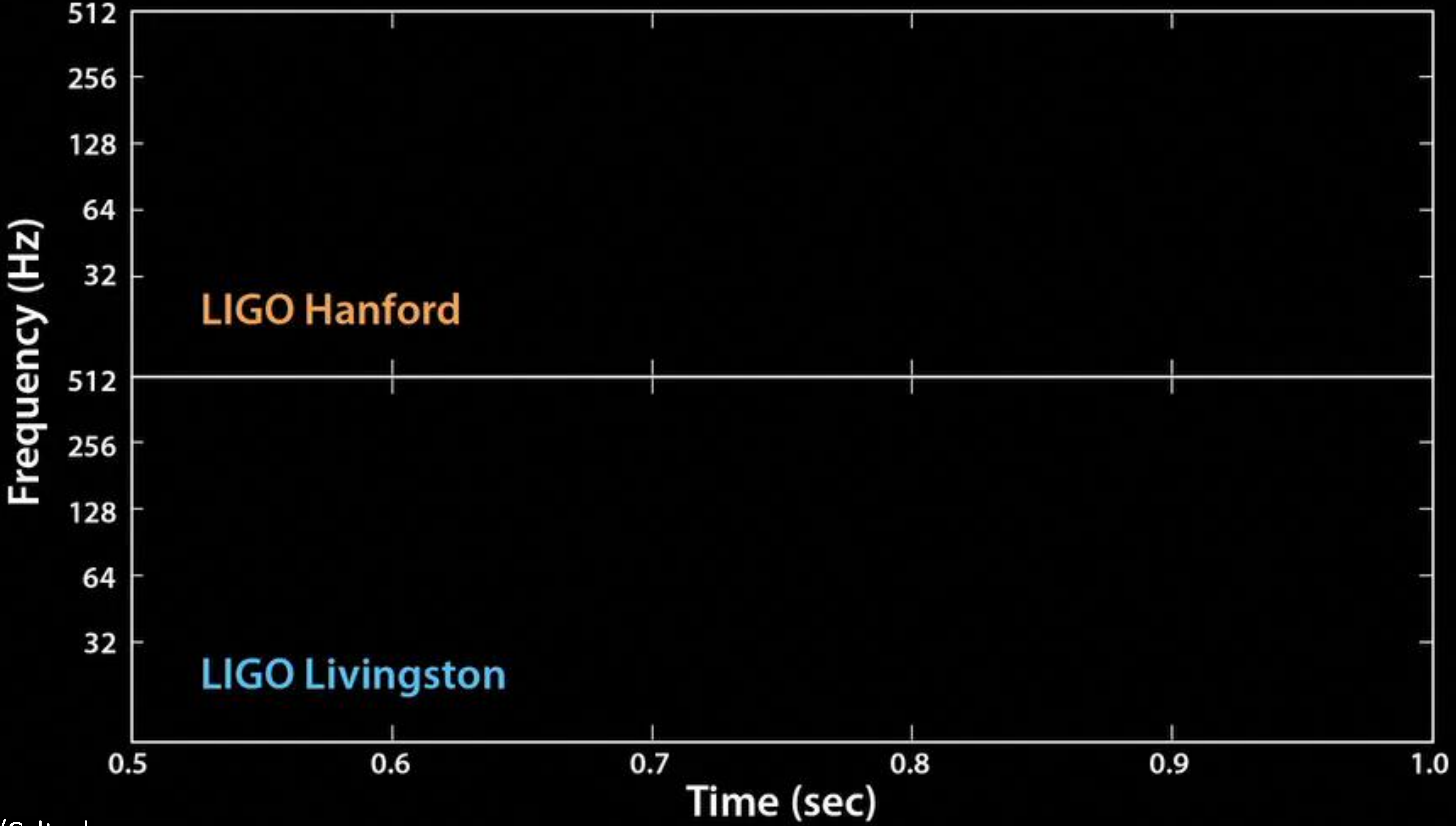
ESA

Current GW detector network (IGWN)

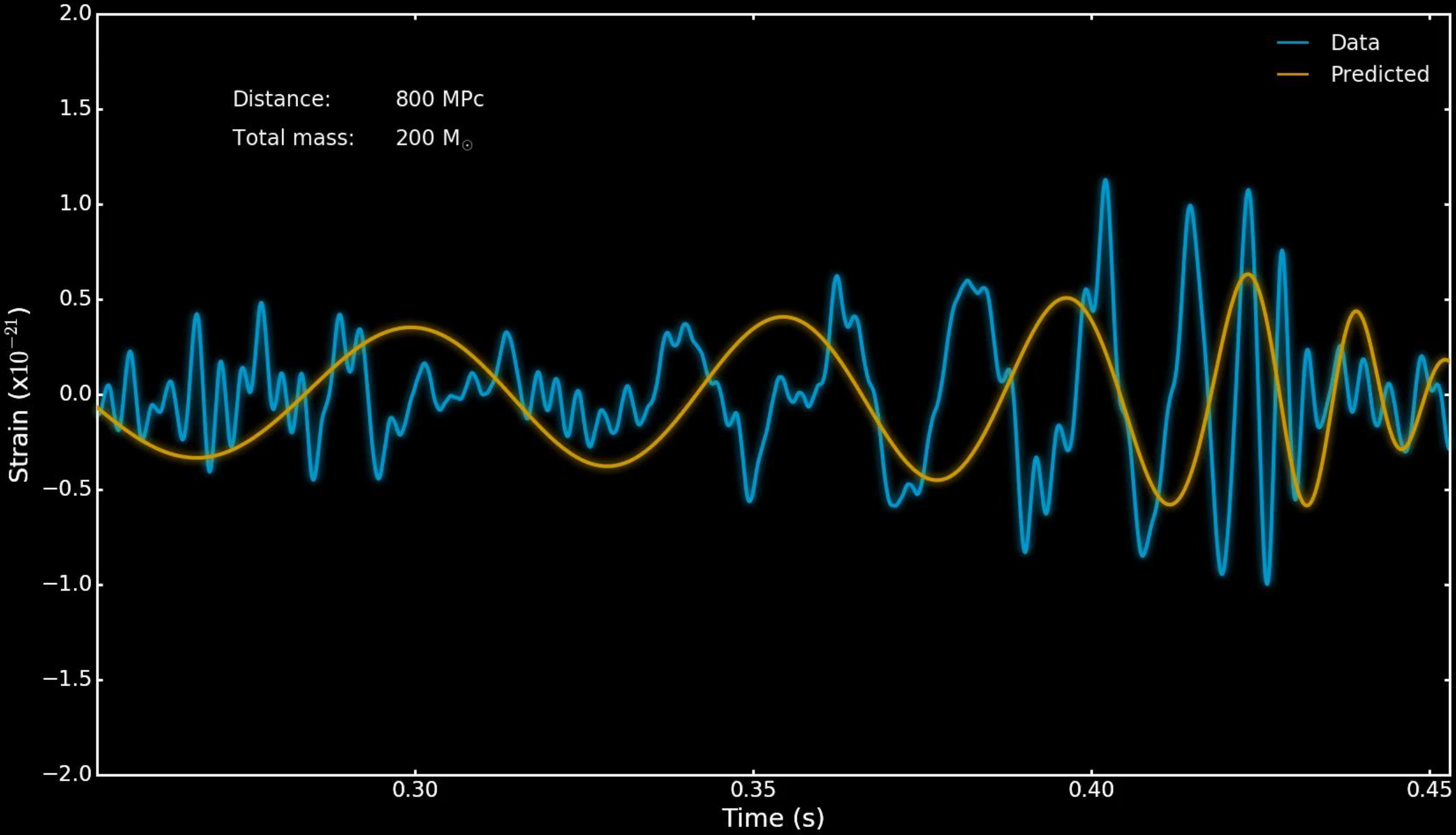


LIGO/Caltech


-0.69s
SXS



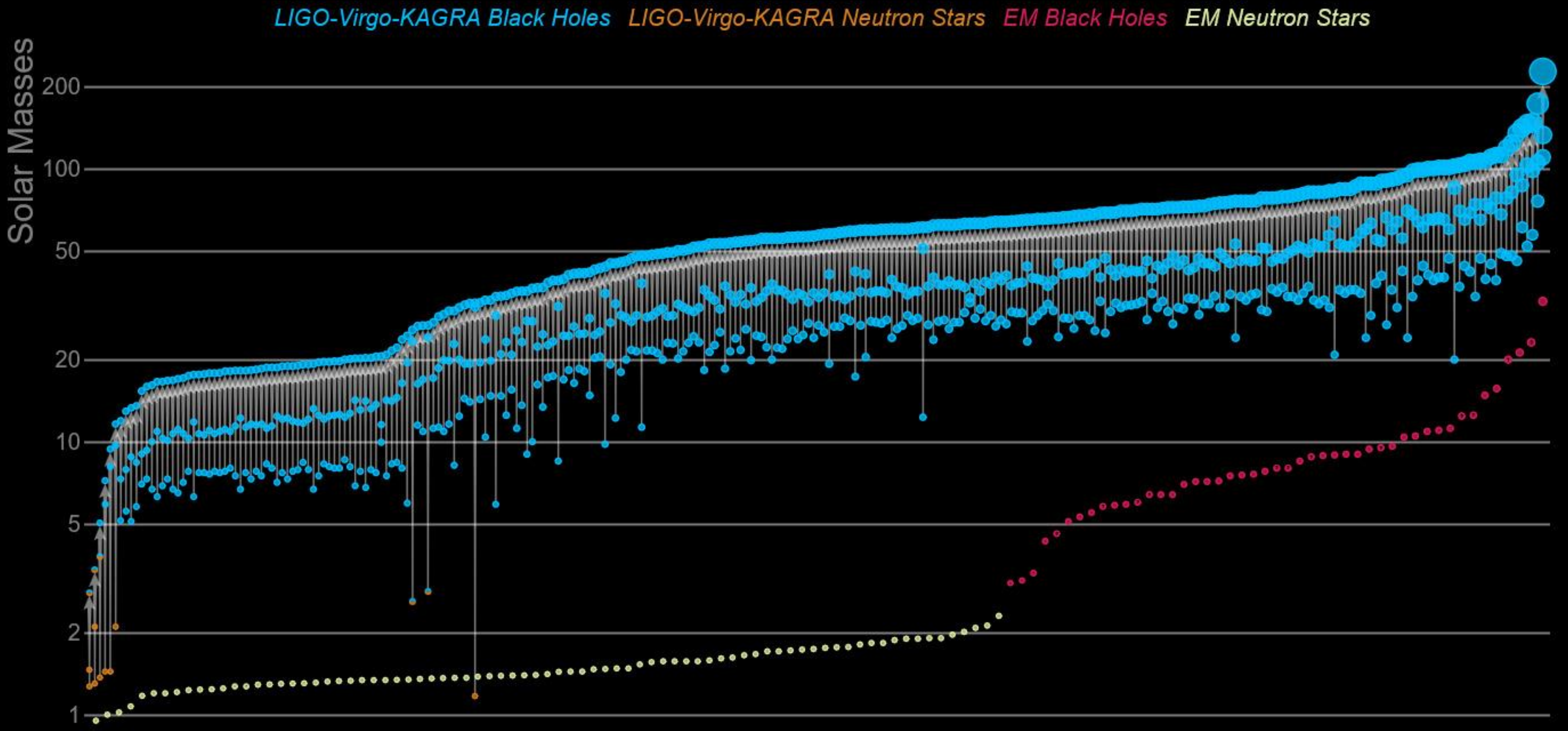
LIGO/Caltech



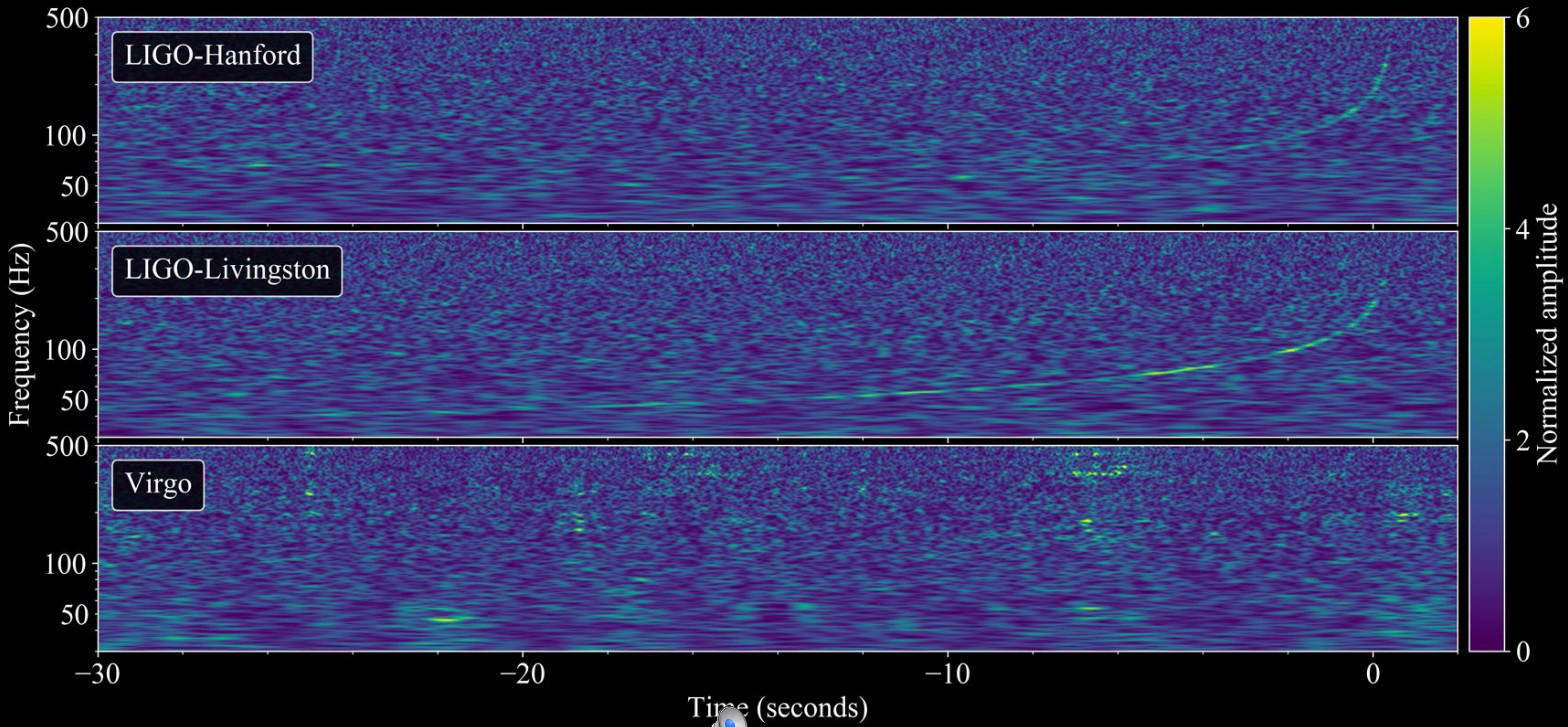
Credit: LIGO/Cardiff University/C.North/M.Hannam


THE NOBEL PRIZE IN PHYSICS 2017
RAINER WEISS
I'm telling you,
This sh*t is POSSIBLE!
KIP THORNE
And this is how a signal would look like.
Whoop!
BARRY BARISH
Let's pick one and MOVE ON guys...
PROPOSAL #2
PROPOSAL #1
PROPOSAL #7
PROPOSAL #10
PROPOSAL 13
PROPOSAL
PROPO
NUTSINEE KIJBUNCHOO © 2017
RONALD DREVER
Don't tell me you've never heard of the POUND-DREVER-HALL TECHNIQUE...
ANTIMATTERWEBCOMICS.COM

Masses in the Stellar Graveyard



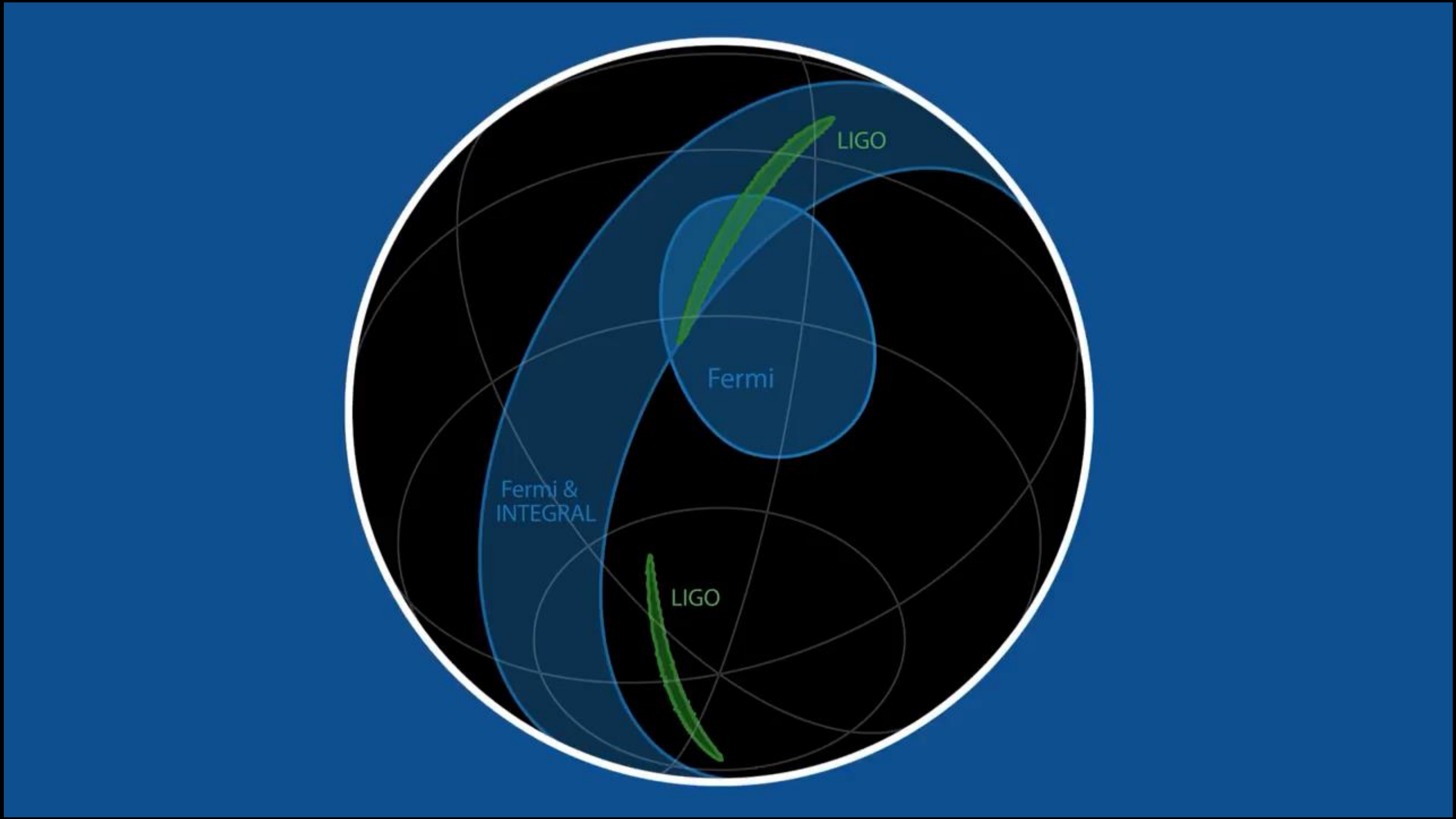
LIGO-Virgo-KAGRA | Aaron Geller | Northwestern



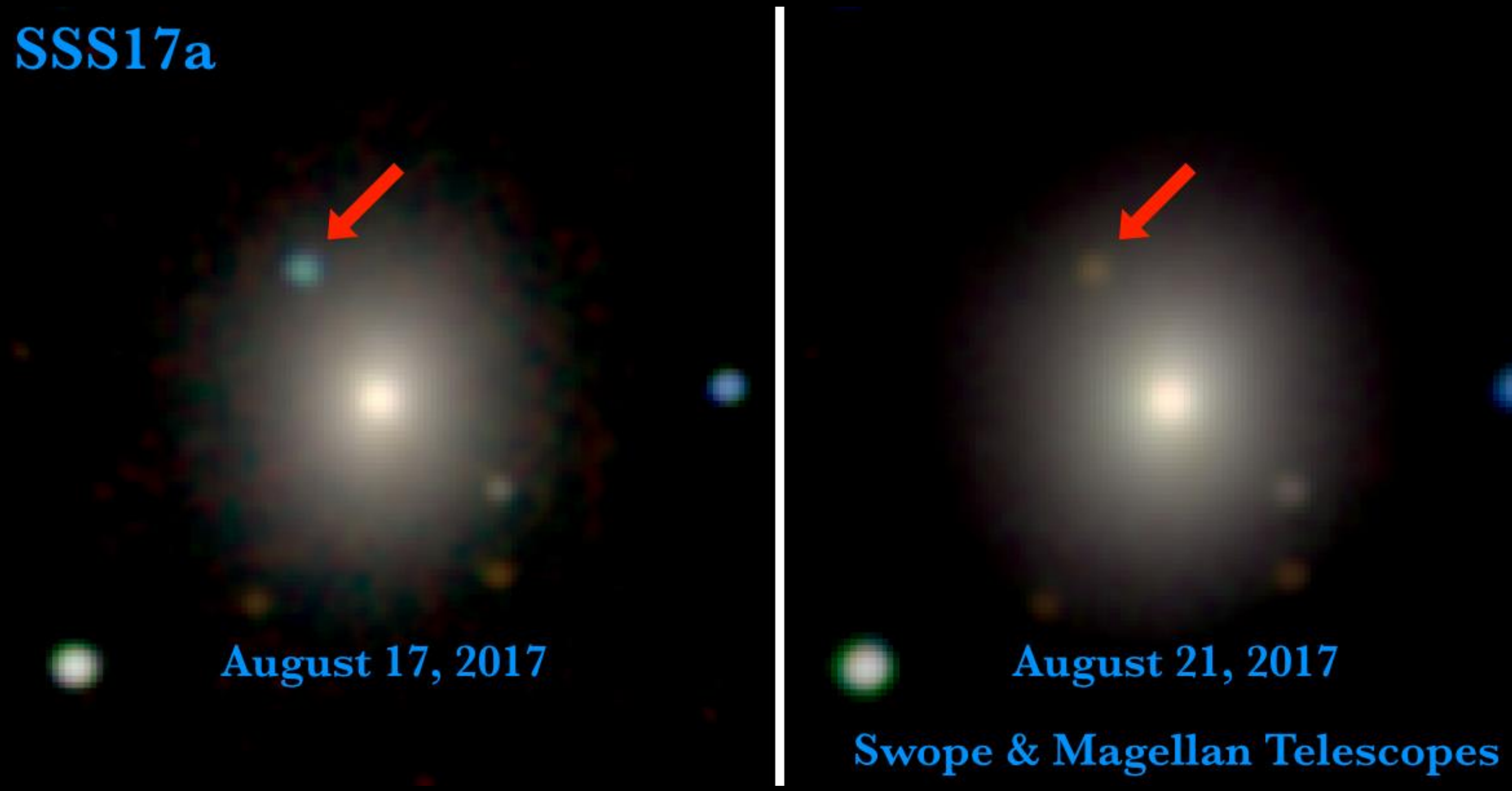
mage:LIGO/Caltech

Audio simulation: Scott Hughes, MI

Sky localization with GWs



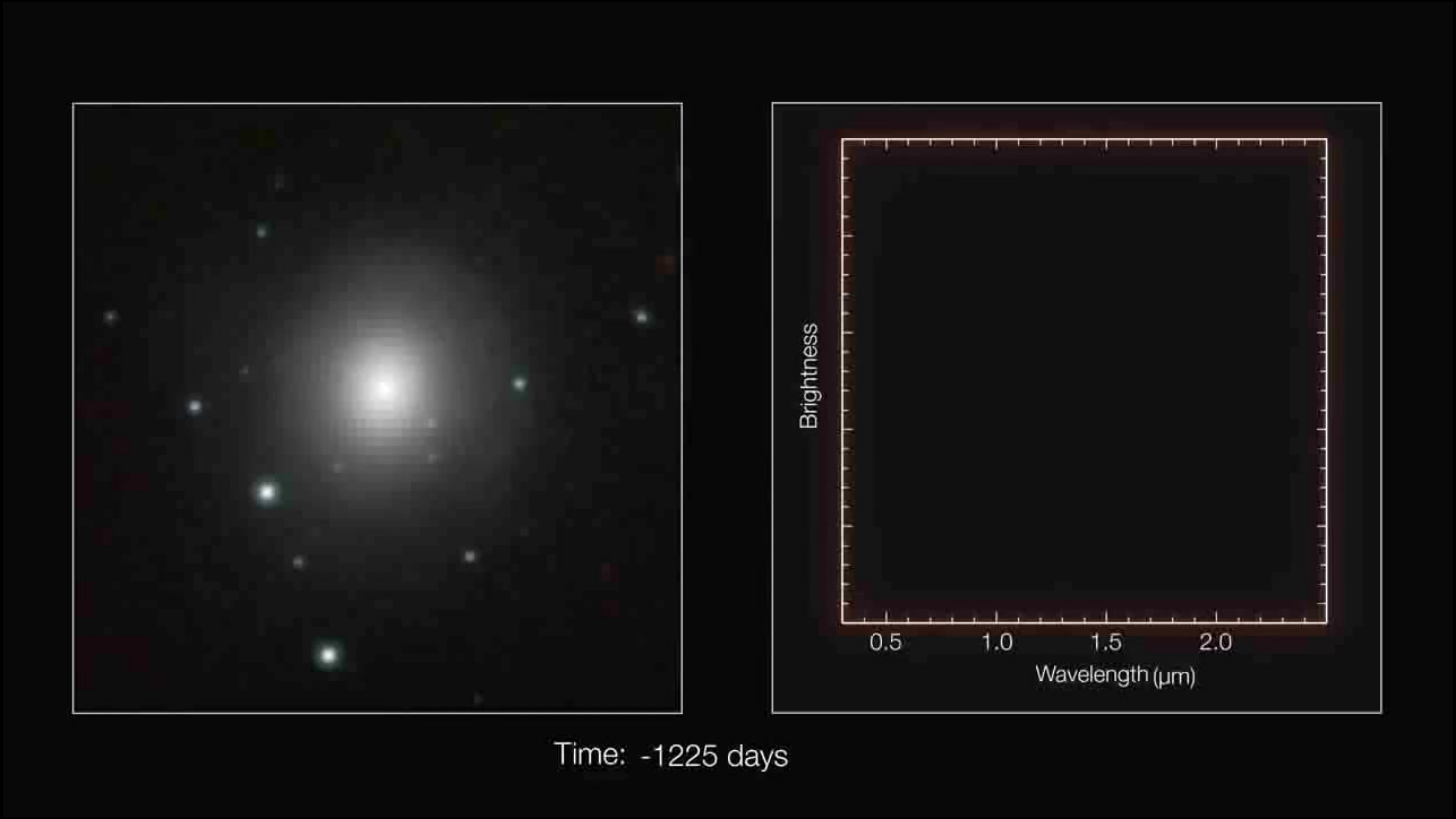
Discovery of an optical counterpart



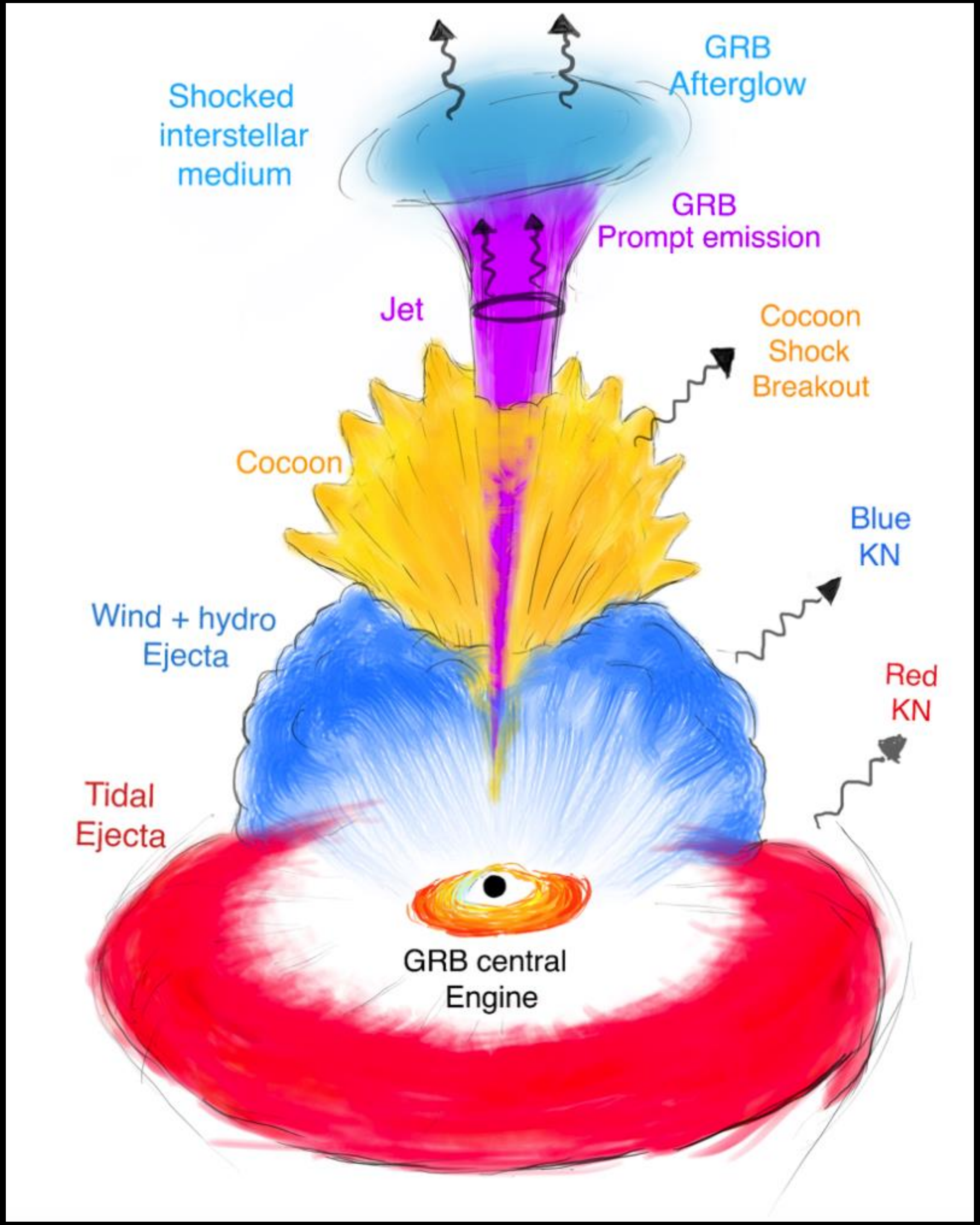
D. Coulter et al. 2017 arXiv 1710.05452

Image: 1M2H/UC Santa Cruz and Carnegie Observatories/Ryan Foley

The first multi-messenger discovery with GWs

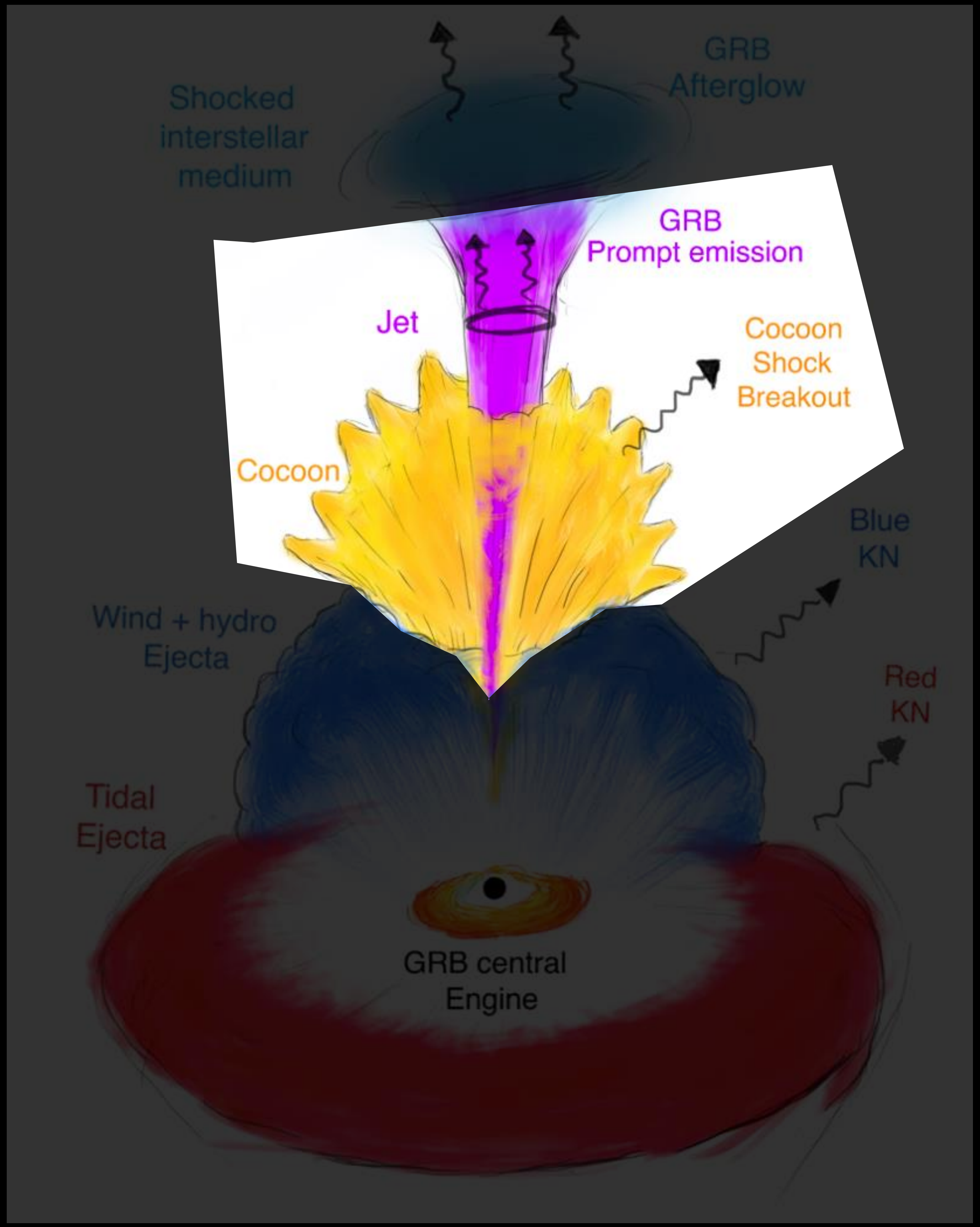


Compact Binary Coalescences



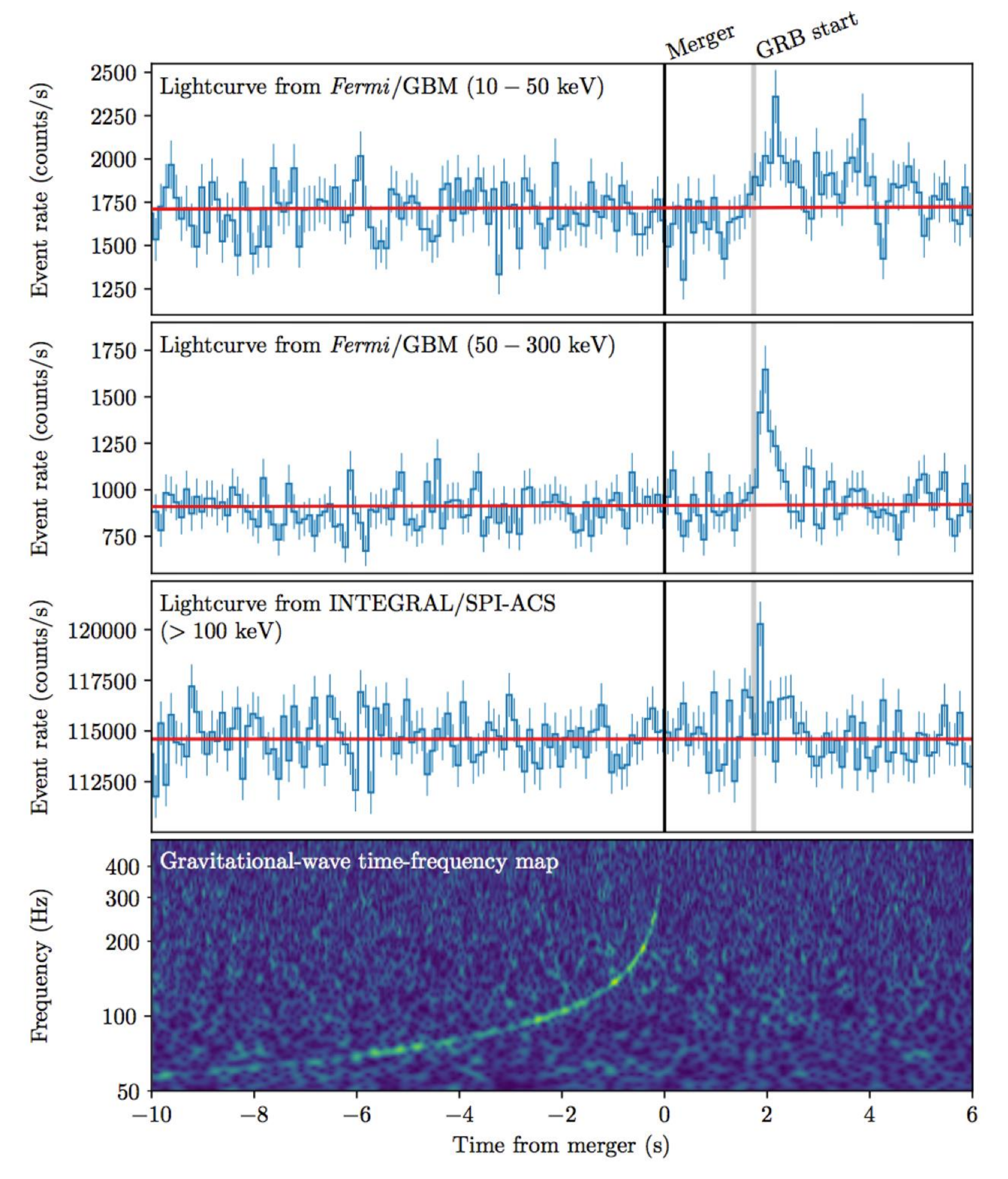
[Ascenzi+, arXiv:2011.04001]

Compact Binary Coalescences

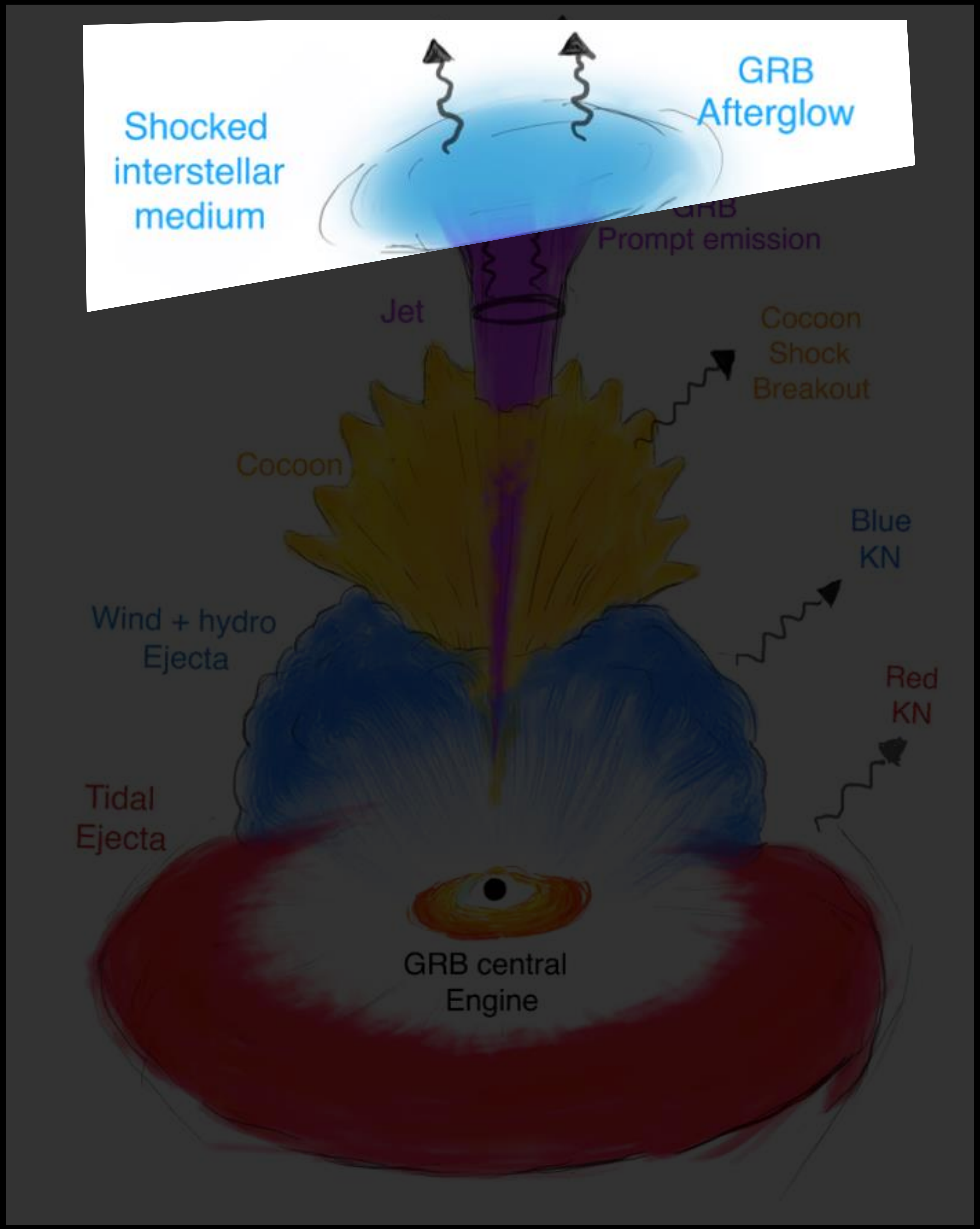


[Ascenzi+, arXiv:2011.04001]

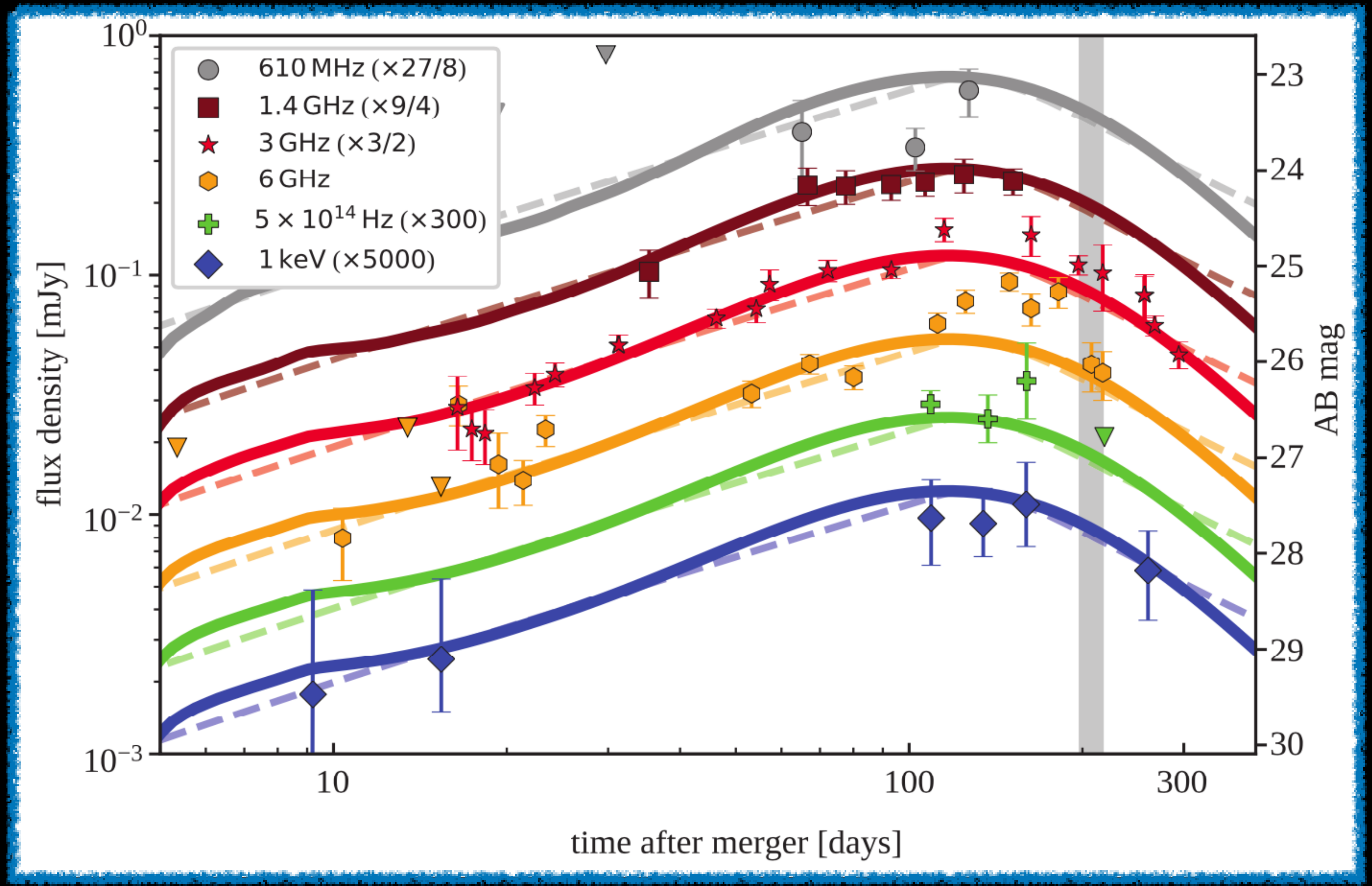
Gamma Ray Burst (GRB)



Compact Binary Coalescences

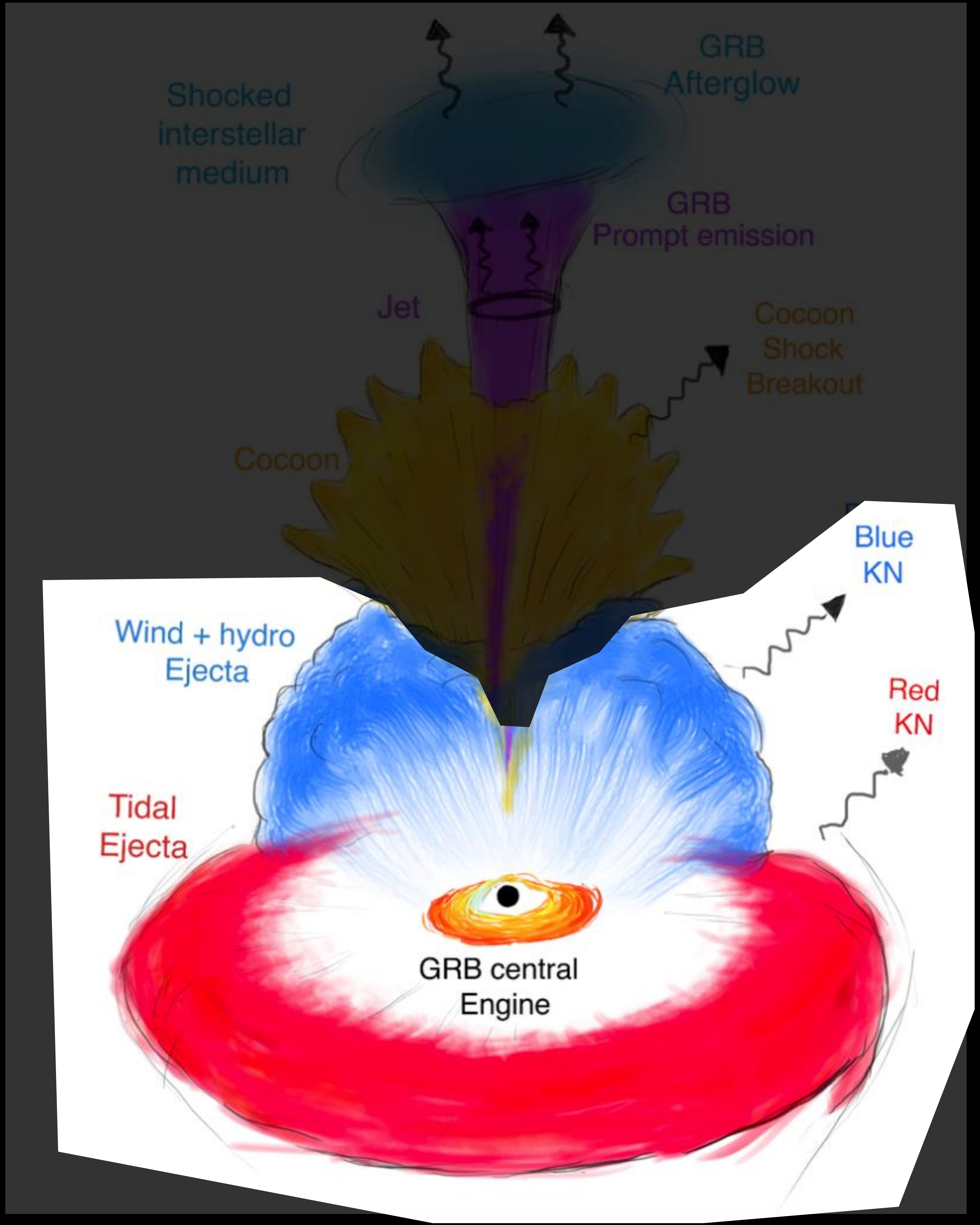


GRB Afterglow



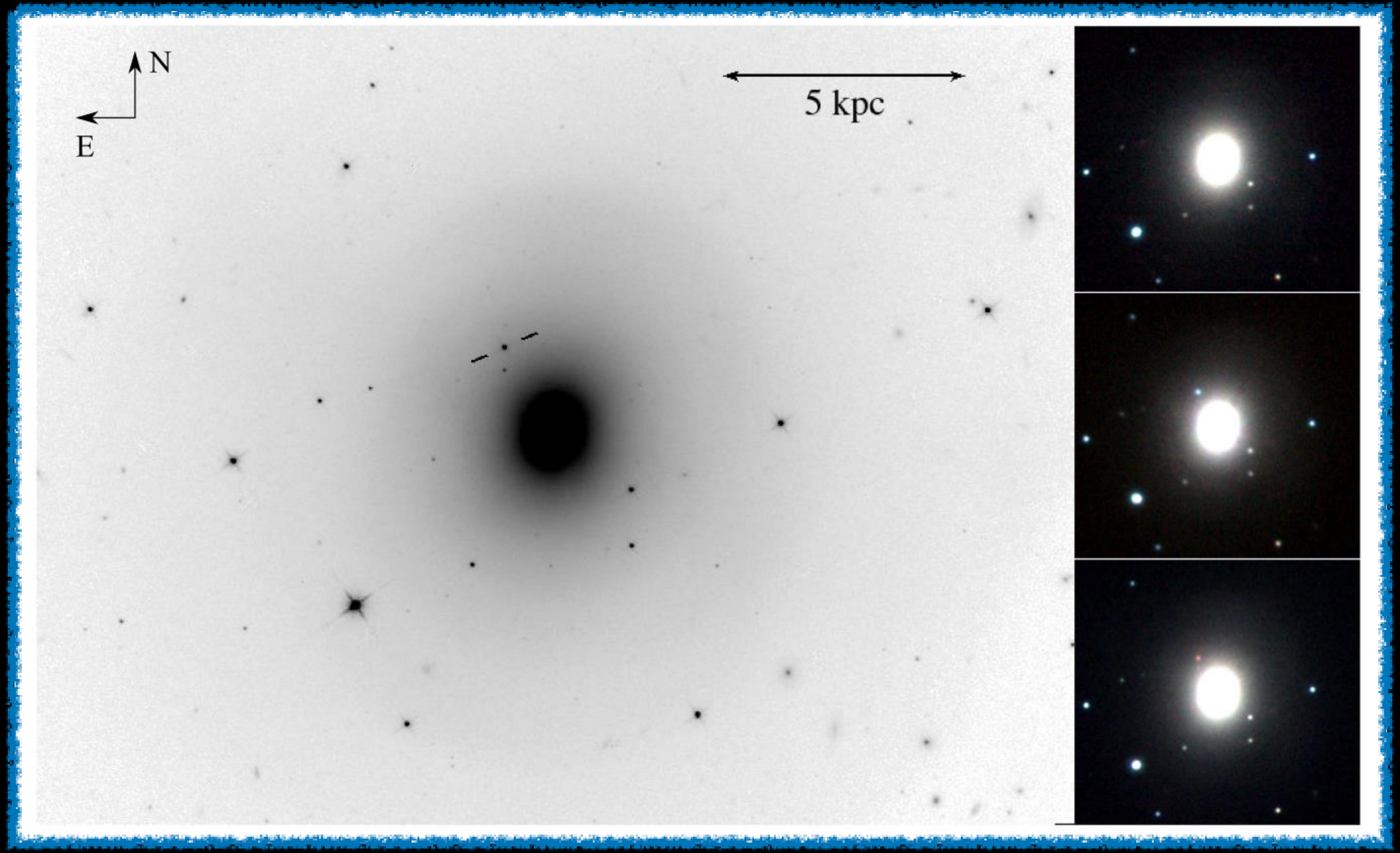
[Ascenzi+, arXiv:2011.04001]

Compact Binary Coalescences



[Ascenzi+, arXiv:2011.04001]

Kilonova

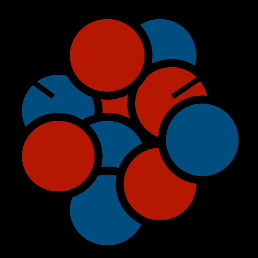


[Tanvir+2017, ApJ]

Kilonovae

“Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei”

[Metzger+2010, MNRAS]

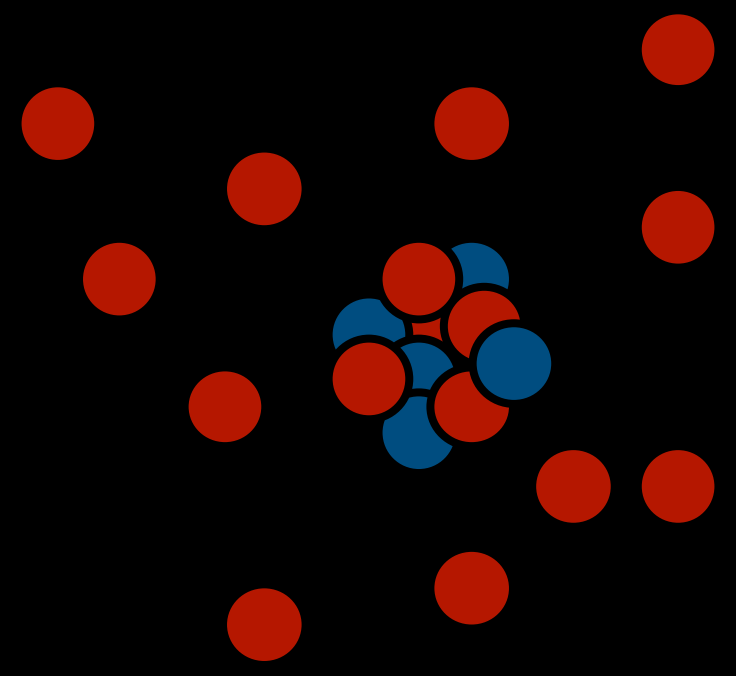


Seed nucleus

Kilonovae

“Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei”

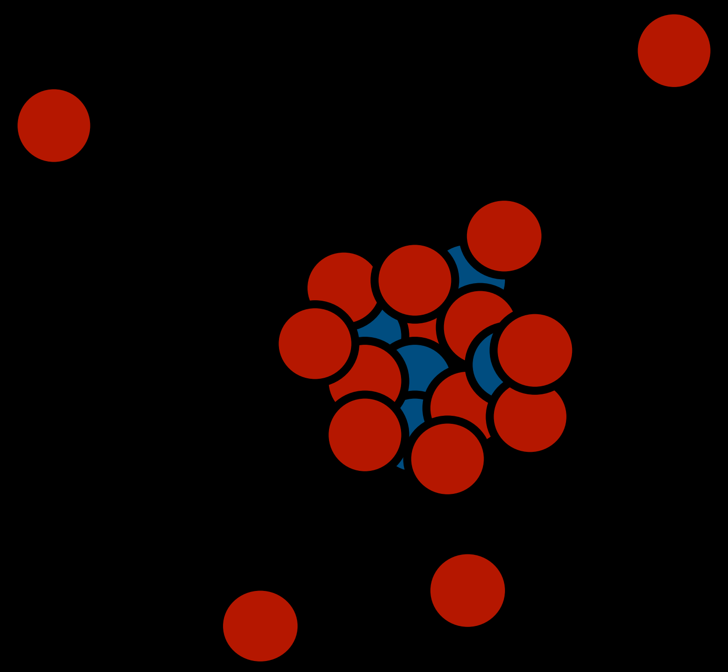
[Metzger+2010, MNRAS]



Kilonovae

“Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei”

[Metzger+2010, MNRAS]



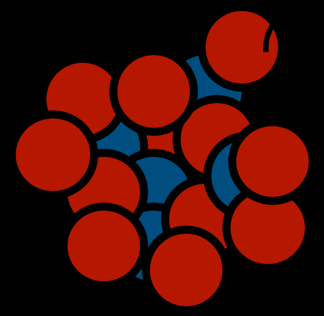
Neutron capture
via r-process

Kilonovae

“Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei”

[Metzger+2010, MNRAS]

—> Increasing wavelength



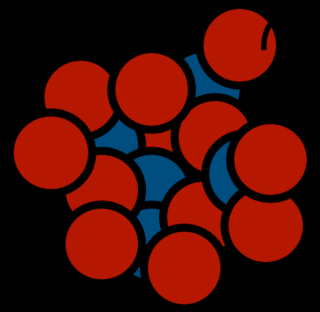
Unstable nuclei -> radioactive decay -> γ photons

Kilonovae

“Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei”

[Metzger+2010, MNRAS]

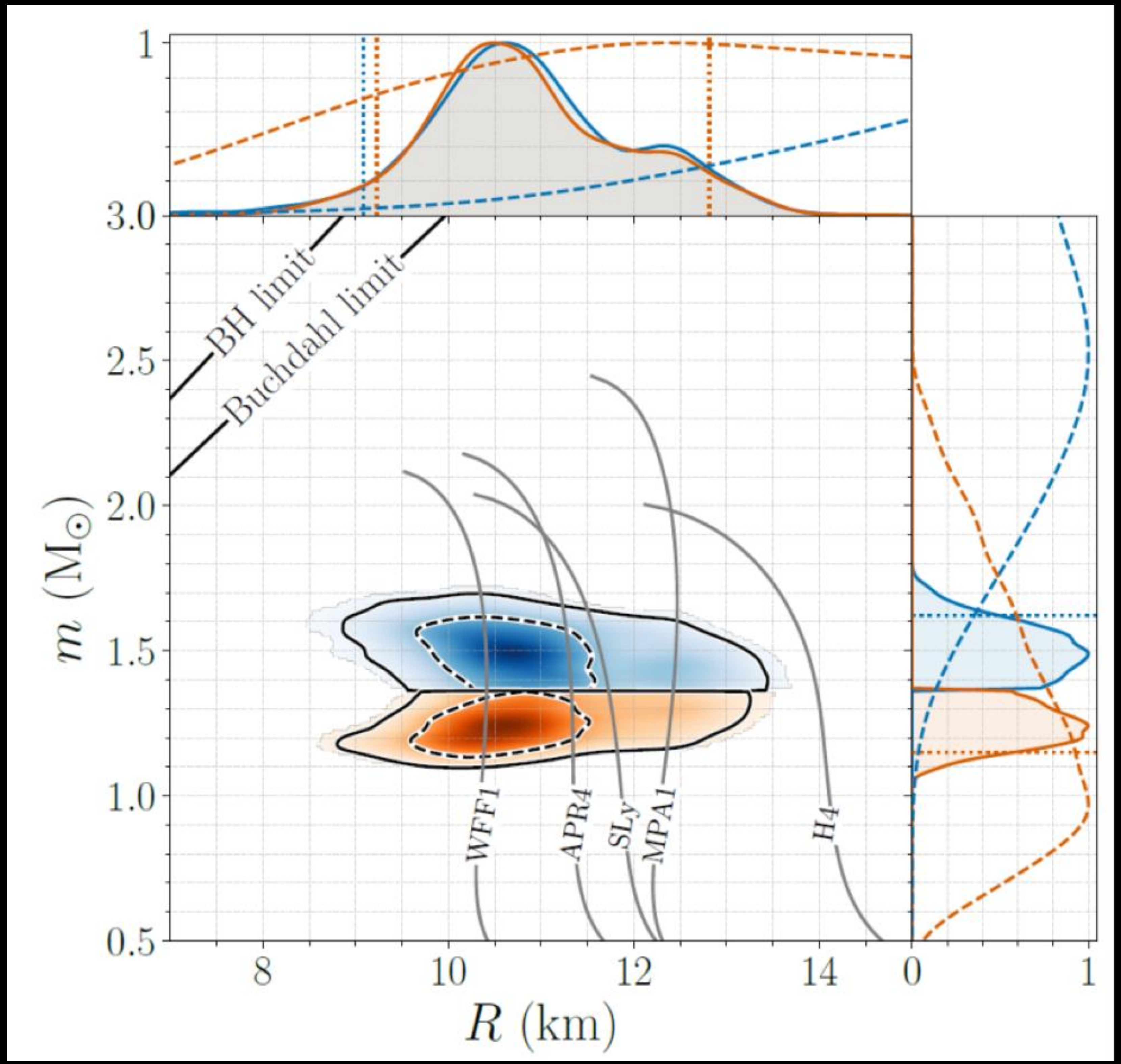
—> Increasing wavelength



T ~ 10 000 K @ 1 day

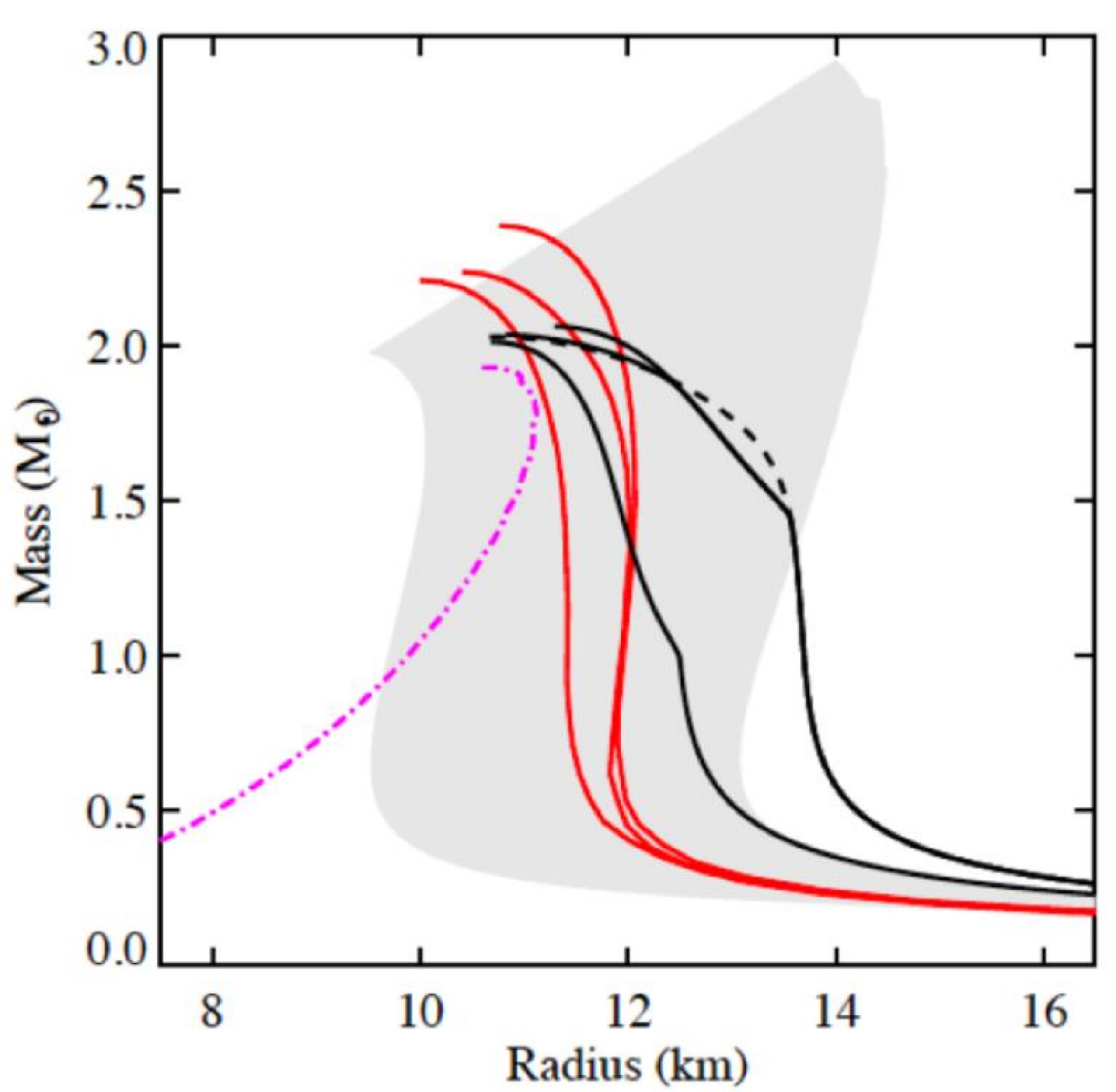
γ photons thermalize within the dense ejected material
->
setting the temperature

The neutron star equation of state



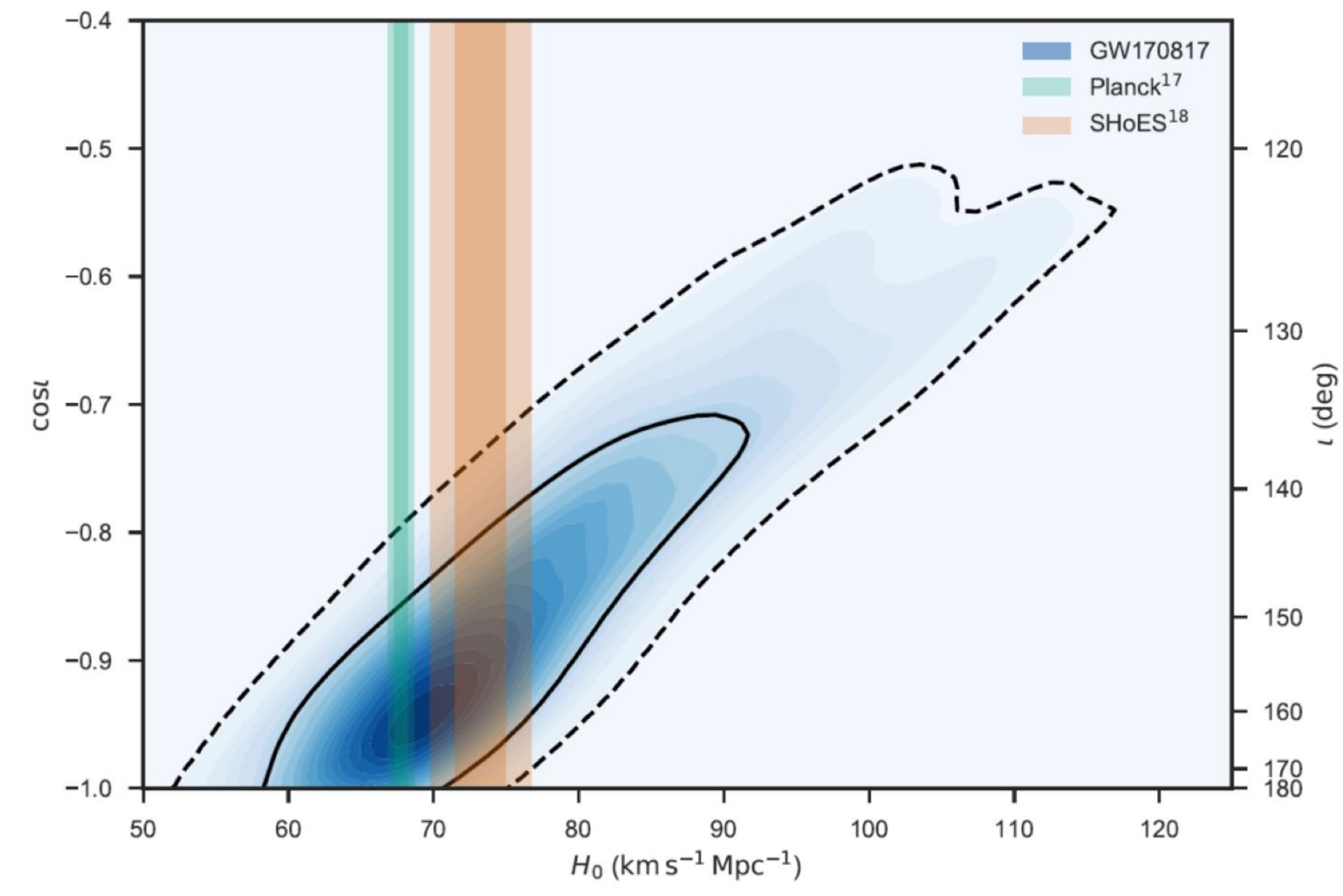
Described by the Equation of State:

- Pressure-density or equivalently mass-radius.
- Numerous models proposed.



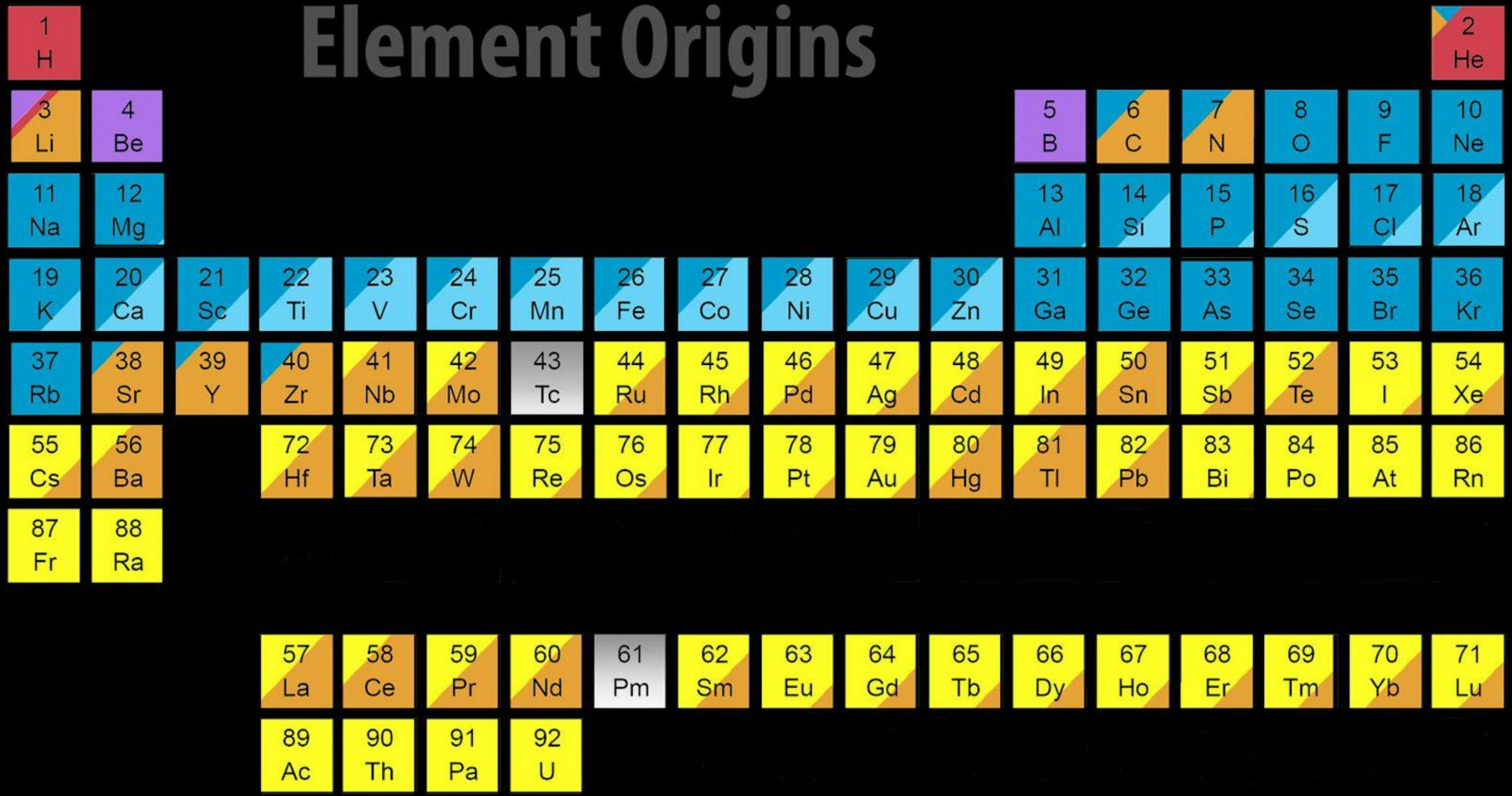
- Stiffness of NS determines how much it deforms under gravity (tidal effects).
- Estimate NS deformations from GW signals.
- Some EoS models are ruled out, others are consistent with observations.

Expansion rate of the Universe



- EM counterpart’s host galaxy helps estimate the redshift of the binary.
- GW signal provides a direct measure of the luminosity distance (modulo distance/inclination degeneracy).
- Independent of the cosmic distance ladder.

Element Origins



Merging Neutron Stars **Exploding Massive Stars** **Big Bang**

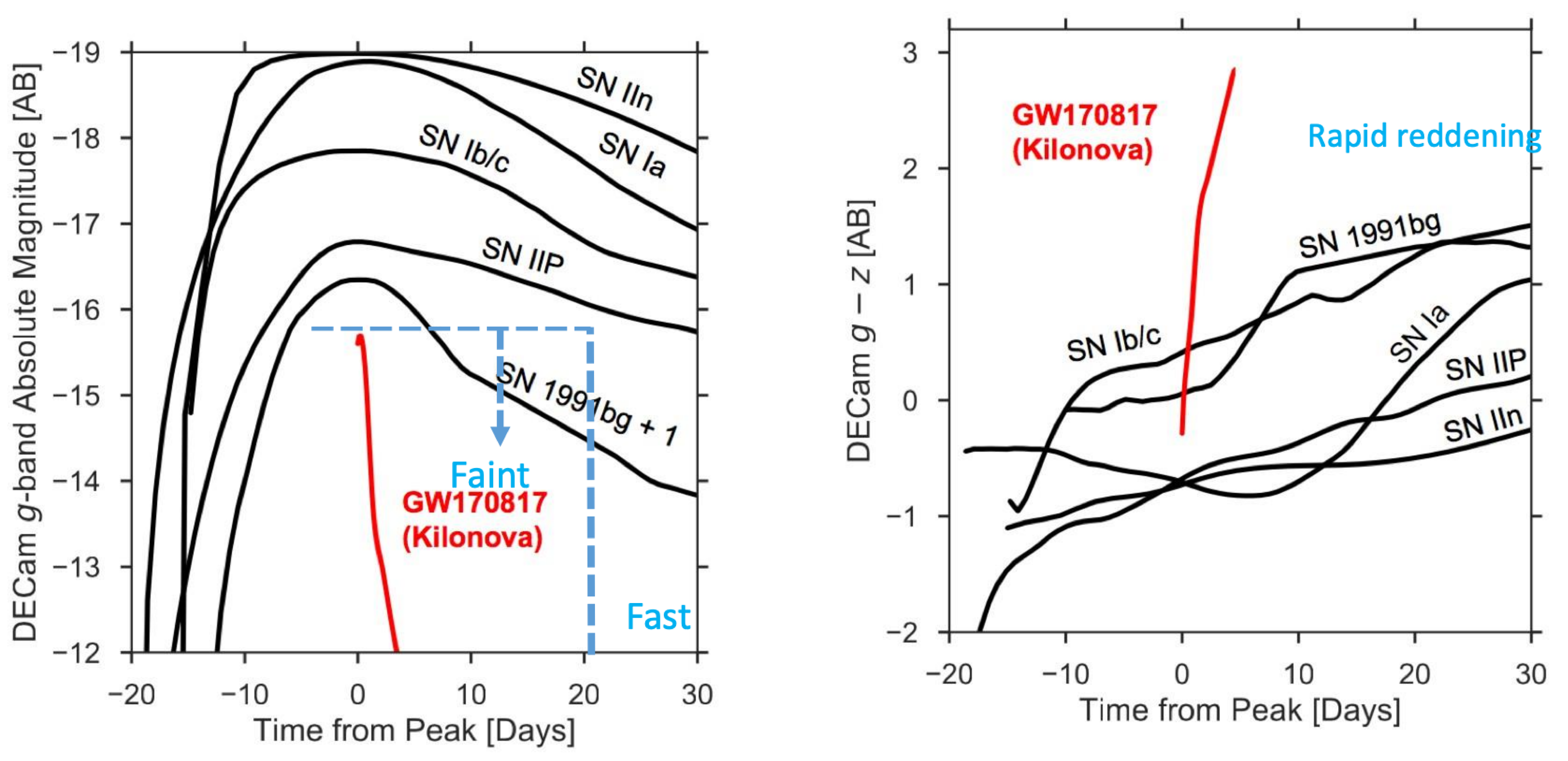
Dying Low Mass Stars **Exploding White Dwarfs** **Cosmic Ray Fission**

Based on graphic created by Jennifer Johnson

So what will it take to find more?

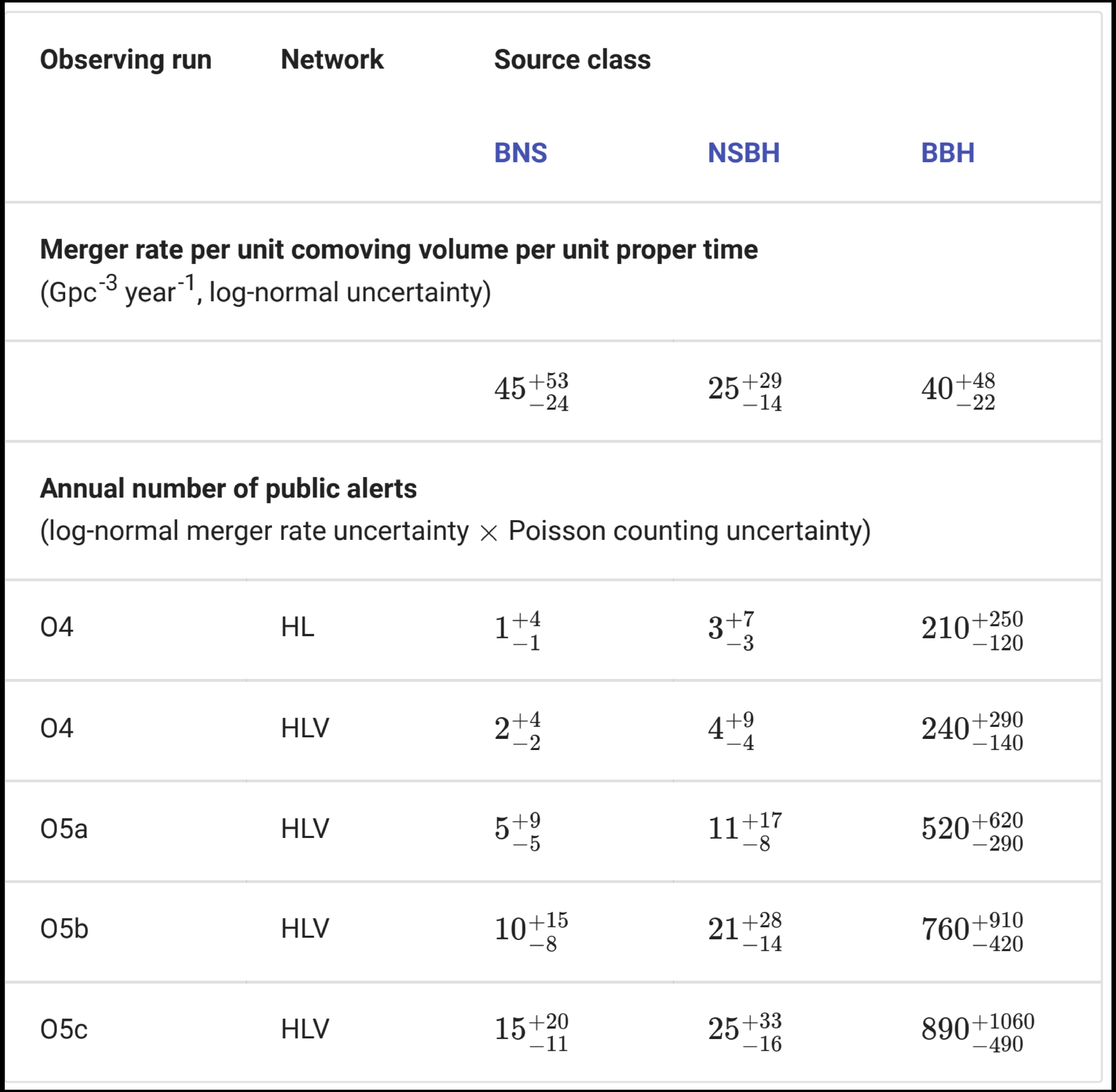
What’s challenging about this?

Kilonovae



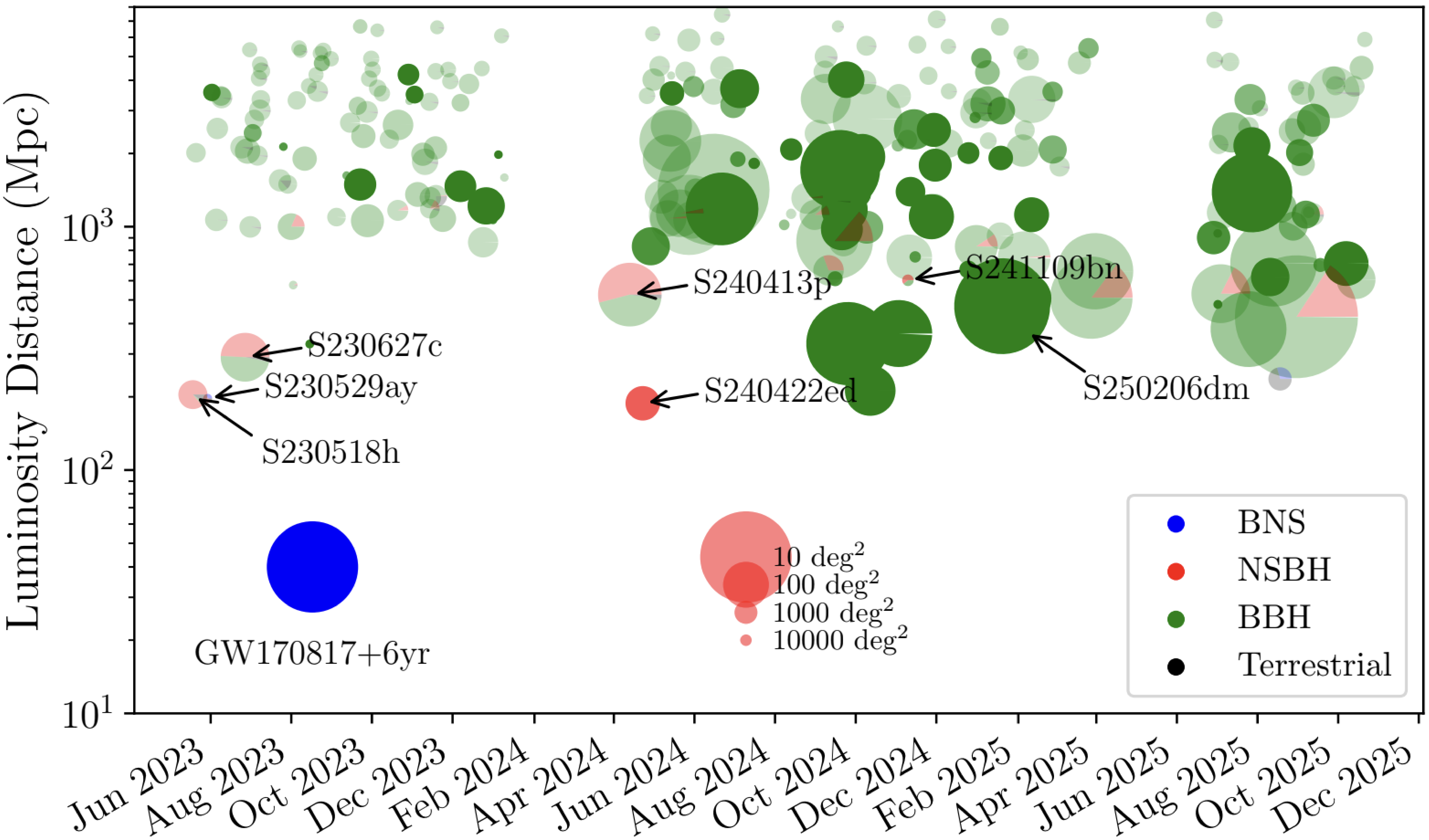
modified from Andreoni+2018, LSST White Paper

Compact Binary Merger Rates



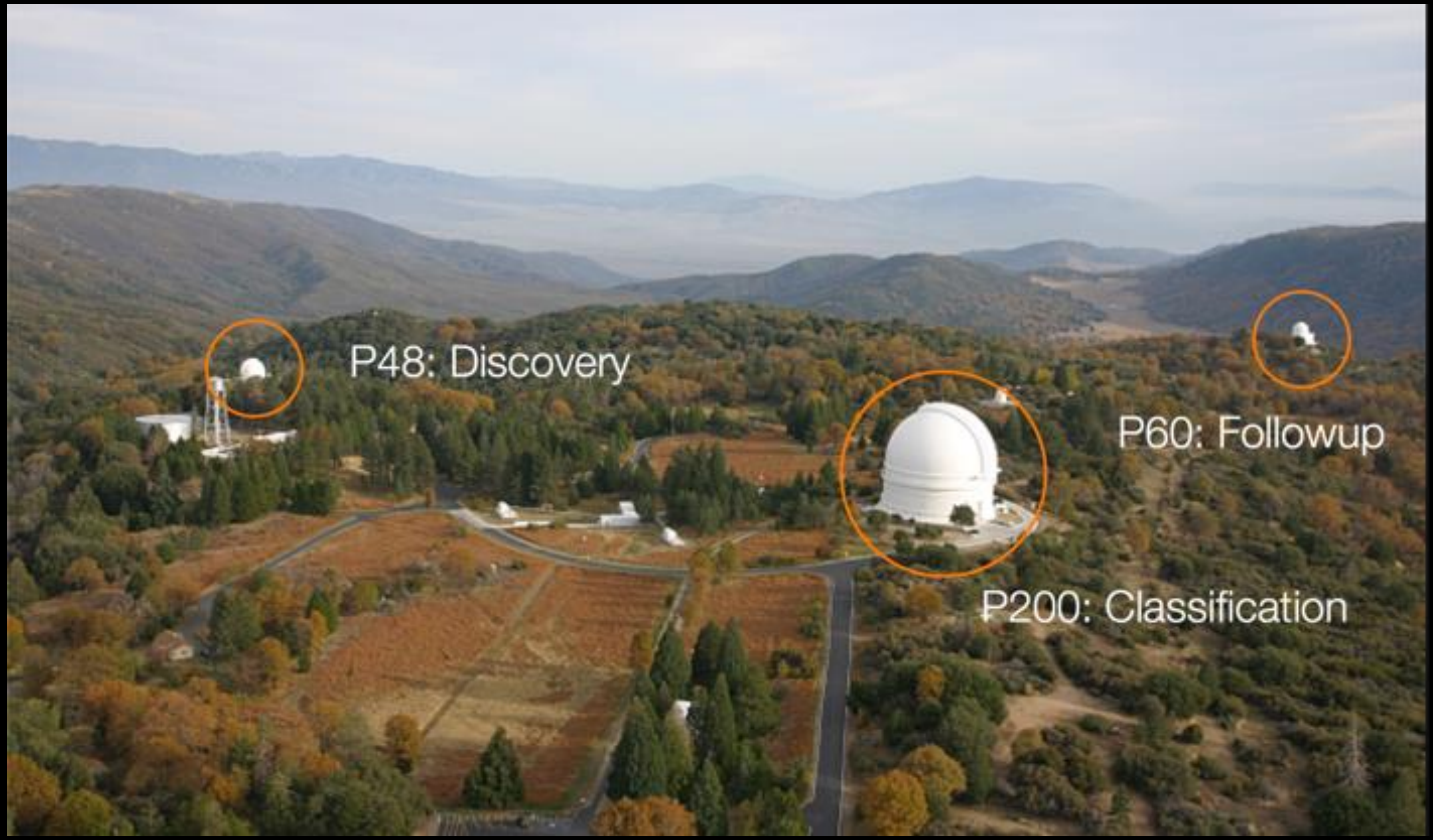
Observing run	Network	Source class BNS	 NSBH	 BBH
Merger rate per unit comoving volume per unit proper time (Gpc^{-3} year^{-1}, log-normal uncertainty)				
		45^{+53}_{-24}	25^{+29}_{-14}	40^{+48}_{-22}
Annual number of public alerts (log-normal merger rate uncertainty × Poisson counting uncertainty)				
O4	HL	1^{+4}_{-1}	3^{+7}_{-3}	210^{+250}_{-120}
O4	HLV	2^{+4}_{-2}	4^{+9}_{-4}	240^{+290}_{-140}
O5a	HLV	5^{+9}_{-5}	11^{+17}_{-8}	520^{+620}_{-290}
O5b	HLV	10^{+15}_{-8}	21^{+28}_{-14}	760^{+910}_{-420}
O5c	HLV	15^{+20}_{-11}	25^{+33}_{-16}	890^{+1060}_{-490}

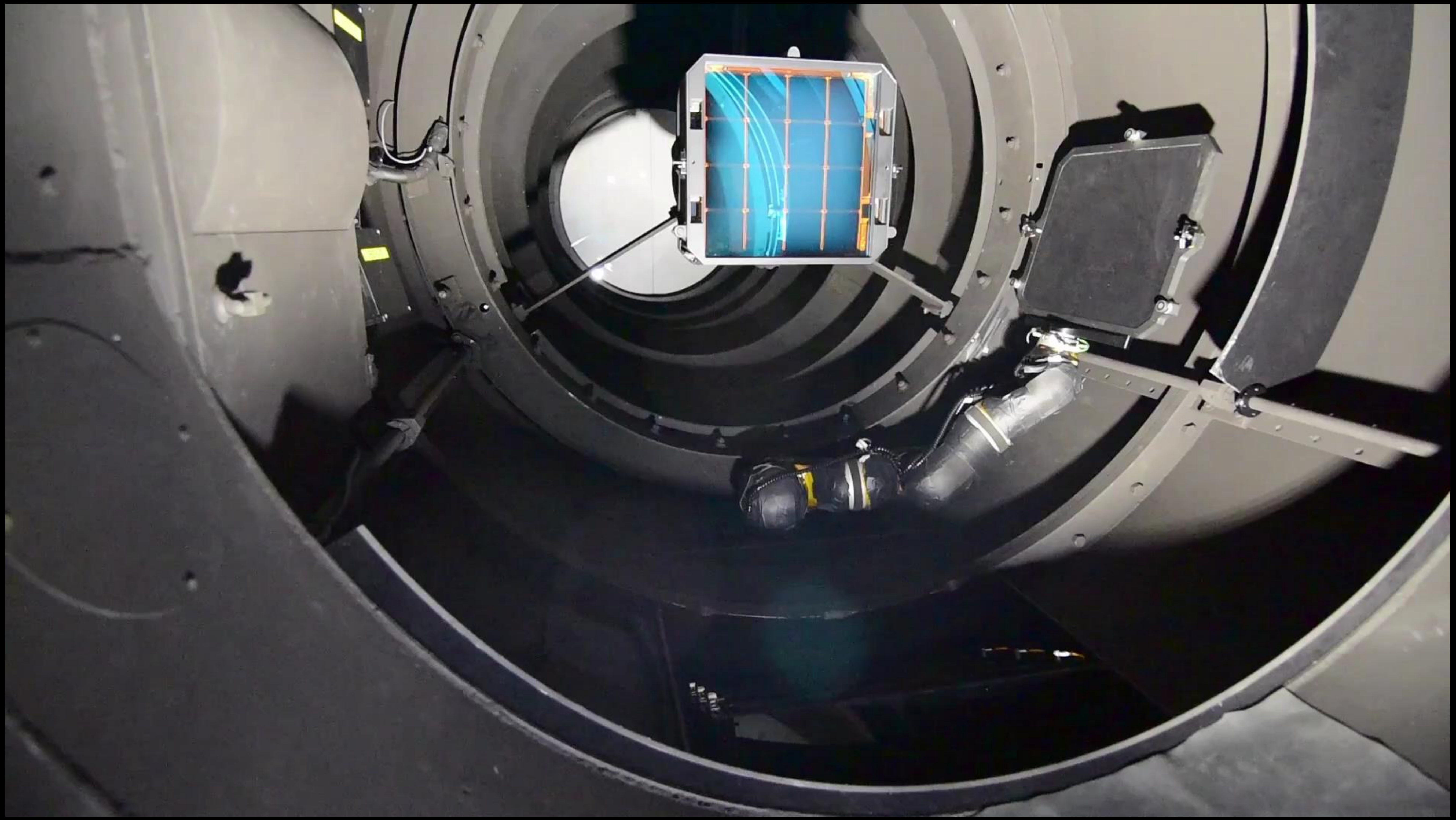
IGWN User Guide


Luminosity Distance (Mpc)
10^3
10^2
10^1
S240413p
S230627c
S230529ay
S230518h
S240422ed
S241109bn
S250206dm
GW170817+6yr
10 deg^2
100 deg^2
1000 deg^2
10000 deg^2
BNS
NSBH
BBH
Terrestrial
Jun 2023
Aug 2023
Oct 2023
Dec 2023
Feb 2024
Apr 2024
Jun 2024
Aug 2024
Oct 2024
Dec 2024
Feb 2025
Apr 2025
Jun 2025
Aug 2025
Oct 2025
Dec 2025

What’s the status quo for follow-up?

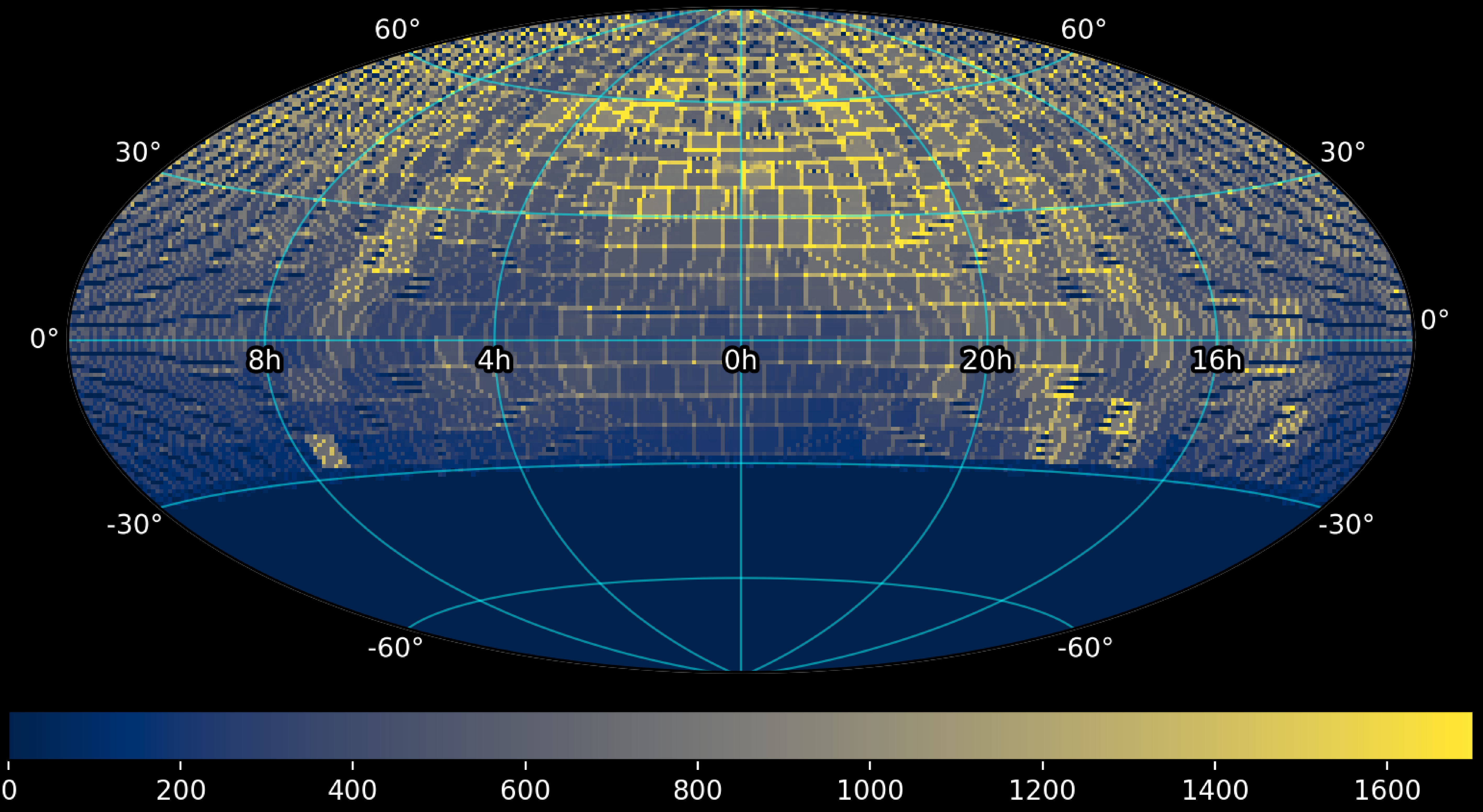
The Zwicky Transient Facility





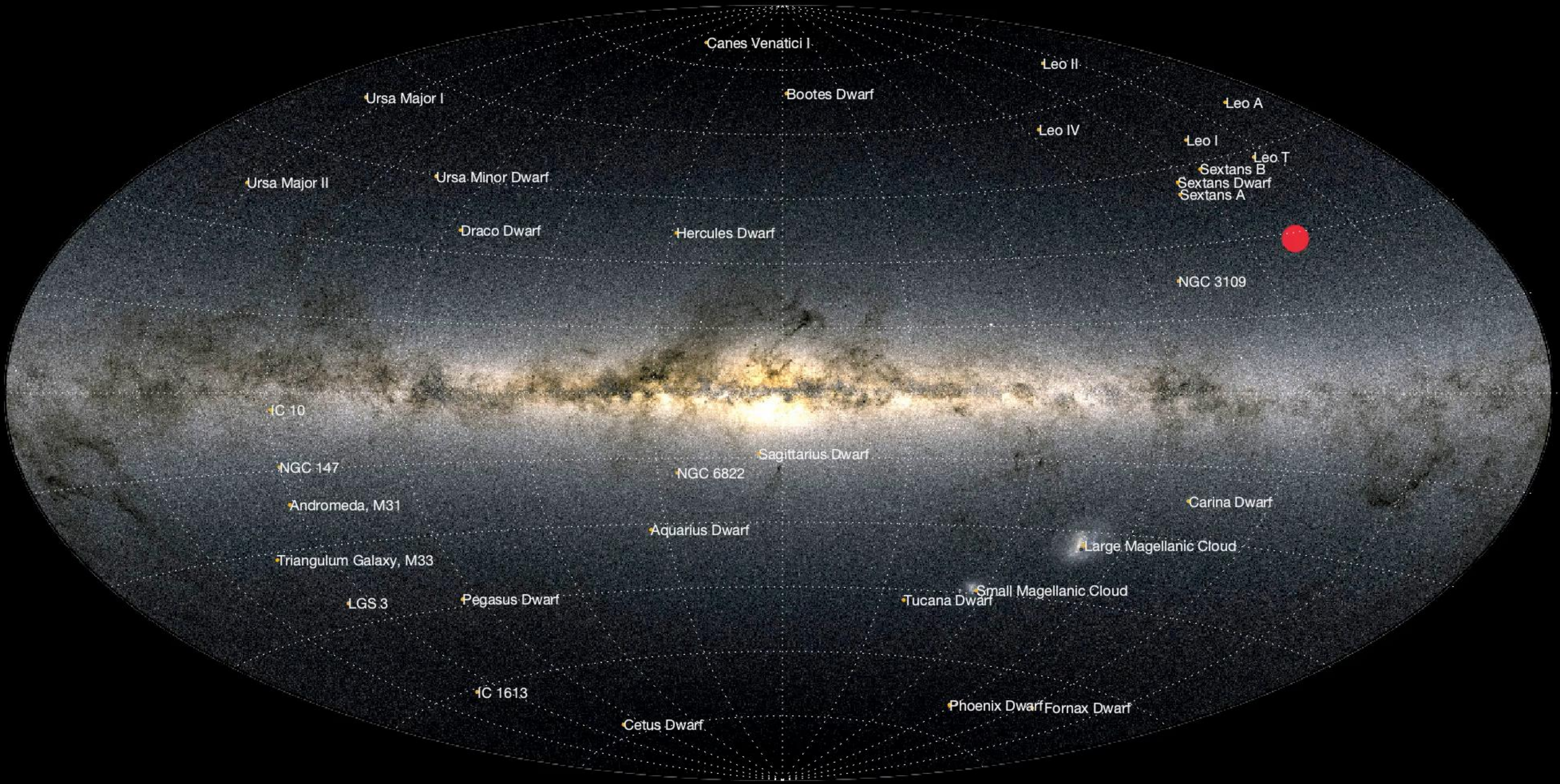
The Zwicky Transient Facility

- 1.4 TB (compressed) of image data per night
- 6 PB over the past six years
- More than 1 million exposures taken – over 1 year of open shutter time
- Total sky area covered is 47 million deg2 – 1.5 – 2x LSST 10-year total
- >787 billion photometric measurements for over 4.72 billion sources
- Over 1 billion sources have more than 50 data points in g and r
- Up to 1 million transient alerts per night
- Over 700 million alerts (56 TB) published



Bright Transient Survey

Supernovae Classified: **1** Date: 21-Feb-2017

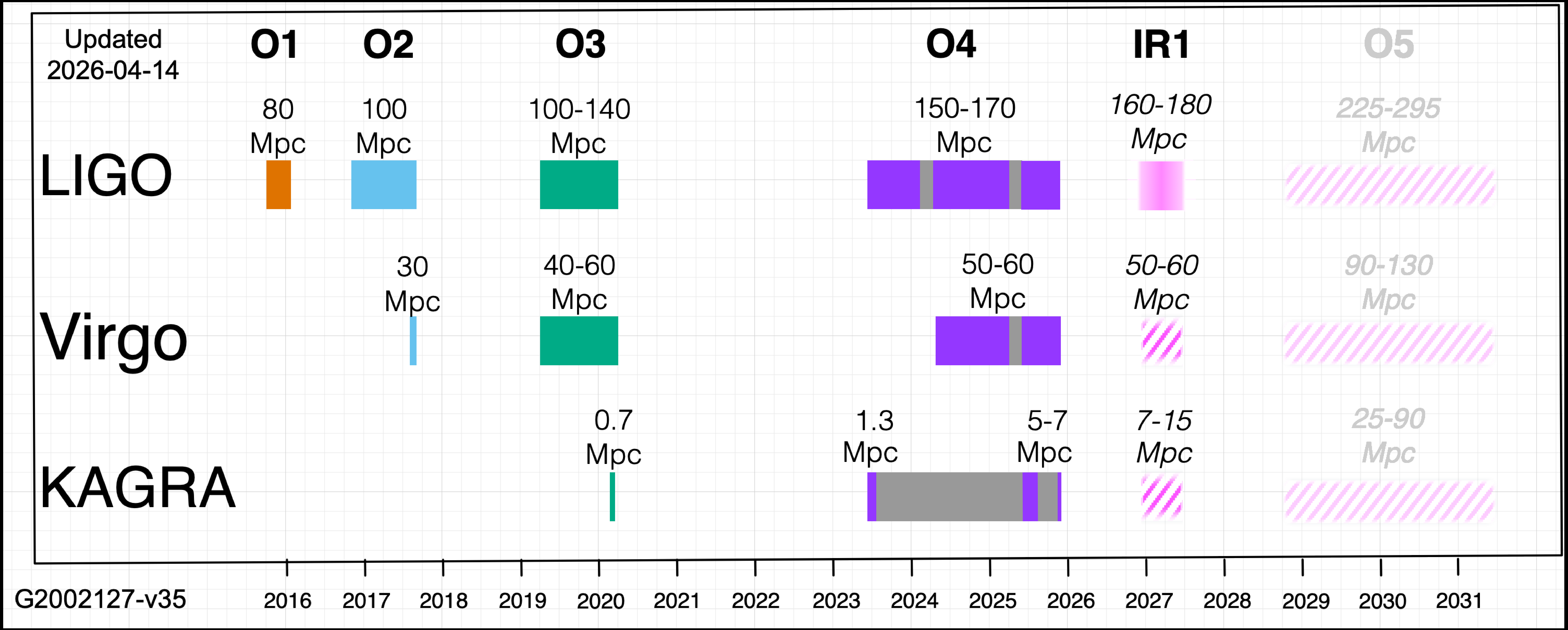


● Exploding White Dwarfs ● Collapsing Massive Stars ● Superluminous Supernovae

Underlying image credit: ESA/Gaia/DPAC

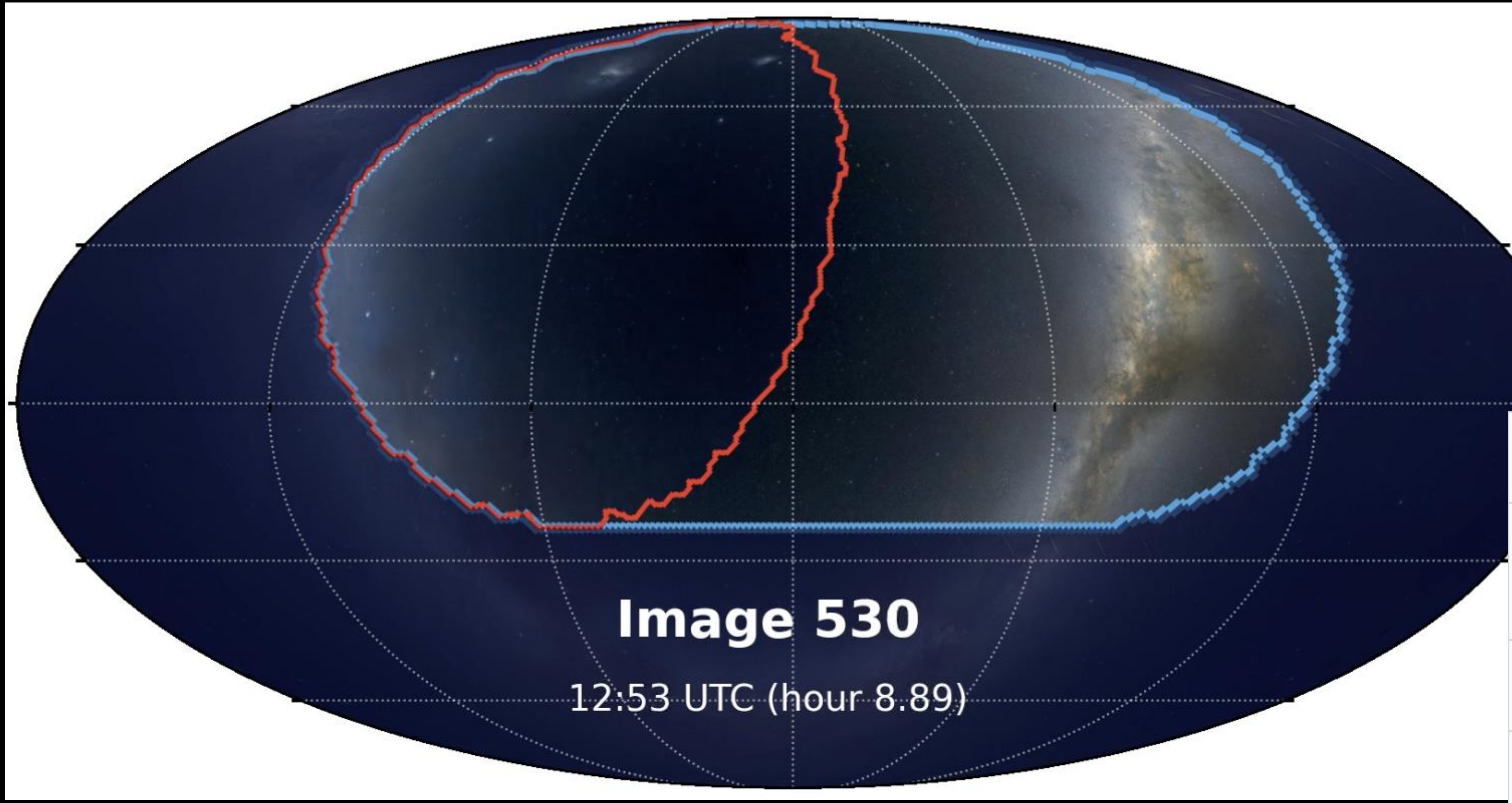
Caltech

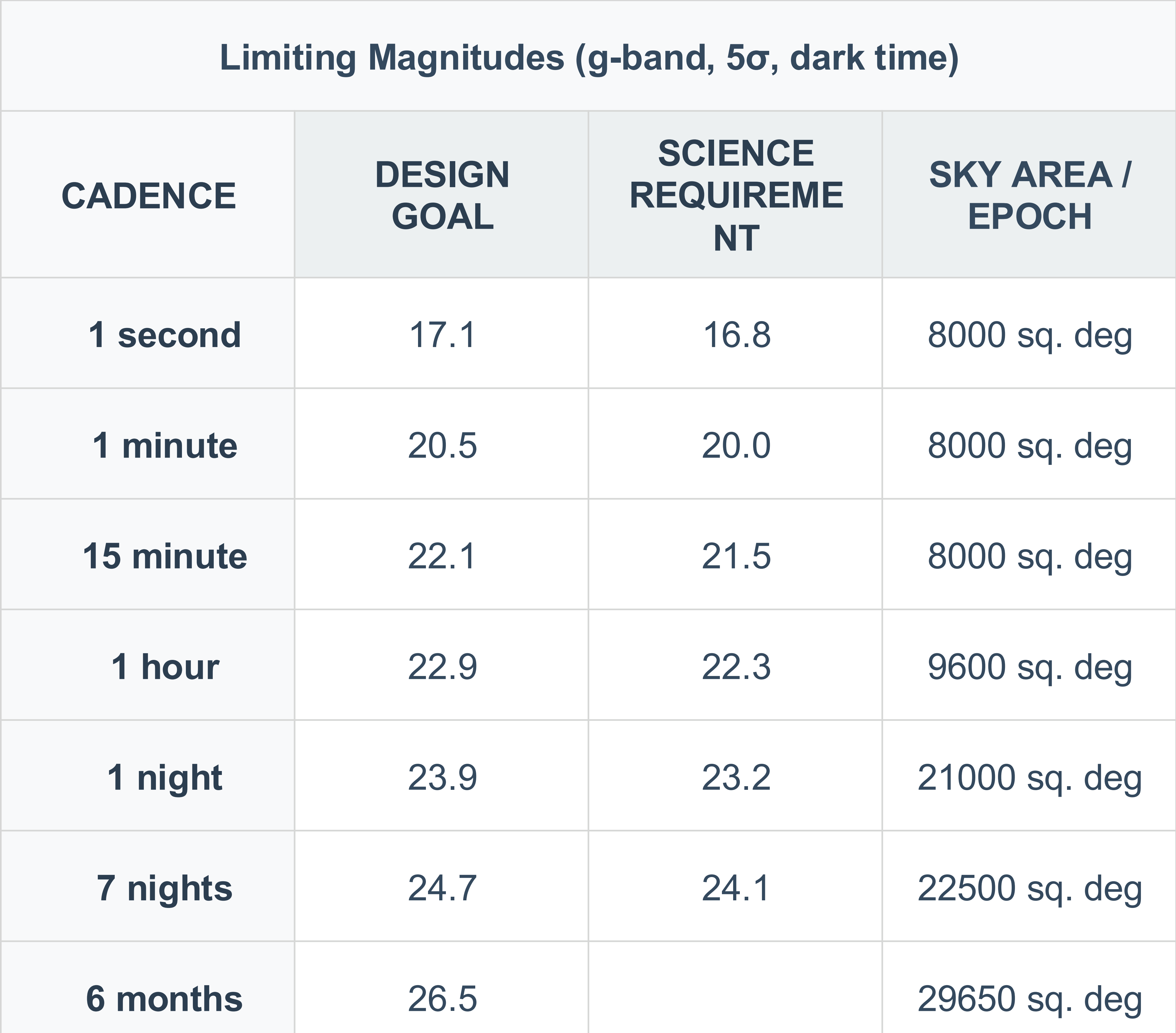
What’s the future for instrumentation?


Updated
2026-04-14
O1
O2
O3
O4
IR1
O5
80 Mpc
100 Mpc
100-140 Mpc
150-170 Mpc
160-180 Mpc
225-295 Mpc
LIGO
30 Mpc
40-60 Mpc
50-60 Mpc
50-60 Mpc
90-130 Mpc
Virgo
0.7 Mpc
1.3 Mpc
5-7 Mpc
7-15 Mpc
25-90 Mpc
KAGRA
G2002127-v35
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031

Rubin Observatory

Argus Array





Limiting Magnitudes (g-band, 5σ, dark time)

CADENCE	DESIGN GOAL	SCIENCE REQUIREMENT	SKY AREA / EPOCH
1 second	17.1	16.8	8000 sq. deg
1 minute	20.5	20.0	8000 sq. deg
15 minute	22.1	21.5	8000 sq. deg
1 hour	22.9	22.3	9600 sq. deg
1 night	23.9	23.2	21000 sq. deg
7 nights	24.7	24.1	22500 sq. deg
6 months	26.5		29650 sq. deg

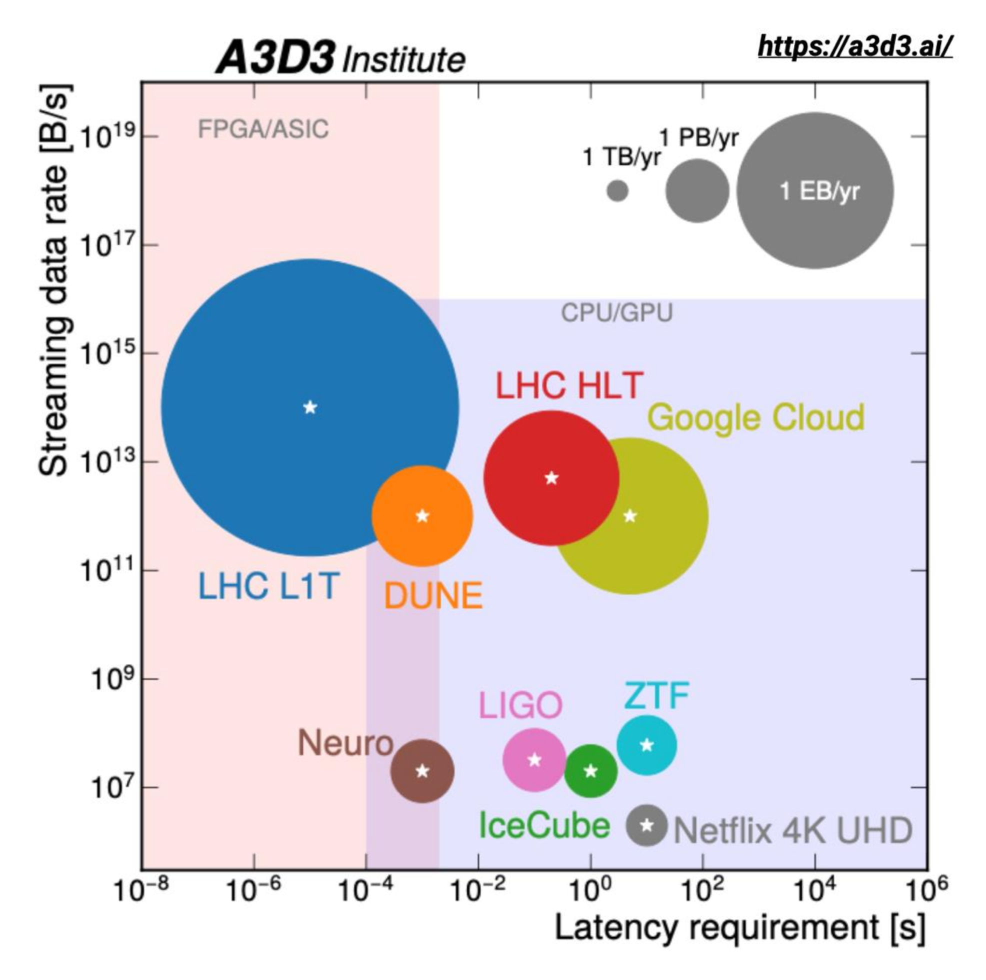
However, with these amazing instruments, another challenge arises…

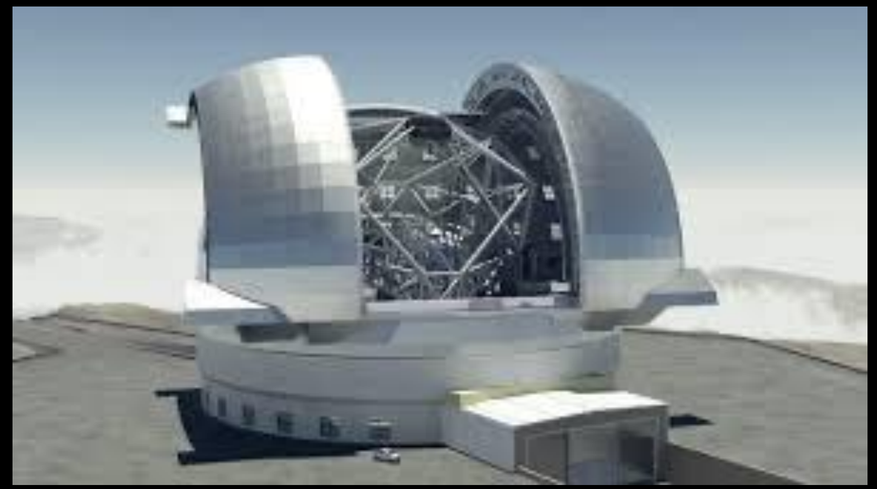
The next generation of surveys and facilities

- LSST (2025): ~20 TB/day => 10 PB/yr
- CSST (2026): ~TBs/day => 10s PB/yr
- Roman Space Telescope (2027): 20 TB/day => 7.3 PB/yr
- ngVLA (2030s): 20 TB/day => 7.3 PB/yr
- SKA (2030s): ~1 PB/day => 300 PB/yr
- ELT (~2030): ~PBs/yr
- DSA-2000 (2028): 3.5 PB/day => 1.3 EB/yr (26 LHCs)
- DUNE (2028): => ~TBs/s => 1.8 EB/yr

For comparison:

- HL-LHC (2029): 700 TB/s => 1 EB/yr
- Facebook: 300 PB of data



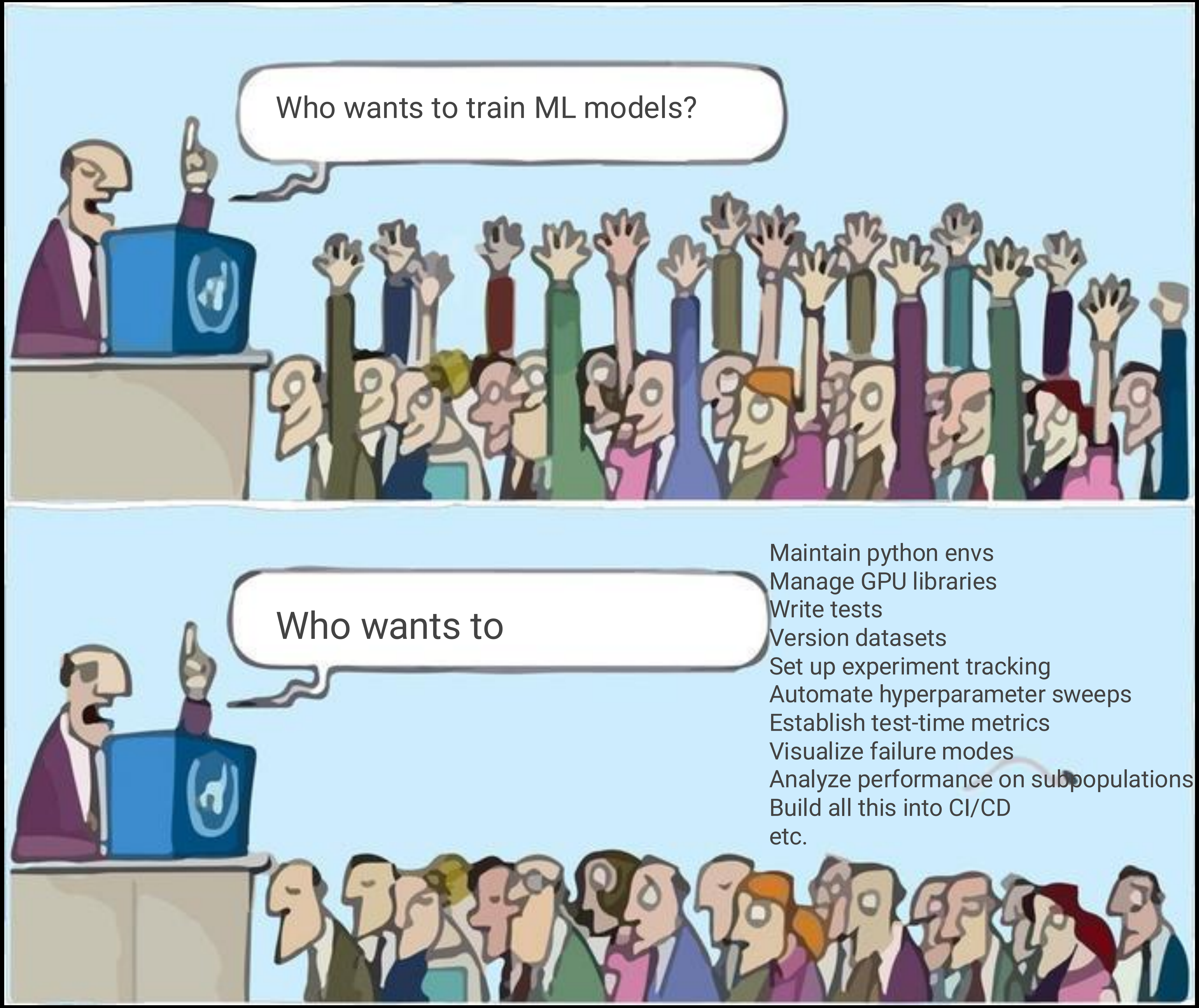


Is data science the answer?

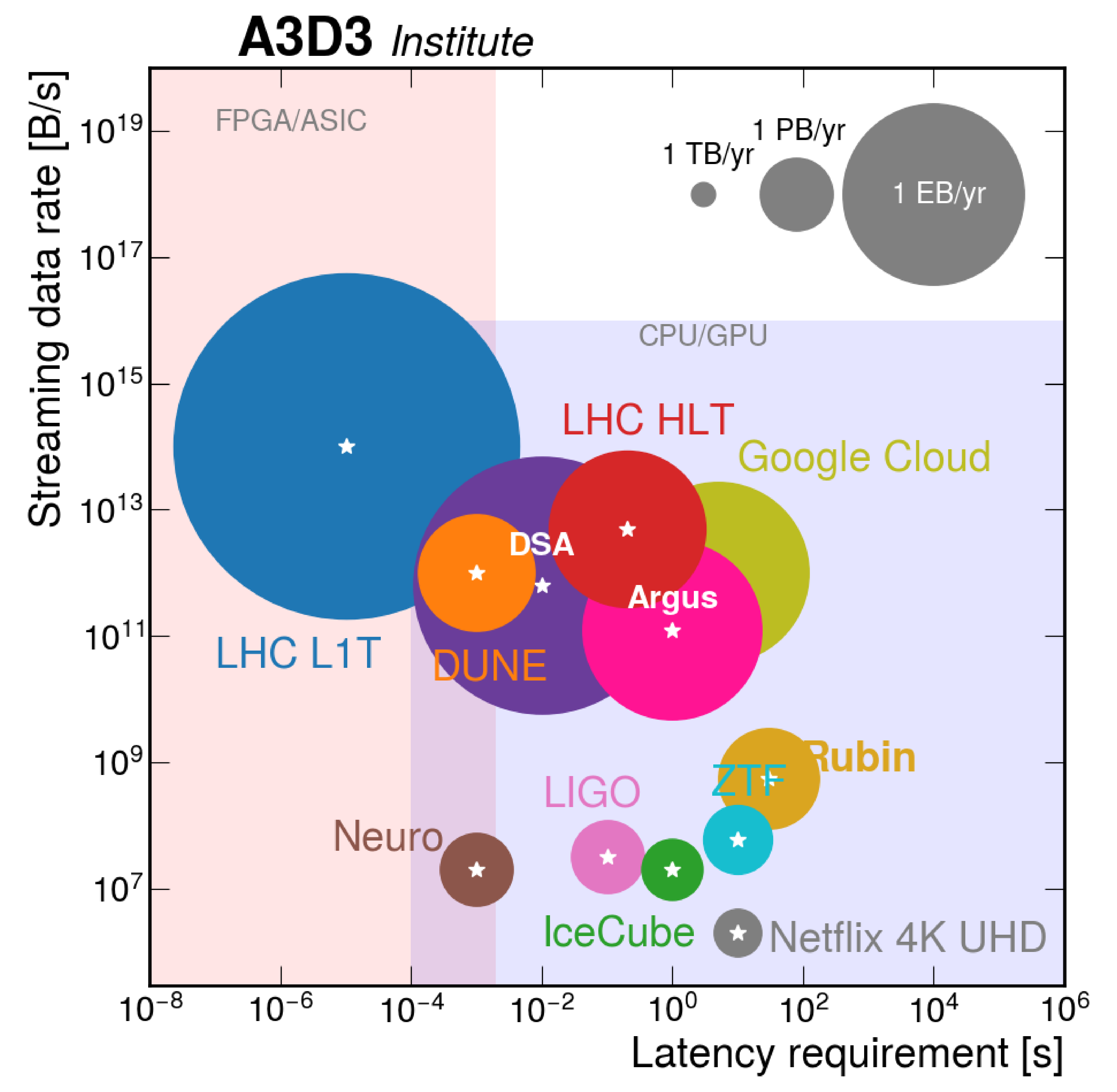
A future of ML in astronomy

- TBs to PBs of data per night produced by facilities with fast low-latency high throughput inferencing models (embedded ML) driving control and decision systems
- Information extracted (optimized representations) and followup decisions made according to a teleological learnt strategy
- Patterns and relationships identified and put into context with other
- If science is defined by continuous differentiable relationships then automated discovery becomes increasingly more effective
- => “You will wake up and your smartphone will tell explain to you what it discovered last night”

Are we having fun yet?



Accelerated Artificial Intelligence Algorithms for Data Driven Discovery



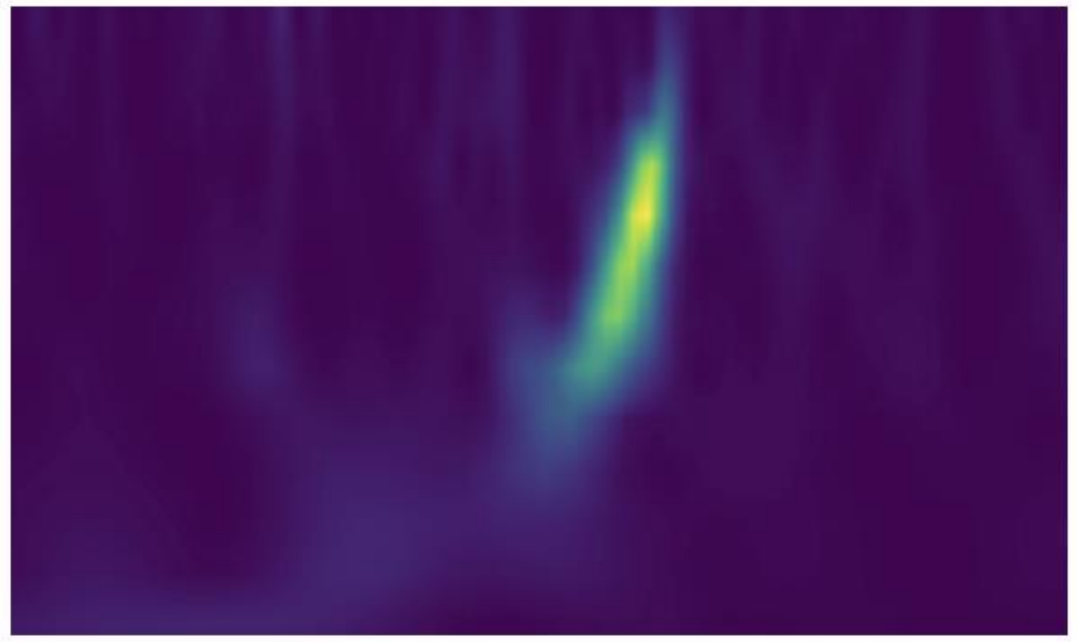
- Integration of three core components: domain scientists, computer scientists and hardware engineers
- Leading a paradigm shift in the application of real-time AI at scale to advance science and engineering discoveries.

One example: gravitational waves

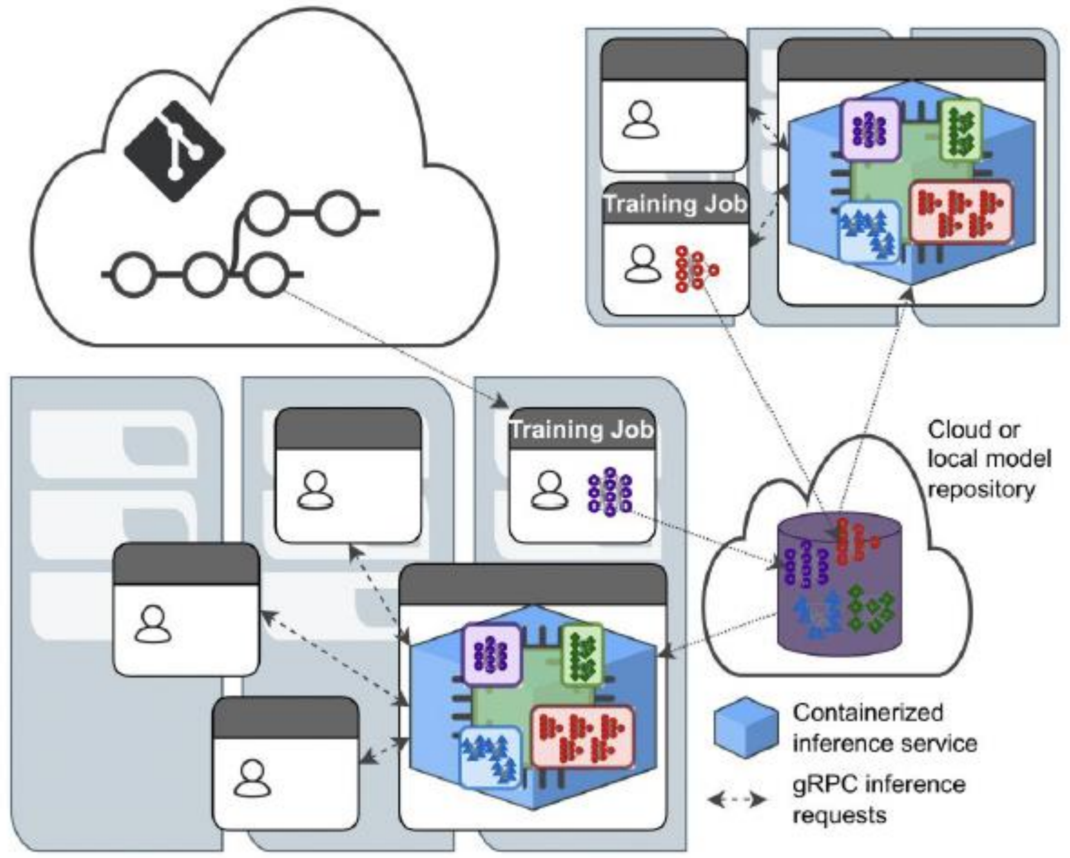
ML4GW

Utilities

ml4gw: Torch-based GW tools



hermes: Simplifying IaaS



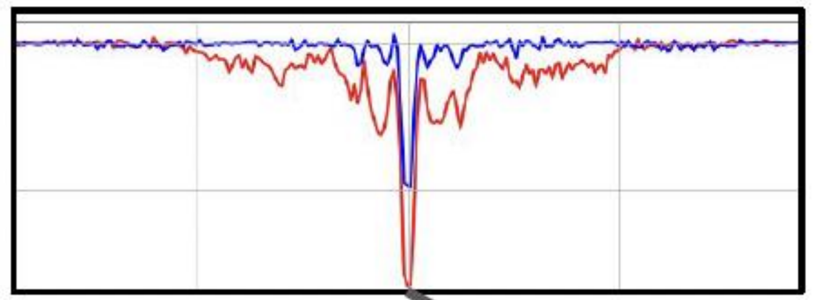
Discovery Pipeline

Noise removal

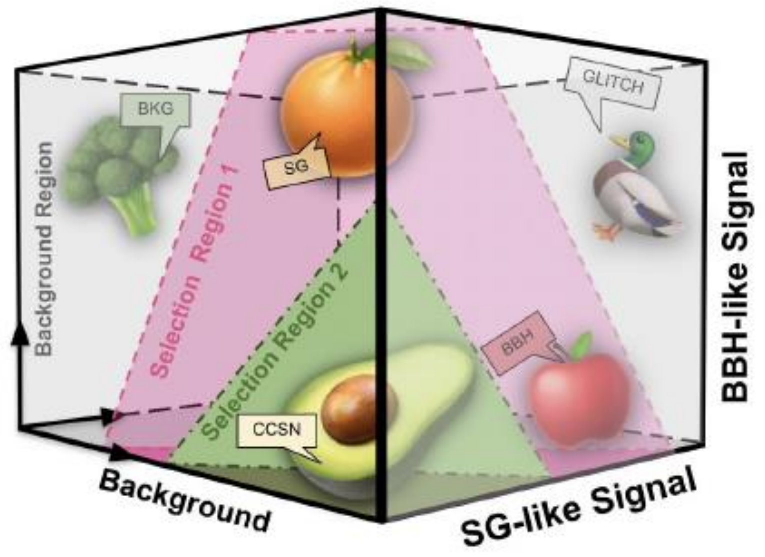
Detection

Parameter Estimation

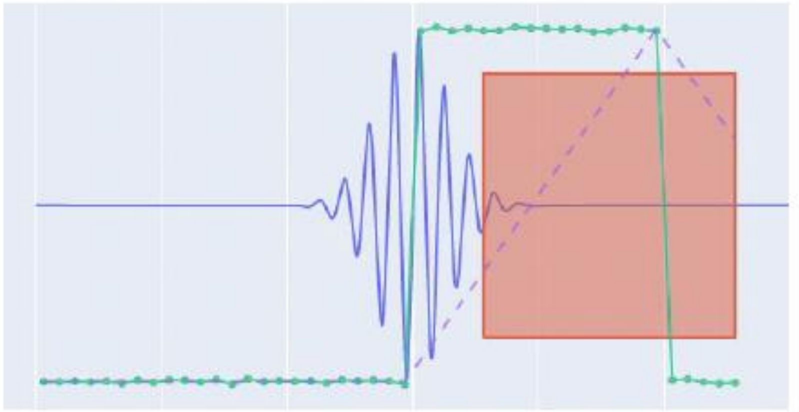
DeepClean



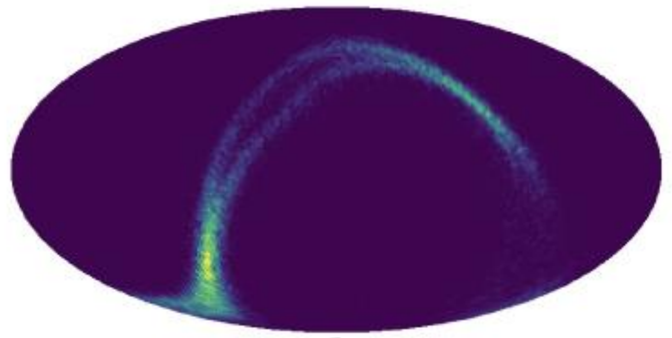
GWAK (Anomalies)



Aframe (CBCs)



AMPLFI



From discovery to a compute problem

Every detection needs a false-alarm rate. In gravitational waves, we compute this using an empirical background

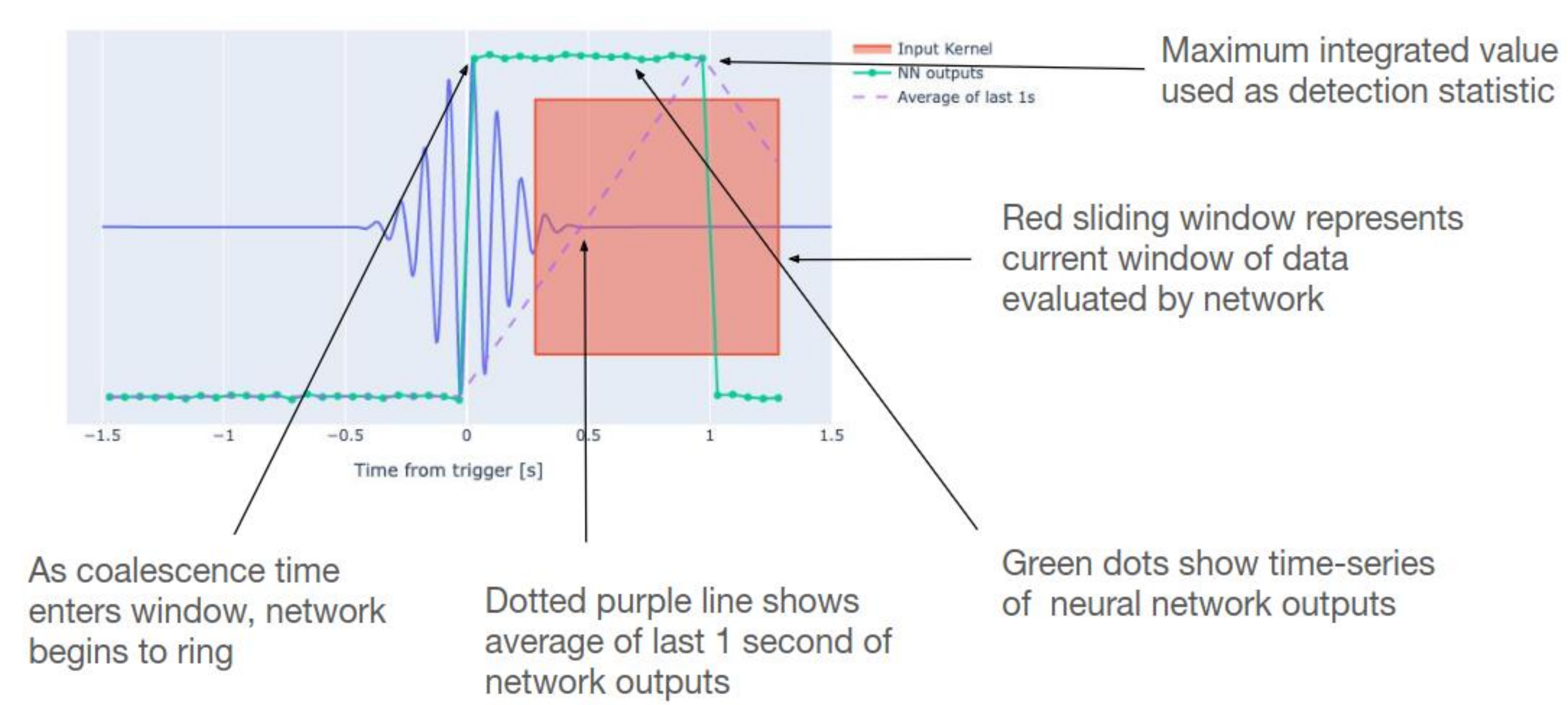
Background = timeslides: O(100) years of it to claim O(years) significance

Plus injection campaigns (ours + collaboration rates-&-populations sets)

None of it is optional. None of it is cheap.

The good — Inference-as-a-Service got us into O4

- Aframe (CBC detection) + AMPLFI (parameter estimation) + DeepClean (denoising)
- **hermes (ML4GW): Torch → ONNX → TensorRT (FP16), served by NVIDIA Triton**
  - Triton inference server on the access point
  - HTCondor jobs do dataloading + fire async inference requests



Maximum integrated value used as detection statistic

Red sliding window represents current window of data evaluated by network

As coalescence time enters window, network begins to ring

Dotted purple line shows average of last 1 second of network outputs

Green dots show time-series of neural network outputs

Aframe detected S250830bp ~5 s before another pipeline — first alert

The good — throughput, when it works

Vanilla single-GPU: ~512 s′/s on one 16 GB V100

  1 yr background ≈ 17 hr → 100 yr ≈ 70 days ('assuming nothing goes wrong')

Local IaaS (Triton): ~3800 s′/s, scales nearly linearly with GPUs

  1 yr background → a couple of hours

~7× per-GPU — and it keeps scaling, if we can feed it GPUs

The bad — 'assuming full access to all available GPUs'

That 3800 s′/s assumes the whole LDG GPU pool — which we never actually get

We contend for ~a dozen GPUs shared across the entire collaboration

Inference is increasingly bottlenecked by data transfer, not compute

- High inference cadence ⇒ heavily overlapping input windows ⇒ redundant I/O

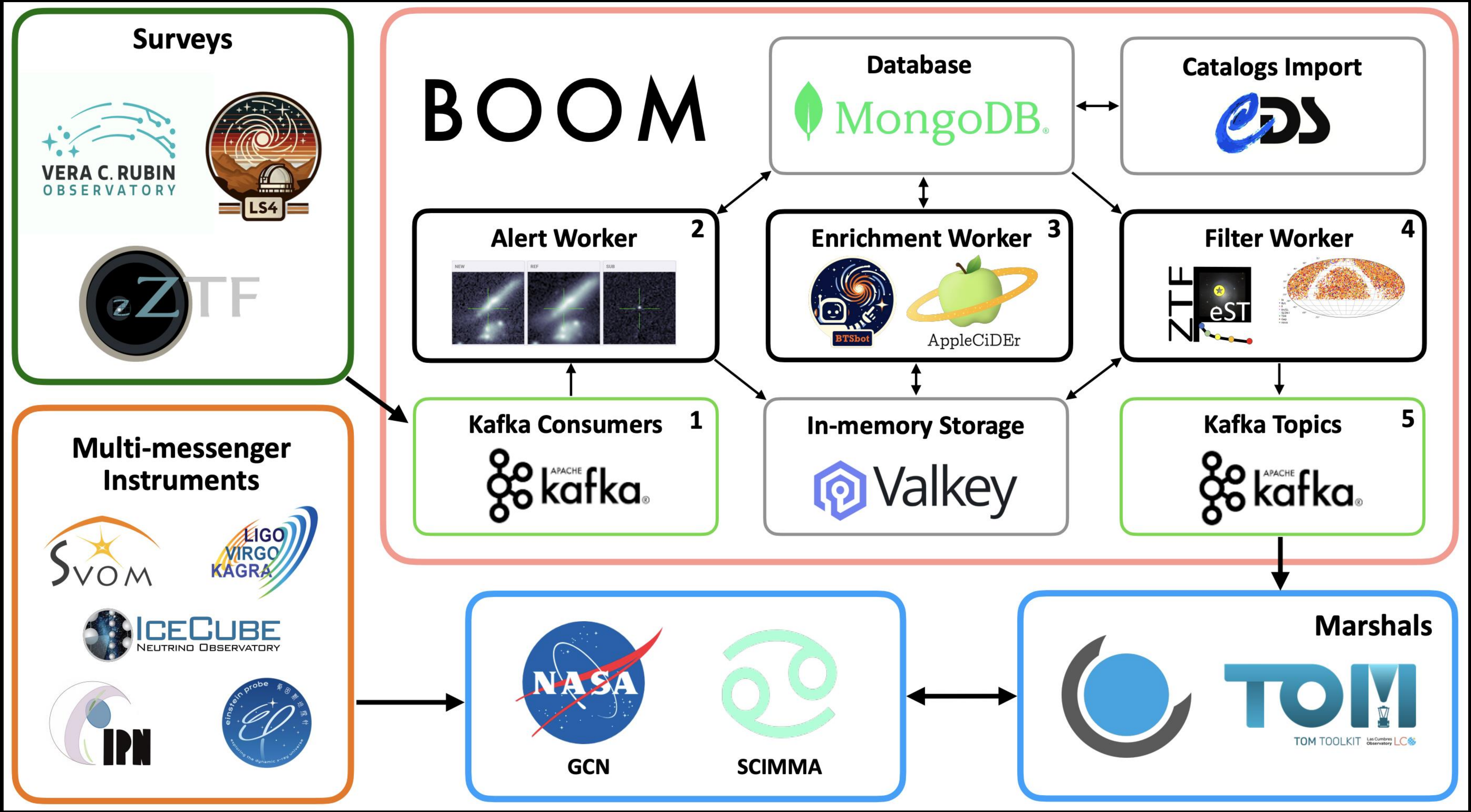
Throughput we can demonstrate ≫ throughput we can sustain

HTC ≠ HPC — matching our workloads to the right resource

- **Latency-critical → needs HPC / dedicated resources**
  - Online alerts: GWCelery, live Aframe / AMPLFI inference
  - Rubin real-time data processing
  - *HTC can't guarantee task priority or time-to-completion — wrong tool here*
- **Throughput, latency-tolerant → HTC / OSPool sweet spot**
  - O(100) yr background timeslides + injection campaigns
  - NMMA forward-modeling on individual interesting sources (kilonova / EM fits)
  - *Embarrassingly parallel; 'best-effort, no priority guarantee' is acceptable here*
  - *NRP (Nautilus k8s GPU hypercluster) — same model: dev/offline, not production LL*
- **Containers: Apptainer images (OSPool standard), not Docker**

Where is the multi-messenger?

The technical ecosystem we support



Surveys

VERA C. RUBIN OBSERVATORY

LS4

ZTF

Multi-messenger Instruments

SVOM

LIGO VIRGO KAGRA

ICECUBE NEUTRINO OBSERVATORY

IPN

einstein probe

BOOM

Database
MongoDB

Catalogs Import
CDS

Alert Worker 2
NEW REF SUB

Enrichment Worker 3
BTSbot
AppleCiDEr

Filter Worker 4
ZTF
eST

Kafka Consumers 1
APACHE kafka

In-memory Storage
Valkey

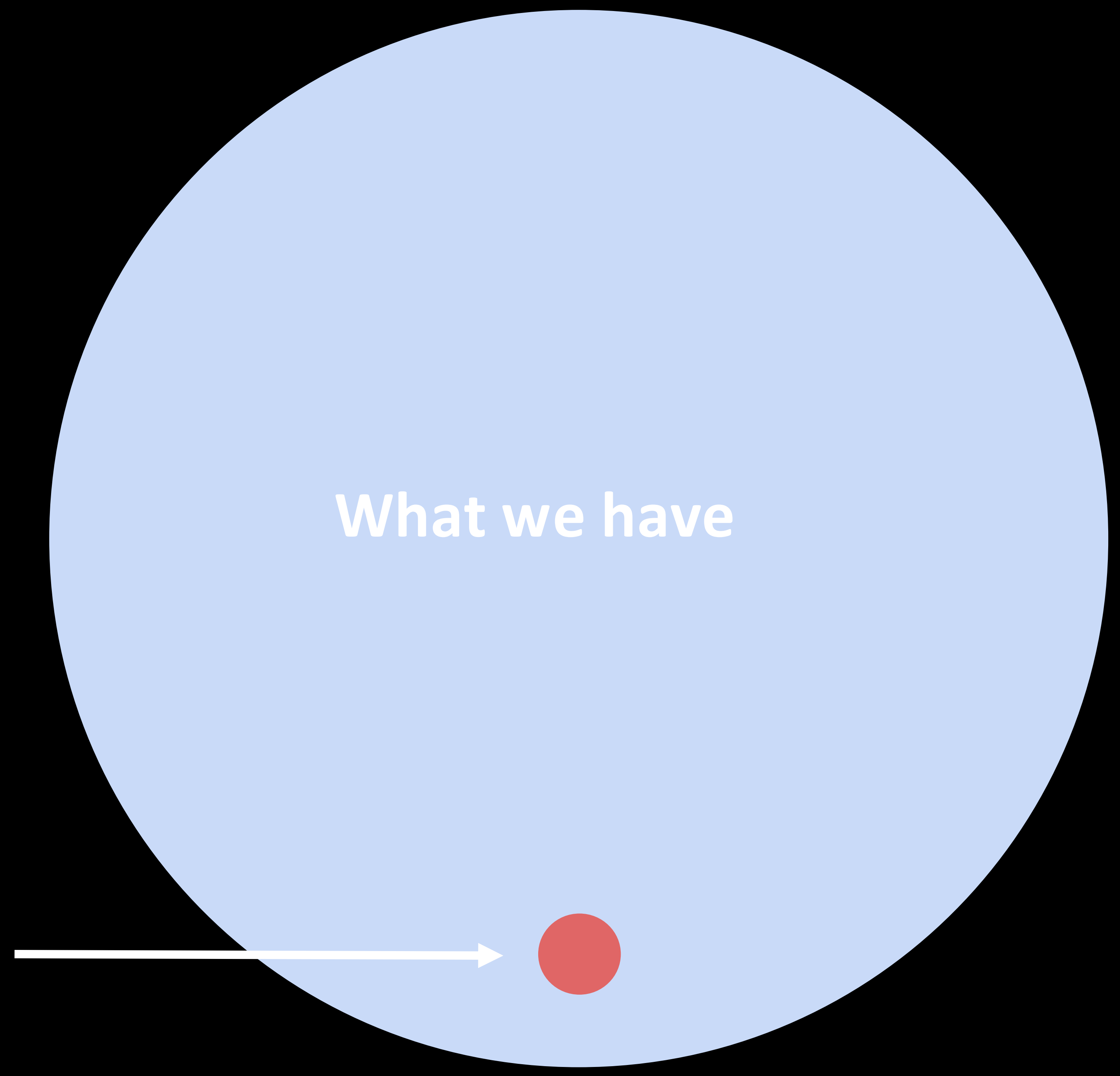
Kafka Topics 5
APACHE kafka

NASA
GCN

SCIMMA

Marshals
TOM
TOM TOOLKIT Las Cumbres Observatory LCO

A ~~Astronomy~~ Big Data* Problem

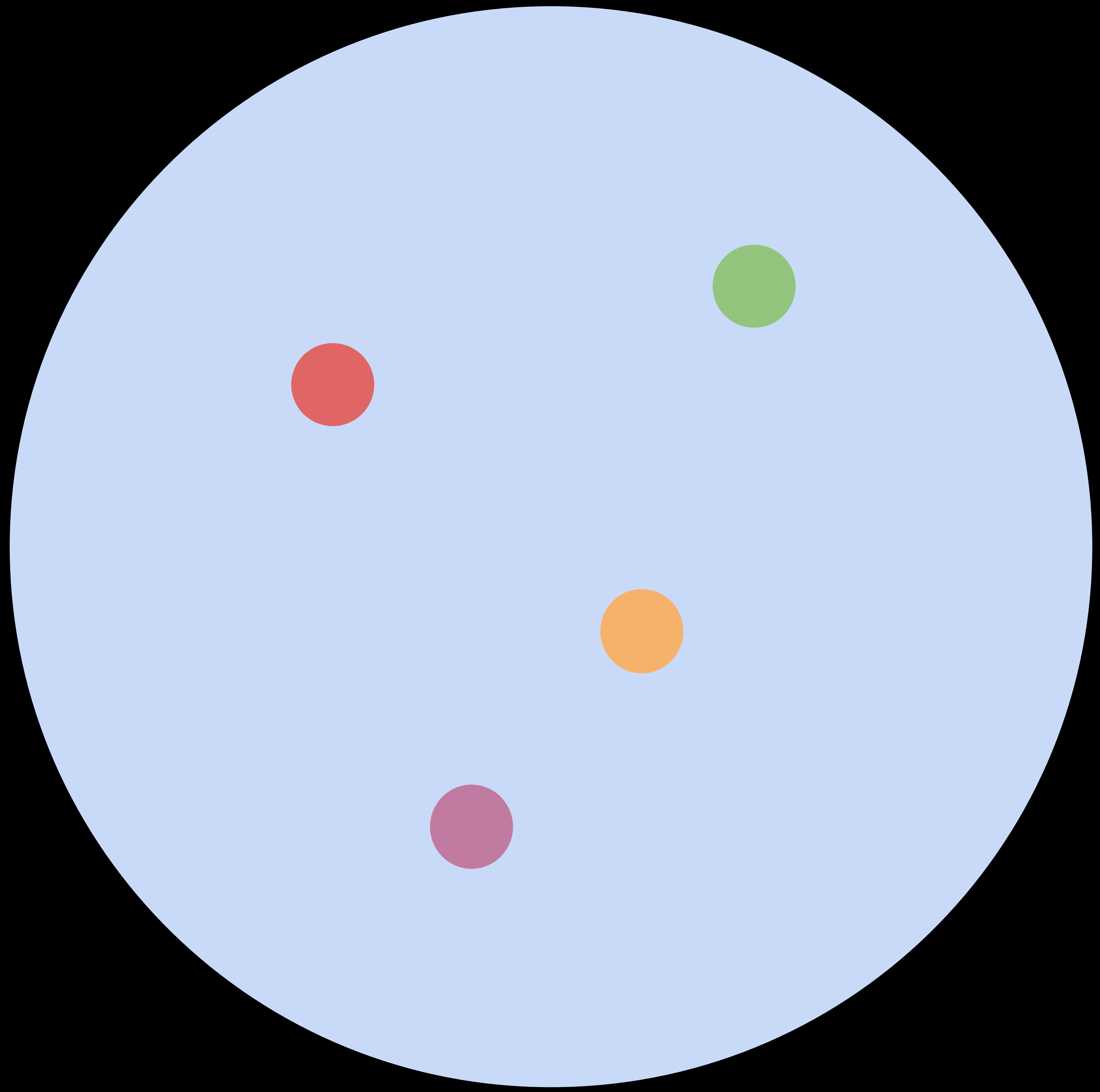


What we have

What we care about

*Rubin: 20 million alerts / night

An Astronomy Problem



- ● What **A** cares about
- ● What **B** cares about
- ● What **C** cares about
- ● What **D** cares about

A Latency Problem

We are looking for…

- Rare objects
- Young objects
- Fast evolving objects
- …

Our solution: BOOM

A custom multi-survey broker written in Rust with performance and interactions with community scientists as the highest priorities.*

* As Daniel mentioned, there are a number of other brokers, that solve various aspects of this problem. It is not so dissimilar to the idea behind there being a number of production LVK analyses.
https://rubinobservatory.org/for-scientists/data-products/alerts-and-brokers

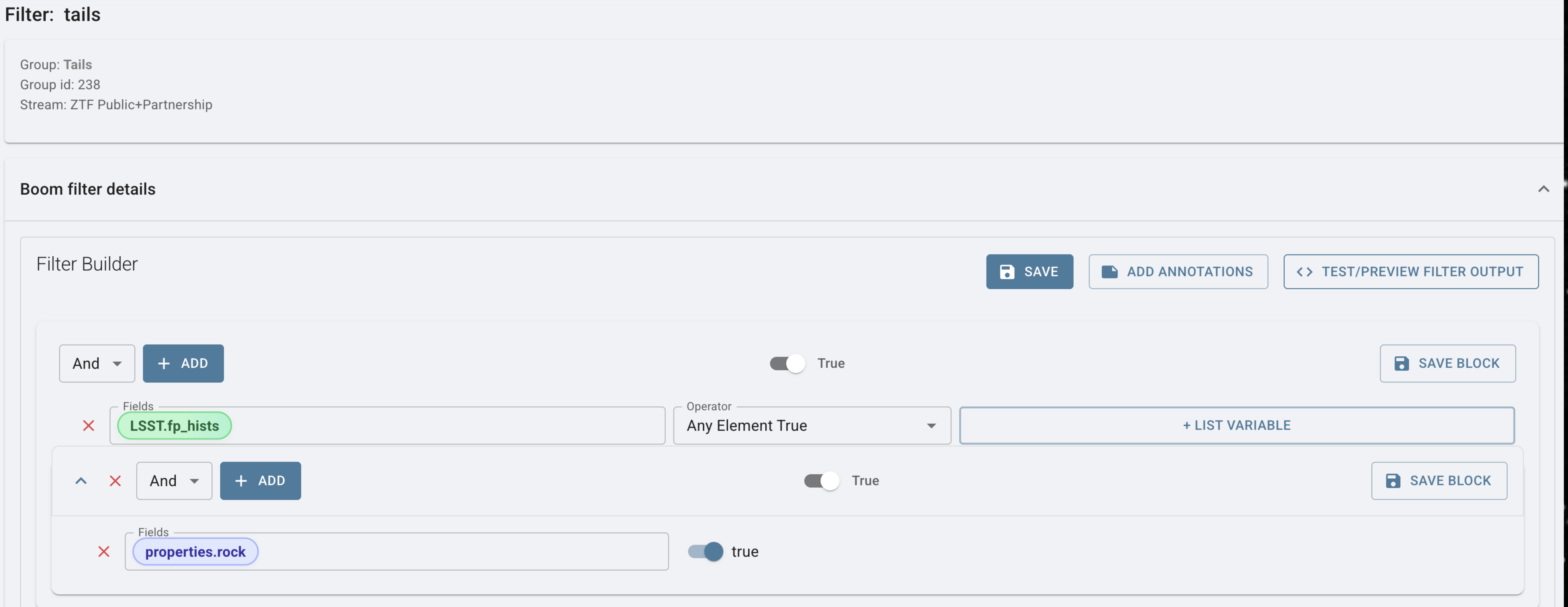
It is kind of like the ~7 real-time pipelines in the LVK, but as if the pipelines were also alert systems and everyone could subscribe to any one.

Trying to impose some structure on chaos

To run in production, we will require that all ML pipelines be exportable in ONNX format.

- *Charnock & Moss (2017)*
- *RAPID (Muthukrishna+2019)*
- *AVOCADO (Boone+2019)*
- *SuperNNova (Möller+2019)*
- *FLEET (Gomez+2020, Gomez+2022)*
- *SuperRAENN (Villar+2020)*
- *Stamp Classifier (Carrasco-Davis+2020)*
- *SCONE (Qu+2021)*
- *ParSNIP (Boone+2021)*
- *TimeModRNN (Pimentel+2023)*
- *First Impressions (Gagliano+2023)*
- *NEEDLE (Sheng+2023)*
- *Superphot+ (de Soto+2024)*
- *BTSBot (Rehemtulla+2024)*
- *More…*

To enable the scientists to help themselves



Filter: tails

Group: Tails
Group id: 238
Stream: ZTF Public+Partnership

Boom filter details

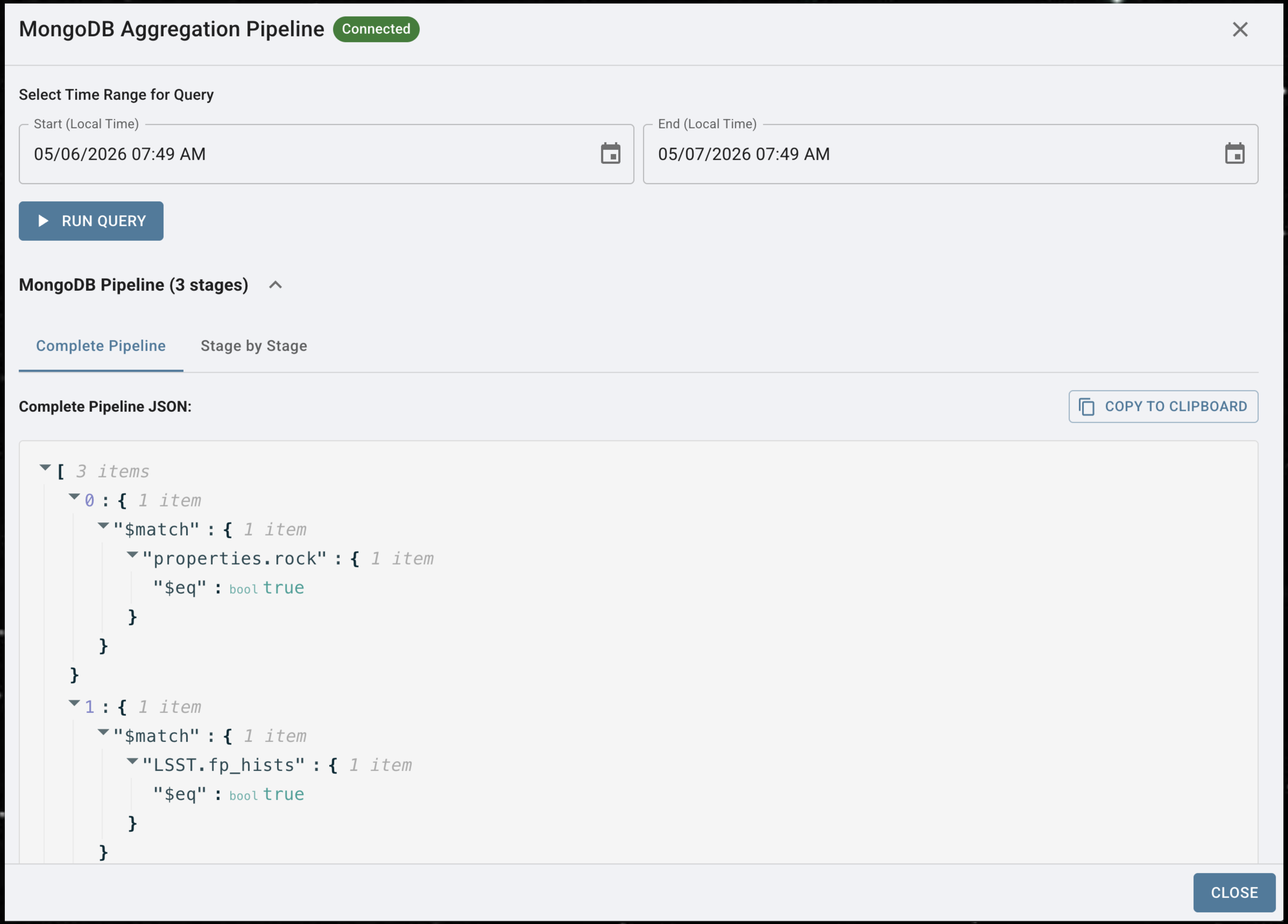
Filter Builder

And — True
Fields: LSST.fp_hists — Operator: Any Element True

And — True
Fields: properties.rock — true

We are building out front-end interfaces so scientists can create filters…

To enable the scientists to help themselves



Select Time Range for Query

Start (Local Time): 05/06/2026 07:49 AM

End (Local Time): 05/07/2026 07:49 AM

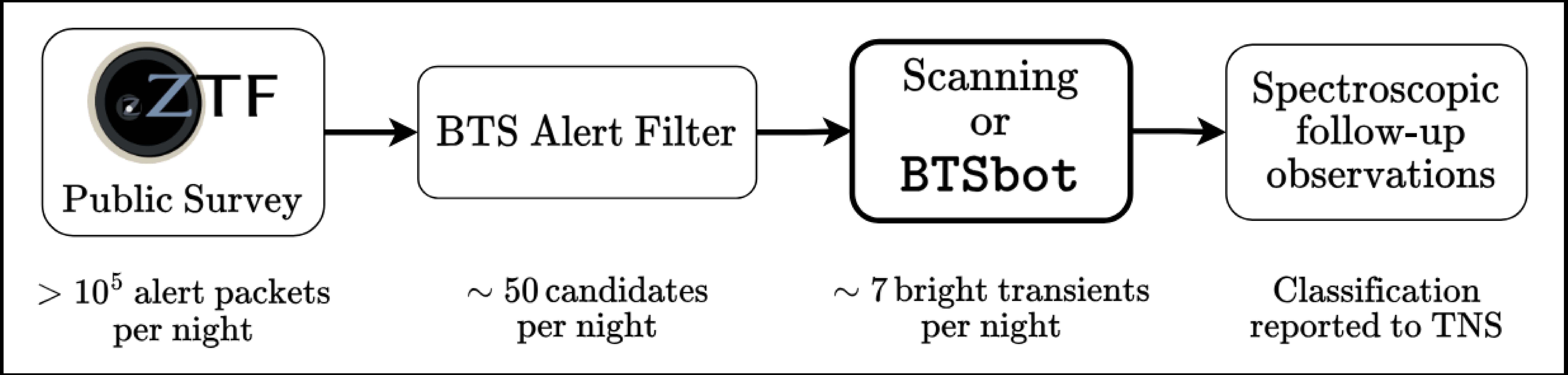
Complete Pipeline JSON:

```
▼[ 3 items
  ▼0 : { 1 item
    ▼"$match" : { 1 item
      ▼"properties.rock" : { 1 item
        "$eq" : bool true
      }
    }
  }
  ▼1 : { 1 item
    ▼"$match" : { 1 item
      ▼"LSST.fp_hists" : { 1 item
        "$eq" : bool true
      }
    }
```

Automating the export of the associated MongoDB queries…

And has been enabling some cool ML workflows…

End-to-end automation



ZTF Public Survey → BTS Alert Filter → Scanning or BTSbot → Spectroscopic follow-up observations

$> 10^5$ alert packets per night

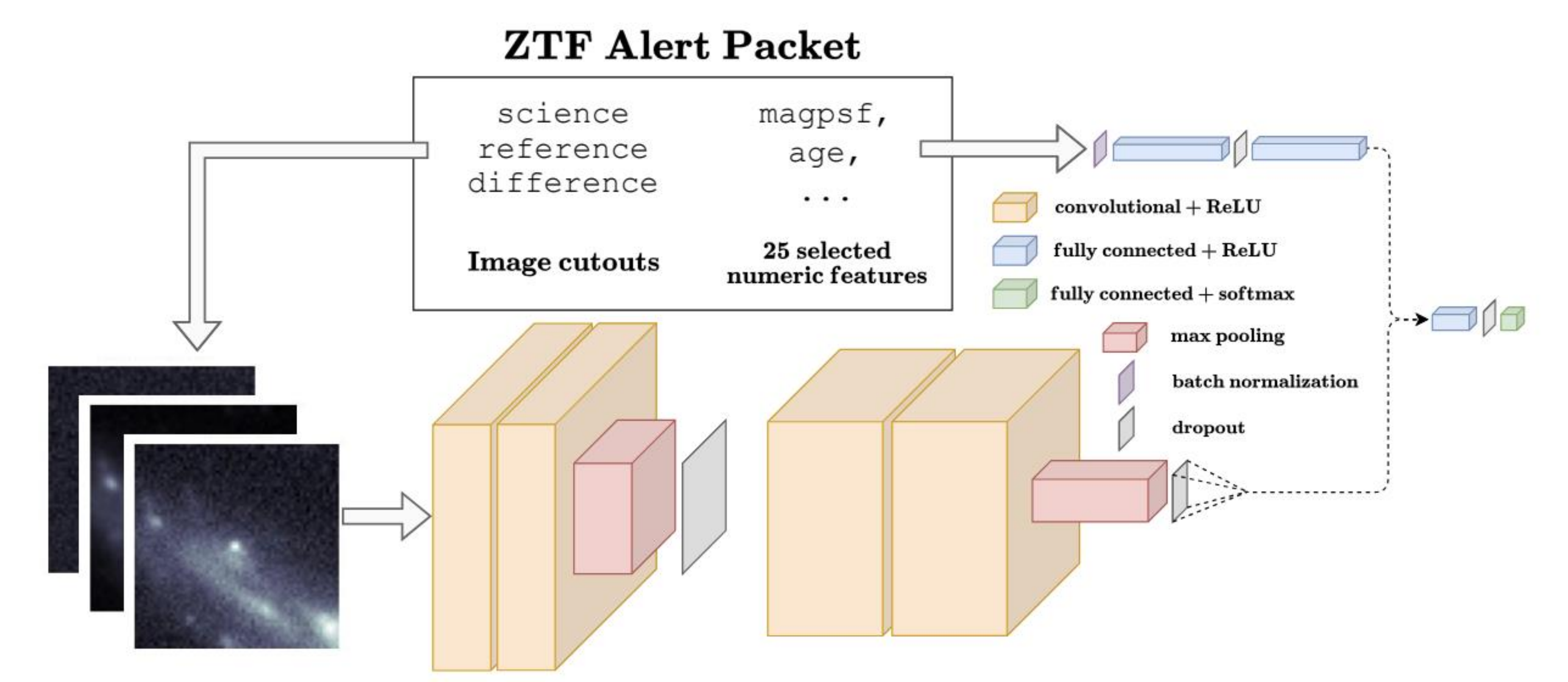
$\sim$ 50 candidates per night

$\sim$ 7 bright transients per night

Classification reported to TNS

BTSbot (Rehumtulla+ 24):

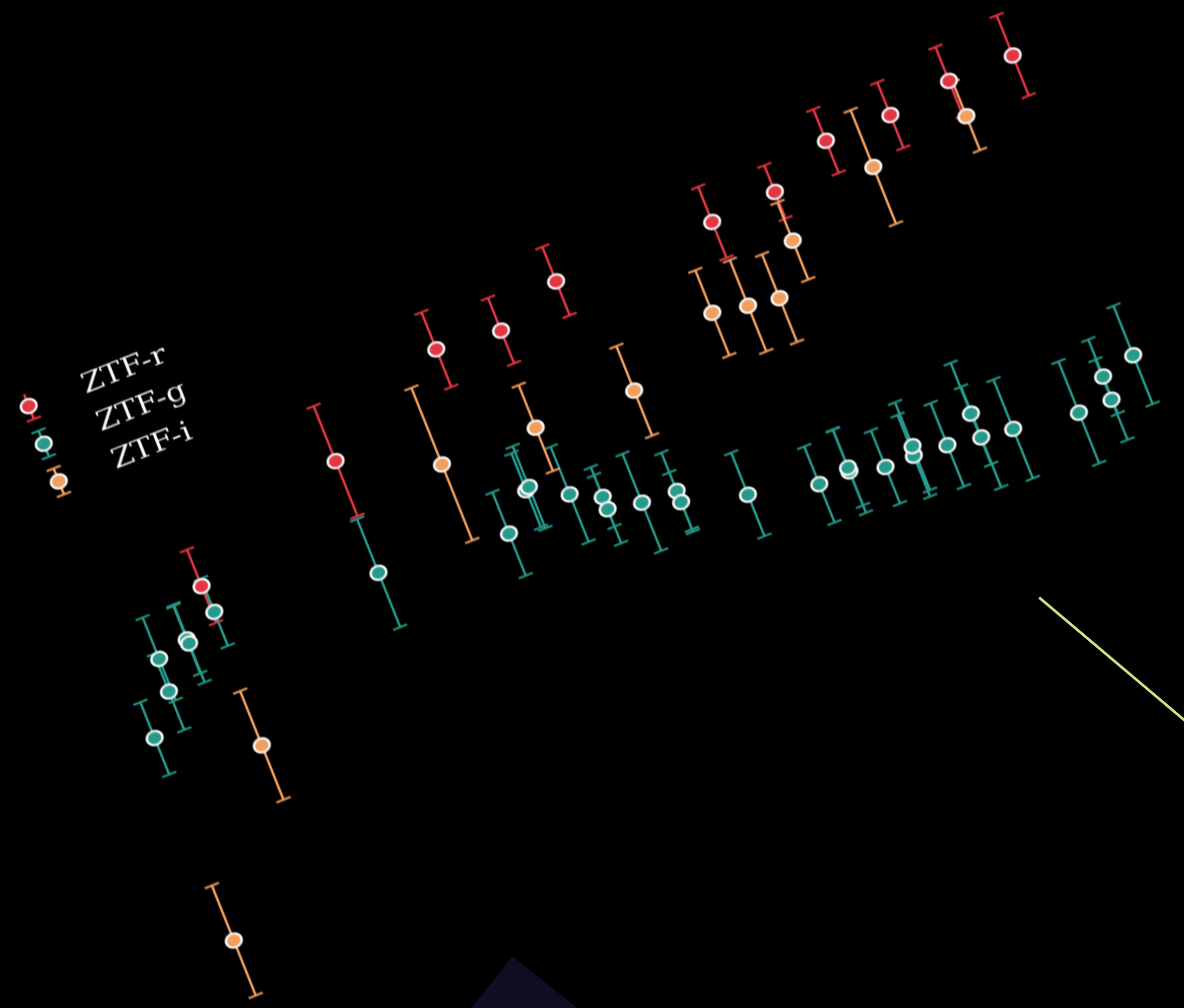
- Automatically submits reports of spectroscopically classified SN Ias to Transient Name Server (TNS):
  - >1000 sources saved by BTSbot
  - >700 SEDM triggers sent
  - >100 fully autonomously classified SN Ia
  - A significant boost in survey efficiency

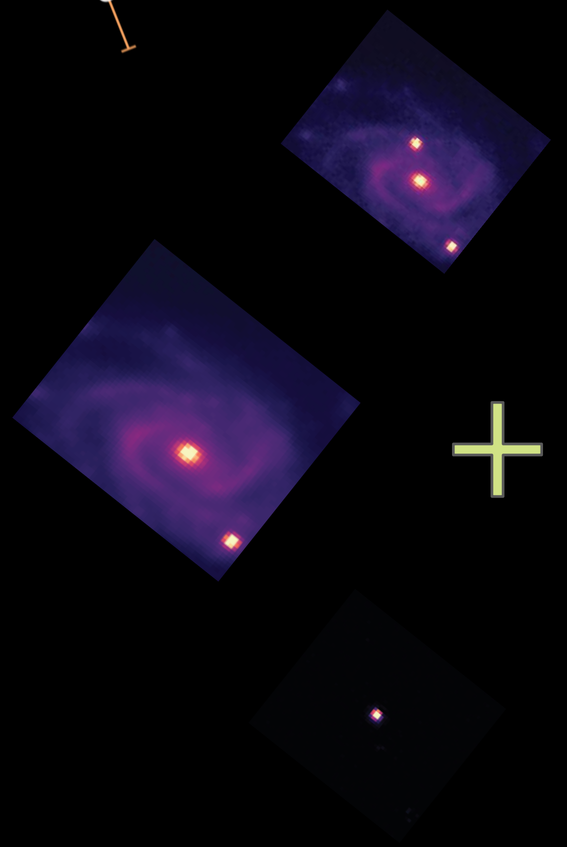


Multi-modal classification at scale

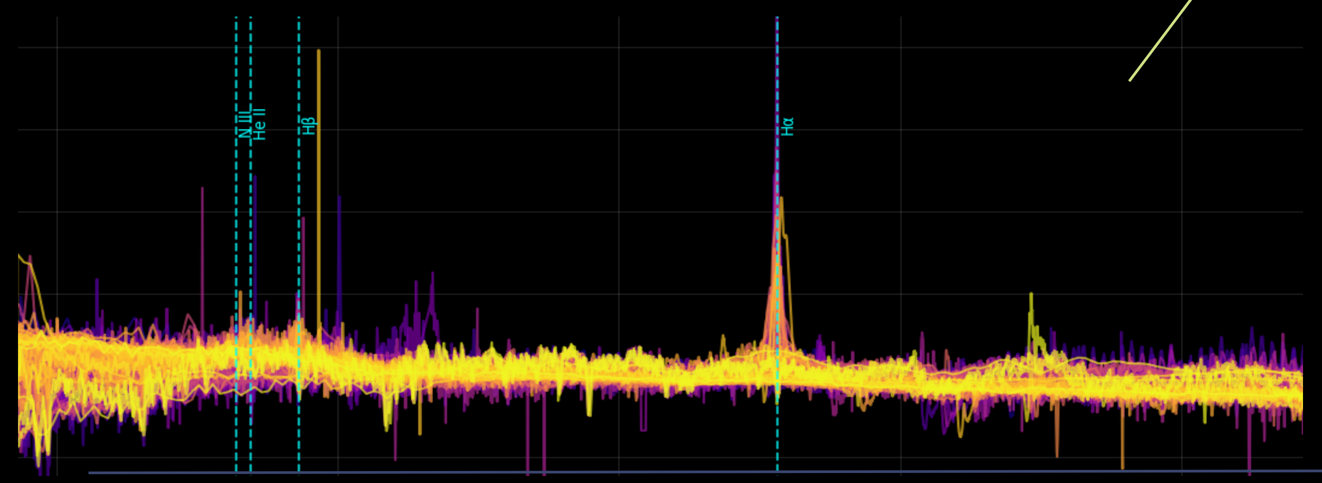
[Junell, Sasli, et al.,2024]

[Xu, Sasli, et al.,2024]





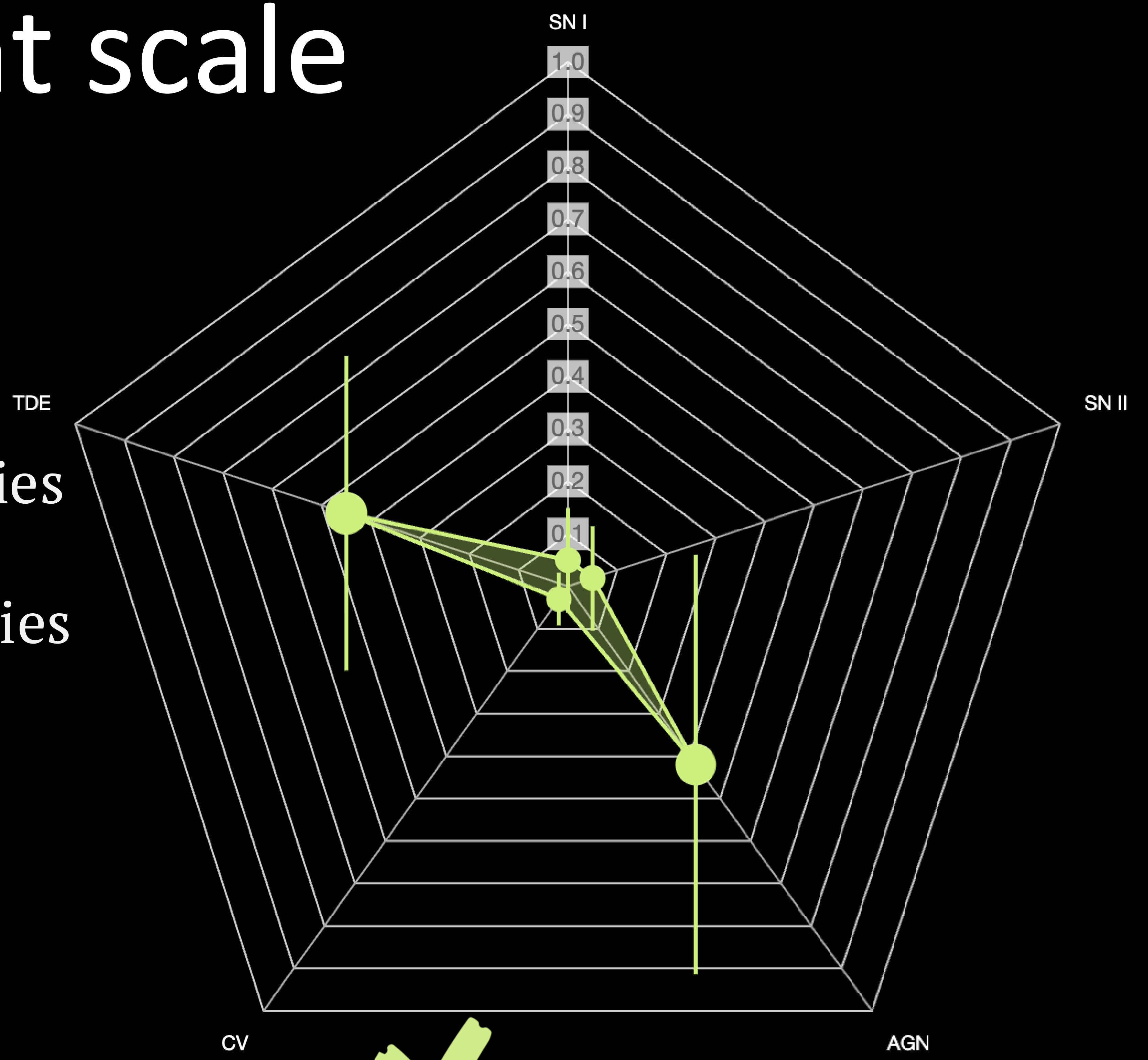
+ f_1 | f_2 | . . . . . . | f_N



AppleCiDEr

=

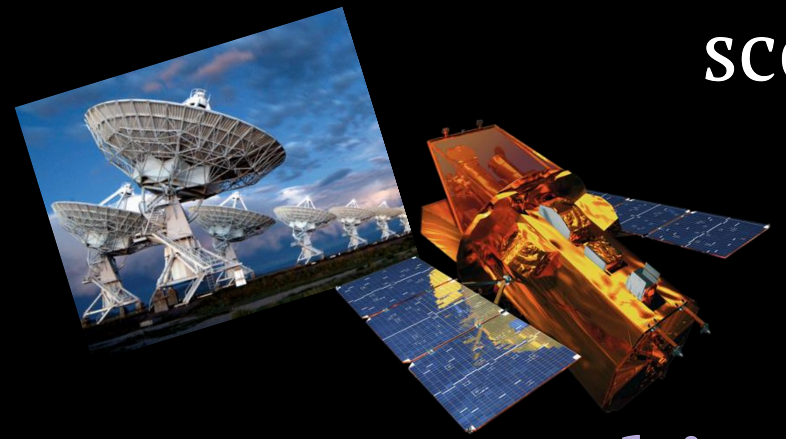
Probabilities
&
uncertainties



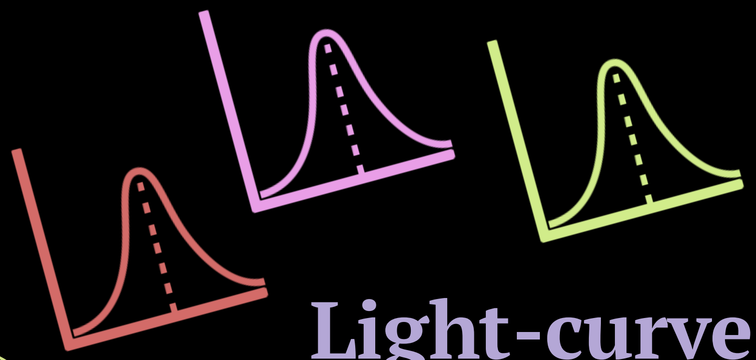
Anomaly
score

Low

High



Multi-wavelength
follow-up



Light-curve
parameter
estimation

github.com/skyportal/applecider

The computing challenge & where we need the community

- **Scale: Rubin delivers ~10M alerts/night — each enriched, classified, filtered in near-real time**

This is the online / latency-critical regime → dedicated 24/7 service infra, not best-effort HTC

- Combine ZTF + Rubin (+ GW / Fermi / SVOM / Einstein Probe) streams without falling behind

Open questions for the community:

Elastic scaling of Kafka consumers + GPU classification at Rubin rates?

Cost-effective per-alert ML inference at ~10M/night?

Sharing / serving embedding spaces across brokers & downstream users?

Hosting a 24/7 low-latency service: dedicated vs NRP k8s vs cloud?

O5 vision: feed the filtered stream into joint GW × EM classification pipelines

Electromagnetic Wave Windows

X-Ray

Optical

Radio

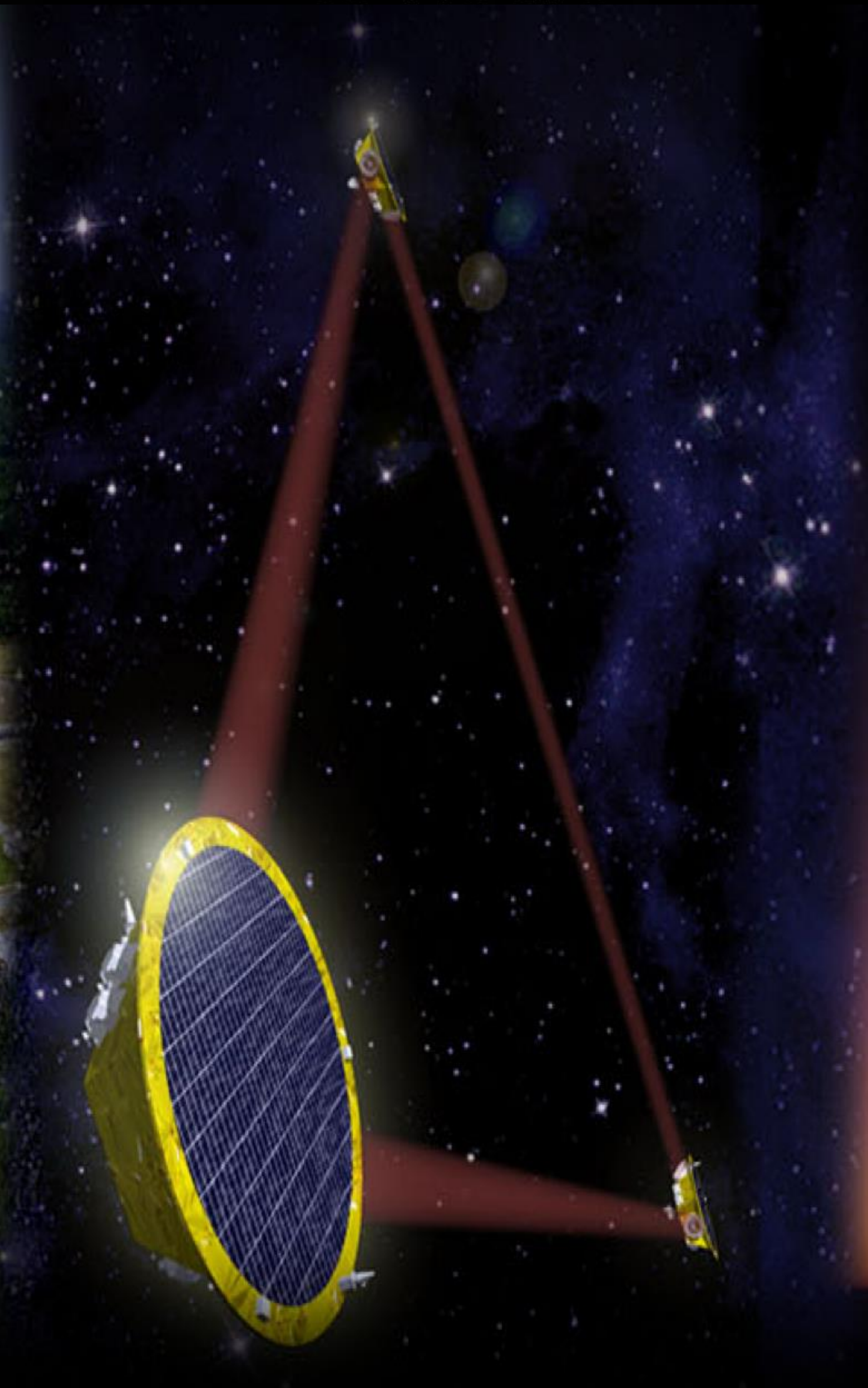
L. Nuttall

Gravitational Wave Periods

Milliseconds

LIGO/Virgo

Minutes
to Hours

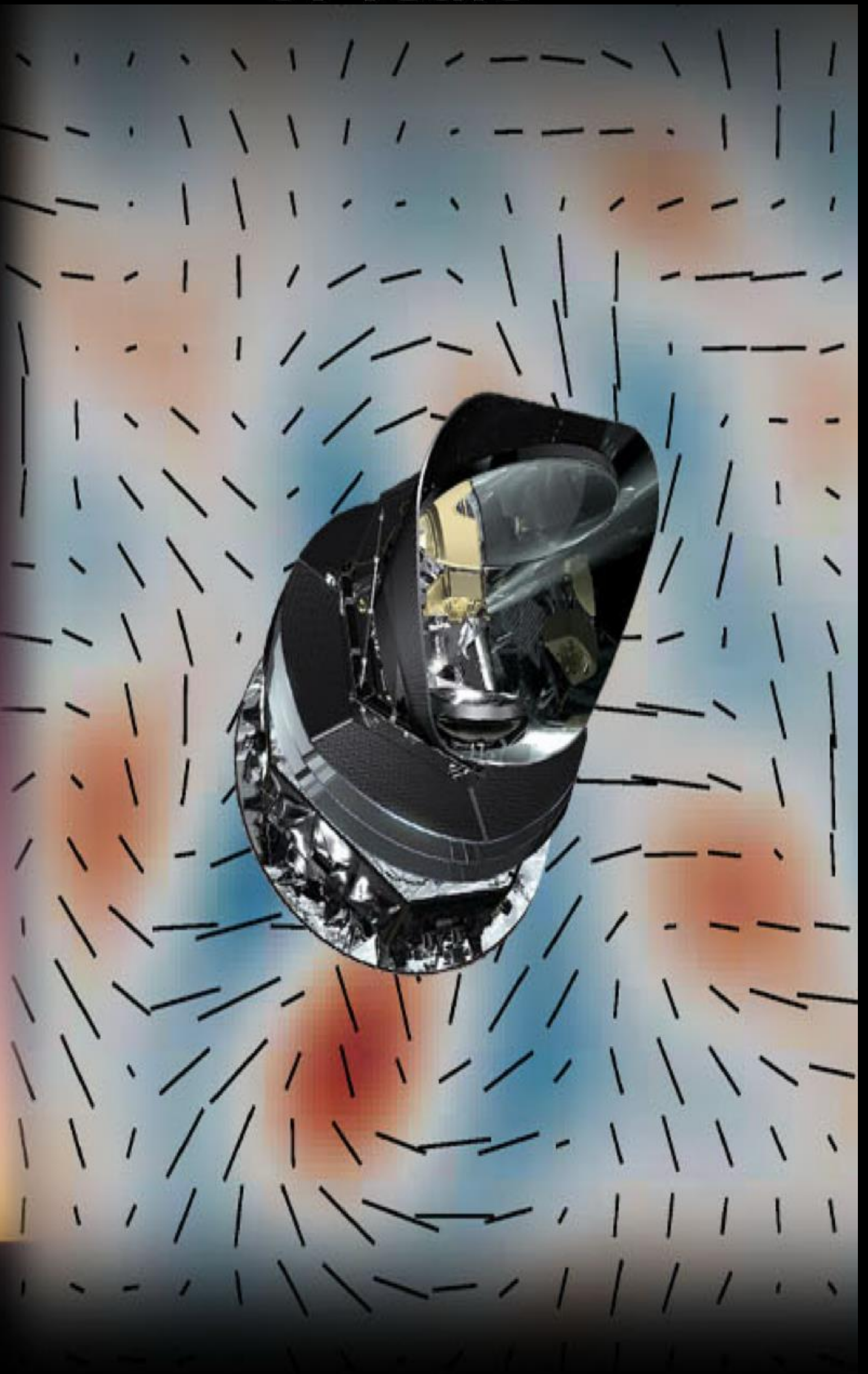


LISA

Years
to Decades

Pulsar timing

Billions
of Years



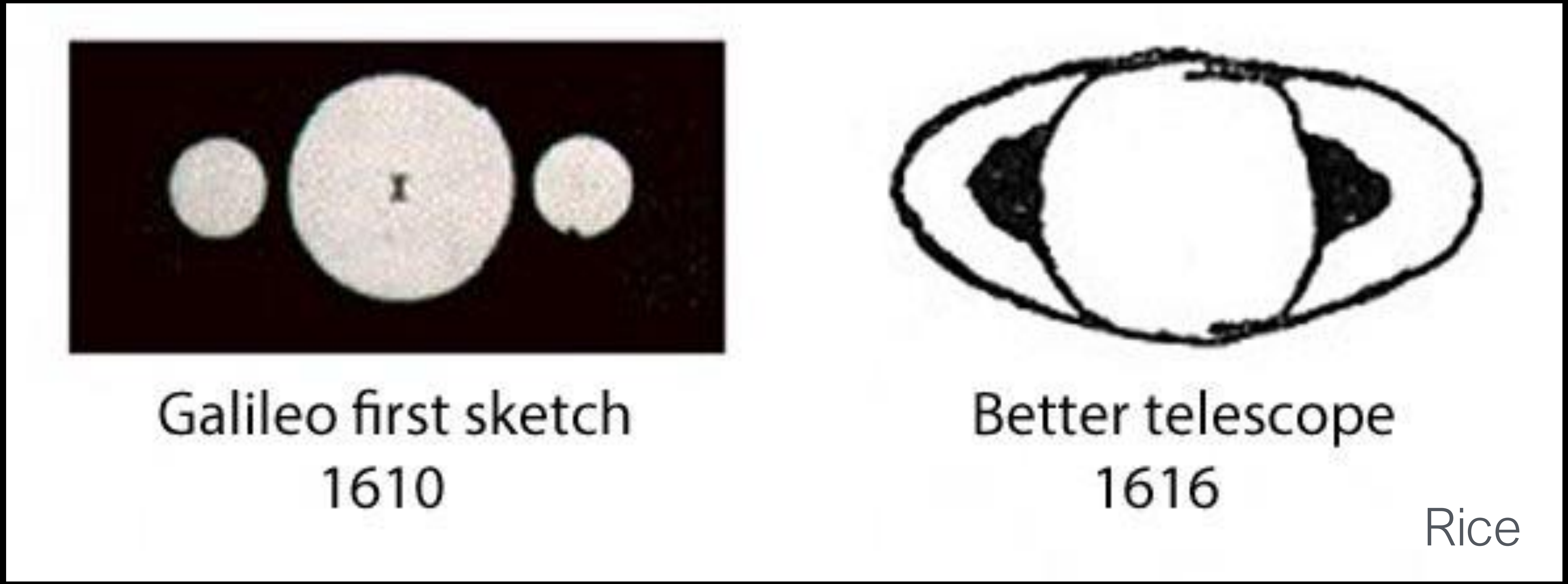
CMB polarization

Illustration - NASA

Illustration -NASA

Illustration

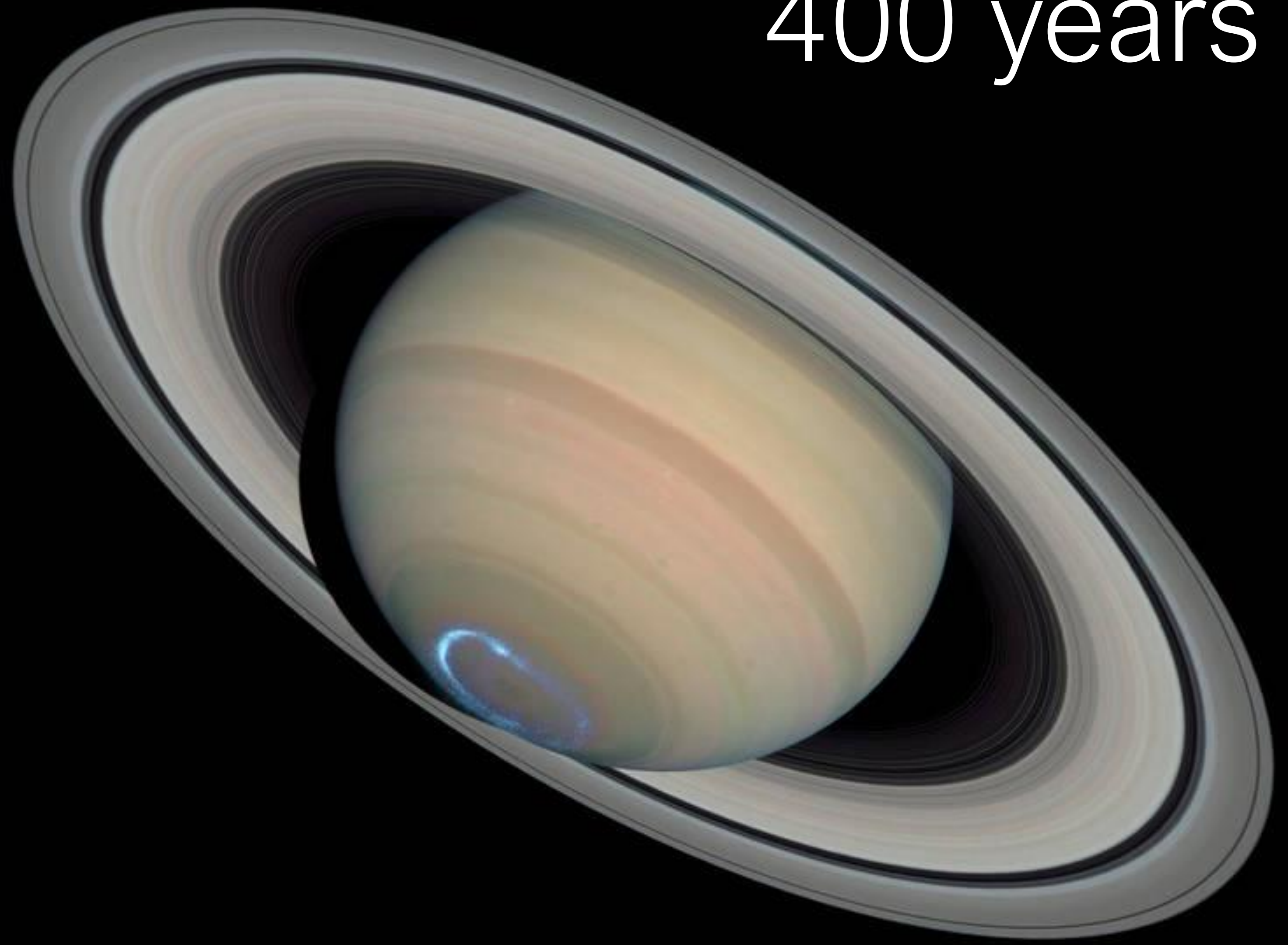
This is just the beginning of gravitational wave astrophysics!

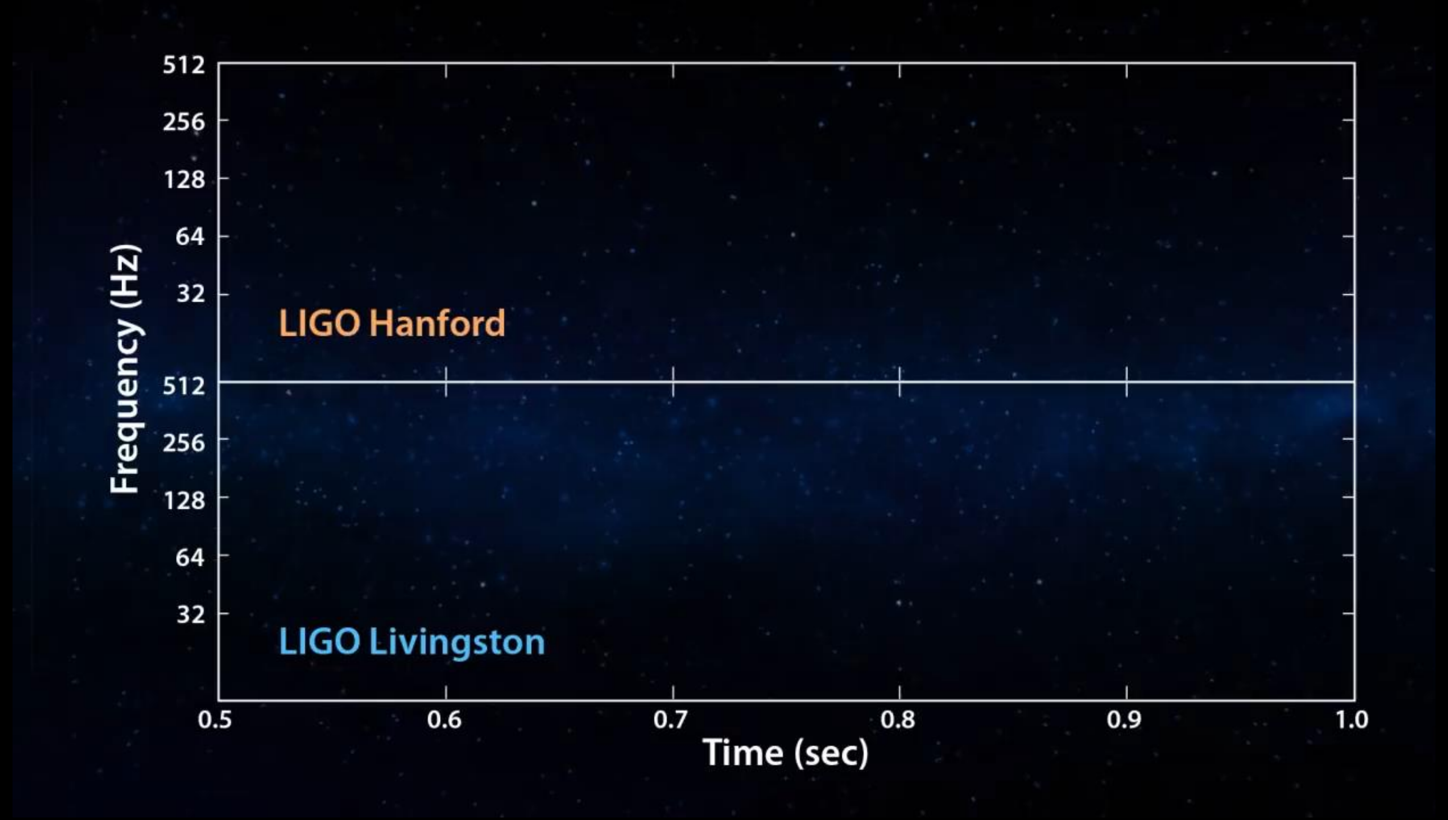

Galileo first sketch
1610

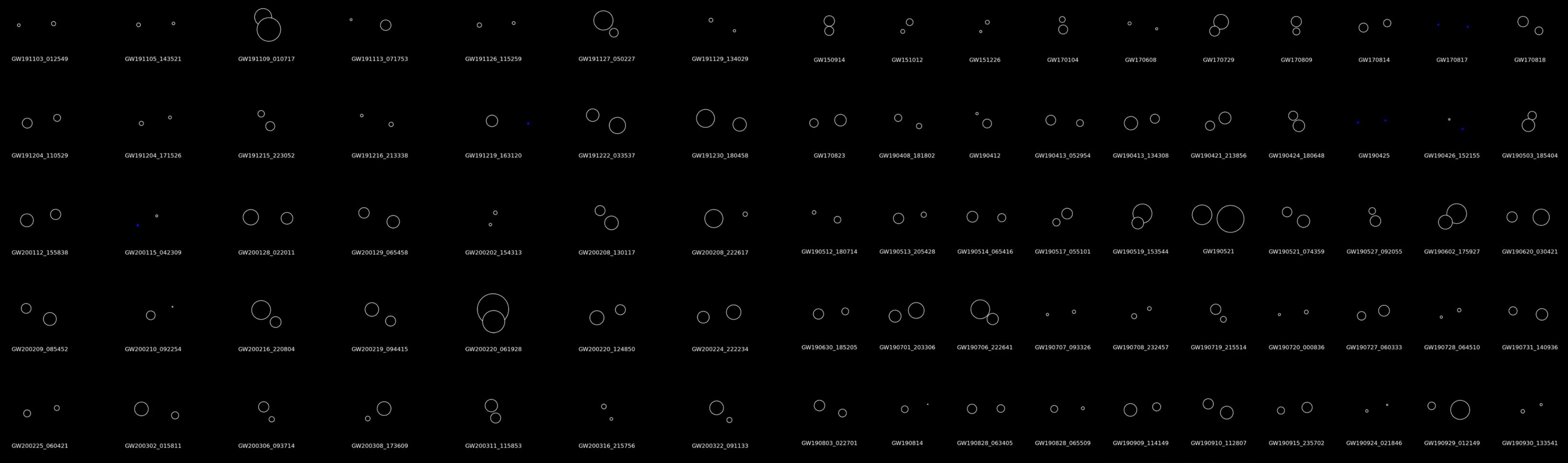
Better telescope
1616

Rice

400 years later…


HST




Zoheyr Doctor / CIERA / LIGO-Virgo Collaboration

Zoheyr Doctor / CIERA / LIGO-Virgo Collaboration

?

Slide by G. Losurdo

Visit the Gravitational Wave Open Science Centre: gwosc.org

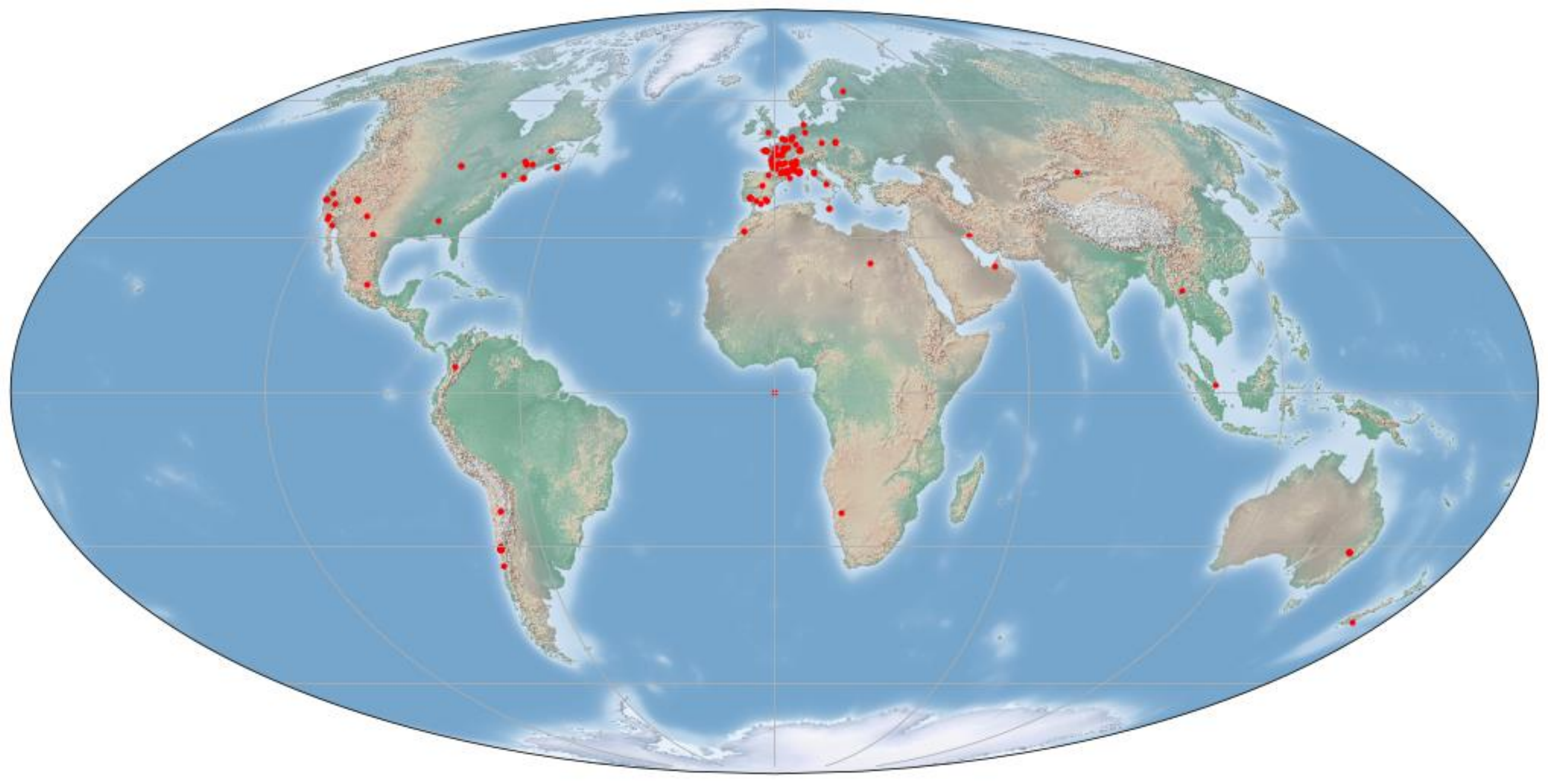
Help us out on GravitySpy.org!

NSF/LIGO/Sonoma State University/A. Simonnet

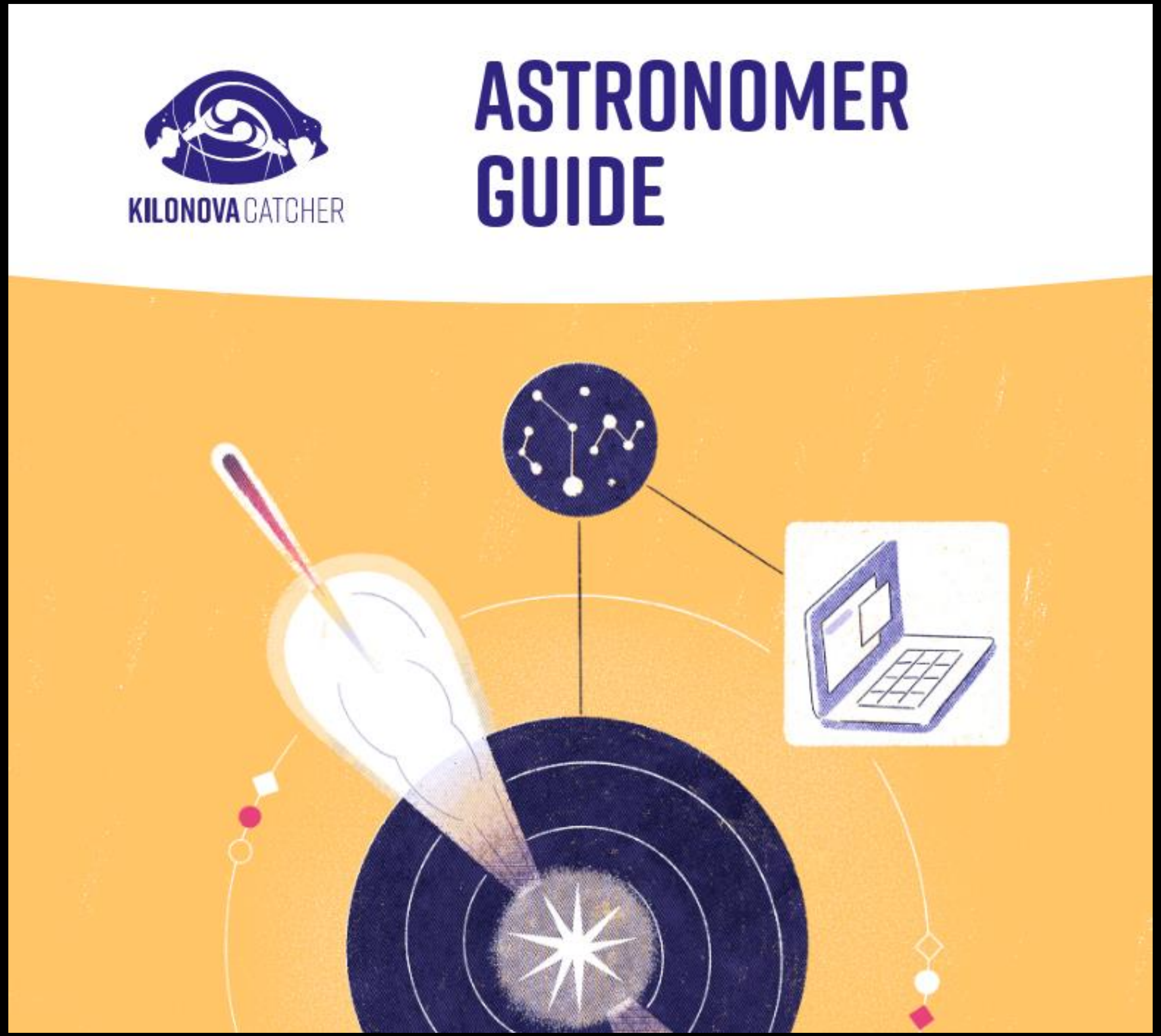
Kilonova Catcher

KILONOVA CATCHER

130 amateur astronomers
World wide



2 publications / yr with citizens as co-authors



KILONOVA CATCHER

ASTRONOMER GUIDE

Operated by GRANDMA

Thank you